

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

School of Science
Department of Physics and Astronomy
Astrophysics and Cosmology Master's Degree

Enhancing the potential of gravitational waves as standard sirens: a statistical analysis

Candidate:
Niccolò Passaleva

Supervisor:
**Prof. Michele Ennio Maria
Moresco**

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

Academic Year 2022/2023

Abstract

The discovery of gravitational waves (GWs) from compact binary coalescences in 2015 unlocked new possibilities for studying the Universe. As their intrinsic loudness can be predicted by General Relativity, GWs provide a direct measurement of the luminosity distance, making them standard sirens. To complement this information and use them as cosmological probes, it is necessary to break the degeneracy between the redshift and the binary masses. In the absence of an electromagnetic counterpart, galaxy catalogues can be used to break this degeneracy, and in this case GWs are referred to as dark sirens.

In this Thesis, I used GW events as dark sirens to constrain parameters of different binary black hole mass function (MF) models. I investigated two mock GW catalogues simulating the current O4 and future O5 observing runs by the LIGO-Virgo-KAGRA network, and explored the capability of discriminating between the different MF models with future data. I implemented a new MF model and a nested sampling-based posterior sampling method in the `CHIMERA` code to estimate Bayesian evidence of different MFs. In addition, I developed code to compute statistical diagnostics such as $\log\mathcal{B}$, DIC, and PPC for model selection. Analysing the GW catalogues with available MFs, I constrained their parameters, compared them, and identified the best-fitting model along with the capability to distinguish them. Simultaneously, I conducted a sizing analysis to estimate computational time for GW data analysis. In the last part, I investigated the potential effects of different catalogue assumptions on results, including errors in galaxy catalogue redshift, GW event data smoothing methods, and varying GW event samples.

This work paves the way for the optimisation of future GW analysis, proposing strategies to maximise the scientific return and computing capabilities and providing forecasts on expected performances achievable with future GW data.

Contents

Abstract	i
Preface	1
1 Introduction	7
1.1 Cosmological framework	7
1.1.1 Basic Equations of Cosmology	8
1.2 The theory behind gravitational waves	12
1.2.1 Gravitational waves propagation	13
1.2.2 Gravitational waves interaction with test masses	14
1.2.3 Gravitational waves generation	16
1.3 Gravitational waves as standard sirens	18
1.3.1 Bright sirens	19
1.3.2 Dark sirens	19
1.3.3 Spectral sirens	20
1.3.4 Cosmological interest	21
1.3.5 Astrophysical interest	22
1.4 State of the art of GW analysis	22
1.4.1 GW150914	23
1.4.2 GW170817	25
1.4.3 GWTC-3	26
1.4.4 Future observing runs: O4 and O5	28
1.4.5 Future detectors	29
2 Methods	33
2.1 Statistical framework	33

2.1.1	Likelihood	34
2.1.2	Selection effects	37
2.2	Black Hole Binaries mass function models	38
2.2.1	Truncated Power Law	38
2.2.2	Broken Power Law	39
2.2.3	Power Law + Gaussian Peak	40
2.2.4	Power Law + 2 Gaussian Peaks	42
2.3	Posterior sampling	43
2.3.1	Affine Invariant Sampling	45
2.3.2	Nested Sampling	46
2.3.3	Assessing the scaling performances of different samplers . . .	49
2.4	CHIMERA	53
3	Data	59
3.1	Catalogues generation	59
3.2	Catalogues characterisation	61
4	Results	65
4.1	The Black-Holes Binaries Mass Function	65
4.1.1	Extending the Binary Black Holes Mass Function Models . .	67
4.1.2	Model selection criteria	67
4.1.3	Constraining the BBH mass function parameters	70
4.1.4	Sizing tests	78
4.1.5	Disentangling different BBH mass function models with O4 and O5 data	80
4.2	Dependence of the results on the assumed catalogs	84
4.2.1	Dependence on z_{err}	85
4.2.2	Dependence on the smoothing of the GW localisation volumes	88
4.2.3	Dependence on the subsampling	89
5	Conclusions and Future Prospects	99
5.1	Main results	101
5.2	Future prospects	106
A	Corner plots	117

Preface

Scientific Framework The discovery of the accelerated expansion of the Universe represents a significant milestone in modern cosmology. To understand which are the underlying mechanisms driving this expansion, over the years, different cosmological probes, such as the Cosmic Microwave Background, Type Ia Supernovae, and Baryon Acoustic Oscillations, have been studied in depth and used. At present, these probes are being pushed to their limits, and impressive constraints have been obtained allowing us to shape the concordance cosmological model. However, the recent achievement of precision at the percentage level has shown tensions between measurements obtained through different methodologies, the main one being the H_0 tension between the values obtained with Type Ia supernovae (Riess et al., 2022) and the CMB (Aghanim et al., 2020). These tensions may be an indication of systematic effects that are not accounted for or of the first proof of new physics to be discovered. Thus, new cosmological probes are now being used in order to confirm or eliminate these tensions and increase the accuracy and robustness of parameter constraints (Moresco et al., 2022).

Since the first direct detection of a gravitational wave (GW) event, i.e. GW150914 (Abbott et al., 2016), and the later one of GW170817 (Abbott et al., 2017b), which opened the era of multimessenger astronomy, GWs have started to be used as alternative cosmological probes (Schutz, 1986; Holz and Hughes, 2005; Dalal et al., 2006). Indeed, since their loudness comes directly from assuming the validity of general relativity, they provide a model-independent estimate of the luminosity distance without the need for further calibrations. This is opposite to what is done with supernovae,

which require a distance ladder calibration. For this reason, GWs emerge as promising *standard sirens*. In particular, compact binary coalescences (CBCs) are amongst the most favourable systems that are expected to produce well-detectable GW signals. However, to obtain a measurement of the luminosity distance $d_L(z)$ from these signals, information on the redshift of the source is needed. The main problem of this approach is that with these signals it is not possible to measure directly the redshift z : high masses at high z produce the same signal as low masses at low z . Therefore, there is a degeneracy between z and the source binary system masses. Resolving this degeneracy is essential for obtaining constraints on various cosmological parameters (such as H_0) and on the characteristics of Binary Black Hole (BBH) populations (such as their mass distribution).

Method To break the degeneracy between the GW parameters, three different approaches have been developed to obtain cosmological constraints from GW observations (Moresco et al., 2022). In these approaches, GWs are identified as *bright*, *dark*, and *spectral* sirens. We refer to them as bright sirens when the GW signal has an associated electromagnetic counterpart. To date, there is only one example, GW170817, which is associated with the detection of a kilonova. This event enabled precise measurement of the source redshift through the observation of the host galaxy spectrum, providing the first GW measurement of H_0 and other constraints on modified gravity theories (Abbott et al., 2017a,b, 2019). Unfortunately, the majority of the GW events detected so far are merging BBH, which are not expected to produce detectable EM signals. For this reason, these events are called dark sirens. In this case, the redshift information can be obtained by cross-correlating the GW signals with galaxy catalogues. In practice, all galaxies within the GW localisation volume are used as potential hosts to obtain the redshift information through a statistical analysis (Del Pozzo, 2012). Finally, another approach relies on GW data alone. Here we refer to them as spectral sirens and the knowledge of the redshift is obtained including prior knowledge of the astrophysical properties of the GW emitters (mass distribution and rate evolution).

In this Thesis, I studied two mock catalogues, one reproducing GW detections with LIGO-Virgo-KAGRA O4 detector network and one with O5 detector network, to explore the constraints that can be obtained from GWs used as dark sirens. In particular, I have studied the constraints on BBH mass function parameters, the possibility or not to find a single preferred model, and the dependence of the constraints on H_0 given different analysis parameters. For this reason, the work has been divided into four steps.

Firstly, I have studied three publicly available samplers: `emcee` (Foreman-Mackey et al., 2013), `zeus` (Karamanis et al., 2021) and `dynesty` (Speagle, 2020), characterising their performances and how they scale with input parameters. Secondly, I used the publicly available Python software `CHIMERA` (Borghini et al., 2024) to constrain and study the source population, at fixed cosmology, by developing the code further to include a not implemented mass function. To perform model comparison, I used Bayesian hierarchical analyses implementing a nested sampling analysis by combining `CHIMERA` with `dynesty`. Then, I compared the results implementing model selection methods, namely Bayesian evidence, Deviance Information Criterion, and Posterior Predictive Checks. Lastly, I explored how selecting different values of galaxies' redshift uncertainty and KDE bandwidth, as well as different subsamples of the O5 dataset, influence the final mass posterior distribution of H_0 .

Scientific Objective The main objectives of this dissertation are the following:

1. Characterisation of three publicly available parameter space samplers studying their differences when constraining parameters of a test model and how they scale when changing some of their parameters.
2. Development and validation of a new module in `CHIMERA`, to include a BBH mass function model that extends the models already included.
3. Development of a new module to perform a nested sampling analysis combining the `dynesty` nested sampler with `CHIMERA`.
4. Estimation and comparison of the mass function parameters from two mock catalogues simulating the future O4 and O5 GW datasets.

5. Comparison of the different mass functions with three robust model comparison criteria, namely Bayesian evidence, Deviance Information Criterion, and Posterior Predictive Checks; aiming to assess the possibility of finding strong evidence of a favoured model with future data accuracies.
6. Study of possible biases and dependences on different parameters and sample selection of the H_0 posterior distribution.

Outlook The structure of the Thesis is summarised here with a brief description of the contents of each Chapter:

- In Chapter 1 the general cosmological framework, in which GWs are embedded, is presented, along with a description of how they can be used as cosmological probes and their analysis state of the art. Specifically, after a first introduction on the concordance cosmological model, the theory behind GWs will be discussed. Then, the methods to use them as standard sirens are presented. In the last part of the Chapter a brief description of the two main GW events, GW150914 and GW170817, will be presented, followed by a recap of the present-day GW dataset and of the future observing runs and detectors.
- In Chapter 2 the statistical framework is described, together with a presentation of BBH mass distributions and methods to sample the parameter space. In particular, a brief recap of the hierarchical Bayesian statistic basis is presented, followed by a more complete picture of Bayesian inference: likelihood and selection effects. After that, four mass function models are described, namely: truncated power law, broken power law, power law plus one Gaussian peak, and power law with two Gaussian peaks. In the last part of the Chapter, the affine invariant and nested sampling methods are described with a description of three samplers.
- In Chapter 3 the data used for the analyses are presented. Specifically, in the first part it is described how the two datasets have been generated, then, it follows their characterisations.

- In Chapter 4 the results will be presented. In the first section, the mass function parameters constraints are reported and discussed along with the computational scaling of these analyses and model comparison and selection. The second section presents the results obtained by studying how the posterior distributions of H_0 are affected by the variation of some parameters and the sample selection.
- In Chapter 5 all the results are summarised along with the future perspectives of this work.

Chapter 1

Introduction

This first chapter introduces the theoretical framework within which my dissertation project unfolds. In the first section, the basic equations describing the general cosmological context are presented. Then, different aspects of Gravitational Waves (GW) are discussed, starting from how they are embedded in the cosmological model up to how they can be used as probes in both cosmology and astrophysics.

1.1 Cosmological framework

Thanks to Einstein's General Relativity theory we have the possibility to model and test our Universe. Over the last century many different models have been proposed, all based on Einstein's theory, but nowadays the community has come up with a concordance model of cosmology (Dodelson and Schmidt, 2003) known as flat Λ CDM. In this model, the Universe is shaped as a flat Euclidean spacetime in accelerated expansion, whose dynamic is determined by the different energy-matter components that permeate it, such as the baryonic matter and radiation. Nowadays, the Universe is dominated by the cosmological constant Λ , associated with Dark Energy, and Cold Dark Matter, which is non-baryonic matter. Even though this model is supported by different probes and observational evidence, such as the cosmic microwave background (CMB) and large-scale structures, many of its ingredients are not known yet. For example, the two dominating components of

the local Universe go beyond the Standard Model of particle physics. In addition, the inflation mechanism, introduced to explain the generation of the initial perturbations that explain the formation of large structures and CMB anisotropies, is largely undetermined.

1.1.1 Basic Equations of Cosmology

The geometry of the Universe can be described by a metric in which, as we are dealing with an expanding Universe, it is introduced the scale factor $a(t)$. Using for convenience polar coordinates, we can express the metric as:

$$g_{\mu,\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & \frac{a^2(t)}{1-\kappa r^2} & 0 & 0 \\ 0 & 0 & a^2(t)r^2 & 0 \\ 0 & 0 & 0 & a^2(t)r^2 \sin^2\theta \end{pmatrix}. \quad (1.1)$$

This is the so-called Friedmann-Lemaitre-Robertson-Walker metric where the scale factor $a(t)$ curves the space part of the spacetime while κ its curvature. In particular, the Universe can be flat ($\kappa = 0$), open ($\kappa = -1$), or closed ($\kappa = 1$). The concordance model selects a flat Universe. The form of this metric arises from the cosmological principle, that implies the homogeneity and isotropy of the Universe (Weinberg, 1972). It is now possible to define the interval, which is a metric invariant, directly from the FLRW metric:

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

Because of the time dependence of the scale factor we can not measure distances in the classical way so we need to define some quantities that represent them (Coles and Lucchin, 2003)(Fig.1.1). The first one is directly related to $a(t)$ and is the proper distance:

$$d_{pr}(t) = a(t)r \quad (1.3)$$

Deriving the previous equation with respect to t , we obtain the Hubble-Lemaitre law:

$$V_r(t) = \frac{\dot{a}(t)}{a(t)} d_{pr}(t), \quad (1.4)$$

from which we recognize the Hubble parameter $H(t)$ and the Hubble constant $H_0(t)$, that is the present-day value of $H(t)$:

$$H(t) = \frac{\dot{a}(t)}{a(t)}, \quad H_0 = \frac{\dot{a}(t_0)}{a(t_0)}. \quad (1.5)$$

The scale factor can be linked to the redshift z with the expression:

$$a(t) = \frac{1}{1+z} \quad (1.6)$$

where for simplicity $a(t_0) = 1$. Another distance used in cosmology is the comoving distance, defined in the comoving frame of reference, which is not sensitive to the expansion of the Universe:

$$d_c = c \int_t^{t_0} \frac{dt'}{a(t')} = c \int_0^z \frac{dz'}{H(z')} = a(t_0)r \quad (1.7)$$

where we have used eq. 1.6 to change the integration variable from dt' to dz' . Eqs. 1.3-1.7 can not be directly used in Cosmology as they depend on time through the expansion factor $a(t)$: a photon leaving a galaxy at $z = 3$ starts travelling with a different value of a with respect to the present time. To solve this problem, some known quantities can be adapted. In astronomy, a classic way to determine the distance of an object is by measuring the angular size $\Delta\theta$ of the object itself knowing its size L . In this way, we obtain the angular diameter distance $d_A = \frac{L}{\Delta\theta}$. In the comoving frame, the size of the object becomes $\frac{L}{a}$ and the distance is given by eq. 1.7 so we have that $\Delta\theta = \frac{L}{a} \frac{1}{d_c}$. Comparing the two expressions, we come up with the cosmological angular diameter distance:

$$d_A = \frac{1}{1+z} d_c \quad (1.8)$$

Another distance definition is the luminosity distance that derives from the expression of the flux $F = \frac{L}{4\pi d_L^2}$. As we are in an expanding Universe the energy of the emitted photons will be smaller today with respect to the emission time by a factor of a , then the energy per unit time passing through a comoving shell, thus the flux, will be smaller by a factor of a^2 . Therefore, comparing the two relations we have the luminosity distance expression:

$$d_L = (1 + z)d_c \quad (1.9)$$

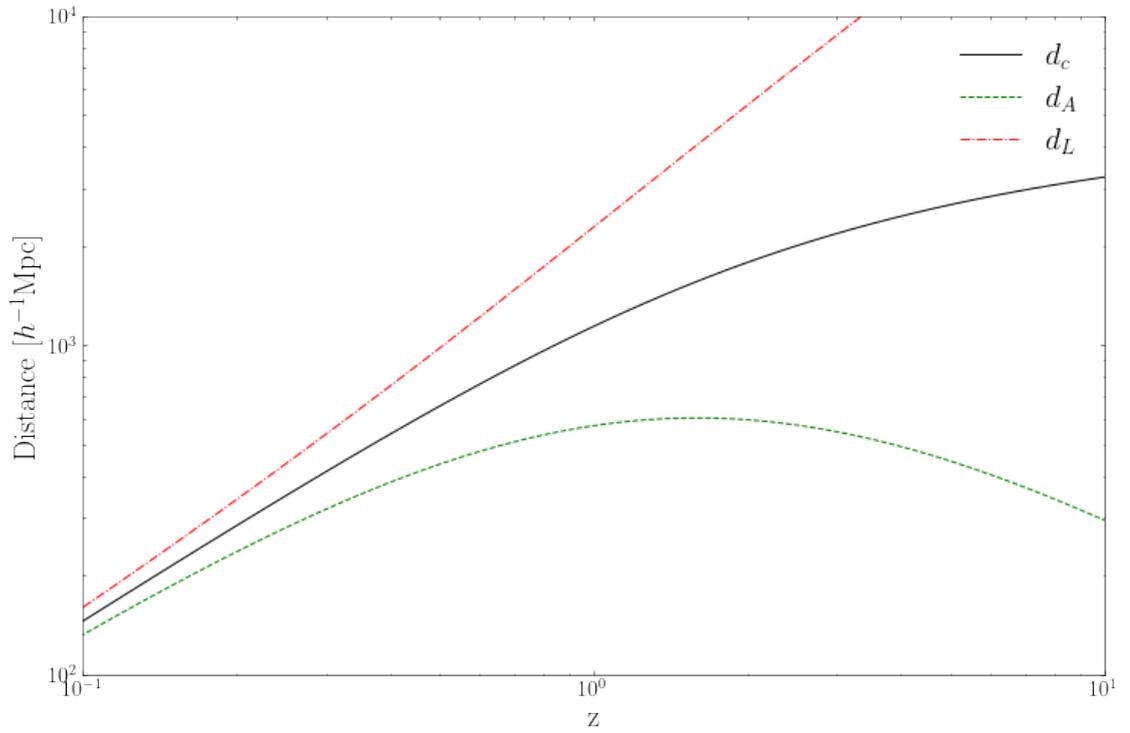


Figure 1.1: The three different kinds of distances as a function of redshift z in the concordance cosmological model: the comoving distance d_c (black line), the angular diameter distance d_A (green line), and the luminosity distance d_L (red line).

The dynamic of the Universe is determined by the Einstein field equations:

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \frac{8\pi G}{c^4}T_{\mu\nu} \quad (1.10)$$

where $R_{\mu\nu}$ is the Ricci Tensor, R is the Ricci scalar and $T_{\mu\nu}$ is the energy-momentum tensor. First of all, I want to focus on $T_{\mu\nu}$ which describes the content of the Universe. In the concordance model, because of the cosmological principle, the Universe is modeled at first order as a perfect fluid (Weinberg, 1972), thus the form of the energy-momentum tensor is forced to be the following one:

$$T^{\mu\nu} = \begin{pmatrix} -\rho c^2 & 0 & 0 & 0 \\ 0 & P & 0 & 0 \\ 0 & 0 & P & 0 \\ 0 & 0 & 0 & P \end{pmatrix}, \quad (1.11)$$

where ρ is the energy density while P is the pressure of the fluid. In the Λ CDM model, both the total pressure and the total energy density include a contribution from the cosmological constant Λ , namely $P_\Lambda = -\rho_\Lambda = -\frac{\Lambda}{8\pi G}$. A conservation law can be written applying the covariant derivative on $T_{\mu\nu}$:

$$\frac{\partial \rho}{\partial t} - \rho \left[-3 \frac{\dot{a}}{a} (1 + w_i) \right] = 0, \quad (1.12)$$

where $w = \frac{P}{\rho}$ takes different values for radiation ($= \frac{1}{3}$), matter ($= 0$), and the cosmological constant Λ ($= -1$). From this expression, the following relation between density and redshift can be obtained:

$$\rho_i(z) = \rho_{i,0} a^{-3(1+w_i)} = \rho_{i,0} (1+z)^{3(1+w_i)}. \quad (1.13)$$

Considering now the left-hand part of eq. 1.10, it is possible to derive the other two main equations of Cosmology, the Friedmann equations:

$$\left(\frac{\dot{a}(t)}{a(t)} \right)^2 = \frac{8\pi G}{3} \rho \quad (1.14)$$

$$\frac{\ddot{a}(t)}{a(t)} = -\frac{4\pi G}{3} (\rho + 3P) \quad (1.15)$$

Eq.1.14 is obtained considering only the time-time components of eq.1.10, while eq.1.15 comes from the space-space components. From eq.1.14 we can define the critical density $\rho_{cr,0} = 3 \frac{H_0^2}{8\pi G}$, that is the quantity that nullifies the equation, from

which in turn we can define the density parameters Ω_r , Ω_m , Ω_Λ as:

$$\Omega_i = \frac{\rho_{cr,0}}{\rho_{i,0}}. \quad (1.16)$$

In the flat Λ CDM model, the sum of the three density parameters is equal to 1 and represents the total energy density of the Universe. By combining the two Friedmann equations with those for H_0 and $\rho_{cr,0}$, we obtain an expression for $H(z)$ as a function of Ω_i :

$$H(z) = H_0[\Omega_r(1+z)^4 + \Omega_m(1+z)^3 + \Omega_\Lambda]^{1/2} \quad (1.17)$$

1.2 The theory behind gravitational waves

From the theoretical point of view, GWs are the solution of the Einstein field equation (eq.1.10), in particular they are a solution of its linearized version. Gravitational waves can be associated to ripples in the space-time that are no less than a very small field perturbing the metric. In the linearized theory, the field equation is expanded around the Minkowski flat space-time metric which is perturbed by this small field:

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \quad |h_{\mu\nu}| \ll 1 \quad (1.18)$$

With this expression of the metric at linear order, the new field equations are:

$$R_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}R = 8\pi GT_{\mu\nu} \quad (1.19)$$

Writing down the full expressions for the linearized Ricci's scalar and tensor and choosing the Lorentz gauge, it is possible to retrieve a simple wave equation (Maggiore, 2007):

$$\begin{cases} \square h_{\mu\nu} = -16\pi GT_{\mu\nu} \\ \partial_\nu h^{\mu\nu} = 0 \end{cases} \quad (1.20)$$

where $\square = \partial^\mu \partial_\mu$ is the D'Alembertian operator. To solve this system of equations the compact-source approximation is assumed, in other terms the source is near the

origin of the coordinate system and the point at which the field is evaluated is at a distance $r = \|\bar{x}\|$ much larger than the extension of the source. Considering also the fact that we are dealing with an isolated system, the generic final expression for a GW is:

$$h^{ij}(ct, \bar{x}) = -\frac{2G}{c^6 r} \left[\frac{d^2 I^{ij}(ct')}{dt^2} \right]_{ct'=ct-r} \quad (1.21)$$

and where i, j indicate the spatial components and $I^{ij}(ct)$ is the quadrupole-momentum tensor of the source. It is important to highlight the fact that h^{ij} is a very small quantity due to the constant $\frac{1}{c^6}$. In vacuum ($\square h_{\mu\nu} = 0$), eq. 1.21 reduces to the simpler equation:

$$h^{\mu\nu}(ct, \bar{x}) = \int A^{\mu\nu}(\bar{k}) e^{ik_\rho x^\rho} d^3\bar{k} \quad (1.22)$$

which is the superposition of plane-wave solutions where $A^{\mu\nu}$ is a 4x4 matrix which defines the amplitude of the GW. The physical solutions corresponding to propagating GWs in an empty space are obtained by taking the real part of $h^{\mu\nu}$. Note that $k = (\frac{\omega}{c}, k_1, k_2, k_3) = (\frac{\omega}{c}, \bar{k})$ is the 4-wavevector, and from the wave equation it results that $|k|$ is equal to 0, this leads to the fact that both phase and group velocities of GWs are equal to c .

1.2.1 Gravitational waves propagation

In order to study the propagation of GWs the assumption of being in an empty space is held, thus a possible solution of eq. 1.20 in vacuum will be $h^{\mu\nu} = A^{\mu\nu} e^{ik_\rho x^\rho}$. In addition to the Lorentz gauge, there is still the freedom to use a further gauge transformation (Hobson et al., 2006) of the type $h'_{\mu\nu} = h_{\mu\nu} - \partial_\mu \xi_\nu - \partial_\nu \xi_\mu$ where ξ^μ satisfies the Lorentz gauge if $\square \xi^\mu = 0$. Wrapping all together this new gauge condition with the Lorentz one and the symmetry of the amplitude matrix, the independent components of a GW propagating in vacuum reduce to just two and they will be the two possible GWs polarizations. This new adopted gauge is called transverse-traceless gauge (TT gauge); indeed if we consider a GW propagating along the x^3 direction only and applying all the above considerations to the plane-

wave solution we obtain that:

$$A_{TT}^{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & h_+ & h_x & 0 \\ 0 & h_x & -h_+ & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad (1.23)$$

$$A_{TT}^{\mu\nu} = h_+ e_+^{\mu\nu} + h_x e_x^{\mu\nu}, \quad e_+^{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad e_x^{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad (1.24)$$

where it is possible to distinguish the plus and the cross polarizations. Notice that, as the name itself of the gauge explains, the trace is null as well as the components not transverse to the propagation's direction. Obviously the TT gauge can be generalized to a generic direction.

1.2.2 Gravitational waves interaction with test masses

Now we have an expression for $h_{\mu\nu}$ thus it is possible to study the effect of the interaction of a GW with test masses. If we take only one free particle at rest it could resemble that the GW does not produce any effect, but this is only because we are choosing the TT gauge. So to study the effect it is essential to consider a set of test particles, in particular to simplify things we pretend them to be all in the $x_1 - x_2$ plane. The spatial distance between them can be expressed through the 4-vector $\xi^i = (\xi^1, \xi^2, 0)$ and after introducing the new quantity $\zeta^i = \xi^i + \frac{1}{2} h_k^i \xi^k$ the physical separation becomes $l^2 = \delta_{ij} \zeta^i \zeta^j$. It is useful to check how the two polarizations behave separately. Let us consider first the plus polarization, with simple computations it can be obtained that:

$$\bar{\zeta} = (\xi^1, \xi^2, 0) + \frac{1}{2} h_+ \cos[k(x^0 - x^3)] (\xi^1, -\xi^2, 0) \quad (1.25)$$

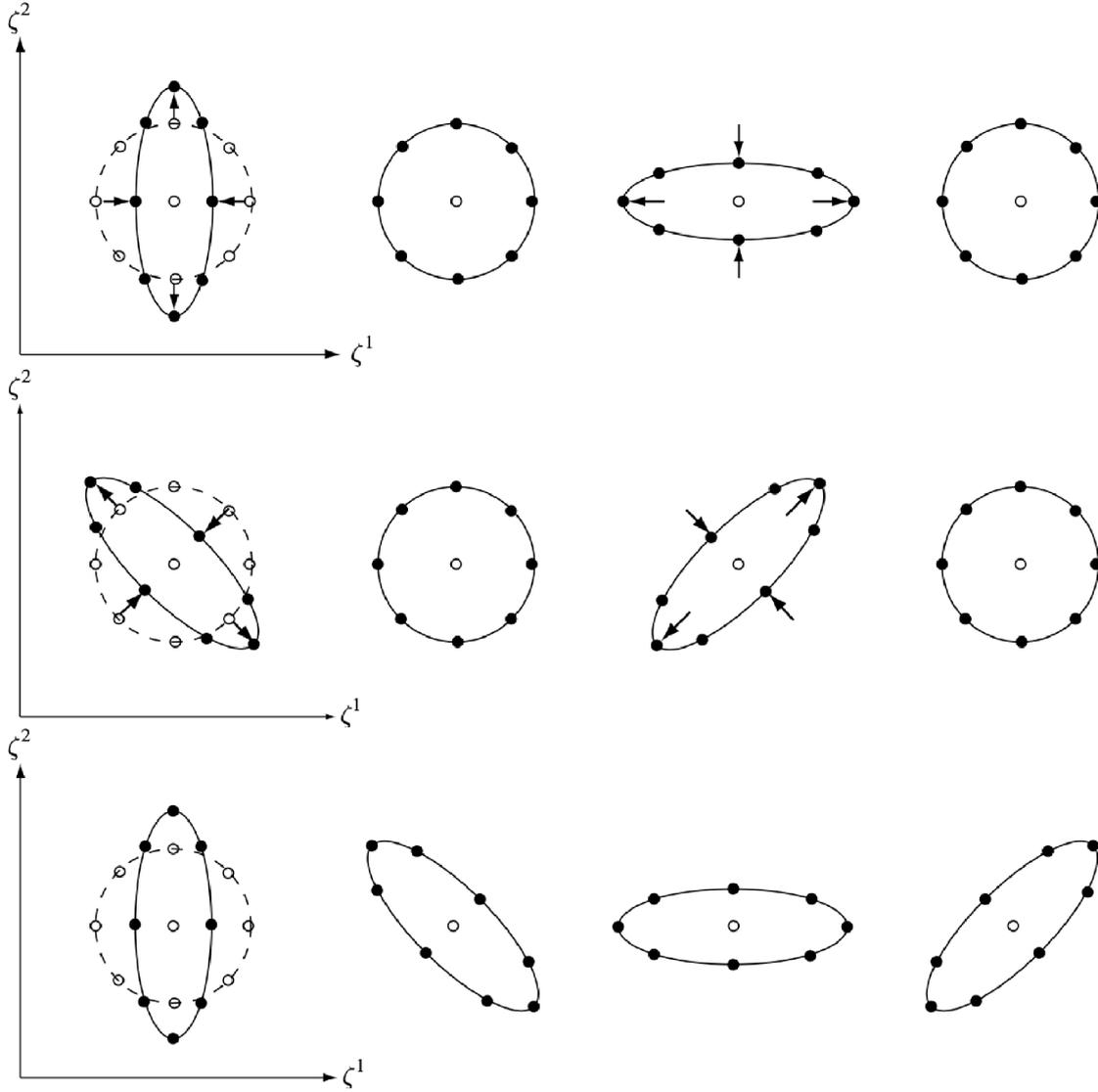


Figure 1.2: The three panels show the interaction of GWs, with three different polarization states, with test masses. In all three panels, the initial configuration of the particles is shown by the open dots and from left to right $k(x^0 - x^3)$ is equal to $2n\pi$, $(2n + \frac{1}{2})\pi$, $(2n + 1)\pi$, $(2n + \frac{3}{2})\pi$. a) In the top panel the GW has $A^{\mu\nu} = ae_+^{\mu\nu}$, i.e. the plus polarization. b) In the central panel the GW has $A^{\mu\nu} = be_x^{\mu\nu}$, i.e. the cross polarization. c) In the bottom panel the GW has $A^{\mu\nu} = a(e_+^{\mu\nu} + ie_x^{\mu\nu})$, i.e. the right-handed circular polarization. Illustrations taken from Hobson et al. (2006)

where the cosine part represents the time variation (top panel Fig. 1.2). On the other hand, the cross polarization produces (central panel Fig. 1.2):

$$\bar{\zeta} = (\xi^1, \xi^2, 0) + \frac{1}{2}h_x \cos[k(x^0 - x^3)](\xi^2, \xi^1, 0) \quad (1.26)$$

The general solution is a combination of the 2 polarizations, for example if h_x is chosen such that $h_x = \pm ih_+$, we obtain the two circular polarizations shown in the bottom panel of Fig. 1.2.

1.2.3 Gravitational waves generation

The generation of a GW is obtained directly from eq.1.21, so what is needed is an explicit expression for h^{ij} . To do this we can simplify things assuming to be in the compact source approximation and in a slow-moving particle regime. This last assumption allows us to express the quadrupole-momentum tensor as the quadrupole-momentum of the matter density distribution:

$$I^{ij} = c^2 \int \rho(ct, \bar{x}) x^i x^j d^3\bar{x} \quad (1.27)$$

Considering a binary system, in which two objects rotate in a circular orbit, it can be shown that the two polarization signals measured by an observer of the two-body system behaves as:

$$\begin{aligned} h_+ &= \frac{4}{r} \left(\frac{GM_c}{c^2} \right)^{\frac{5}{3}} \left(\frac{\pi f_{GW}}{c} \right)^{\frac{2}{3}} \left(\frac{1 + \cos^2(i)}{2} \right) \cos(2\pi f_{GW}t) \\ h_x &= \frac{4}{r} \left(\frac{GM_c}{c^2} \right)^{\frac{5}{3}} \left(\frac{\pi f_{GW}}{c} \right)^{\frac{2}{3}} \cos(i) \sin(2\pi f_{GW}t) \end{aligned} \quad (1.28)$$

where i is the angles formed between the direction of the observer and the normal to the orbital plane; f_{GW} is the frequency of the GW and it corresponds to two times the frequency of the binary system; M_c is called chirp mass and corresponds to $\mu^{\frac{3}{5}} m_{tot}^{\frac{2}{5}}$ with μ and m_{tot} the reduced and total mass of the rotating system. To be more accurate an inspiraling binary system should be considered, indeed the system loses energy because of the GW emission. In light of this, the quasi-circular orbit approximation can be introduced too, therefore both the separation and the angular frequency of the rotating system are time-dependent, in particular the last one increases as the first one gets smaller and smaller. It is necessary now to

introduce the quantity:

$$\Phi(t) = 2\pi \int_{t_0}^t f_{GW}(t) dt \quad (1.29)$$

that enters in the trigonometric functions of eq:1.28 that then become:

$$\begin{aligned} h_+ &= \frac{4}{r} \left(\frac{GM_c}{c^2} \right)^{\frac{5}{3}} \left(\frac{\pi f_{GW}(t)}{c} \right)^{\frac{2}{3}} \left(\frac{1 + \cos^2(i)}{2} \right) \cos(\Phi(t)) \\ h_x &= \frac{4}{r} \left(\frac{GM_c}{c^2} \right)^{\frac{5}{3}} \left(\frac{\pi f_{GW}(t)}{c} \right)^{\frac{2}{3}} \cos(i) \sin(\Phi(t)) \end{aligned} \quad (1.30)$$

It is possible to observe the typical shape of a GW generated by an inspiral binary if we move from the standard time coordinate t to the so-called time to coalescence $\tau = t_{coal} - t$; indeed rewriting eq.1.30 with the dependence to τ it can be seen that both frequency and amplitude increase while approaching t_{coal} Fig.1.3. As described at the beginning of this chapter we want to embed the propagation

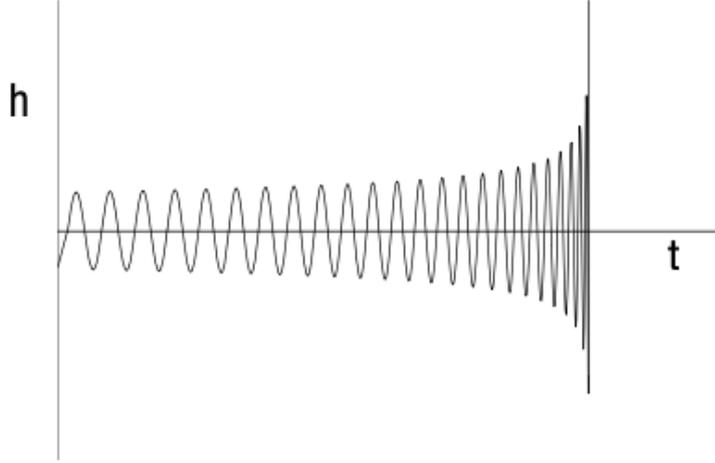


Figure 1.3: The time evolution of the GW amplitude in the inspiral phase of a binary system. Illustration taken from Maggiore (2007).

of a GW, produced by a compact binary system, in the Λ CDM model. Eq.1.30 are expressed assuming that the binary system is close to the observer, but now we want to see what happens if we move the GW source at cosmological distance

and consider the expansion of the Universe. First, the physical distance r has to be modified in $a(t_0)r$. Then, due to the fact that at cosmological distances the observed frequency is not the same of the emitted one ($f_{GW}^{obs} = \frac{f_{GW}^{em}}{1+z}$) a dependence on the redshift has to be introduced. It is worth thinking that there is no reason that h_+ and h_x remain decoupled even in an expanding Universe, it can be demonstrated that they actually don't mix. Grouping all the new terms and remembering eq. 1.9, we obtain the expressions for GWs propagating in a Λ CDM Universe:

$$\begin{aligned} h_+ &= \frac{4}{d_L} \left(\frac{G\mathcal{M}_c}{c^2} \right)^{\frac{5}{3}} \left(\frac{\pi f_{GW}^{obs}(t)}{c} \right)^{\frac{2}{3}} \left(\frac{1 + \cos^2(i)}{2} \right) \cos(\Phi(t)) \\ h_x &= \frac{4}{d_L} \left(\frac{G\mathcal{M}_c}{c^2} \right)^{\frac{5}{3}} \left(\frac{\pi f_{GW}^{obs}(t)}{c} \right)^{\frac{2}{3}} \cos(i) \sin(\Phi(t)) \end{aligned} \quad (1.31)$$

where $\mathcal{M}_c = M_c(1+z)$. The typical amplitude of these signals is roughly $\mathcal{O}(10^{-21})$.

1.3 Gravitational waves as standard sirens

The measurement of the GW signals is opening new doors both in the field of astrophysics and cosmology. In the works of Schutz (1986); Holz and Hughes (2005); Dalal et al. (2006) it is enlightened how GW signals coming from merging black holes and neutron stars can be used as cosmological probes, in particular because they can be used in the same way as standard candles. Indeed it can be easily seen from eq. 1.31 that by measuring both the GW amplitude and frequency, it is possible to obtain the distance d_L of the source without any additional calibration. The main problem of this approach is that the signal is degenerate with the redshift z : high masses at high redshifts produce the same signal as low masses at low redshifts. Therefore, it is impossible to obtain both redshift and luminosity distance from GW data alone. Nevertheless, if the z information is somehow available cosmological parameters can be constrained thanks to the distance-redshift relation in eq. 1.9. Due to the fact that the loudness of these kinds of sources comes directly from assuming the validity of general relativity, they provide a model-independent estimate of the luminosity distance, hence they have been defined as

standard sirens. Moreover GW signal depends on d_L as $\frac{1}{d_L}$ instead of $\frac{1}{d_L^2}$, like the majority of electromagnetic signals, thus they can be used to probe higher redshifts. However, recovering the redshift of the source producing the detected GW is not a simple task, therefore three different approaches (Moresco et al., 2022) have been developed to maximise the usage of the continuously increasing GW catalogues, which nowadays extend up to $d_L \gtrsim 5$ Gpc and will soon go beyond 10 Gpc, allowing to probe the expansion history out to $z > 1$.

1.3.1 Bright sirens

The straightforward case, but also the less probable to happen, is the case when multimessenger observations provide an electromagnetic counterpart to the GW signal giving the possibility to precisely assign a host galaxy to the event. In this eventuality, the GW event is said to be a bright siren. In this case, the redshift of the source is obtained from the host galaxy spectrum and combined with the estimate of the luminosity distance obtained from the GW signal. This is the optimal case but nowadays even the rarest one, as just $\frac{1}{90}$ of events can provide this opportunity. Indeed, in order to produce an EM counterpart the emitting event must involve a neutron star whose merging with another neutron star or a black hole produces a gamma-ray burst (GRB) and/or a kilonova. The only example of a bright siren is the GW170817 event, which was associated with the detection of the GRB 170817A. This event has an outstanding importance as it opened the doors to multimessenger astronomy providing the first standard siren measurement of the Hubble constant and other constraints on modified gravity theories (Abbott et al., 2017a,b, 2019).

1.3.2 Dark sirens

Unfortunately, the majority of the detected GW events were emitted by the merging of Binary Black Hole (BBH), thus any kind of EM signal could be produced. In this case, the events are called dark sirens.

Without an EM counterpart, one may think of identifying a unique host galaxy localising the GW event in the sky and this would be essentially equal to a bright

siren. Unfortunately, the detected events have a too large localisation area to follow this approach but a similar procedure can be used taking advantage of galaxy catalogues. Galaxies within the GW localisation area are considered possible hosts of the event and then, thanks to Bayesian statistics, as described in Del Pozzo (2012), used to obtain the correct values of cosmological parameters. Obviously, the smaller the localisation area, the more informative will be the inferred values. Spectroscopic galaxy catalogues are not very often available at the typical distances of GW events, but photometric catalogues can still be useful when they overlap with the GW skymap.

A second path that can be followed when there are no complete galaxy catalogues available, but the detected event is well localised, is to match the spatial clustering of GW sources as a function of the luminosity distance with the clustering of galaxies as a function of the redshift. The matching between these two spatial clustering measurements is cosmological dependent and can be used to constrain cosmological parameters. Finally, another extension of the statistical dark siren method is to use prior knowledge on the merger redshift distribution, derived from measurements of the star formation rate and time delay distribution of binary mergers, that has to be compared against the distribution inferred from GW data.

1.3.3 Spectral sirens

The last proposed procedure relies only on GW data, without the need for any other dataset that provides information on redshift. The knowledge on z is extracted directly from some relevant features in the source population, in particular the BBH one as they are the most detected but also binary neutron stars are a promising and useful population. In this case, GW events are referred to as spectral sirens (Ezquiaga and Holz, 2022). At present time, the strongest feature of this approach is the steep drop-off in the black hole mass distribution, observed at $\sim 40 - 65M_{\odot}$, interpreted as the imprint of the pair-instability mechanism, which may be accompanied by a pile-up of black hole immediately below the gap at $\gtrsim 35M_{\odot}$. Due to the expected independence on z of the pair-instability physics, the location of such feature can be inferred together with the $z - d_L$ relation by observing a set of GWs that provide the redshifted mass distribution as a func-

tion of luminosity distance. This is not the only feature studied and used, other possible properties can be for example the spin distribution or the rate evolution.

Nowadays this method is not able to produce competitive constraints on cosmological parameters. However, population models are improving with the increasing number of data, leading to a better calibration of these features and therefore to be considered robust cosmological probes.

1.3.4 Cosmological interest

As mentioned at the beginning of this section, there is a strong relation between the study of GWs and the estimation of cosmological parameters thanks to the independence of the luminosity distance estimate from any cosmological model. The parameter that is currently under the lens is H_0 because of the interest in the H_0 tension between the value $H_0 = (67.27 \pm 0.60)$ km/s/Mpc obtained studying the CMB (Aghanim et al., 2020), so at high redshift, and $H_0 = (73.04 \pm 1.04)$ km/s/Mpc based on the Supernovae calibrated by Cepheids (Riess et al., 2022), therefore at low z and because the present day measurements are mainly in the low redshift regime which allows to simplify eq.1.9 in:

$$d_L = \frac{cz}{H_0} + \mathcal{O}(z^2) \quad (1.32)$$

Until now all the results obtained with the three different methods including the one extracted from the bright siren GW170817 ($H_0 = 70_{-8}^{+12}$ km/s/Mpc) can not solve the tension as they are too broad (Abdalla et al., 2022). It is obviously expected that with future detectors and measurements, the constraints will be more precise; for example, it has been predicted that around 50 additional observations of bright sirens are needed to discriminate between CMB and SN measurements of H_0 with a precision of 1–2%. With GW standard sirens it is also possible to probe the Universe at cosmological redshifts and as a consequence to be sensitive to other parameters besides the Hubble constant such as Ω_m and $w(z)$. Another important opportunity is to test general relativity and put new constraints on modified gravity theories.

1.3.5 Astrophysical interest

The detection of GW events is not a cosmological interest only, but it has many implications in the astrophysical framework too, in particular in the investigations on GW sources. From the last observing run it has been possible to study in more depth the binaries population and their features (Abbott et al., 2023b). The most studied binaries are the black hole ones (BBH) as they are the sources with the most detected events. It has been possible to draw two conclusions about their mass distribution. The first one is that the distribution itself is not a simple power law but there is evidence, obtained with different analysis, suggested by overdensities in the chirp mass distribution, of the presence of a peak at masses around $35M_{\odot}$ and one between 8 and $10M_{\odot}$. The second one is that the results obtained by the observed data provide inconclusive evidence of the population gap between 50 and $120M_{\odot}$ due to pair-instability SN predicted by stellar evolution models. The evolution of the merger rate density of BBH has been constrained to be at least an increasing power-law as a function of redshift. The black hole population is not the only one that is studied even neutron star-black hole (NSBH) and neutron star (BNS) binaries cover an important role. In addition, combining results from both populations has challenged the inferred neutron star mass distribution with a peak obtained from radio observations. Last but not least more detailed studies on the final stages of supernovae explosions have been carried on because of the scarcity of observations between the heaviest neutron star and the lightest black hole leading to two main different models that have to be validated by future data. One model predicts the gap between $2-5M_{\odot}$ due to a rapid timescale ($\sim 10\text{ms}$ instead of $\sim 100\text{ms}$) SN instability (Belczynski et al., 2012); the second model instead fills the gap predicting a delayed timescale ($\sim 200\text{ms}$) due to an accretion of the proto-neutron star before the explosion (Zevin et al., 2020).

1.4 State of the art of GW analysis

Achieving the current state of GWs detections was not a straightforward journey because both of the poor knowledge of the topic and the weakness of the signals. The path towards present-day detectors and analysis started in 1966 with the res-

onant detector of Joe Weber (Weber, 1960) which however never detected any signal. Then, the community in parallel moved to the solution of interferometric detectors; the first one was TAMA300 located in Japan which took data between 1999 and 2004, but with a too-low sensitivity. The same sensitivity problem was found in the German detector GEO600, which started to take data in 2003. It has been now updated and it is still working but just as a testing facility. At present time, the majority of resources are spent on interferometers; the three main running detectors are LIGO, Virgo, and KAGRA.

LIGO is located in the USA and it is composed of two identical interferometers placed 3002 km far from each other, one in Livingston, Louisiana, and one in Hanford, Washington. Virgo is the Italian interferometer placed near Pisa, while KAGRA is the newest one and it is placed under a mountain in Japan. The advantage of having many instruments around the world is that data can be combined giving a higher probability that detected events are really GW signals and not just noise, and moreover to better localise the event in the sky.

The first evidence of the existence of GW is an indirect one. In 1974, Hulse and Taylor detected the binary pulsar PSR 1913+16 for which they were able to compute the shrinking of its orbit and to compare it with the value predicted by general relativity (Hulse and Taylor, 1975). In order to have the first direct evidence we had to wait until 2015. In the following paragraphs, I describe the first direct detections and the future perspective of GW detection and analysis.

1.4.1 GW150914

The first-ever direct detection of a GW is GW150914 (Abbott et al., 2016) that has been detected on September 14, 2015 thanks to the two interferometers of the LIGO observatory located at Hanford and Livingston. The signal started at a detected frequency $f_{GW} \simeq 30 Hz$, followed by eight cycles of the inspiral phase characterised by a steady increase in frequency and amplitude. The GW signal reached its maximum amplitude at $f_{GW} \simeq 144 Hz$, at which the merging occurred, followed by the ringdown phase. After a complete statistical analysis, it was concluded that this event, which was only subsequently classified as GW150914, was produced by two BHs with initial masses $m_1 \sim 36 M_{\odot}$ and $m_2 \sim 29 M_{\odot}$, that

coalesced at a luminosity distance of $dL \sim 440 \text{ Mpc}$ to form a $62M_{\odot}$ black hole. Assuming the cosmological parameters from Planck (Aghanim et al., 2020), it is obtained that the source is located at $z \simeq 0.09$.

The properties obtained from the analysis of this event give us the opportunity to further investigate the characteristics of compact binary coalescences and in particular of BBH systems, to which GW150914 belongs. This is particularly important since no BBH system had ever been discovered before the observation of GW150914, as it is not expected to produce an electromagnetic counterpart. One of the most important pieces of information about the black hole-black hole binary population obtained from GW150914 was the first observation of two BHs with masses of $\sim 30M_{\odot}$ and the birth of a black hole with a mass of $\geq 60M_{\odot}$. The novelty of this result is related to the fact that the only BHs with stellar mass that had been observed until then mostly had masses in the range of $5 - 10M_{\odot}$, with some exceptions, but always with masses smaller than $20M_{\odot}$. Due to the need to observe electromagnetic emission, these observations were exclusively X-ray binaries. Therefore, it is clear that the discovery of GW150914 has significantly extended the mass range of stellar-mass BHs.

The observation of GW150914 provided evidence that BBH systems can form and merge at a detectable rate within the time scale of the universe. The observation of such an event was not obvious, because a black hole-black hole system might not form at all, due to some processes that could effectively suppress the two main formation mechanisms:

- The BBH arises from the development of an isolated binary system. In this case, the suppression mechanism is associated with the interaction between a first black hole and a companion star that undergoes massive expansion during its post-main sequence phases. In particular, due to the extended envelope of the star, the binary system enters a phase of joint envelope evolution in which the black hole moves through the envelope of its companion. This motion within the envelope generates a certain friction that leads to the merger of the two objects before the formation of the second black hole, thus blocking the possibility of the formation of a BBH system.
- The BBH is formed thanks to the dynamic interaction between pre-existing

isolated BHs in a dense cluster. In this case, the main suppression mechanism is due to the dynamic ejection of central objects as a result of three-body interactions.

1.4.2 GW170817

In section 1.3.1 I have already mentioned the event GW170817. It has been detected on August 17, 2017 by both the LIGO and Virgo interferometers. It has been associated with a neutron star-neutron star coalescence, the first ever neutron star-neutron star merger detected. As mentioned before, a short γ -ray burst was observed independently by Fermi-GBM and was later associated with the same merger event. The signal reached the detectors at a frequency of $f_{GW} \simeq 30Hz$ and reached its maximum amplitude at $f_{GW} \simeq 400Hz$.

After a complete statistical analysis, it was possible to deduce the astrophysical properties of the compact objects responsible for the emission. In particular, it was found that the masses of the two objects composing the binary system are $m_1 \sim 1.48M_{\odot}$ and $m_2 \sim 1.27M_{\odot}$, respectively, while the luminosity distance $dL \sim 40$ Mpc, corresponding to a redshift $z \simeq 0.008$.

The masses obtained, which are much smaller than those of GW150914, were the first evidence that this event was generated by two NSs, and not a neutron star-black hole binary system or a black hole-black hole merger. Furthermore, the discovery of the electromagnetic counterpart was a piece of additional evidence that GW170817 was a binary neutron star (BNS) merger. The fact that the event was detected by three interferometers enables the localisation of the event within a sky region of $\sim 28deg^2$, just for comparison, the localisation area of GW150914 is $\sim 600deg^2$. Soon after this observation, the Fermi-GBM telescope detected a short γ -ray burst (GRB170817A) coming from the same region identified for GW170817. The time interval between the GW detection and the GRB observation was (Maggiore, 2018):

$$\Delta t = 1.734 \pm 0.054s \quad (1.33)$$

which is in accordance with the theoretical time difference expected for such events. In fact, the most accredited theory believes that GRBs are produced by internal

processes, such as shocks or magnetic reconnection, in an ultra-relativistic and highly collimated jet, which just follows from the merger of the compact objects. Moreover, the Fermi-GBM observation, combined with other optical/IR observations, identified GRB 170917A at a location on the sky with right ascension $\alpha(J2000) = 13^h09^m48^s.085 \pm 0.018$ and declination $\delta(J2000) = -23^\circ22'53''.343 \pm 0.218$ from the centre of the galaxy NGC 4993. Thanks to the observation by MUSE/VLT, it was possible to measure the redshift of the galaxy: $z = 0.009783$, corresponding to a luminosity distance $dL = 40.4 \pm 3.4\text{Mpc}$ (Hjorth et al., 2017), which is consistent with the GW estimate, therefore confirming that GW170817 occurred in NGC 4993. Therefore, the detection of a short GRB spatially and temporally associated with GW170817 gave us the first proof that GRBs can be produced during a BNSs merger event. Furthermore, thanks to the independent measurement of the redshift it was possible to use the event as a bright siren and to infer cosmological parameters. In particular, the inferred Hubble constant is $H_0 = 70 \pm 12\text{km/s/Mpc}$.

These results that followed the detection of just one GW event gave birth to multi-messenger astrophysics and showed how powerful it can be.

1.4.3 GWTC-3

Every GW instrument doesn't work continuously but there are time windows in which observational runs are planned. This strategy has been chosen in order to allow improvements of the used technologies aimed to achieve always better sensitivities. With time windows, each collaboration (LIGO, Virgo, KAGRA) knows when to stop and restart the interferometers, preserving the simultaneous data taking. For each run of observations, defined with capital O and a progressive number (e.g. O3), a new catalogue of transient GW signals is produced updating the previous one.

At the time in which I am writing, interferometers have just started the fourth observing run O4, and the most comprehensive set of GW observations presented to date is the Gravitational Wave Transient Catalogue 3 (GWTC-3) (Abbott et al., 2021), which collects signals detected in the second part of the third run (O3b), i.e. in the period between 1 November, 2019 and 27 March, 2020. In the previous

catalogue, i.e. O2, as well as for the majority of the time of O3, only LIGO and Virgo were collecting data (Fig.1.4). The duty cycles for the three interferometers in this period were: 79% (115.7 days) for LIGO Hanford, 79% (115.5 days) for LIGO Livingston, and 76% (111.3 days) for Virgo. Furthermore, for 96.6% of the time (142.0 days) at least one interferometer was observing, while for 85.3% (125.5 days) at least two detectors were observing. While KAGRA joined only for the final part of the run by completing a two-week observation run combined with GEO 600.

The catalogue contains 90 possible GW candidates from compact binaries coalescence of which 35 were added from O3b. Events have been chosen with a probability to be an astrophysical source greater than 50%. The previous catalogues, i.e. GWTC-1 (O1+O2) in which there were both GW150914 and GW170817, GWTC-2.1(O3a), contained respectively 11 and 55 candidates. It is therefore possible to notice a strong increase of GW candidates, thanks to the improved sensitivity of the detectors network. Indeed, by virtue of the improvements to LIGO and Virgo not only did the sensitivity get better but also it was possible to probe the Universe up to BNS ranges, i.e. the average distance at which a fiducial $1.4M_{\odot} + 1.4M_{\odot}$ BNS could be detected with a signal-to-noise ratio (SNR) of 8, is ~ 140 Mpc with LIGO and up to ~ 50 Mpc with Virgo, while during the first observing runs the farthest we could go was ~ 100 and ~ 30 Mpc, respectively. All the 35 new events introduced in GWTC-3 are signals that have been evaluated to come from BBHs mergers with the exception of 2 of them, GW200105_162426 and GW200115_042309, which are consistent with originating from neutron star–black hole binaries (NSBHs) as they have an estimated secondary mass $m_2 \simeq 1.9M_{\odot}$ and $m_2 \simeq 1.5M_{\odot}$ (Abbott et al., 2021), respectively, that are within the mass range of known NSs, and primary masses well above this range. Another event worth mentioning is GW200210_09224, because it has an estimated secondary mass of $m_2 \simeq 2.83M_{\odot}$ which is right above the NSs mass range, and therefore it could be either the lightest black hole or the heaviest neutron star ever detected.

1.4.4 Future observing runs: O4 and O5

The network of detectors is now in the O4 phase. This observing run started on May 24, 2023 and it is planned to end on January 2025. Until March 2024 only LIGO Hanford and LIGO Livingston will collect data, and then Virgo will join the run too. In the meantime, KAGRA has been stopped at the beginning of summer 2023 and will restart in spring 2024 and then it is planned to stop again in order to allow its sensitivity to get higher than 10 Mpc (BNS range). In this run, LIGO is expected to reach a BNS range greater than 160 Mpc. Due to the reached high sensitivity, already in the testing phase there have been alerts of possible candidates.

After O4 another observing run will be launched, called O5, it is planned to start at the beginning of 2027 and to stop in 2030 (Fig.1.4). Collaborations have already started to plan strong enhancements to push further the reduction of noise to reach ~ 3300 Mpc of BNS range. With these improvements, it is expected that detectors will be crowded with real GW signals and, thanks to the cooperation between the 4 interferometers, to increase the rate of multi-messenger sources.

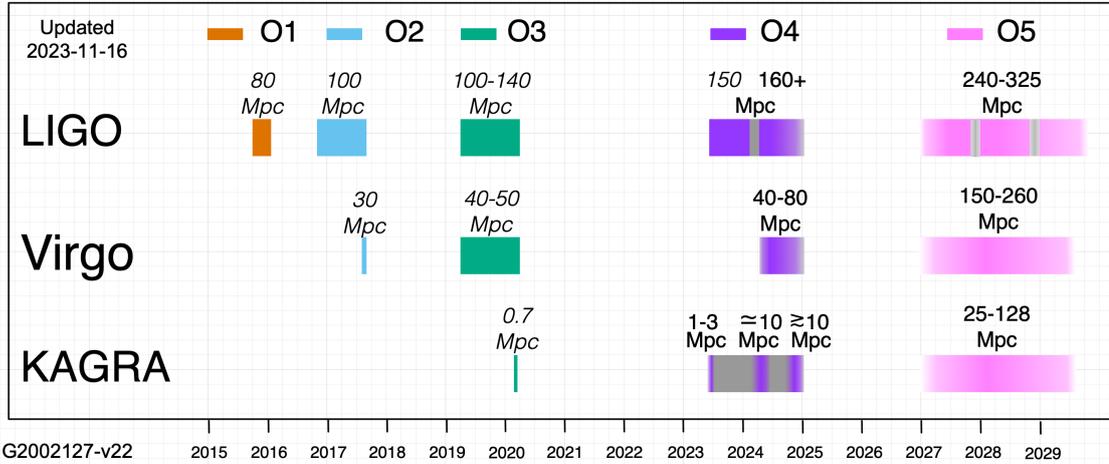


Figure 1.4: Past and future observing runs of the LVK collaboration. The coloured regions show the periods in which data have been and will be collected, while the grey regions indicate the periods in which the instruments will be under maintenance. The distances reported over every observing run are the BNS range for a single detector SNR threshold of 8. Illustration taken from <https://observing.docs.ligo.org/plan/>.

1.4.5 Future detectors

The existing detectors have intrinsic limitations, such as their arms length or even their location, therefore the gains on noise reduction are limited. To go beyond the current facilities, it is necessary to push the limit of all the technologies, significantly increase the arms length, and explore new solutions. To move in this direction we are approaching the era of the 3rd generation of GW observatories that are already being planned and that will constitute a new global 3G network for GW detection and observation. There are three planned and under discussion detectors: two ground-based interferometers, i.e. Einstein Telescope (ET) (Punturo et al., 2010) and Cosmic Explorer (CE) (Evans et al., 2021), and one that will operate in space, i.e. Laser Interferometer Space Antenna (LISA) (Amaro-Seoane et al., 2017). Actually, it exists also a Chinese counterpart of LISA which is called Taiji (Luo et al., 2020).

ET is going to be built in Europe, even if the location and the configuration are under debate. In particular, a deep study has been carried on in Branchesi et al. (2023) to define what are the scientific advantages and disadvantages of one configuration over the others. In general, it is planned to have at least two detectors with arms length of 15 km and each of them to have 2 interferometers, one for high and one for low frequencies.

CE the US 3G interferometer. It is based on the current LIGO concept but it will be 10 times longer and 10 times more sensitive.

There are fundamental facility and technology differences between CE and ET, but there is a strong overlap for everything related to astrophysics, data analysis, calibration, and computing. It is predicted that with these two observatories it will be possible to approach $z \sim 100$ and therefore to study BBH with masses of about $30M_{\odot}$ (Fig.1.5).

LISA will be a space-based GW observatory. It is an ESA and NASA project that will consist of three spacecraft separated by 2.5 million km in a triangular formation, following the Earth in its orbit around the Sun.

The synergy between these three detectors will open the possibility to probe the era of the formation of the first astronomical objects with unmatched statistics and SNR. In addition to this, the merger events would be monitored from the

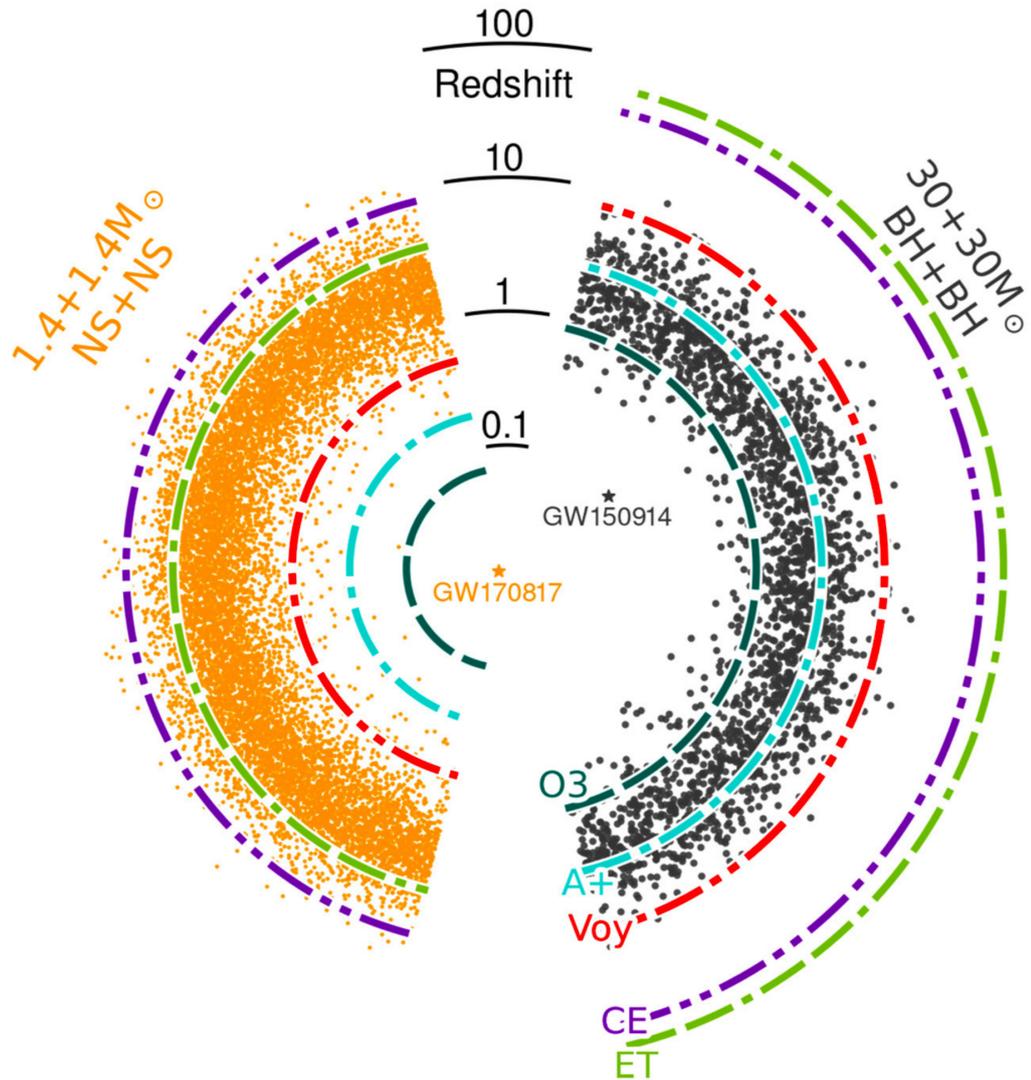


Figure 1.5: Redshift horizon at which a detection with SNR= 8 could be made with current and future ground-based interferometers for $1.4M_{\odot}$ neutron stars binaries (orange dots on the left) and $30M_{\odot}$ black holes binaries (black dots on the right). The violet and green lines represent the forecast horizons for Cosmic Explorer and Einstein Telescope, respectively. Illustration adapted from Evans et al. (2021).

early stages of the merging, opening the possibility to foresee and plan in advance follow-up observations with EM facilities.

Chapter 2

Methods

This chapter describes the data analysis process. In the first part, the statistical framework and its implementation in the pipeline are presented. Then, the astrophysical models that have been used in the inference. The full analysis has been carried out using CHIMERA (Combined Hierarchical Inference Model for Electromagnetic and gRavitational Wave Analysis), a novel Python code for the joint analysis of GW transient catalogues and galaxy catalogues (Borghi et al., 2024), allowing one to carry on cosmological studies combining GWs and catalogues of potential host galaxies. The workflow of CHIMERA will be briefly described in the last section of this Chapter.

2.1 Statistical framework

The first aim of this work is to study different BBH mass models and potential biases that may arise from their misinterpretation. This is obtained by adopting a hierarchical Bayesian inference approach, as this method allows one to simultaneously perform parameter estimation and model selection (Adams et al., 2012; Thrane and Talbot, 2019). Bayesian inference is now widely used in the gravitational wave field as there is a direct link between data and models. For example, it is used to reconstruct the sky localisation of a GW event, to determine the Hubble constant, and to study the formation mechanism of black hole binaries. The term hierarchical arises because the analysis has two levels: at the highest level,

there is the space of models being considered, while at the lower level, there are the parameters of the models themselves. The general problem is that we have a population of events, which are described by a set of event-level parameters $\bar{\theta}$ (i.e. those appearing in eq.1.31), from which we want to determine the set of hyper-parameters $\bar{\lambda}$ that describe the source population, so that the number density of sources follows the equation:

$$\frac{dN}{d\theta}(\bar{\lambda}) = N p_{pop}(\bar{\theta}|\bar{\lambda}) \quad (2.1)$$

where p_{pop} is the distribution of the properties of the individual event called population function and it is usually modeled analytically as we will see later (Mandel et al., 2019). In CHIMERA, the population-level parameters are divided into three different sets: BH mass function parameters $\bar{\lambda}_m$; BH rate parameters $\bar{\lambda}_r$; cosmological parameters $\bar{\lambda}_c$; while the event-level parameters are $\theta = \{d_L, \hat{\Omega}, m_1, m_2\}$, where d_L is the luminosity distance to the source, $\hat{\Omega}$ is the sky localisation, and are $m_{1,2}$ the binary masses. A definite and unique value of $\bar{\lambda}$ cannot be obtained; the best that can be done is to obtain a posterior probability which, in the usual Bayesian formalism, is given by:

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

where $\{\bar{d}_i\}$ are the measured data, $\pi(\bar{\lambda})$ is the prior probability that contains the previous knowledge on $\bar{\lambda}$, $p(\{\bar{d}_i\}|\bar{\lambda})$ is the likelihood function \mathcal{L} , i.e. the probability to obtain the data given the parameters. The term $p(\{\bar{d}_i\})$ is the so-called evidence and is later denoted as \mathcal{Z} . This quantity is no less than the likelihood marginalised over all the parameters and therefore it is just a normalisation factor that is not considered in parameter inference; however, when it comes to model comparison, it is a crucial quantity that needs to be measured and I will describe later how.

2.1.1 Likelihood

The core part of statistical analysis is the construction and evaluation of the likelihood. Assume to have a set of N_{ev} of independent GW events $d^{GW} = \{\bar{d}_i^{GW}\}$ from

which we can measure the luminosity distance, the total likelihood is proportional to the product of all single events' likelihood times the population function:

$$\mathcal{L}(d^{GW}|\bar{\lambda}) = p(d^{GW}|\bar{\lambda}) \propto \frac{1}{\xi(\bar{\lambda})^{N_{ev}}} \prod_{i=1}^{N_{ev}} \int p(d_i^{GW}|\theta_i, \bar{\lambda}_c) p_{pop}(\theta_i|\bar{\lambda}) d\theta_i \quad (2.3)$$

the term $\xi(\bar{\lambda})^{N_{ev}}$ is introduced with the purpose of taking into account the selection bias; the meaning of this term will be described in the next section. Notice that the single event probability depends on the cosmological hyperparameters, since the data contain information on the parameters in the detector frame θ^{det} , and thus to work in the source frame, parameter conversions are required. The luminosity distance d_L is given by eq.1.9 while $m_{1,2}^{det} = m_{1,2}(1+z)$.

Let us now see how the different components can be expressed. The single event likelihood $p(d_i^{GW}|\theta_i, \bar{\lambda}_c)$, following the Bayes theorem, can be expressed as:

$$p(d_i^{GW}|\theta_i, \bar{\lambda}_c) \propto \frac{p(\theta_i|d_i^{GW}, \bar{\lambda}_c)}{\pi(\theta_i)} \quad (2.4)$$

where $p(\theta_i|d_i^{GW}, \bar{\lambda}_c)$ is obtained from $p(\theta_i^{det}|d_i^{GW}, \bar{\lambda}_c)$ which is the distribution that provides the sets of event level parameters samples, while $\pi(\theta_i)$ is the prior probability.

The population function can also be divided into the product of two other distributions:

$$p_{pop} = p(m_1, m_2|\bar{\lambda}_m) p(z, \hat{\Omega}|\bar{\lambda}_z, \bar{\lambda}_c) \quad (2.5)$$

this expression can be used only under the assumption that $p(m_1, m_2|\bar{\lambda}_m)$ does not evolve over time. This last probability is the distribution of m_1 and m_2 given a well-shaped mass function of the sources that will be described in detail in section 2.2. The other term can be expressed as the product of two other distributions:

$$p(z, \hat{\Omega}|\bar{\lambda}_z, \bar{\lambda}_c) \propto p_{gal}(z, \hat{\Omega}|\bar{\lambda}_z) p_{rate}(z|\bar{\lambda}_z) \quad (2.6)$$

p_{gal} is the probability to find a galaxy at redshift z and RA, DEC coordinates $\hat{\Omega}$. This probability can be expressed as the weighted sum of two other distributions:

$$p_{gal}(z, \hat{\Omega}|\bar{\lambda}_c) = f_{\mathcal{R}}p_{cat}(z, \hat{\Omega}|\bar{\lambda}_c) + (1 - f_{\mathcal{R}})p_{miss}(z, \hat{\Omega}|\bar{\lambda}_c) \quad (2.7)$$

The first one, p_{cat} , is the probability distribution built from galaxies in the catalogue while p_{miss} is introduced to encode in the likelihood the completeness of the catalogue itself and includes information on both the number of missing galaxies as a function of redshift and sky position and how these missing galaxies are distributed (Gair et al., 2023; Borghi et al., 2024).

p_{rate} instead is the probability that at redshift z there is a galaxy hosting a GW event. This distribution is as follows:

$$p_{rate}(z|\bar{\lambda}_z) \propto \frac{\psi(z, \bar{\lambda}_z)}{(1+z)} \quad (2.8)$$

where the denominator is introduced for the conversion between the source and detector frame while the numerator represents the merger rate evolution of compact objects with redshift. In my work $\psi(z, \bar{\lambda}_z)$ is modelled following the parametrisation proposed in Madau and Dickinson (2014):

$$\psi(z, \bar{\lambda}_z) = \frac{(1+z)^\gamma}{1 + \left(\frac{1+z}{1+z_p}\right)^{\gamma+\kappa}} \quad (2.9)$$

Wrapping up all the pieces, i.e. eq.2.4-2.8, the final expression of the full likelihood is:

$$p(d^{GW}|\bar{\lambda}) \propto \frac{1}{\xi(\bar{\lambda})^{N_{ev}}} \prod_{i=1}^{N_{ev}} \int dz d\hat{\Omega} \mathcal{K}_{GW,i}(z, \hat{\Omega}|\bar{\lambda}_c, \bar{\lambda}_m) p_{gal}(z, \hat{\Omega}|\bar{\lambda}_c) \frac{\psi(z, \bar{\lambda}_z)}{(1+z)} \quad (2.10)$$

$$\mathcal{K}_{GW,i}(z, \hat{\Omega}|\bar{\lambda}_c, \bar{\lambda}_m) \equiv \int dm_1 dm_2 \frac{p(z, m_1, m_2, \hat{\Omega}|d_i^{GW}, \bar{\lambda}_c)}{\pi(d_L)\pi(m_1^{det})\pi(m_2^{det})} \frac{p(m_1, m_2|\bar{\lambda}_m)}{\frac{dd_L}{dz}(z; \bar{\lambda}_c)(1+z)^2} \quad (2.11)$$

where the $(1+z)^2$ and $\frac{dd_L}{dz}(z; \bar{\lambda}_c) = \frac{d_L}{1+z} + \frac{c(1+z)}{H(z, \bar{\lambda}_c)}$ terms are the Jacobian factors coming from the passage from detector to source frame. It is worth to say that in

CHIMERA the prior probabilities included in eq.2.4 are assumed to be flat for m_1 and m_2 , $\pi(m_{1,2}) = 1$, while for luminosity distance is imposed to be $\pi(d_L) \propto d_L^2$.

2.1.2 Selection effects

To have a complete view of the likelihood, it is key to characterise and understand the meaning of the normalization factor $\xi(\bar{\lambda})$. As I said at the beginning of the previous section, this term is fundamental to provide a proper posterior distribution, as in general some events are easier to observe than others due to intrinsic properties or instrument limitations. If this bias is not included in the analysis, it would eventually lead to a wrong evaluation of uncertainties and a wrong shaping of the posteriors. The selection effect can be accounted for by introducing a detection probability p_{det} , thus the selection function can be expressed as (Mandel et al., 2019):

$$\xi(\lambda) = \int d\theta p_{det}(\theta, \bar{\lambda}_c) p(m_1, m_2 | \bar{\lambda}_m) p_{gal}(z, \hat{\Omega} | \bar{\lambda}_c) \frac{\psi(z; \bar{\lambda}_z)}{1+z} \quad (2.12)$$

Considering a dataset of GW events the detectability is considered deterministic, therefore an event is considered detectable if the data exceeds a certain threshold r_{thr} . In the context of this work, the threshold is the signal-to-noise ratio (SNR) of the event, thus p_{det} is essentially the likelihood distribution of observed SNRs and can be expressed as:

$$p_{det}(\theta, \bar{\lambda}_c) = \int_{SNR_{(d^{GW})} > r_{thr}} p(d^{GW} | \theta, \bar{\lambda}_c) dd^{GW} \quad (2.13)$$

A GW event has an intrinsic SNR due to the properties of the system emitting the signal, for example the amplitude of the signal itself has a strong dependence on the compact objects' masses, but also the inclination and the sky localisation affect the SNR. Moreover, there are stochastic fluctuations in the interferometers that may change the intrinsic SNR which is then different from the observed one. We must also consider that the present-day interferometers are sensitive only to a certain range of GW signal frequencies and this is a bias that must be considered when inferring population distributions.

2.2 Black Hole Binaries mass function models

A crucial part for using GWs as standard sirens is the modeling of the p_{pop} probability (eq.2.5) and in particular of the piece that encodes information on the compact object masses distributions: $p(m_1, m_2 | \bar{\lambda}_m)$. With the increasing amount of data from the recent observing runs, it has been seen, as I have already mentioned in section 1.3.5, that the derived black hole mass function is not a simple power-law but it shows different features. Many models have been proposed to fit the data; in my work, I have analysed those mentioned in the article of the LIGO-Virgo Collaboration about the population properties of compact objects derived from GWTC-2 (Abbott et al., 2021), used in Abbott et al. (2023a) for a cosmological analysis with the GWTC-3 data. These models are the Truncated Power-Law (TPL), the Broken Power Law (BPL), the Power Law + 1 Gaussian Peak (PLP), and the Power Law + 2 Gaussian Peaks (PL2P). The TPL, PLP and PL2P models were already present in the **CHIMERA** code and validated with the public available code **MGCosmoPop** (Mancarella et al., 2022b; Mancarella and Genoud-Prachex, 2022), therefore, I had to expand the model library of **CHIMERA** implementing the BPL model. In general, all of the following models are built as the product of a primary and a secondary mass probability distributions.

2.2.1 Truncated Power Law

This model is the simplest one. The primary mass distribution is a power law with a sharp cut-off at the lower and upper ends. The full model is completed by the secondary mass distribution, which is modelled as a simple power law.

$$p(m_1 | \alpha, m_l, m_h) \propto \begin{cases} m_1^{-\alpha} & \text{if } m_l < m_1 < m_h \\ 0 & \text{otherwise} \end{cases} \quad (2.14)$$

$$p(m_2 | \beta, m_l, m_1) \propto \begin{cases} m_2^\beta & \text{if } m_l < m_2 < m_1 \\ 0 & \text{otherwise} \end{cases} \quad (2.15)$$

Thus the hyperparameters $\bar{\lambda}_m$ for the TPL are:

- m_l , the minimum mass of the power law component;

- m_h , the maximum mass of the power law component;
- α , the spectral index of the m_1 distribution;
- β , the spectral index of the m_2 distribution.

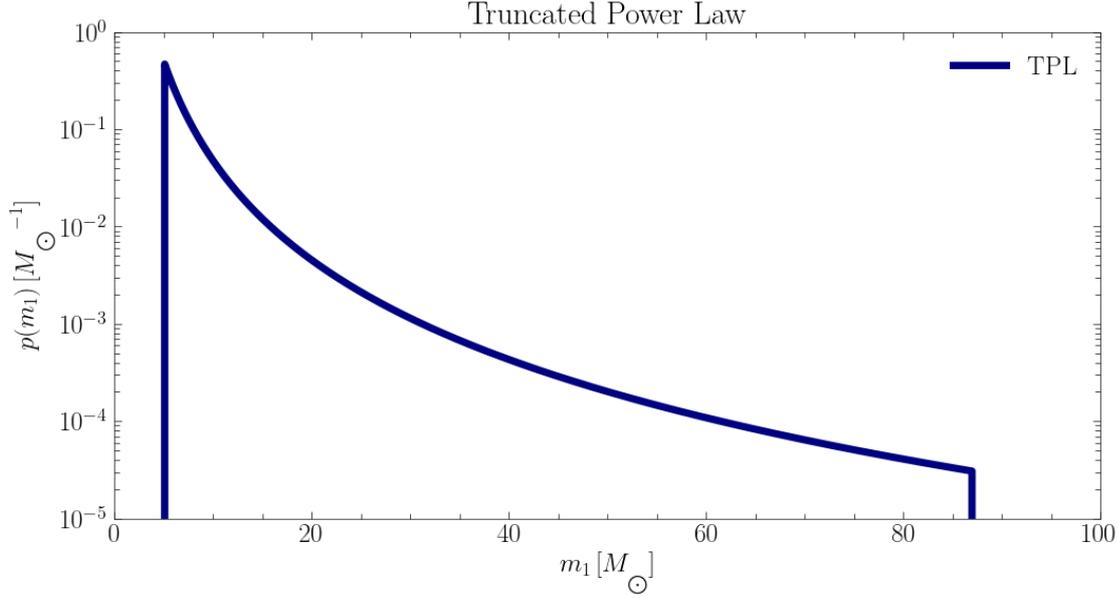


Figure 2.1: Truncated Power Law model for the primary mass distribution.

2.2.2 Broken Power Law

The BPL model is an extension of the TPL one. The primary mass distribution follows a broken power law instead of a simple power law to reproduce a possible tapering at high masses due to the pair-instability super-novae mass gap. In addition, it employs a smoothing function to prevent a sharp cut-off at low masses.

$$p(m_1|\alpha_1, \alpha_2, \delta_m, m_l, m_h, b) \propto \begin{cases} m_1^{-\alpha_1} S(m_1|m_l, \delta_m) & \text{if } m_l < m_1 < m_{break} \\ m_1^{-\alpha_2} S(m_1|m_l, \delta_m) & \text{if } m_{break} < m_1 < m_h \\ 0 & \text{otherwise} \end{cases} \quad (2.16)$$

where $m_{break} = m_l + b(m_h - m_l)$ represents the mass at which there is the break in the spectral index while $S(m|m_l, \delta_m)$ is the smoothing function:

$$S(m|m_l, \delta_m) = \begin{cases} 0 & \text{if } m < m_l \\ \left[\exp\left(\frac{\delta_m}{m-m_l} + \frac{\delta_m}{m-m_l-\delta_m}\right) + 1 \right]^{-1} & \text{if } m_l < m < m_l + \delta_m \\ 0 & \text{if } m \geq m_l + \delta_m \end{cases} \quad (2.17)$$

The secondary mass distribution still follows a power law shape but with the addition of the smoothing function:

$$p(m_2|\beta, m_l, m_1) \propto \begin{cases} m_2^\beta S(m_2|m_l, \delta_m) & \text{if } m_l < m_2 < m_1 \\ 0 & \text{otherwise} \end{cases} \quad (2.18)$$

Thus the hyperparameters $\bar{\lambda}_m$ for the BPL are:

- m_l , the minimum mass of the power law component;
- m_h , the maximum mass of the power law component;
- α_1 , slope of the m_1 distribution for masses $< m_{break}$;
- α_2 , slope of the m_1 distribution for masses $> m_{break}$;
- β , slope of the m_2 distribution;
- δ_m , range of mass tapering on the lower end of the mass distribution;
- b , fraction of the way between m_l and m_h at which the primary mass distribution undergoes a break.

2.2.3 Power Law + Gaussian Peak

This model is another extension of the truncated mass function, it has a Gaussian peak feature that tries to take into account the empirical pile-up of BBH events with primary mass m_1 at $\simeq 35M_\odot$. This concentration of events could be due to the gap caused by pair-instability supernovae that should be located at masses right after the pile-up (Talbot and Thrane, 2018). Consequently, the m_1 distribution is

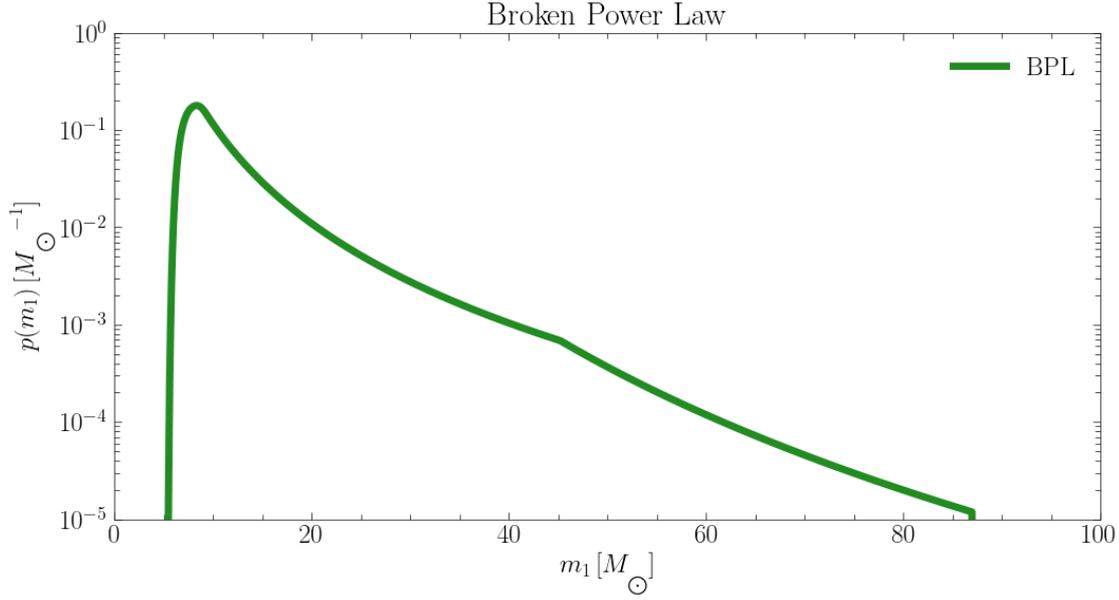


Figure 2.2: Broken Power Law model for the primary mass distribution.

shaped as follows:

$$\begin{aligned}
 p(m_1 | \lambda_p, \alpha, \delta_m, m_l, m_h, \mu_g, \sigma_g) &= \\
 &= [(1 - \lambda_p) \mathfrak{P}(m_1 | -\alpha, m_h) + \lambda_p G(m_1 | \mu_g, \sigma_g)] S(m_1 | m_l, \delta_m)
 \end{aligned} \tag{2.19}$$

where \mathfrak{P} is a normalized truncated power law distribution as in eq.2.14; meanwhile G is a Gaussian distribution with mean μ_g and width σ_g . Lastly, S is the same smoothing function as in eq.2.17. The secondary mass distribution follows the same distribution of eq.2.18. Combining all of these components the resulting hyperparameters $\bar{\lambda}_m$ for the PLP model are:

- m_l , the minimum mass of the power law component;
- m_h , the maximum mass of the power law component;
- λ_p , is a mixing fraction determining the relative prevalence of mergers in T and G, i.e. the fraction of BBH systems in the Gaussian component. By definition, its value is included between 0 and 1;
- α , slope of the m_1 power law distribution;
- β , slope of the m_2 distribution;

- δ_m , range of mass tapering on the lower end of the mass distribution;
- μ_g , mean of the Gaussian component;
- σ_g , width of the Gaussian component.

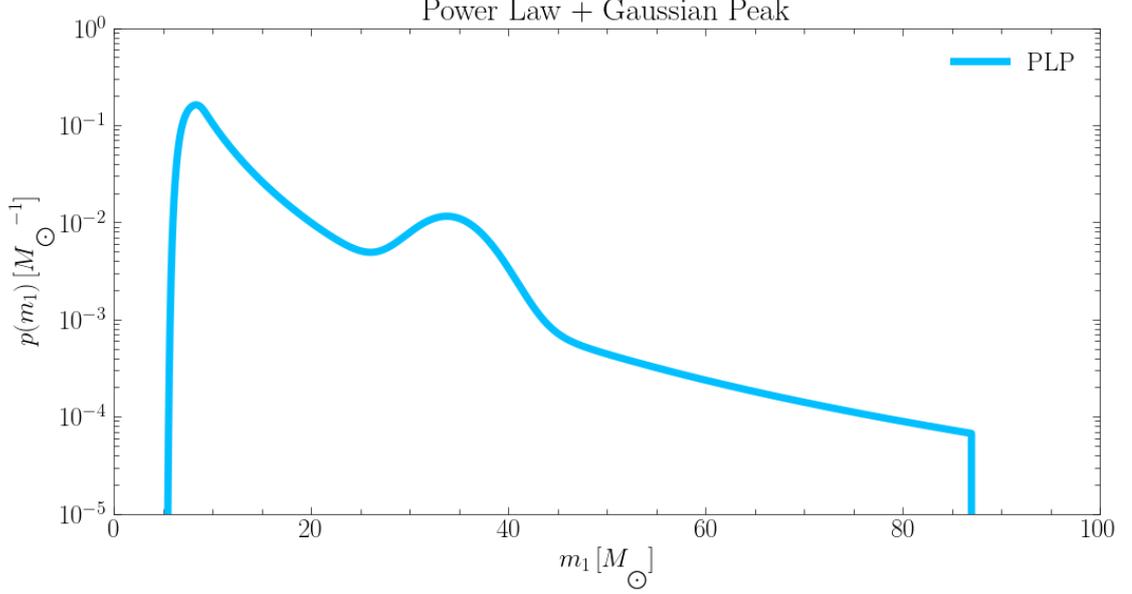


Figure 2.3: Power Law + Gaussian peak model for the primary mass distribution.

2.2.4 Power Law + 2 Gaussian Peaks

This model is basically the same as the PLP one, but has an additional Gaussian feature in the primary mass distribution. This second peak is introduced to reproduce another possible sub-population. In Abbott et al. (2021) it was introduced to investigate a possible concentration of events at high masses, but in Abbott et al. (2023b) it has been seen that it best fits the data if moved to lower masses and supports the findings of the peak-like feature at $\sim 10M_\odot$ in the mass distribution. Thus eq.2.19 is modified as:

$$\begin{aligned}
 p(m_1 | \lambda_p, \lambda_1, \alpha, \delta_m, m_l, m_h, \mu_1, \sigma_1, \mu_2, \sigma_2) &= \\
 &= [(1 - \lambda_p) \mathfrak{P}(m_1 | -\alpha, m_h) + \lambda_p \lambda_1 G(m_1 | \mu_1, \sigma_1) + \lambda_p (1 - \lambda_1) G(m_1 | \mu_2, \sigma_2)] S(m_1 | m_l, \delta_m)
 \end{aligned}
 \tag{2.20}$$

where we have the two Gaussian components and the parameters λ_p and λ_1 represents respectively the fraction of mergers in any G component and the fraction under $G(m_1|\mu_1, \sigma_1)$, hence the hyperparameters $\bar{\lambda}_m$ of the PL2P model are:

- m_l , the minimum mass of the power law component;
- m_h , the maximum mass of the power law component;
- λ_p , is a mixing fraction determining the relative prevalence of mergers in T and G ;
- λ_1 , is a mixing fraction determining the relative prevalence of mergers in $G(m_1|\mu_1, \sigma_1)$ and $G(m_1|\mu_2, \sigma_2)$;
- α , slope of the m_1 power law distribution;
- β , slope of the m_2 distribution;
- δ_m , range of mass tapering on the lower end of the mass distribution;
- μ_1 , mean of the first Gaussian component;
- σ_1 , width of the first Gaussian component;
- μ_2 , mean of the second Gaussian component;
- σ_2 , width of the second Gaussian component;

2.3 Posterior sampling

In the previous sections, I have presented how the likelihood, hence the posterior is built and which are the parameters that I want to estimate. Constraints on these parameters can be obtained theoretically by finding the maximum of the likelihood analytically and performing integrals over the likelihood to obtain the proper marginalised error bars. Unfortunately, the complexity of 2.10 does not allow one to proceed in this way, and doing a brute force estimation on a grid of points is unfeasible, hence I need to rely on algorithms that, given any posterior, return points, which from now on I will call samples, in the hyperparameters space. These samples exhibit statistical independence from each other and adhere to a distribution consistent with the posterior. These methods, commonly known as

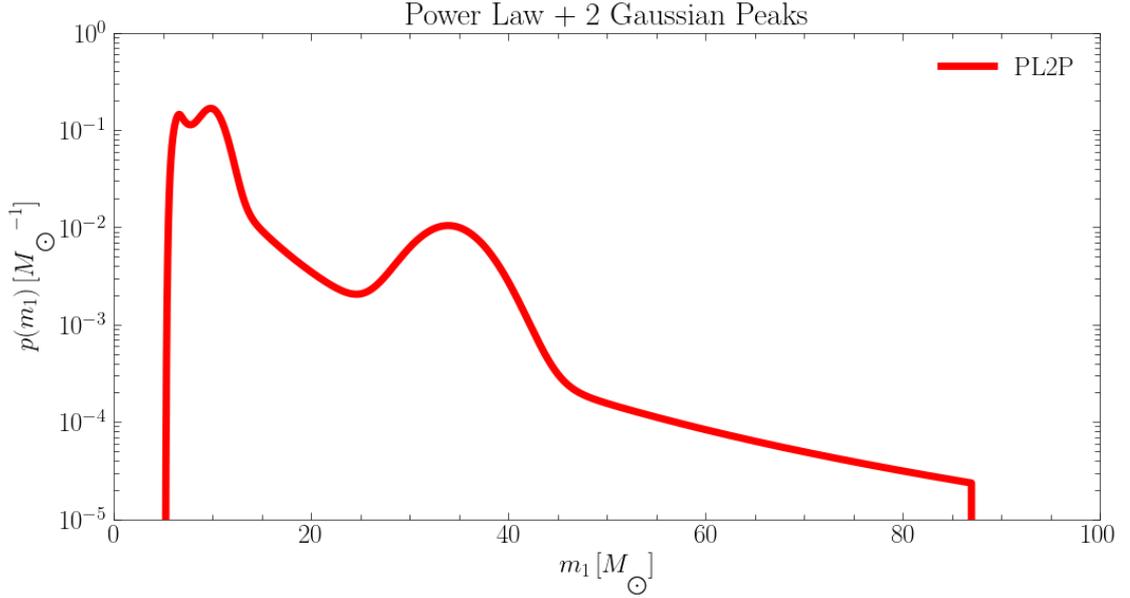


Figure 2.4: Power Law + 2 Gaussian peaks model for the primary mass distribution.

Monte Carlo Markov Chains (MCMC), sample the parameter space with random steps; Monte Carlo refers to the fact that there is a random number generator, while Markov Chains indicate that to generate a sample λ' only the previous one λ is used as input so that the algorithm does not have memory of the previous samples. There are different implementations that have been developed to sample the parameter space; in my work, I have mainly used nested sampling but also affine invariant sampling. In general, all the methods are required to satisfy a condition called detailed balance:

$$P(\lambda)K(\lambda'|\lambda) = P(\lambda')K(\lambda|\lambda') \quad (2.21)$$

where K is the transition probability. This requirement ensures that, sample after sample, the algorithm continues to return points from the correct distribution (Dodelson and Schmidt, 2003; Goodman and Weare, 2010).

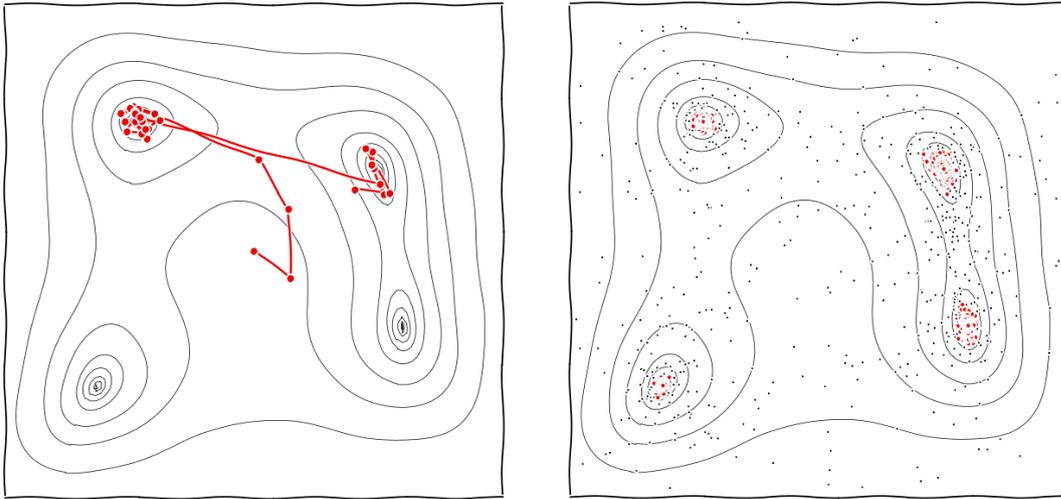


Figure 2.5: The two panels represent a schematic representation of different MCMC methods. In the left panel, it is represented an example of affine invariant sampling. In the right panel, it is represented an example of nested sampling. Illustration taken from Handley (2022)

2.3.1 Affine Invariant Sampling

The affine invariant sampling is a variation of the standard Metropolis-Hasting MCMC algorithm. What distinguishes this kind of sampler from the other methods is that this sampler is invariant under affine transformations, that is, if we consider a variable x with probability density $p(x)$ and perform the transformation from \mathbb{R}^N to \mathbb{R}^N $y = Ax + b$ then the variable y will have probability density $p(y) = p(Ax + b) \propto p(x)$. This means that an MCMC algorithm is affine invariant if the probability $K(x'|x)$ is proportional to $K(y'|y)$.

The simplest algorithm based on the affine invariant sampling rules is the one informally called "stretch move".

What follows is an outline of the procedure (Goodman and Weare, 2010):

- The first step of this algorithm is to draw from the parameter space the so-called ensemble sampler that is a set of n_{walk} walkers $X = (x_1, \dots, x_{n_{walk}})$. Each walker x_k is in \mathbb{R}^N where N is the number of parameters.

- One step $X(t) \rightarrow X(t+1)$ of the procedure consists of updating the entire ensemble. The walkers are updated one at a time, and for every single walker x_k uses the position of all the other walkers, which form the complementary ensemble sampler $X_{[k]}(t) = (x_1(t+1), \dots, x_{k-1}(t+1), x_{k+1}(t), \dots, x_{n_{walker}}(t))$. The single walker is moved accordingly to the stretch move (Fig.2.6): x_k is updated using a random complementary walker x_j , $x_j \in X_{[k]}$, $j \neq k$, with the following proposal:

$$x_k(t) \rightarrow y = x_j + Z(x_k(t) - x_j) \quad (2.22)$$

where Z is a scaling random variable drawn from the density distribution $g(z)$:

$$g(z) = \begin{cases} \frac{1}{\sqrt{z}} & \text{if } z \in [\frac{1}{a}, a] \\ 0 & \text{otherwise} \end{cases} \quad (2.23)$$

where $a > 1$ is the stretch factor parameter that can be adjusted to improve performance.

- This definition of $g(z)$ satisfies the condition $g(\frac{1}{z}) = zg(z)$ and therefore the proposal 2.22 is symmetric, hence the detailed balance condition for the move $x_k(t+1) = y$ is satisfied using as acceptance probability the following one:

$$\min \left\{ 1, Z^{N-1} \frac{p(y)}{p(x_k(t))} \right\}; \quad (2.24)$$

if instead the move is rejected, then $x_k(t+1) = x_k(t)$.

This procedure is repeated for all walkers and until a certain stopping criterion is reached. In the literature, many different algorithms of affine invariant sampling can be found that can be better suited for particular models, i.e. parameter space.

2.3.2 Nested Sampling

The nested sampling algorithm is a Monte Carlo integration technique developed with the main purpose of providing an estimation of the Bayesian evidence (Skilling, 2004). As it has already said in the introduction of section 2.1

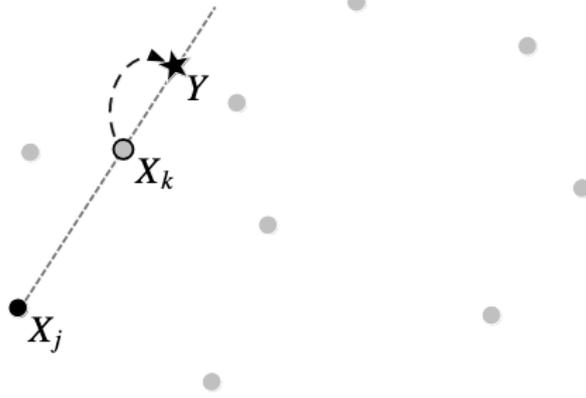


Figure 2.6: A stretch move. The light dots represent the walkers not participating in this move. The proposal is generated by stretching along the straight line connecting x_j to x_k . Illustration taken from Goodman and Weare (2010).

the evidence is the probability at the denominator of eq.2.2 and it is the likelihood marginalised over all the parameters:

$$\mathcal{Z} = p(\{\bar{d}_i\}) = \int \mathcal{L}(\{\bar{d}_i\}|\bar{\lambda})\pi(\bar{\lambda})d\bar{\lambda} \quad (2.25)$$

Obviously, this algorithm simultaneously estimates the posterior parameters. The focus of nested sampling is to solve the multidimensional integral of eq.2.25. It is important to note that the integrand, from eq.2.2, can be expressed as:

$$\mathcal{L}(\{\bar{d}_i\}|\bar{\lambda}) \cdot \pi(\bar{\lambda}) = \mathcal{Z} \cdot p(\bar{\lambda}|\{\bar{d}_i\}) \quad (2.26)$$

so basically is decomposed in a magnitude \mathcal{Z} and a shape $p(\bar{\lambda}|\{\bar{d}_i\})$ which is the posterior. The method (Skilling, 2006; Ashton et al., 2022) is based on the simplification of the traditional Riemann-style integration in which the integration space is decomposed into small volume elements, and then the result is obtained by summing over them:

$$\mathcal{Z} = \int \mathcal{L}(\{\bar{d}_i\}|\bar{\lambda})\pi(\bar{\lambda})d\bar{\lambda} = \lim_{|\Delta\lambda| \rightarrow 0} \sum \mathcal{L}(\{\bar{d}_i\}|\bar{\lambda})\pi(\bar{\lambda})\Delta\lambda \quad (2.27)$$

Due to the infeasibility of such computation in the case of a multi-dimensional space with smaller and smaller volume elements, the approach followed in the nested sampling algorithm is to combine volumes in which the integrand is almost constant (Fig.2.7).

This is reached through an integral transformation (for a formal proof, see box2 of Ashton et al. 2022) and in the end the integral to compute is:

$$\mathcal{Z} = \int_0^1 \mathcal{L}(X)dX = \sum \mathcal{L}(X)\Delta X \quad (2.28)$$

where X is the prior volume (or prior mass) defined as:

$$X(\mathcal{L}^*) = \int_{\mathcal{L} > \mathcal{L}^*} \pi(\bar{\lambda})d\bar{\lambda} \quad (2.29)$$

where \mathcal{L}^* is the iso-likelihood contour. In practical terms the nested sampling

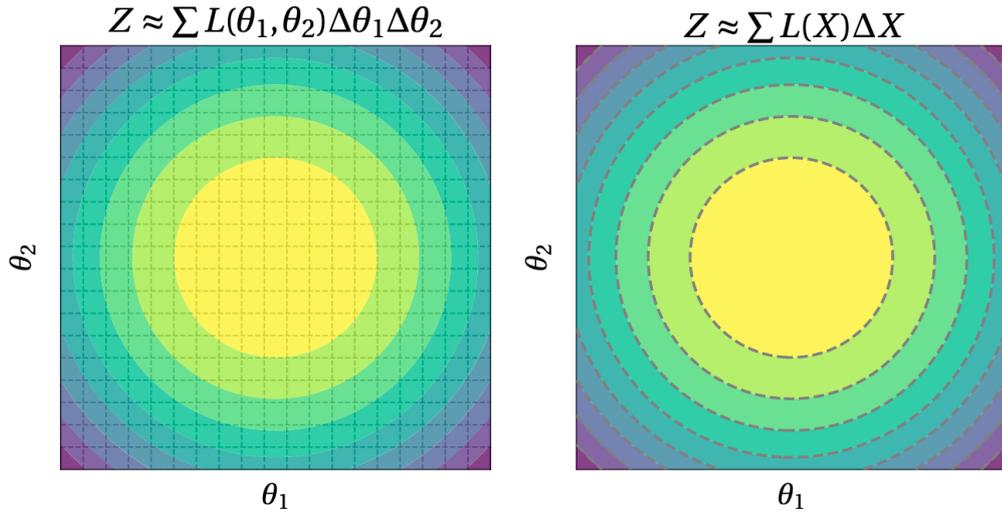


Figure 2.7: In both panels it is represented the same two-dimensional parameter space (the θ parameters correspond to our λ parameters). The colours represent the regions with the same likelihood. The left panel represents the decomposition of the space in small cubes, while in the right panel the decomposition is performed using volumes containing a constant likelihood as done with nested sampling. Illustration taken from Ashton et al. (2022).

algorithm follows the subsequent steps (Fig.2.8):

- The volume is initialized to $X = 1$ while the integral to $\mathcal{Z} = 0$. A factor t , called compression factor, is chosen.
- An ensemble of n_{live} samples is drawn from the prior and their likelihood is computed.
- The smallest one (\mathcal{L}^*) is discarded and the remaining samples are distributed in the volume compressed by t .
- Hence a new sample is drawn from the following prior:

$$\pi(\bar{\lambda})^* = \begin{cases} \pi(\bar{\lambda}) & \text{if } \mathcal{L}(\{\bar{d}_i\}|\bar{\lambda}) > \mathcal{L}^* \\ 0 & \text{otherwise} \end{cases} \quad (2.30)$$

- The estimation of \mathcal{Z} is updated to $\mathcal{Z} = \mathcal{Z} + \mathcal{L}^*(1 - t)X$.
- The volume is now compressed.

All these steps are repeated until a criterion stop is reached. The integral is updated by adding the product of the final volume with the average likelihood among the remaining samples. As presented before, nested sampling allows one to obtain the posterior of the parameters too, indeed at every iteration of the algorithm a sample of the posterior can be easily computed by applying the Bayes theorem, i.e., eq.2.2.

2.3.3 Assessing the scaling performances of different samplers

The algorithms described before in sections 2.3.1 and 2.3.2 have been implemented in many different codes, e.g. for nested sampling see Table 2 of Ashton et al. (2022). Three of the main used in the fields of Astrophysics and Cosmology, and in particular in the GW field, are: **emcee** (Foreman-Mackey et al., 2013), **zeus** (Karamanis et al., 2021), **dynesty** (Koposov et al., 2023; Speagle, 2020). All three of them are Python-based codes, the first and the second one are two different employments of the affine invariant sampling algorithm while the last one is an implementation of the nested sampling procedure. **emcee** and **zeus** differ from each other because

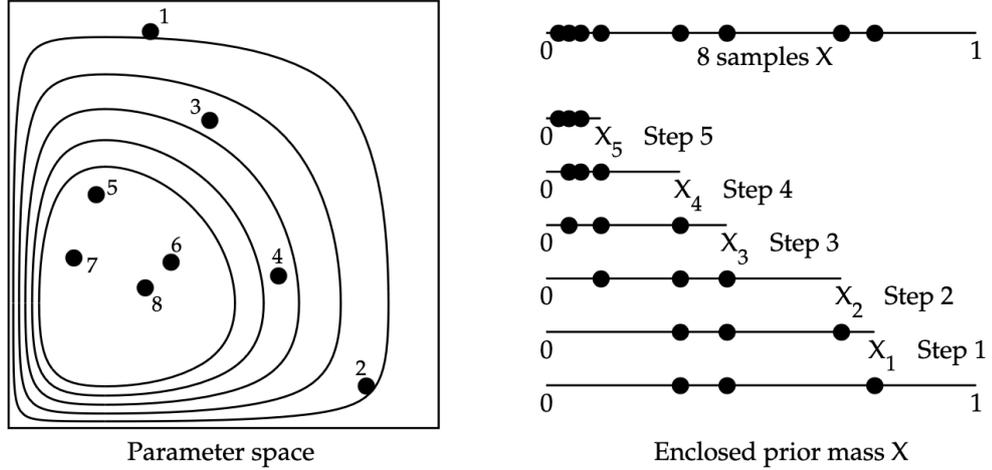


Figure 2.8: Scheme of the nested procedure. In the left panel, it can be observed how the volume X shrinks around the points that are not discarded. The right panel shows a one-dimensional representation which allows us to see at every step the elimination and substitution of the lowest likelihood points. Illustration taken from Skilling (2006).

of the different algorithms of sampling; while **emcee** uses the affine invariant sampling method described in Goodman and Weare (2010), **zeus** is an application of an algorithm called slice sampling (Karamanis and Beutler, 2021) that basically, differently from the procedure described in section 2.3.1, doesn't sample from the entire parameter space but from slices, along a single dimension, defined by the probability density function.

Before implementing the usage of these three samplers with CHIMERA I have compared the results obtained sampling the parameter space of the model of $H(z)$ with the $w_0 - w_a$ parameterisation (Linder, 2003) of the equation of state for dark energy:

$$H(z) = H_0 \sqrt{\Omega_m(1+z)^3 + \Omega_\Lambda(1+z)^{3(1+w_0+w_a)} e^{-3w_a \frac{z}{z+1}} + (1 - \Omega_m - \Omega_\Lambda)(1+z)^2}. \quad (2.31)$$

Then, using the same model as a reference, I have performed some scaling tests of all three codes: in the case of **emcee** and **zeus** I have investigated how they scale changing the number of walkers and the number of iterations, while in the

case of **dynesty** the only parameter that was worth changing was the number of live points which is directly linked to the number of iterations needed to reach the stopping criterion. For these tests I have used a dataset of $H(z)$ values taken from different works with cosmic chronometers (see Tab.2.1).

From Tab.2.2 it can be seen that the three samplers return posterior values consistent with each other as they are all within the 1σ confidence level of the other two. This can also be observed in the corner plot of Fig.2.9 where the distributions produced by the three samplers are overlapped. It can be noticed that in general the results obtained with **dynesty** have a larger 1σ error and that the 1-dimensional distributions are less peaked with respect to the other two. This can be attributed to the fact that **dynesty** is an evidence-oriented sampler that leads to larger posterior errors, however in spite of this they are still satisfying. It can be concluded therefore that if one is interested in both posterior estimation and evidence of the studied model, he can rely on the usage of **dynesty** still taking into consideration that, with affine invariant samplers, better posterior parameters constraints can be obtained.

Concerning the scaling tests, comparing the affine invariant samplers with the nested sampler is more challenging because of the absence of common parameters for tuning. Consequently, I opted to perform two distinct scaling tests: one involving only the **emcee** and **zeus** samplers, with a fixed number of algorithm iterations while varying the number of employed walkers; the other varying the number of iterations, facilitating comparison among the three samplers. It is important to note that for **emcee** and **zeus**, the number of iterations can be predetermined, while with **dynesty**, this quantity is directly linked to the number of live points used: a higher count of live points necessitates more iterations. Accordingly, in this test, I initially executed the code using nested sampling, increasing the number of live points, and subsequently used the corresponding iteration counts for the affine invariant sampler runs fixing the number of walkers to 32. In the top panel of Fig.2.10 it can be seen that, as a general trend, both the running times of **emcee** and **zeus**, with the number of iterations fixed to 5000, increase with the number of employed workers as expected. Moreover, although in Karamanis et al.

Table 2.1: $H(z)$ measurements (in units of $[km/Mpc/s]$) and their errors at the corresponding z . This dataset can be downloaded from https://apps.difa.unibo.it/files/people/Str957-cluster/astro/CC_data/

z	$H(z)$	$errH(z)$	Reference
0.07	69	19.6	(Zhang et al., 2014)
0.09	69	11.9991	(Simon et al., 2005)
0.12	68.6	26.2	(Zhang et al., 2014)
0.17	83	8.00037	(Simon et al., 2005)
0.1791	74.91	3.8069262	(Moresco et al., 2012)
0.1993	74.96	4.9001352	(Moresco et al., 2012)
0.2	72.9	29.6	(Zhang et al., 2014)
0.27	77	13.9986	(Simon et al., 2005)
0.28	88.8	36.6	(Zhang et al., 2014)
0.3519	82.78	13.94843	(Moresco et al., 2012)
0.3802	83	13.54	(Moresco et al., 2016)
0.4	95	16.9955	(Simon et al., 2005)
0.4004	76.97	10.18	(Moresco et al., 2016)
0.4247	87.08	11.24	(Moresco et al., 2016)
0.4497	92.78	12.9	(Moresco et al., 2016)
0.4783	80.91	9.044	(Moresco et al., 2016)
0.47	89.0	49.6	(Ratsimbazafy et al., 2017)
0.48	97	62.0024	(Stern et al., 2010)
0.5929	103.8	12.49752	(Moresco et al., 2012)
0.6797	91.6	7.961872	(Moresco et al., 2012)
0.75	98.8	33.6	(Borghi et al., 2022)
0.7812	104.5	12.19515	(Moresco et al., 2012)
0.8754	125.1	16.70085	(Moresco et al., 2012)
0.88	90	39.996	(Stern et al., 2010)
0.9	117	23.0022	(Simon et al., 2005)
1.037	153.7	19.6736	(Moresco et al., 2012)
1.3	168	17.0016	(Simon et al., 2005)
1.363	160	33.58	(Moresco, 2015)
1.43	177	18.0009	(Simon et al., 2005)
1.53	140	14	(Simon et al., 2005)
1.75	202	39.996	(Simon et al., 2005)
1.965	186.5	50.43	(Moresco, 2015)

Table 2.2: $H(z)$ model posterior parameters, eq.2.31, and their 1σ confidence levels obtained with the three samplers.

Sampler	H_0	Ω_m	Ω_Λ	w_0	w_a
emcee	$68.7^{+8.0}_{-6.2}$	$0.22^{+0.15}_{-0.12}$	$0.49^{+0.24}_{-0.26}$	$-1.58^{+0.82}_{-0.91}$	-0.6 ± 1.7
zeus	$68.8^{+8.4}_{-6.3}$	$0.21^{+0.15}_{-0.11}$	$0.49^{+0.23}_{-0.26}$	$-1.60^{+0.86}_{-0.92}$	$-0.7^{+1.7}_{-1.6}$
dynesty	$67.9^{+12.2}_{-8.8}$	$0.28^{+0.31}_{-0.16}$	0.49 ± 0.29	$-1.44^{+0.98}_{-1.01}$	$-0.7^{+1.8}_{-1.6}$

(2021) it is reported that in problems related to Baryon Acoustic Oscillation and exoplanet radial velocity fitting **zeus** is way faster than **emcee**, my tests show that in this case of study the latter is approximately 5 times faster than the first one. These two trends for the affine invariant samplers can be seen also in the bottom panel of Fig.2.10 where the running time is plotted against the number of iterations, in addition in this plot it is possible to observe the scaling of **dynesty**: its running time seems to be quite constant or at least slowly growing up, with fluctuations, as the amount of live points accretes. Hence in general the nested sampler is slower than **emcee**, especially with few iterations, but it is faster than **zeus**.

These tests not only provide me an overview of the performances of these three samplers but with them I also had the opportunity to create a code structure, that after I have implemented in the analysis code of CHIMERA, that allows selecting which of these three sampler used for the analysis and that automatically returns the sampling time, some preliminary plot and the evidence in the case of dynesty. Last but not least they also help the decision of which sampler, between **emcee** and **zeus**, to use for the analysis in Borghi et al. (2024).

2.4 CHIMERA

Speaking in practical terms, all the aforementioned equations have to be encoded in a programming language in order to be used for an actual analysis. This is what has been done in Borghi et al. (2024) where CHIMERA is presented. As already said at the beginning of this Chapter, CHIMERA is a new Python code designed to fit cosmological and astrophysical population parameters implementing bright, dark, and spectral sirens methods, hence by also using information from galaxy cata-

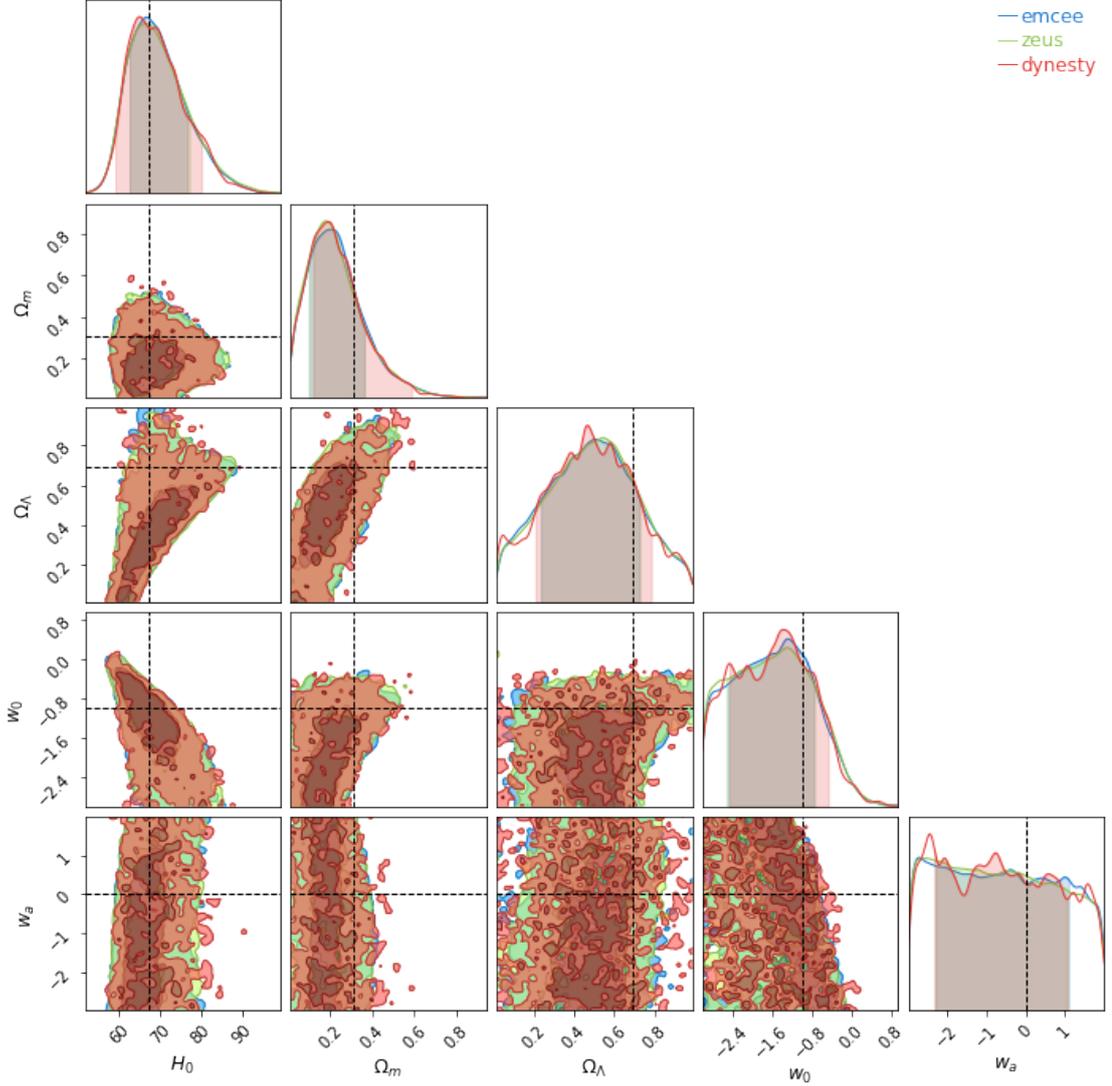


Figure 2.9: This corner plot shows the 2-dimensional and 1-dimensional posterior distributions of the $H(z)$ model posterior parameters, eq.2.31, obtained with the three samplers. The blue regions are those obtained with **emcee**, the green regions are those obtained with **zeus** and the red ones are those obtained with **dynesty**. The black dashed lines represent the true values of the parameters taken from Aghanim et al. (2020).

logues. The main goal of this code is to compute efficiently the result of eq. 2.10. The code flowchart is displayed in Fig. 2.11. CHIMERA is built around two core modules Likelihood.py and Bias.py, that are in the violet boxes in Fig. 2.11, which

are respectively devoted to the computation of the integrals in eq. 2.10 and to the computation of the selection bias term $\xi(\bar{\lambda})$. The Likelihood.py module basically stores all the population models, i.e. mass, rate, and spin models, and the GW and galaxy data. In parallel, the Bias.py module stores all the population models and the GW injection catalogue data, which are simulated GW events. Finally, following eq. 2.3, the complete likelihood is computed. This involves initially computing the log-likelihood for each event and subsequently subtracting the log-bias term, which is computed once and multiplied by the number of events.

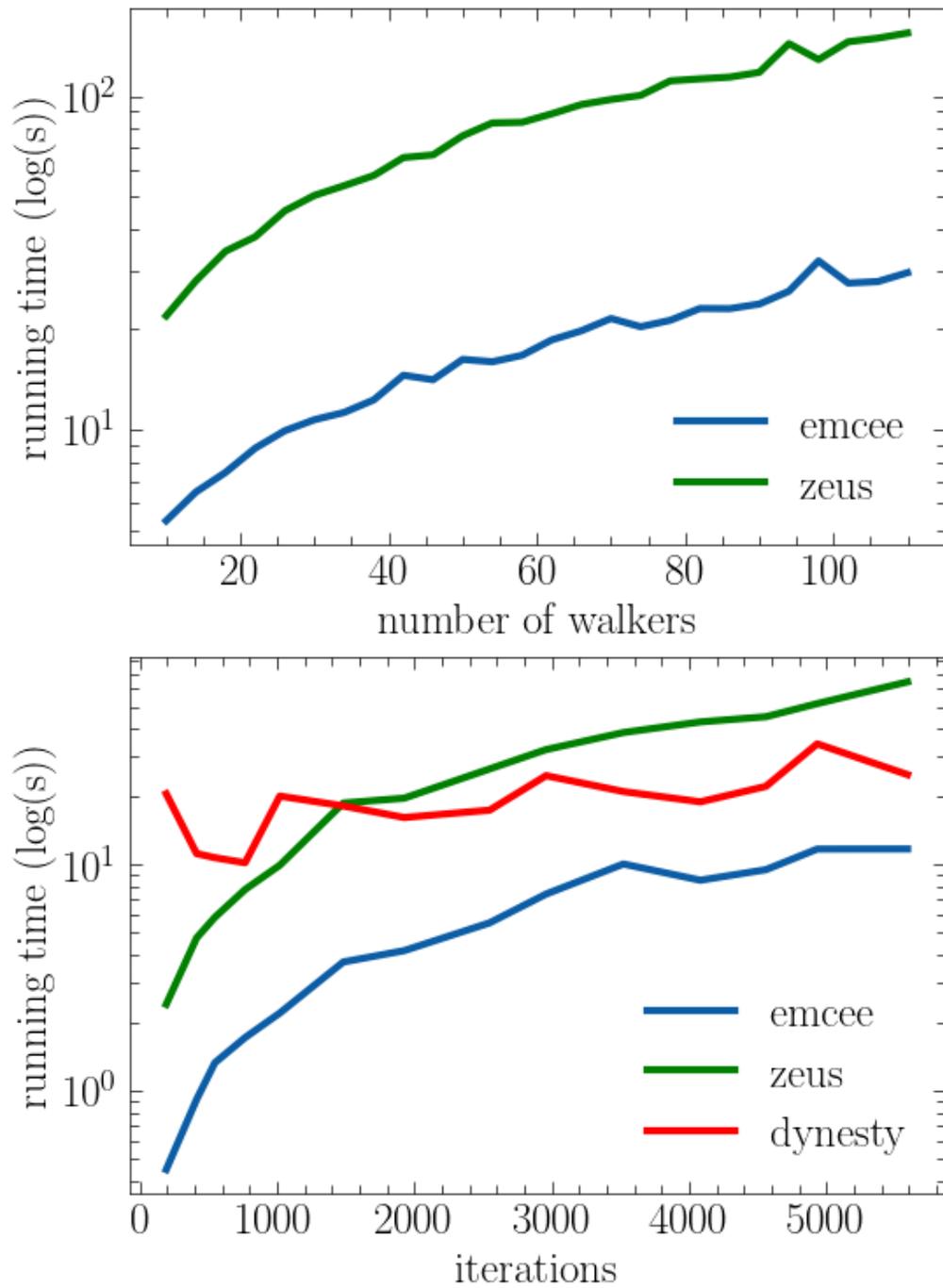


Figure 2.10: Top panel: scaling of **emcee**, blue line, and **zeus**, green line, as a function of the number of walkers. Bottom panel: scaling of **emcee**, blue line, **zeus**, green line, and **dynesty**, red line as a function of the number of iterations.

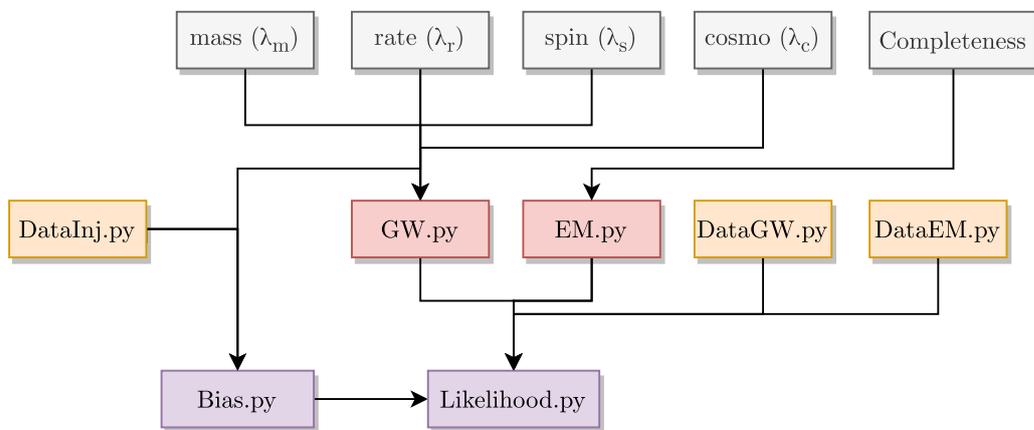


Figure 2.11: The flowchart of CHIMERA. The main modules include functions related to the full likelihood computation (violet), computation of probabilities (red), data file I/O (yellow), and model functions (gray). Illustration taken from Borghi et al. (2024).

Chapter 3

Data

This Chapter introduces the GW data and the galaxy catalogue used in the hierarchical Bayesian analyses of the binary black holes mass functions and the studies on the parameters' dependencies of the H_0 posteriors performed with 1-dimensional analyses. The GW data used are not those collected by the LVK Collaboration so far, but are mock data generated to validate the CHIMERA code and forecast the constraints on cosmological and population parameters from the ongoing O4 and future O5 runs. In the first part, the generation of the two GW event mock catalogues and the characteristics of the mock galaxy catalogue are described. The second section summarises the main features of the two GW catalogues.

3.1 Catalogues generation

The primary goal of CHIMERA is to utilise GW events as Dark and Spectral sirens. Therefore, it is essential to possess not only a catalogue of GW events but also one of galaxies; thus, in particular, in Borghi et al. (2024) the code has been tested using a mock galaxy catalogue created from the MICE Grand Challenge light-cone simulation v2 (Fosalba et al., 2015; Crocce et al., 2015; Fosalba et al., 2014; Carretero et al., 2014) which covers one octant of the sky. The same galaxy catalogue is adopted in the following analyses. Not the entire MICE galaxy catalogue has been used, but only a subsample, obtained by imposing a uniform in comoving volume distribution and selecting only those galaxies with stellar masses greater than

$M_*/M_\odot > 10^{10.5}$. This cut in mass is reasonable as merger events are expected to occur in massive galaxies. Indeed, the binary merger rate is strictly related to the stellar mass. Due to the fact that we want to derive constraints using GWs as dark sirens in combination with galaxy catalogues, the GW catalogues need to be built in such a way that every event has an associated host galaxy in the galaxy catalogue. Therefore, from this parent sample the sources of GW events have been sampled using a weight proportional to the merger rate parametrisation described in eq.2.9. In particular, the assumed λ_z hyperparameters are those of Tab.3.1. In

γ	κ	z_p
2.7	3	2

Table 3.1: Fiducial values of the merger rate parametrisation used to select the potential hosts of the GW events from the parent sample.

this way, a total of approximately 4×10^5 potential hosts have been identified. To each potential host, we assign a GW merger event in which the binary source frame properties, such as the masses $m_{1,2}$ are drawn from a fiducial population distribution. Then, these quantities are converted into detector frame ones by assuming the same fiducial Λ CDM model with $H_0=70$ and $\Omega_{m,0}=0.25$ used in the MICEv2 simulation. The masses are assigned using as mass distribution the PLP model described in section 2.2.3 that has been employed with the hyperparameters λ_m obtained in Abbott et al. (2023b) (see Tab.3.2).

m_l	m_h	λ_p	α	β	δ_m	μ_g	σ_g
5.1	87	0.039	3.4	1.1	4.8	34	3.6

Table 3.2: Fiducial parameters of the PLP mass distribution used to assign masses to the GW events.

Then, the observed signal, the corresponding SNR, and the parameter estimation (PE) samples have been simulated using the **GWFAST** pipeline (Iacovelli et al., 2022a,b). Two different network configurations have been selected to do that, one for the O4-like catalogue and the other for O5-like one (from now on they will be denoted just as O4 and O5). The first configuration used the two LIGO interferometers, Virgo, and KAGRA detectors, while the second configuration, in addition

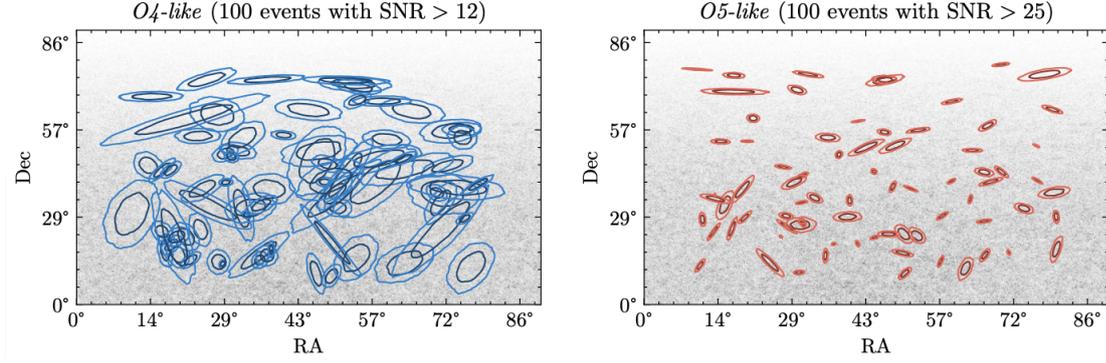


Figure 3.1: Sky localisation area at 1σ and 2σ of the best 100 events of the O4 catalogue (left panel) and O5 catalogue (right panel) overlapped to the map of the potential host galaxies (gray dots) extracted from MICEv2. Illustration taken from Borghi et al. (2024).

to the aforementioned, also included the planned LIGO interferometer in India. In both cases, public-available (<https://dcc.ligo.org/LIGO-T2000012/public>) sensitivity curves created by the LIGO-Virgo-KAGRA (LVK) collaboration have been used: one for the first configuration with expected sensitivities of O4, one for the second configuration with expected sensitivities of O5. Each catalogue contains 5000 PE posterior samples for each event and the associated signal-to-noise ratio. Not all the events are used in the analyses, but only a subsample of them. Indeed, only the best 100 events that can be observed over approximately one year, are chosen. The sub-catalogues are obtained by selecting those events that have a $\text{SNR} > 12$ for the O4 catalogue and a $\text{SNR} > 25$ for the O5 catalogue. The different cuts in SNR are due to the different network sensitivities. The choice to select only 100 events is mainly due to computational reasons. Indeed, the likelihood of eq. 2.10 is computationally expensive, thus, it is necessary to select a number of events of this order of magnitude.

3.2 Catalogues characterisation

In Fig.3.1 it can be seen how the GW events are distributed in the RA-DEC coordinate system.

The redshift range of both O4 and O5 events is roughly the same, $0.05 \lesssim z \lesssim 1$

(see Tab.3.3). The events are similarly distributed in redshift even if the two sen-

	z_{min}	z_{max}
O4-like	0.07	0.84
O5-like	0.10	0.97

Table 3.3: Maximum and minimum values of the event redshifts of the two catalogues. These values correspond to the maximum/minimum median of the 5000 samples from each event.

sitivities are different. In particular, the O5 configuration is expected to detect events at higher redshift than the O4 one. However, the similar redshift distribution can be explained by the different cuts in SNR described in the previous section. The improvement in signal detection with the O5 network can be translated into improved localisation capabilities. This is due to both the addition of a fifth interferometer and the increased sensitivities of those already existing. Fig.3.1 shows indeed that the sky localisation areas are much smaller in the O5 catalogue (right panel).

In the improved network, there are 10 times more events with $SNR > 50$ and the maximum SNR is 3 times greater than the O4 case (see Tab.3.4). This is directly connected to the SNR cuts that, in the case of O5 allow to observe better than in the O4 case events at the same redshift. The best event in the O4 catalogue is an outlier of the general distribution of events, and it can be thought of as a statistical fluctuation. Nevertheless, it is still a possible detection of this network configuration. The same can be said for the two O5 events with $SNR > 100$.

	SNR_{max}	n_{ev}	$SNR > 25$	n_{ev}	$SNR > 50$	$SNR > 100$
O4-like	54.03		12		1	–
O5-like	142.79		100		12	2

Table 3.4: Maximum signal-to-noise ratio and number of events with $SNR > 25$, $SNR > 50$ and $SNR > 100$ for the two catalogues.

A tracker of an event's localisation goodness is the number of galaxies per localisation volume ($N_{gal,vol}$), where the localisation volume is obtained from the sky localisation $\hat{\Omega}(RA, DEC)$ and the luminosity distance uncertainty δd_L . In fact,

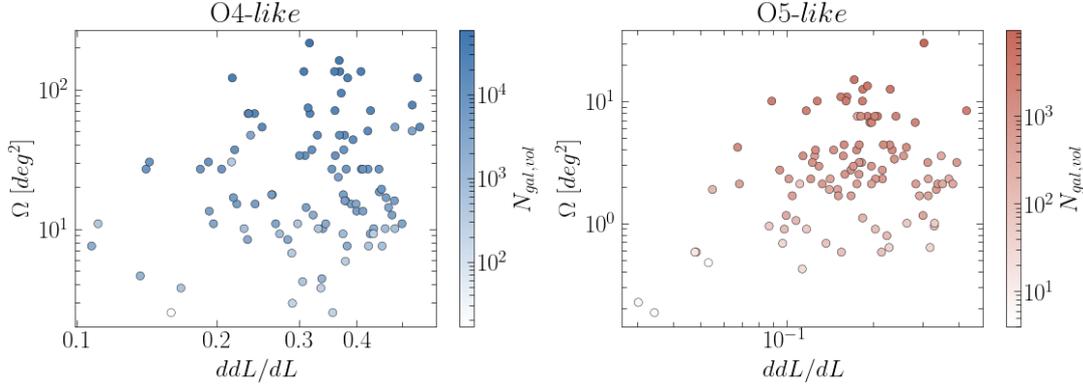


Figure 3.2: Distribution of the relative uncertainty on the luminosity distance and sky localisation area as a function of the number of galaxies per localisation volume ($N_{gal,vol}$) for the O4 (top panel) and O5 (bottom panel) catalogues.

as the relative uncertainties of both $\hat{\Omega}$ and d_L increase, $N_{gal,vol}$ also increase. Note that in the O4 case (left panel of Fig.3.2), both the axis and the colormap have values that are an order of magnitude higher than those of the right panel, i.e. O5 events. As can be qualitatively seen from Fig.3.3, the events included in the O4 catalogue (blue histogram) have a number of potential host galaxies equal to, or even an order of magnitude higher than the worst cases of the O5 catalogue (red histogram). In particular, the worst localised event of O4 has 6 times the number of galaxies per localisation volume of the worst one detected by the improved network (see Tab.3.5). The O5 catalogue has three events with $N_{gal,vol} < 10$, while the best O4 event has $N_{gal,vol} = 17$. Finally, it is worth noticing that while only the 40% of events detected with the enhanced network have more than 500 galaxies in their localisation volume, in the O4-like catalogue this percentage increases up to 91%. This remarkable reduction of possible hosts is of primary importance for the usage of GW events as dark sirens, as discussed in sec.1.3.2.

	$N_{gal,vol}^{min}$	$N_{gal,vol}^{max}$	$n_{ev} N_{gal,vol} > 50$	$n_{ev} N_{gal,vol} > 500$
O4-like	17	59531	99	91
O5-like	4	9727	92	40

Table 3.5: Maximum and minimum number of galaxies per localisation volume ($N_{gal,vol}$) of the events of both catalogues and number of events with $N_{gal,vol} > 50$ and $N_{gal,vol} > 500$ for both catalogues.

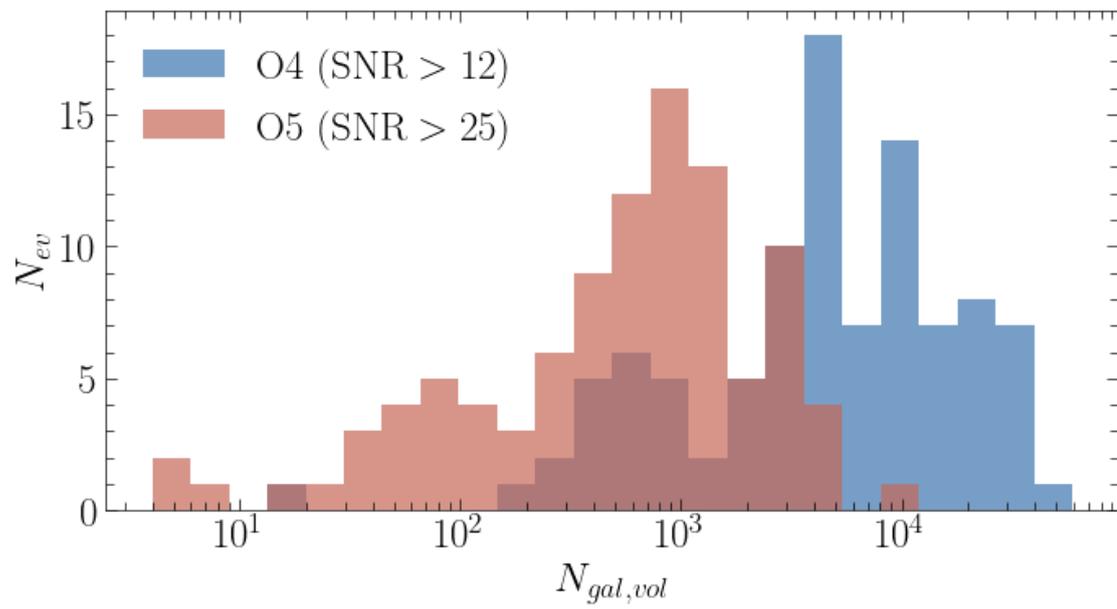


Figure 3.3: Distribution of the GW events as a function of the number of galaxies per localisation volume for both catalogues: O4 (blue); O5 (red).

Chapter 4

Results

In Sect. 1.3 it has been described how GWs can be used as cosmological probes; then in Sect. 2.1 their statistical treatment for this purpose has been discussed. In this Chapter, the results obtained with dark sirens analyses through the GW and galaxy data described in Chapter 3 will be presented, exploiting the `dynesty` nested sampler discussed in Sect. 2.3.3.

The Chapter is divided in two main parts: one in which the results obtained in the analyses of the BBH mass functions are presented and discussed; one in which, instead, the dependencies of the H_0 posterior distribution on the catalogues assumptions are presented and discussed. In particular, in Sect. 4.1 the codes implementations for the analyses and the analyses' results are displayed; in Sect. 4.2 the considerations on the H_0 posterior distribution obtained by varying several parameters are shown off.

4.1 The Black-Holes Binaries Mass Function

To achieve even more precise results for cosmological parameters like H_0 , we need not only additional observations, but also a better understanding of the properties of the astrophysical sources of these GW events. Since most of the signals we detect are believed to come from BBH mergers, it is of paramount importance to study the BBH population in its entirety. For example, we need to study how the component masses are distributed. This information is intrinsically related to

their formation mechanisms. Eventually, a feature in the mass distribution (e.g., break, peak, or change in slope) can be used to obtain cosmological information, see Ezquiaga and Holz (2022).

In the literature, there are no established theoretical models that describe the mass distribution of BBH systems. Thus, in order to study their properties and to use them as dark and spectral sirens it is important to empirically find the model that best fits the available data. With the increasing number of detected GW events, it is and will be possible to distinguish more substructures underlying the general shape. Many different models have been proposed to reproduce the data distributions in the best possible way, and they have already been presented in Sect. 2.2. In the work of Abbott et al. (2021) it is shown that using the O2 catalogue the TPL model is highly disfavoured, while the BPL is only marginally disfavoured with respect to the power-law with one or two Gaussian peaks models. This was also confirmed in Abbott et al. (2023b). Currently, the most favoured model is the PLP, which includes a Gaussian feature peaking at around $34M_{\odot}$. However, the PL2P model is also being considered, as there are hints supporting the presence of a second peak around $10M_{\odot}$. Therefore, it is important to understand how much the future datasets will be able to constrain these mass functions, as with GWTC3 it was possible to obtain parameter constraints with $\sim 60\%$ of accuracy that does not allow one to have strong evidence of the existence of this peak, and hence the best fitting model. Thus, I will explore the data described in Chapter 3 combining different actions:

- extend the CHIMERA mass functions library implementing the BPL model;
- develop a code structure that allows to use CHIMERA with a Bayesian nested sampler;
- run the code with different configurations assessing its performances and results;
- develop a code that allows to use model comparison criteria.

CHIMERA In particular, the different code runs have been performed fixing all the rate and cosmological parameters. This has been done as the parameter under

study are the mass hyperparameters displayed in Sect. 2.2 and to save some computational time.

4.1.1 Extending the Binary Black Holes Mass Function Models

Before starting the analyses, I had to expand the mass functions library of CHIMERA including also the Broken Power Law (BPL) (2.2.2), which was not present. Since the structure of the mass function modules is standardised, I had to implement only the primary mass distribution as both the smoothing function and the secondary mass distribution are the same as the PLP. To follow the code structure and notation of the already present mass functions, I defined all the probability functions in logarithmic form. This choice is computationally motivated and it is made in order to deal with smaller values. Thus, firstly I defined the primary mass distribution using the definition of eq. 2.16. In order to ensure a smooth transition between the two power-laws it is necessary to include the term $m_{break}^{(-\alpha_1+\alpha_2)}$ in the $m_{break} < m_1 < m_h$ regime which connects the two parts. Then, I defined the normalisation function of $p(m_1)$, which basically is its integral in the range of masses between m_l and m_h . Finally, I summed the log-probability of the primary and secondary mass distributions with the respective normalisations and defined the full probability distribution function of the BPL model by taking the exponential of this sum.

I have then tested and validated the new model by comparing its shape with those already present, an example is present in Fig. 2.2.

4.1.2 Model selection criteria

In the field of Astrophysics, a common problem is selecting a single model that best represents a dataset, among many others. Indeed, different models may all return a reasonable fit but they can be formally different, for example with a different number of parameters, or even based on completely different theoretical assumptions. To compare different models, many statistical methods have been developed over the years, with the ability of correctly weighting the evidence of

every considered model. In this dissertation, I have taken into account just three of them, namely: Bayesian evidence (Trotta, 2008), the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002) and the Posterior Predictive Check (PPC) (Gelman et al., 2013). To use them I have developed a code that gets the results obtained with different models and applies one or more of these three methods returning afterwards the desired comparison metric. Actually, the PPC is not a real metric to evaluate the goodness of fit, but it is a powerful visual check extensively implemented also in the LIGO-Virgo-Kagra analysis.

Bayesian evidence

I have already introduced the Bayesian evidence in Sect. 2.3.2. It comes directly from Bayesian statistics, and is thus the preferred criterion to use as it speaks in terms of model probability. Indeed, it is widely used in the field of gravitational waves, e.g. for the selection of the model that best reproduces the detected waveform (Thrane and Talbot, 2019). In general, we can directly compare the evidence values of different models, since the higher the value, the better the model reproduces the data (Liddle et al., 2006; Liddle, 2007). To compare two models, it is usually computed the so-called Bayes factor, which is the ratio of the two model evidences:

$$\mathcal{B} = \frac{\mathcal{Z}_1}{\mathcal{Z}_2}, \quad (4.1)$$

The value \mathcal{B} directly provides the information on which model is favoured: if $\mathcal{B} > 1$, model 1 is preferred to model 2, and vice versa for $\mathcal{B} < 1$.

Deviance Information Criterion

The Deviance Information Criterion (DIC) is a method for comparing models, specifically developed for Bayesian hierarchical models. It uses a quantity called Bayesian complexity (Kunz et al., 2006), p_D , which is a measure of the number of parameters that the data can constrain. It is defined as:

$$p_D = \overline{D(\theta)} - D(\bar{\theta}) \quad (4.2)$$

where $D(\theta) = -2\ln\mathcal{L}(\theta)$. The DIC is then defined as:

$$DIC = \overline{D(\theta)} + p_D = -2 \left[\overline{\ln\mathcal{L}(\theta)} - \text{var}(\ln\mathcal{L}(\theta)) \right] \quad (4.3)$$

where *var* indicates the variance. To compare the DICs of two models, it is used the difference of the two values $\Delta DIC = DIC_A - DIC_B$, where a positive value indicates a preference for model B, since a lower DIC indicates a favoured model (Edelman et al., 2022). Even though this criterion has some limitations (see Spiegelhalter et al. 2014), it provides a fast and quite reliable evaluation of the goodness of fit for models. More importantly, it does not require the computation of the Bayesian evidence \mathcal{Z} , therefore it can be obtained directly from affine invariant posterior samples (Liddle, 2007).

Posterior Predictive Checks

The Posterior Predictive Checks are a visual way to evaluate models allowing to check if the inferred population is consistent with the observed one (Edelman et al., 2022), even though they don't provide a quantitative way to compare them. The predicted population is obtained through the so-called injections. They are simulated events with associated detector frame event-level parameters θ . Starting from a set of simulated events, with parameters that cover the whole parameter-space they are selected simulating their detectability with a given detector network. To generate the plots of the Posterior Predictive Checks I created a code performing the following steps:

- One random sample is drawn from the hyperparameters' posteriors.
- The θ parameters of both the observed and predicted events are converted from detector to source frame.
- One sample of the event-level parameters is selected for each of the N_{events} events using a rejection sampling procedure which uses the following weights:

$$w = \frac{p_{pop}}{\pi(\theta)} \quad (4.4)$$

where p_{pop} is the one described in eq. 2.5 and $\pi(\theta)$ are the redshift, primary and secondary masses priors.

- N_{events} injections are selected using the same rejection sampling of the previous step.
- All the previous steps are repeated for a total of N_{draws} , thus in the end each observed event will have a total of N_{draws} samples and N_{draws} sets of N_{events} predicted observations will be obtained from the injections.

In this way, both the observed and the predicted events are reweighted to the inferred population model. Then to compare the two sets of data the cumulative distribution functions (CDFs) of every N_{draws} are computed. Then the 90% confidence interval (CI) is computed and plotted against the primary mass distribution. In this way, a model represent a good representation of the data if the 90% CI of the observed CDFs is entirely enclosed inside the one of the predicted CDFs.

4.1.3 Constraining the BBH mass function parameters

In this Section, the results obtained are presented. As a first step, I integrated the current version of CHIMERA in a code that defines the likelihood, priors and posterior, and sample the posterior with dynesty. I setup dynesty to run with 500 live points using the "multi" bounding method which is indicated as the most adaptive one. I ran it in the Static nested sampling modality in order to give more relevance to the model Bayesian evidence estimation with respect to the model's parameters. Furthermore, I set it up in order to be executed in parallel using the computational resources from the parallel computing cluster of the Open Physics Hub (<https://apps.difa.unibo.it/wiki/oph:cluster>) at the Physics and Astronomy Department in Bologna.

The parameters of the models considered are described in Sect. 2.2. The cosmological and rate hyperparameters have been fixed because I noticed, with a first run, that a full nested sampling analysis requires an elevated computational time. Nevertheless, we do not expect this to significantly impact our result, since it is a common assumption also in the literature (Abbott et al., 2023b).

The uncertainty on the galaxies' redshift have been fixed to $z_{err} = 0.001$, which in

Borghi et al. (2024) is referred to as spectroscopic redshift (z_{spec}). It is the accuracy that it is expected to be reached with the next large spectroscopic surveys that will be performed for example with the ESA mission Euclid (Laureijs et al., 2011) or the Dark Energy Spectroscopic Instrument (DESI) (Collaboration et al., 2016). However, in Sect. 4.2.1 it will be explored in detail how this assumption affects the results, and how a different redshift uncertainty propagates to the cosmological parameters derived.

The priors used for the $\bar{\lambda}_m$ were all assumed to be flat and are displayed in Tab. 4.1.

Parameter	m_l	m_h	α	α_1	α_2	
Range	$\mathcal{U}(2,50)$	$\mathcal{U}(50,200)$	$\mathcal{U}(1.5,12)$	$\mathcal{U}(1.5,12)$	$\mathcal{U}(1.5,12)$	
Parameter	β	δ_m	b	λ_p	λ_1	
Range	$\mathcal{U}(-4,12)$	$\mathcal{U}(0.01,10)$	$\mathcal{U}(0,1)$	$\mathcal{U}(0.01,0.99)$	$\mathcal{U}(0.1,0.99)$	
Parameter	μ_g	σ_g	μ_1	σ_1	μ_2	σ_2
Range	$\mathcal{U}(2,50)$	$\mathcal{U}(0.4,10)$	$\mathcal{U}(2,50)$	$\mathcal{U}(0.4,10)$	$\mathcal{U}(5,100)$	$\mathcal{U}(0.4,10)$

Table 4.1: Prior choices of the mass functions' hyperparameters.

To compare the different models we consider only the common parameters between them, namely: α , β , m_l and m_h . Moreover, along with m_l and m_h parameters, I also derived m_5 and m_{95} , which are the 5% and 95% percentile of the mass distribution. The percentiles were obtained from the cumulative distribution functions of the primary mass probability distribution. This choice was made to have a consistent quantity to compare, as m_l and m_h do not always have the same meaning for all the models.

In the following sections, the median values of all the parameters along with their 68% (1σ) CI are presented. The obtained primary mass distributions are also displayed. The corner plots with the two and one-dimensional distributions are shown in Appendix A. This work was also crucial to assess the performance of the nested sampling code, and perform a sizing analysis presented in the following. In this context, I underline that the analyses of the BPL, PLP and PL2P functions with the O5 dataset and of the PL2P function with the O4 data are not converged

yet, despite the long computing power assigned; thus, for these models partial parameter constraints are reported.

Constraints on the Truncated Power Law model

The first mass distribution presented is the Truncated Power Law (TPL). The results obtained by fitting the two mock catalogues data with this model are summarised in Tab. 4.2 and shown in Fig. A.1. In Fig. 4.1, the marginalised prob-

	α	β	m_l	m_h	m_5	m_{95}
O4	$3.05^{+0.14}_{-0.16}$	$3.19^{+0.58}_{-0.97}$	$6.13^{+0.21}_{-0.29}$	$75.5^{+4.4}_{-2.6}$	$6.28^{+0.21}_{-0.30}$	$24.8^{+2.1}_{-2.2}$
O5	$2.90^{+0.11}_{-0.19}$	$1.97^{+0.63}_{-0.41}$	$6.79^{+0.21}_{-0.28}$	$85.6^{+4.0}_{-2.2}$	$6.96^{+0.21}_{-0.27}$	$29.6^{+4.0}_{-1.8}$

Table 4.2: Parameter constraints of the TPL model plus the computed m_5 and m_{95} values. The results are presented with their median values and the 68% CI values.

abilities of the primary mass, obtained with the values in Tab. 4.2, are shown. The first thing that we observe is that, as expected, the TPL model is not a good representation of the fiducial mass function, because it cannot represent the over-density of objects at mass $\sim 30M_{\odot}$. However, it is capable of reproducing well the overall shape. It is possible to observe, from both Fig. 4.1 and the parameters in Tab. 4.2, that the mass function obtained from the O4 data better reproduces the slope of the primary mass distribution with respect to the one obtained with O5 that is a bit shallower. The other way around the O5 mass function has the mass upper limit consistent with the fiducial one. It is also worth nothing that, despite the improved accuracy between O4 and O5 data, we do not observe an improvement in the parameters' constraint. This is also partly expected, since the constraining power in this case is driven by the statistics of BBH event (to be able to reproduce in detail the shape of the MF), and we decided to keep the number of events equal between O4 and O5.

Constraints on the Broken Power Law model

The second mass function analysed is the BPL. The median and the 1σ interval values are collected in Tab. 4.3. The distributions of these results are shown in

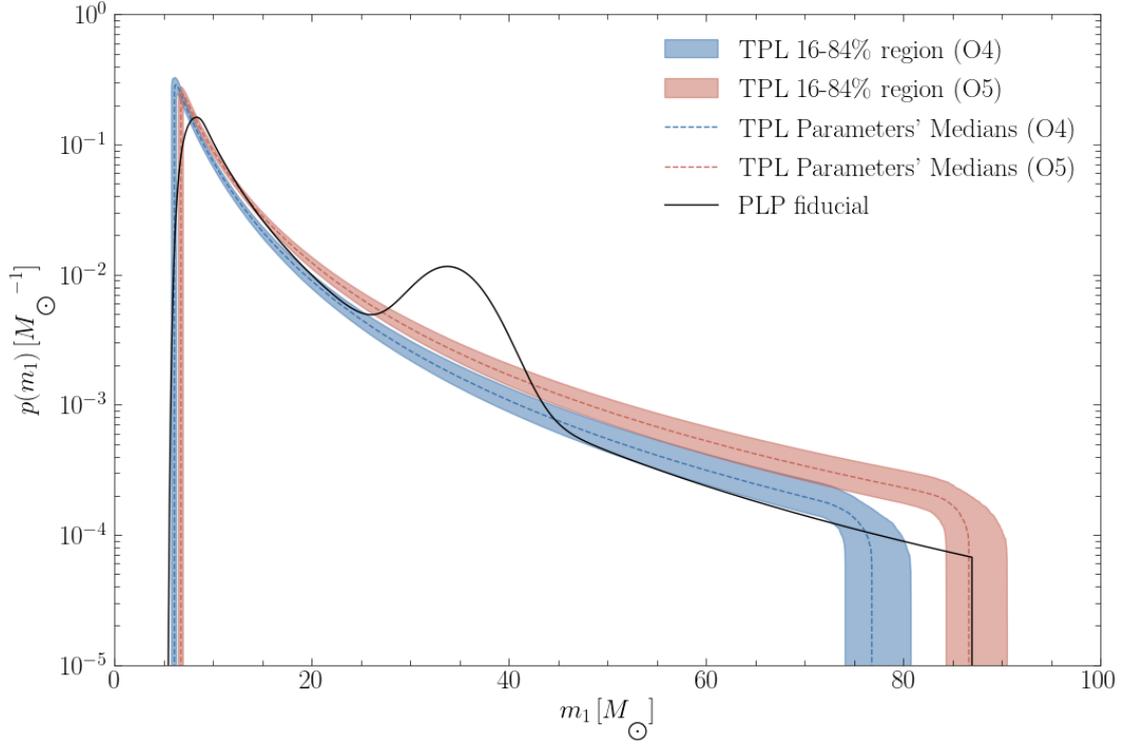


Figure 4.1: Probability distribution function of the TPL primary mass for the two catalogues obtained with the parameters in Tab. 4.2. The shaded bands represent the CI at 68%, while the solid lines represent the median values.

Fig. A.2. In Fig. 4.2, the marginalised probabilities of the primary mass, obtained

	α_1	α_2	b	β	δ_m
O4	$2.21^{+0.25}_{-0.27}$	$8.4^{+1.5}_{-1.2}$	$0.282^{+0.115}_{-0.076}$	$2.63^{+0.78}_{-0.73}$	$4.8^{+3.1}_{-2.9}$
O5	$2.07^{+0.30}_{-0.29}$	$5.99^{+0.44}_{-0.77}$	$0.341^{+0.053}_{-0.124}$	$1.62^{+0.68}_{-0.23}$	$3.4^{+3.6}_{-1.8}$
	m_l	m_h	m_5	m_{95}	
O4	$4.1^{+1.1}_{-1.2}$	127^{+45}_{-36}	$6.14^{+0.47}_{-0.62}$	$32.5^{+1.9}_{-2.3}$	
O5	$5.49^{+0.49}_{-1.80}$	104^{+52}_{-16}	$7.02^{+0.42}_{-0.58}$	$37.0^{+2.5}_{-4.5}$	

Table 4.3: Parameter constraints of the BPL model plus the computed m_5 and m_{95} values. The results are presented with their median values and the 68% CI values.

with the values in Tab. 4.3, are shown. In this case, it can be noticed that the change in the slope of the power law allows for more flexibility than the TPL

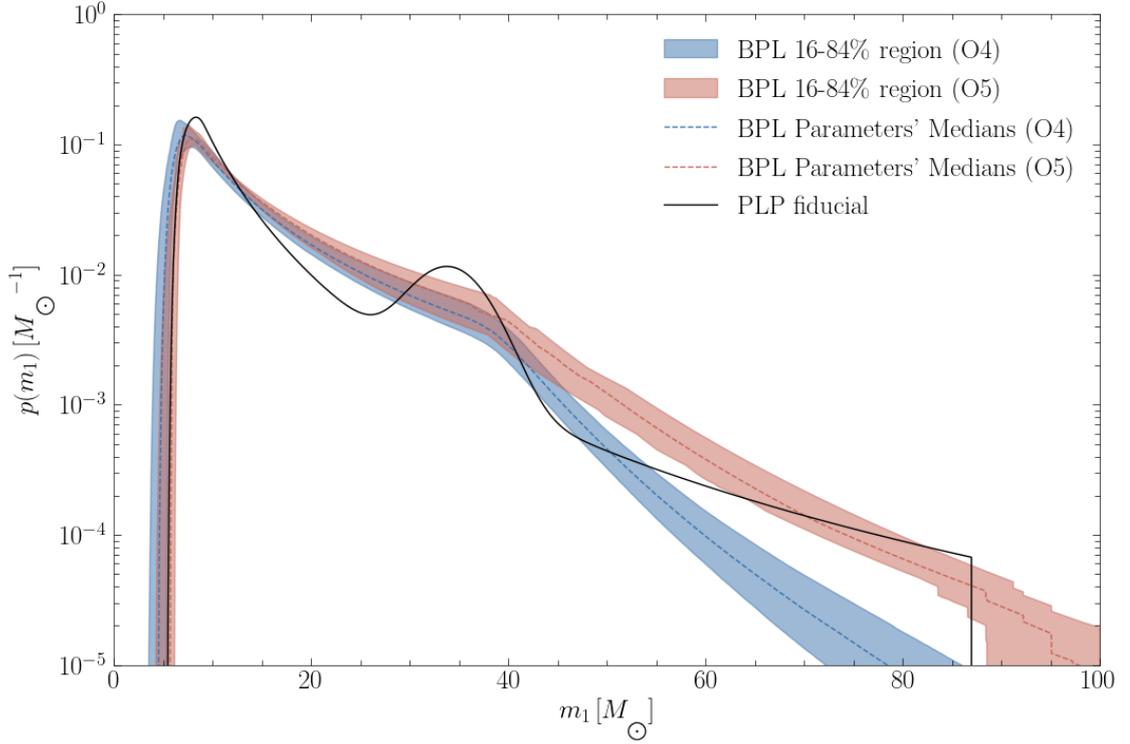


Figure 4.2: Probability distribution function of the BPL primary mass for the two catalogues obtained with the parameters in Tab. 4.3. The shaded bands represent the CI at 68%, while the solid lines represent the median values.

model, and seems to mimic the Gaussian peak. This, however, at the expense of not reproducing the real distribution slope. The m_{95} values are consistent with each other but they do not reconstruct well the steep down of the mass function at high mass values.

Constraints on the Power Law + Gaussian Peak model

The third model studied is the Power Law plus one Gaussian Peak (PLP), which corresponds to the fiducial distribution used to construct the mock catalogues. The median parameters and their 68% CI are reported in Tab. 4.4 while their distributions are in Fig. A.3. In Fig. 4.3, the marginalised probabilities of the primary mass, obtained with the values in Tab. 4.4, are shown. As the PLP is the model used as the fiducial one for the creation of the two mock catalogues, it reproduces in details the fiducial mass function, shown with the black line in

	λ_p	α	β	δ_m	μ_g	σ_g
O4	$(19.3^{+16.1}_{-6.5}) \times 10^{-3}$	$3.90^{+0.32}_{-0.29}$	$2.16^{+0.79}_{-0.68}$	$7.3^{+1.6}_{-2.5}$	$33.1^{+1.9}_{-3.1}$	$5.2^{+2.3}_{-1.7}$
O5	0.041 ± 0.022	$3.42^{+0.13}_{-0.18}$	$1.81^{+0.59}_{-0.18}$	$3.9^{+4.9}_{-1.6}$	$34.93^{+0.72}_{-0.86}$	$4.00^{+1.41}_{-0.90}$
	m_l	m_h	m_5	m_{95}		
O4	$4.33^{+0.56}_{-0.50}$	102^{+62}_{-20}	$6.51^{+0.37}_{-0.42}$	$34.1^{+1.6}_{-2.0}$		
O5	$5.33^{+0.76}_{-1.65}$	114^{+25}_{-22}	6.88 ± 0.52	$36.31^{+0.64}_{-0.91}$		

Table 4.4: Parameter constraints of the PLP model plus the computed m_5 and m_{95} values. The results are presented with their median values and the 68% CI values.

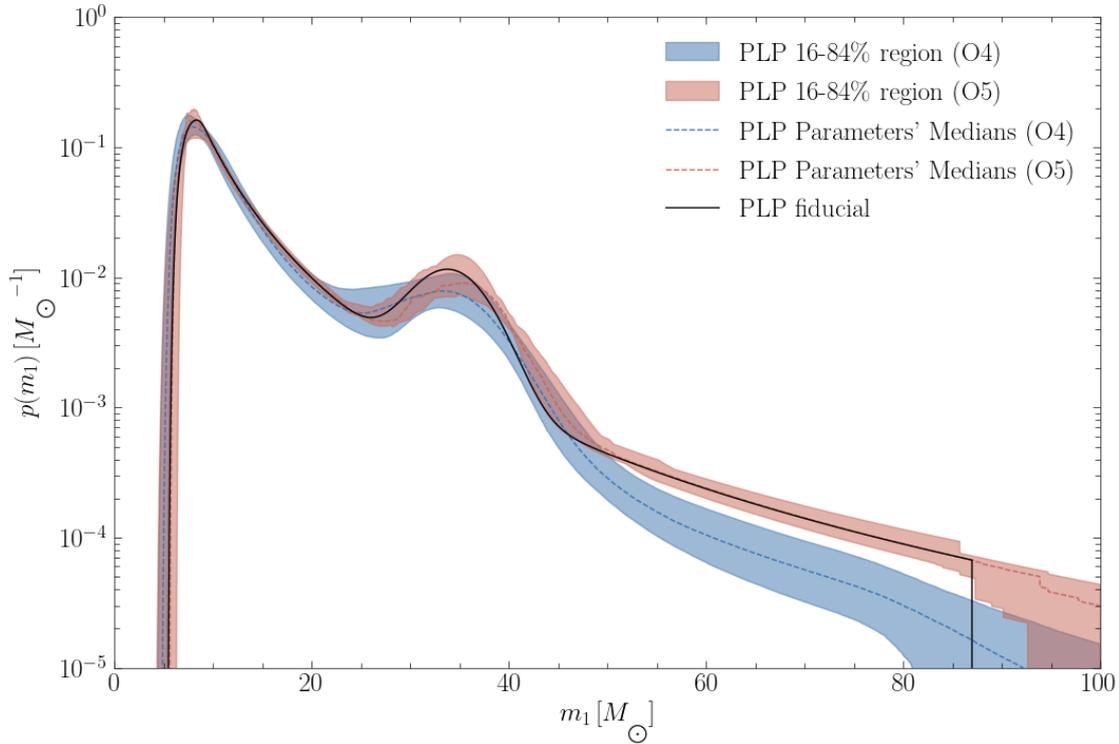


Figure 4.3: Probability distribution function of the PLP primary mass for the two catalogues obtained with the parameters in Tab. 4.4. The shaded bands represent the CI at 68%, while the solid lines represent the median values.

Fig. 4.3, as expected. Furthermore, it is possible to observe an improvement of the fit using the O5 data, even if the analysis has not yet converged. Indeed with this dataset, all the parameters are consistent with the fiducial ones. Moreover, excluding δ_m and m_l distributions which remain flat, the 1σ CIs get narrower

with respect to those obtained with O4: this is very promising since it shows the relative improvement that we are expecting just with 100 BBH events from O4 to O5, where the accuracy on the determination of the Gaussian peak increase by a factor 3. Comparing the m_5 and m_{95} values it is possible to say that they are consistent with the TPL and BPL values.

Constraints on the Power Law + 2 Gaussian Peaks model

The last analysed mass distribution function is the power law plus two Gaussian peaks. The distributions of the parameters are shown in Fig. A.4, while the median values and their 1σ error are collected in Tab 4.5. In Fig. 4.4, the marginalised

	λ_p	λ_1	α	β	δ_m
O4	$0.120^{+0.335}_{-0.087}$	$0.53^{+0.32}_{-0.35}$	3.77 ± 0.51	$2.35^{+1.19}_{-0.89}$	$5.3^{+3.1}_{-3.2}$
O5	$0.100^{+0.218}_{-0.058}$	$0.54^{+0.34}_{-0.35}$	$3.29^{+0.40}_{-0.39}$	$1.77^{+0.87}_{-0.67}$	$3.4^{+2.8}_{-2.1}$
	μ_1	σ_1	μ_2	σ_2	
O4	$25.9^{+8.4}_{-18.3}$	$5.7^{+2.9}_{-2.6}$	$28.1^{+6.2}_{-18.6}$	$6.0^{+2.7}_{-2.5}$	
O5	$32.4^{+3.4}_{-24.1}$	$4.7^{+2.3}_{-1.9}$	$33.4^{+3.1}_{-23.5}$	$4.6^{+2.6}_{-1.7}$	
	m_l	m_h	m_5	m_{95}	
O4	$4.48^{+0.85}_{-0.97}$	118^{+54}_{-32}	$6.39^{+0.58}_{-0.68}$	$33.7^{+2.3}_{-3.0}$	
O5	$5.46^{+0.81}_{-1.39}$	115^{+54}_{-25}	$6.91^{+0.47}_{-0.67}$	$36.7^{+2.0}_{-2.2}$	

Table 4.5: Parameter constraints of the PL2P model plus the computed m_5 and m_{95} values. The results are presented with their median values and the 68% CI values.

probabilities of the primary mass, obtained with the values in Tab. 4.5, are shown. Looking at both Fig. 4.4 and the results obtained for μ_1 and μ_2 , it is possible to notice that the flexibility of this model allows the fit to reproduce the fiducial model, producing two degenerate peaks at the same position of the fiducial one. This is very reassuring, since it suggests that this model might also be sufficiently sensitive to detect further deviation in the distribution of the data. Providing a prior that covers the the lower half part of the masses range for the μ_1 distribution and almost all the range for the μ_2 distribution, a second weak peak centred at a mass near the lower mass limit arises. Actually, I have also performed an analysis

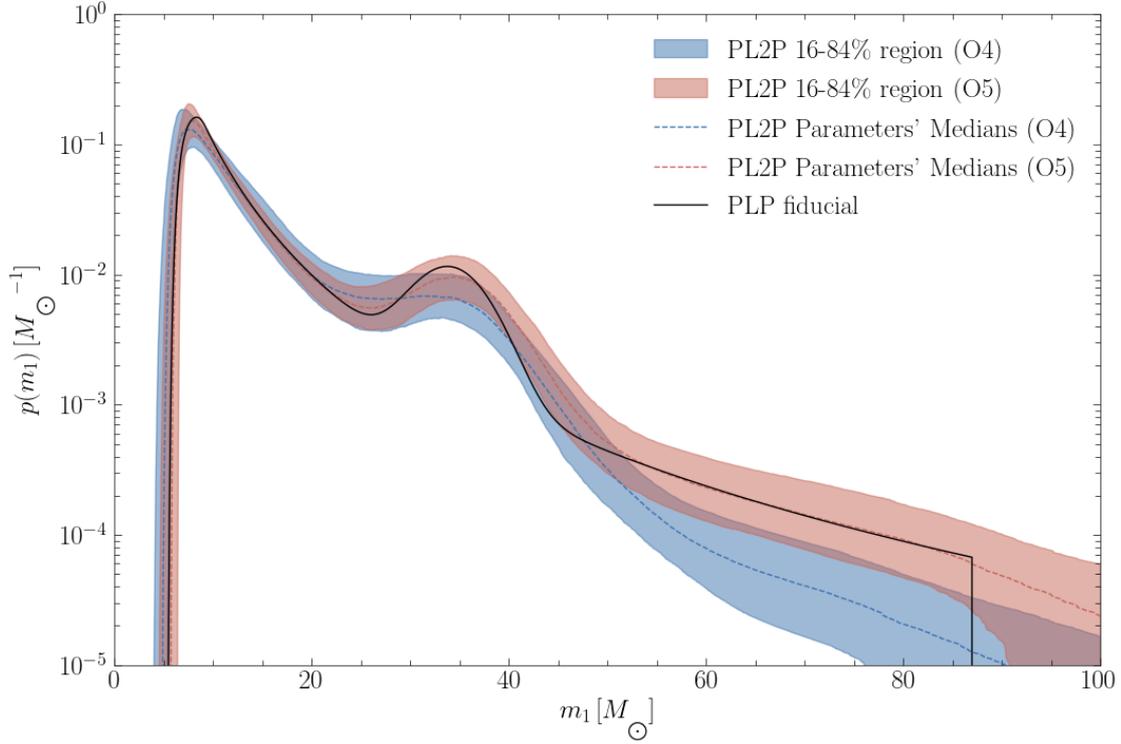


Figure 4.4: Probability distribution function of the PL2P primary mass for the two catalogues obtained with the parameters in Tab. 4.5. The shaded bands represent the CI at 68%, while the solid lines represent the median values.

on this model, for both catalogues with a more restrictive prior on the second Gaussian peak, centred around $10M_{\odot}$ as suggested in Abbott et al. (2021) and done in Mancarella et al. (2022a). This has been tested to see if informing more the posterior on the position of the two peaks would retrieve some artificial structure, questioning hence the constraints obtained in Abbott et al. (2023b). What I have found is that the double peak in the μ_1 and μ_2 distributions disappears and μ_2 corresponds to m_l . This means that the likelihood implemented in CHIMERA, see eq. 2.10, which is the same used in Abbott et al. (2023b), it is not affected by systematic effects that generate non real substructures. We can thus conclude that the second peak found in Abbott et al. (2023b) is a real substructure and does not arise from systematic errors.

Both the analyses with the O4 data and O5 data are not finished yet; at the moment, it is not possible to observe strong differences between the two catalogues

constraints. Comparing the m_5 and m_{95} values it is possible to say that they are consistent with those of the BPL and PLP .

Comparing the performances of different posterior samplers

To characterise the `dynesty` nested sampler implemented in `CHIMERA`, I compared its results and performances against those obtained with the `emcee` sampler in the work of Borghi et al. (2024), considering the same PLP mass distribution. In Fig. 4.5 the marginalised probabilities of the primary mass parameters obtained using the two samplers, for both the mock catalogues, are shown. In the corner plots in Fig. A.5 and Fig. A.5 the parameter distributions are compared.

Looking at all the plots and the data, it is possible to conclude that in the case of O4 the results are highly consistent with each other. The parameters obtained with `dynesty` are all within the 1σ CI of those obtained with `emcee`. The only parameter that has a slightly different distribution is the power-law slope of the secondary mass distribution β . In both cases the fiducial value is outside the 1σ CI of the retrieved values. Moreover, computing their difference as the quadrature-sum of their sigma values, the obtained value differs of $\sim 1\sigma$.

In the case of the O5 results, it is not possible to perform a direct comparison as the analysis with `dynesty` has not reached the converging criteria yet. Nevertheless, looking at Fig. 4.5 it seems that the obtained partial constraints are in agreement with those of the full `emcee` analysis. Moreover, Fig. A.6 suggests that the partial distributions are starting to recover the `emcee` ones for most of the parameters.

4.1.4 Sizing tests

While studying the different models, I kept track of the running times of all the analyses to check the feasibility of these kinds of analyses with `CHIMERA` with the upcoming catalogues of gravitational waves events and galaxies.

The computational times in Tab. 4.6 are reported as CPU time, obtained by dividing the total amount of sampling time by the number of CPUs used, and CPU Hours, obtained by multiplying the hours of sampling time by the number of CPUs used. It can be seen immediately that, as expected, models with higher complexity and a larger number of parameters need more computational time to

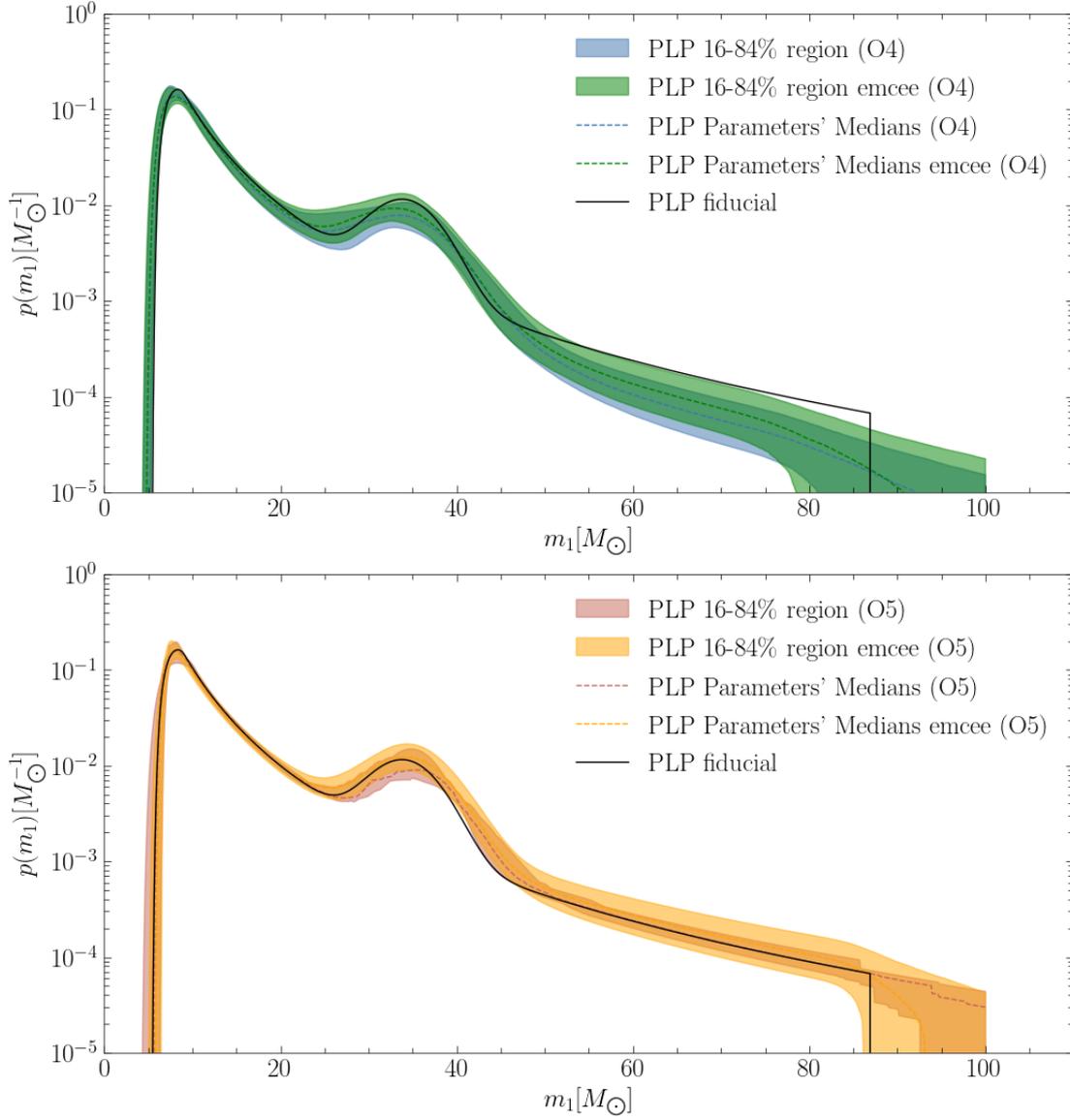


Figure 4.5: The PLP primary mass function obtained with *dynesty* and *emcee*. In the top panel are represented the results obtained with the O4 catalogue; in blue the *dynesty* constraints, in green the *emcee* ones. In the bottom panel are represented the results obtained with the O5 catalogue; in red the *dynesty* constraints, in orange the *emcee* ones.

reach the convergence of the nested sampling algorithm. At the time of writing, some of the analyses are not finished yet, as reported in Tab. 4.6 with the $>$ sign. Moreover, I also noticed that for some cases from a certain point on the sampler

O4	CPU time [s/n_{CPU}]	CPU hours [$h \times n_{CPU}$]	O5	CPU time [s/n_{CPU}]	CPU hours [$h \times n_{CPU}$]
TPL	290.3	1260	TPL	32729.2	2327.36
BPL	1332.0	5781.25	BPL	>12441.6	>54000
PLP	10462.3	29062	PLP	>88992.0	>247200
PL2P	>57369.6	>249000	PL2P	>58752.0	>255000

Table 4.6: Computational time of every model reported as CPU time and CPU hours, in the left table in the case of the O4 catalogue and in the right table in the case of the O5 catalogue. The data have dimensions [s/n_{CPU}] and [$h \times n_{CPU}$]. The models that have CPU time and CPU hours with the $>$ sign are still running at the time of writing.

was not evolving anymore. The reason for that is still under investigation. Even though a priori the analyses with the O5 data are thought to be faster than those with the O4 dataset, due to the higher accuracy in the gravitational wave events localisation, it seems that, at least with `dynesty`, they necessitate much more computational time with respect to their corresponding O4 analyses. This is may be caused by the fact that the used `dynesty`'s settings do not perform well with event posteriors as narrow as those of O5.

4.1.5 Disentangling different BBH mass function models with O4 and O5 data

Thanks to the usage of `dynesty`, I managed to obtain the Bayesian evidence for every mass function model. This allows me to compare and assess which model the data prefers. Furthermore, knowing the log-likelihood of every sample allows me to compute the DIC of the models providing another way to compare them. Finally, I estimated the PPCs as a further term of comparison between them. For each of this methods I have developed a code which is aimed to be integrated in CHIMERA.

It is important to remember that the data catalogues have been built using the PLP model as fiducial distribution, which is hence expected to be the best model. The Bayes factors are summarised in Tab. 4.7, the DIC values in Tab. 4.8, while the PPCs are shown for O4 in Fig. 4.6 and for O5 in Fig. 4.7. In Tab. 4.7, three

different values are displayed: the first is the logarithm of the Bayesian evidence estimated with *dynesty*, the second and the third are the Bayes factor and its logarithm, that are computed using as reference model the one with the lowest Bayesian evidence, i.e. the BPL for O4 and the PLP for O5. In Tab. 4.8 two values for each model are displayed: the absolute value of the DIC (first column), and the difference computed as reference value the lowest one, i.e. the PL2P for O4 and the PLP for O5 (second column). In Fig. 4.6 and Fig. 4.7 are shown four panels representing the PPCs of each mass function. The mass function is specified in the title of the panel along with the respective logarithm of the Bayes factor ($\log(\mathcal{B})$) and the Δ DIC. The dark coloured bands are the 90% CI of the observed population's CDF, while the light coloured bands are the 90% CI of the predicted population's CDF.

Model (O4)	$\log(z)$	\mathcal{B}	$\log(\mathcal{B})$
TPL	-375.55	3.72×10^{-8}	-17.11
BPL	-358.44	1	0
PLP	-358.96	0.60	-0.51
PL2P	-359.67	0.29	-1.22
Model (O5)	$\log(z)$	\mathcal{B}	$\log(\mathcal{B})$
TPL	-170.51	1.32×10^{-7}	-15.84
BPL	-164.49	5.44×10^{-5}	-9.82
PLP	-154.67	1	0
PL2P	-156.37	0.18	-1.7

Table 4.7: Evidences and Bayes Factors. The three columns represent respectively the logarithm value of the Bayesian evidence estimated with *dynesty*, the Bayes factor, and its logarithm value.

Model (O4)	DIC	Δ DIC	Model (O5)	DIC	Δ DIC
TPL	735.42	-68.68	TPL	325.64	-52.17
BPL	694.90	-28.16	BPL	284.31	-10.84
PLP	691.60	-24.86	PLP	273.47	0
PL2P	666.74	0	PL2P	283.21	-9.74

Table 4.8: DIC values for the different models, where Δ DIC is defined as the difference between the considered model and the one with the smallest DIC.

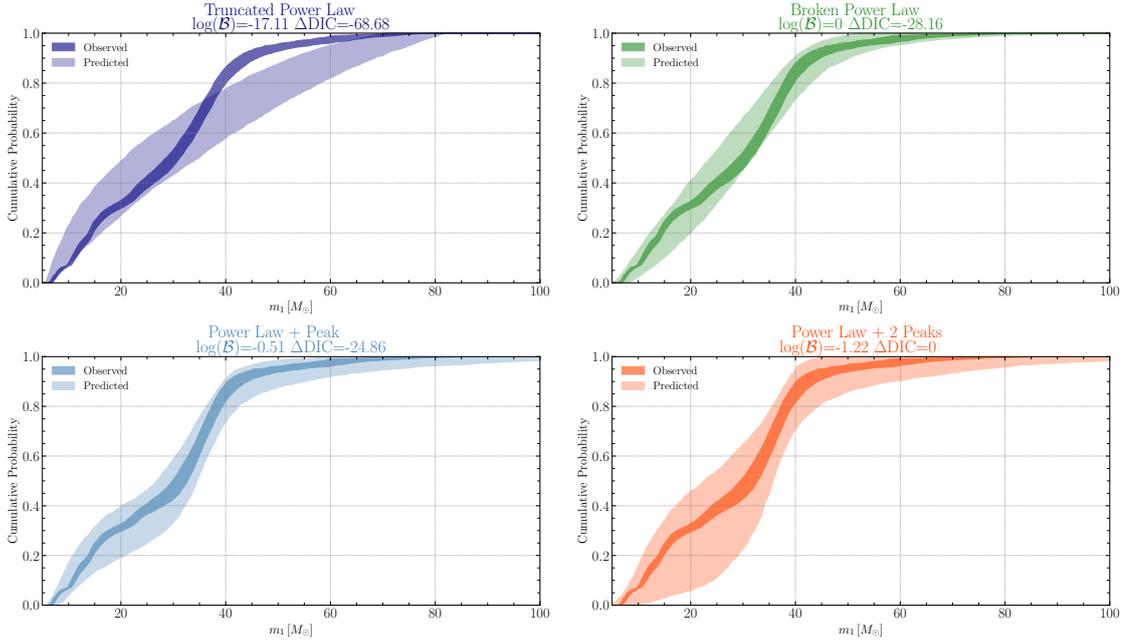


Figure 4.6: Posterior Predictive Checks of the four studied mass functions with the O4 data. The dark coloured bands represents the 90% CI of the observed population’s CDF, the light coloured bands represents the 90% CI of the predicted population’s CDF. In the panels’ titles are displayed the logarithm of the Bayes factor and the ΔDIC . In the top left panel is shown the TPL’s PPC; in the top right panel is shown the BPL’s PPC; In the bottom left panel is shown the PLP’s PPC; in the bottom right panel is shown the PL2P’s PPC.

In the case of the O5 catalogue, all the model selection criteria agree with each other saying that the best fitting model is the power law plus Gaussian peak, as expected. Indeed, both the Bayesian evidence and the DIC are the lowest values among those of the other models and in the PPC the observed curve is well contained in the expected one. It is interesting to notice that the PL2P presents a very small difference with respect to the PLP: this is expected since, as discussed in Sect. 4.1.3, also that model basically converged to the PLP. On the other hand, O5 data allows to significantly discard the BPL and TPL models, since they show a very low Bayes factor and a $\Delta\text{DIC} < -10$ that in the Jeffreys’ scale is rated as a ”decisive” evidence against the model with higher criterion value (Liddle, 2007). Furthermore, their PPCs show that the 90% CI of the observed population’s CDF is not contained and reproduced by the predicted population’s one.

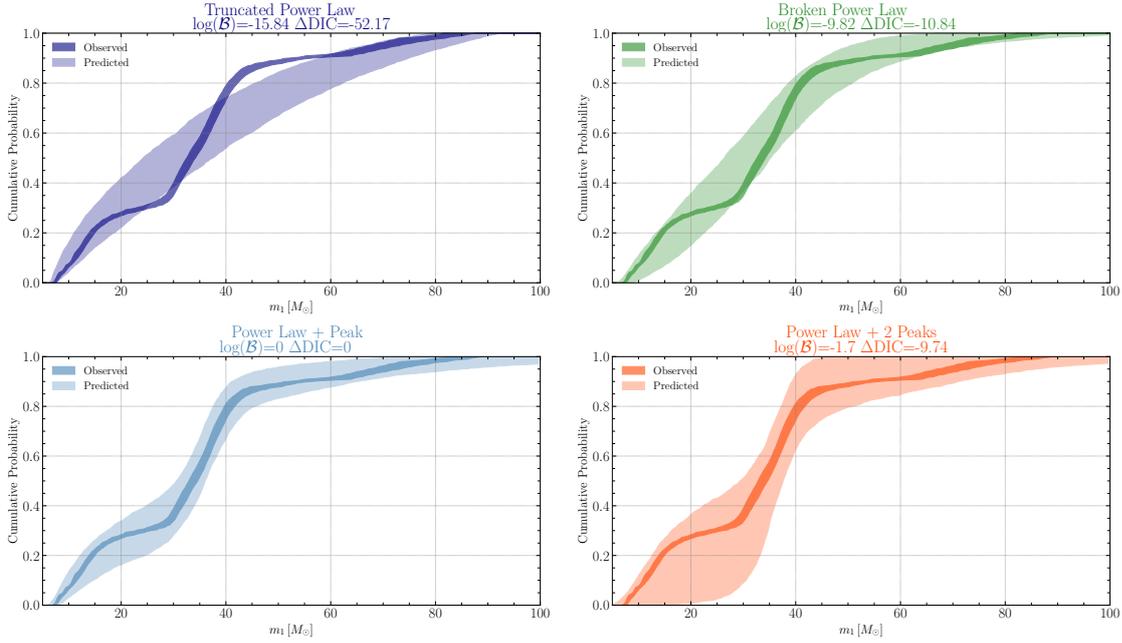


Figure 4.7: Posterior Predictive Checks of the four studied mass functions with the O4 data. The dark coloured bands represents the 90% CI of the observed population’s CDF, the light coloured bands represents the 90% CI of the predicted population’s CDF. In the panels’ titles are displayed the logarithm of the Bayes factor and the Δ DIC. In the top left panel is shown the TPL’s PPC; in the top right panel is shown the BPL’s PPC; In the bottom left panel is shown the PLP’s PPC; in the bottom right panel is shown the PL2P’s PPC.

This is extremely interesting since, at the moment, we have only used 100 high SNR BBH in O5, while the analysis could be extended to a significantly higher statistic. This represents a significant step forward with respect to the analysis of GWTC3 (Abbott et al., 2023a), since with current data only the TPL model could be robustly discarded. On the other hand, for the O4 catalogue the indications of the various criteria implemented are not in agreement. In particular, both for the DIC there is a preference for the PL2P model, while according to the Bayesian evidence, the BPL is slightly favoured with respect to the PLP and the PL2P. This is strange, since looking at the PPCs we could conclude, oppositely to the Bayes factor, that the BPL is not the favoured model. This issue is still under investigation. Instead, it is possible to observe that for both the PLP and PL2P models the inferred population well reproduce the observed one indicating that

both are good model even if the PLP seems slightly better as the predicted band is more close to the observed one with respect to the PL2P. This suggests that it is very likely that we will need data more GW events at the sensitivity of O4 or data with the O5 accuracy to be able to have clear hints of which is the best BBH mass function. In general, it is possible to confirm that the TPL can be safely and significantly disfavoured, as already said in Abbott et al. (2021) and Abbott et al. (2023b).

4.2 Assessing the dependence of the results on the catalogues assumptions

In the previous Sect., it was explored what future GW data will be able to tell us, in terms of our capability to discriminate between different BBH mass functions.

Here, I perform a more detailed analysis on how particular assumptions or conditions of our galaxy and GW catalogs might impact on the results.

With the upcoming gravitational wave data and galaxy surveys, it is important to find a way to optimise data usage and to understand a priori possible biases in the analyses performed with CHIMERA. With this aim, I have in particular studied how results are affected by varying several parameters, namely:

- the error on the redshift z_{err} ;
- the bandwidth of the 3D weighted KDE used to compute the integral in eq. 2.11.
- the number of GW events considered, applying a cut based on the number of galaxies per localisation volume (i.e. on their quality);

Given that in Sect. 4.1.4 I assessed that the time needed for a full MCMC analysis would be prohibitively large for this task, it was decided to only perform a one-dimensional analysis, fixing each time all parameters except one. This is equivalent to slicing the posterior instead of marginalising over the other parameters, but for the purpose of this exercise this will give us some differential results that could be useful for future potential follow-up analyses.

In the following sections, I discuss the various findings in detail.

4.2.1 Dependence on z_{err}

Intending to use gravitational waves as dark sirens, it is important to know how much the constraints on the parameters improve or degrade depending on the redshift accuracy of the galaxy catalogue. In Borghi et al. (2024), this has been explored by using only two different z_{err} , one referred to as spectroscopic ($z_{spec} = 0.001$) and one referred to as photometric ($z_{phot} = 0.05$). In the analysis of Sect. 4.1.3, all the results have been derived in the best-case scenario of a spectroscopic redshift.

Here, I explore how the constraints on H_0 change using six different values of z_{err} , keeping z_{spec} as the best uncertainty and going above the value of z_{phot} . I then used CHIMERA and derived the constraints on H_0 considering both the O4-like and O5-like catalogues and a galaxy catalogue whose redshift errors have been degraded on the grid $z_{err} = [0.001, 0.1]$. To derive the posterior distributions of H_0 in these cases, all parameters except H_0 were fixed at their fiducial values, while H_0 varied in the range $60 < H_0 < 150$ [km/s/Mpc]. The resulting posterior distributions are shown in Fig. 4.8 and 4.9. While in Fig. 4.10 the estimated values with their error bars and the percentage errors against the z_{err} values are plotted. In Fig. 4.8 the posterior distributions obtained with the same catalogue are represented in the same panel with different colours for different redshift accuracy. It is also plotted the fiducial value with a black line to better observe the deviations from it. In Fig. 4.9 the same distributions are shown, but with different z_{err} in different panels in order to compare the posterior obtained with the two catalogues. The two panels in Fig. 4.10 represent respectively the estimated H_0 posterior values with their error bars and their percentage errors plotted against the z_{err} values for both catalogues. The error bars represent the 68% CI of the resulting values. The percentage errors are computed by dividing the difference between the 84th and the 16th percentile values by two times the median value, see eq. 4.5.

$$err H_0 = \frac{H_0^{84th} - H_0^{16th}}{2H_0^{median}} \quad (4.5)$$

It is interesting to notice that, both for O4-like and O5-like configurations, for very large redshift errors a bias appears in the determination of H_0 , shifting it to

larger values. From Fig. 4.8 and 4.9 it can be seen that up to a certain threshold ($z_{err} < 0.02$) the H_0 constraint is mostly unbiased (as also obtained in Borghi et al. 2024) and with a lower uncertainty, but then both the value and the error drift to higher values.

I investigated the possible causes of this bias, trying to perturb the redshift catalogue according to the redshift uncertainty as done in Borghi et al. (2024), but it persists in all the analyses. Qualitatively, the cause of this effect could be traced to how the redshift distribution is treated in the code when it is convolved with an expected "uniform in comoving volume" distribution (see eq. 2.6), but this needs further investigations. Indeed, in the case of larger values of z_{err} we approach the spectral siren case, that in Borghi et al. (2024) is shown to produce H_0 distributions biased towards values smaller than the fiducial one. In general,

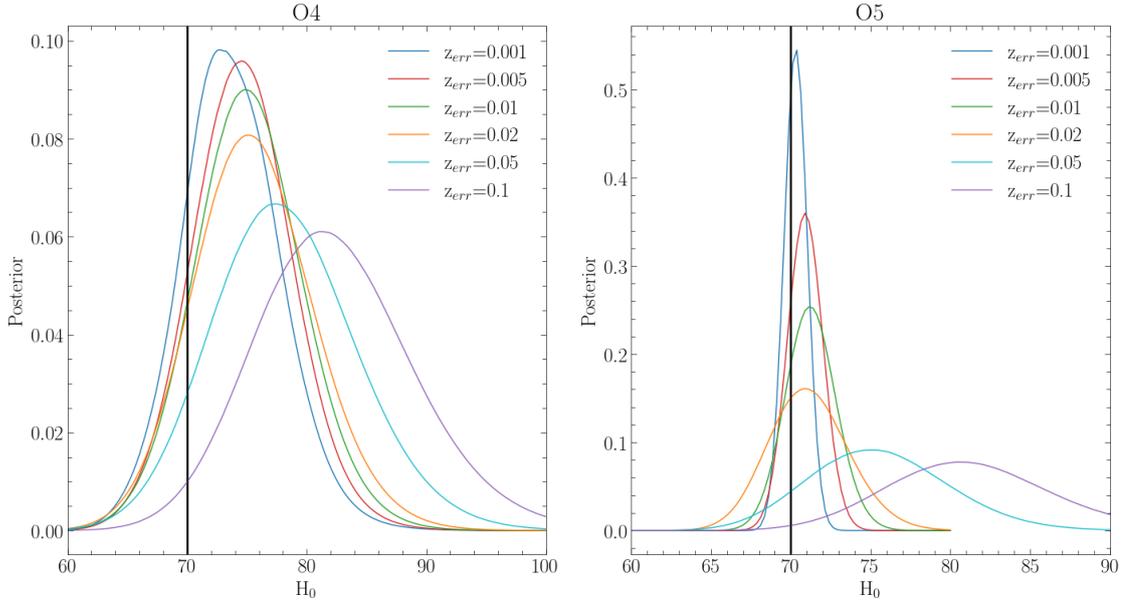


Figure 4.8: H_0 posterior distributions. Different colours represent distributions obtained with the different values of z_{err} . The black vertical lines indicate the fiducial values of the two catalogues, i.e. $H_0 = 70$ [km/s/Mpc]. The distributions obtained from the O4 catalogue are represented on the left panel. In the right panel are represented the distributions obtained from the O5 catalogue.

the posterior distributions obtained with the O5 catalogue show a smaller effect than their equivalents obtained with O4 data, remaining closer to the fiducial value

4.2. DEPENDENCE OF THE RESULTS ON THE ASSUMED CATALOGS 87

and having narrower uncertainties, as it can be seen from Fig. 4.8 and Fig. 4.9. This is clearer from Fig. 4.10 where it can be immediately seen that the best

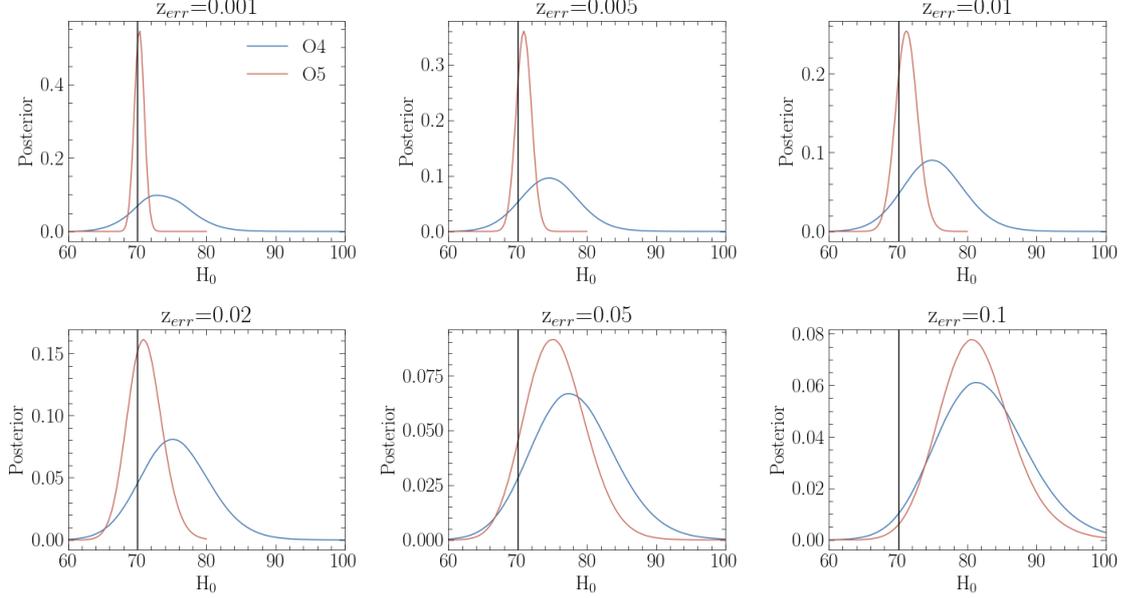


Figure 4.9: H_0 posterior distributions. Every panel shows the result obtained with both O4 and O5 using different z_{err} . The black vertical lines indicate the true value, i.e. $H_0 = 70$ [km/s/Mpc].

values obtained from O4 are higher than those of O5. It can be noticed that the O5 results are less sensitive to the worsening of z_{err} even if with error values greater than $z_{err} = 0.02$ the values steeply increase. Indeed, with $z_{err} < z_{phot}$ the values remain constant around $H_0 \simeq 71$ [km/s/Mpc], while with the worst redshift accuracy $z_{err} = 0.1$ the value rise up to $H_0 \simeq 81$ [km/s/Mpc], being consistent with the O4 value. Looking instead at the percentage errors, it can be observed that with the best redshift accuracy, in the case of O5, it is reached a 1% accuracy on the H_0 estimation, which is $\sim 4.5\%$ lower than the one obtained with O4. This discrepancy remains until $z_{err} = z_{phot}$ is reached and this difference drops down to just 2%. Moreover, comparing these results with those obtained with the MCMC analyses of Borghi et al. (2024) it is possible to observe that with the best uncertainty on z the values are consistent with each other, while in the case of z_{phot} the 1-dimensional results have a lower percentage error. This can be attributed to the fact that in the 1-dimensional analyses all the hyperparameters except H_0 are

fixed to their fiducial values.

4.2.2 Dependence on the smoothing of the GW localisation volumes

The bandwidth of a Gaussian KDE is a free parameter of strong importance for the determination of the data probability density distribution. Indeed, it determines the smoothness of the distribution as it basically represents the standard deviation of the Gaussian curves assigned to every data sample. The two extreme cases are when it is equal to 0 or to ∞ : in the first case, the estimation consists of a summation of n delta functions positioned at the coordinates of the examined samples, while in the second case a Gaussian centred in the mean of the data samples is obtained. Thus using a too small bandwidth the obtained probability distribution will show too many artefacts, while using a too large bandwidth all the substructures are cancelled away. I probed how the H_0 distribution changes using seven different values of the bandwidth, namely 0.01, 0.05, 0.1, 0.3, 0.5, 1.0, and 1.5. The central value, 0.3, is the one used in Borghi et al. (2024). In Fig. 4.11 and Fig. 4.12 it is shown how the different bandwidths influence the shape of the estimated probability distributions. In each panel is represented the event samples distribution in the form of a histogram over which are overlapped the respective 3D KDE estimated distribution, one for each pixel in the localisation area. Every panel corresponds to a different value of the KDE bandwidth. It is possible to see that with low values of the bandwidth, the event samples distribution is not sampled properly as the estimated distributions have too high narrow peaks. On the other hand, with high values of the bandwidth, the real distribution is over smoothed by the KDE, again returning a broader (and wrong) estimated distribution. It can be seen that for the O5 events the best value among those studied here is 0.3, the one used in Borghi et al. (2024); for O4, instead, both 0.3 and 0.5 seem to reproduce well the data distribution. Thus, in general, there may be not a single best value for the KDE bandwidth as different bandwidth values may be the best suited for different events.

From Fig. 4.13 it can be observed that in the case of the O5 analysis what changes the most is the fact that the distributions are significantly narrower; in particular,

with the two lowest values of the bandwidth, the distributions result highly peaked. On the other hand, looking at the O4 results it can be observed that with the three lowest bandwidth values, the H_0 probability distribution is highly disturbed. In addition, increasing the bandwidth the distributions are centred around values closer and closer to the fiducial one. Inspecting Fig. 4.14, we can also see that all the best values obtained in O5 are consistent with each other, while those of O4, as already commented, move away from the fiducial value. On the other hand, as expected, the percentage errors increase in both cases; this is due to the enlargement of the posterior distribution that is directly connected to the oversmoothing with high values of the bandwidth. Considering only the posteriors of H_0 , it is not clear which is the best bandwidth value to choose since, with the exception of the lowest values, the distributions are well shaped. Thus it is important to see how, actually, the estimated distributions with different bandwidths reproduce the data samples. In general, with this two GW catalogues the selection of 0.3 as KDE bandwidth value for the analysis is justified but, all in all, it is necessary to optimise a similar analysis in the context of the future GW analyses.

4.2.3 Dependence on the subsampling

With the upcoming observation runs we expect to have an always increasing number of gravitational waves events, with 3G network we expect to have $O(10^5)$ events (see Branchesi et al. (2023)). Therefore, in order to optimise the use of the facilities for the possible analyses, it is important to understand what kind of data is driving the results and whether it is possible to reduce the computational time by discarding part of the less informative events, potentially with minimal or no impact on the analysis.

To explore this, I have studied different subsamples of the O5 catalogue, which I recall comprises only 100 events, considering the two regimes of z_{spec} ($z_{err} = 0.001$) and z_{phot} ($z_{err} = 0.05$). It was decided to explore different cuts depending on the number of galaxies per localisation volume $N_{gal,vol}$, which is also a proxy of the quality of the data. Other cuts were also considered (based on the SNR of the GW event or on the accuracy of the determination of the luminosity distance), but we preferred this one since it is directly linked to the best GW event that can be

obtained, and we wanted to explore if the "golden events" with few galaxies in the localisation volume were the ones dominating the constraining power. The different cuts, with the corresponding number of GW events associated, are summarised in Tab. 4.9. In the case of the spectroscopic redshift, it can be seen from the left panel

Sample cuts	N_{events}
$N_{gal,vol} < 10$	3
$N_{gal,vol} < 40$	8
$N_{gal,vol} < 70$	12
$40 < N_{gal,vol} < 500$	32
$N_{gal,vol} < 100$	17
$N_{gal,vol} < 500$	40
$N_{gal,vol} < 1000$	66

Table 4.9: Subsample used for this analysis and their total number of GW events.

of Fig. 4.15 that all the values are consistent with each other and in particular with the fiducial value and the full MCMC analysis of Borghi et al. (2024). It can be noticed that keeping only the few best-localised events does not improve the results. On the other hand, looking at the subsample with $40 < N_{gal,vol} < 500$, it can be said that excluding both the best and worst localised produces a worse result, suggesting that very well localised events give a significant contribute to the parameter estimation. Nevertheless, as it can be observed in the right panel of Fig. 4.15, even discarding the best events, those with less than 40 galaxies in the localisation volume, the uncertainty on H_0 is not worse than the one obtained when considering them.

This is an important result because it suggests a potential strategy for future analysis. In particular, discarding the GW events with a number of galaxies in the localisation volume larger than 10^3 does not have a significant impact on the analysis, suggesting that one could focus only on the best events. On the other hand, the result obtained in the case $40 < N_{gal,vol} < 500$ suggests that even if the "golden events" with just a few galaxies in the localisation volume were not observable, still the cosmological constraints can be retrieved reducing only slightly the accuracy. It is important to remark that this conclusion is valid only if one is interested in the cosmological parameters as, for example, in the case of the mass function parameters more are the considered events better are the resulting

constraints.

In the case of the photometric redshift, it can be observed from Fig. 4.16 that it is more relevant to have a sufficient number of events instead of a few well-localised events. Studying also the complementary subsample, I have seen confirmation of this result. This can be attributed to the fact that in the case of poor accuracy on the galaxies' redshift, the information is dominated by the gravitational wave events number and the information on the source population such as the BBH mass function.

Finally, comparing these results with the two obtained with the full MCMC analysis of Borghi et al. (2024) we can conclude that in the z_{spec} case the results uncertainties are compatible, while in the case of z_{phot} the constraints on H_0 are better in my analysis. This can be attributed to the fact that among all the fixed parameters there is especially μ_g which, as it can be seen in figure 8 of Borghi et al. (2024), has a strong degeneracy with H_0 that is broken by assigning to it its fiducial value.

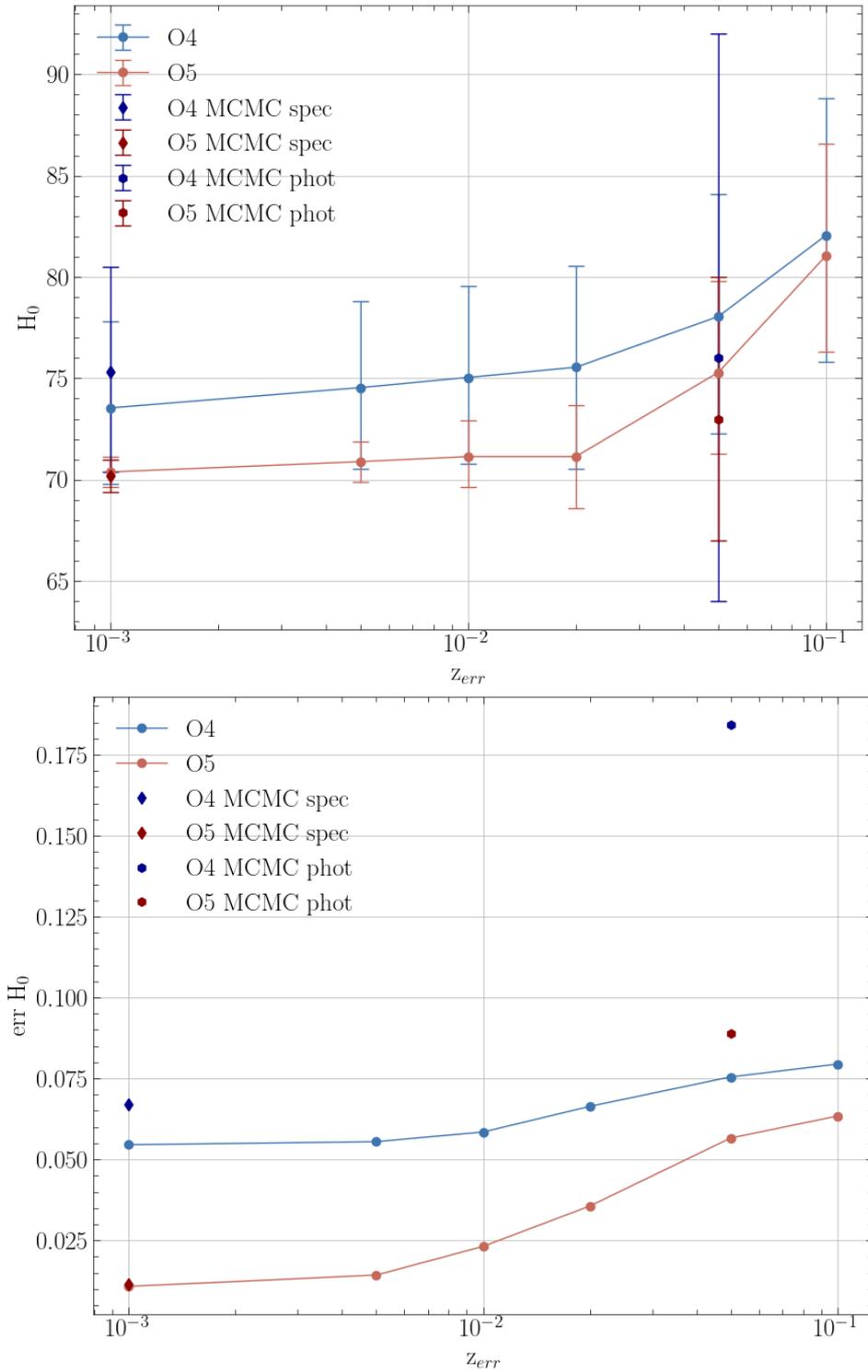


Figure 4.10: In the top panel are represented the median values of H_0 with their errorbars obtained with different z_{err} . The single dark points represent the values obtained with the complete MCMC analyses of Borghi et al. (2024). In the bottom panel are represented the percentage errors of the values in the top panel. In both panels the blue points are obtained from the O4 dataset while the red ones from the O5 dataset.

4.2. DEPENDENCE OF THE RESULTS ON THE ASSUMED CATALOGS 93

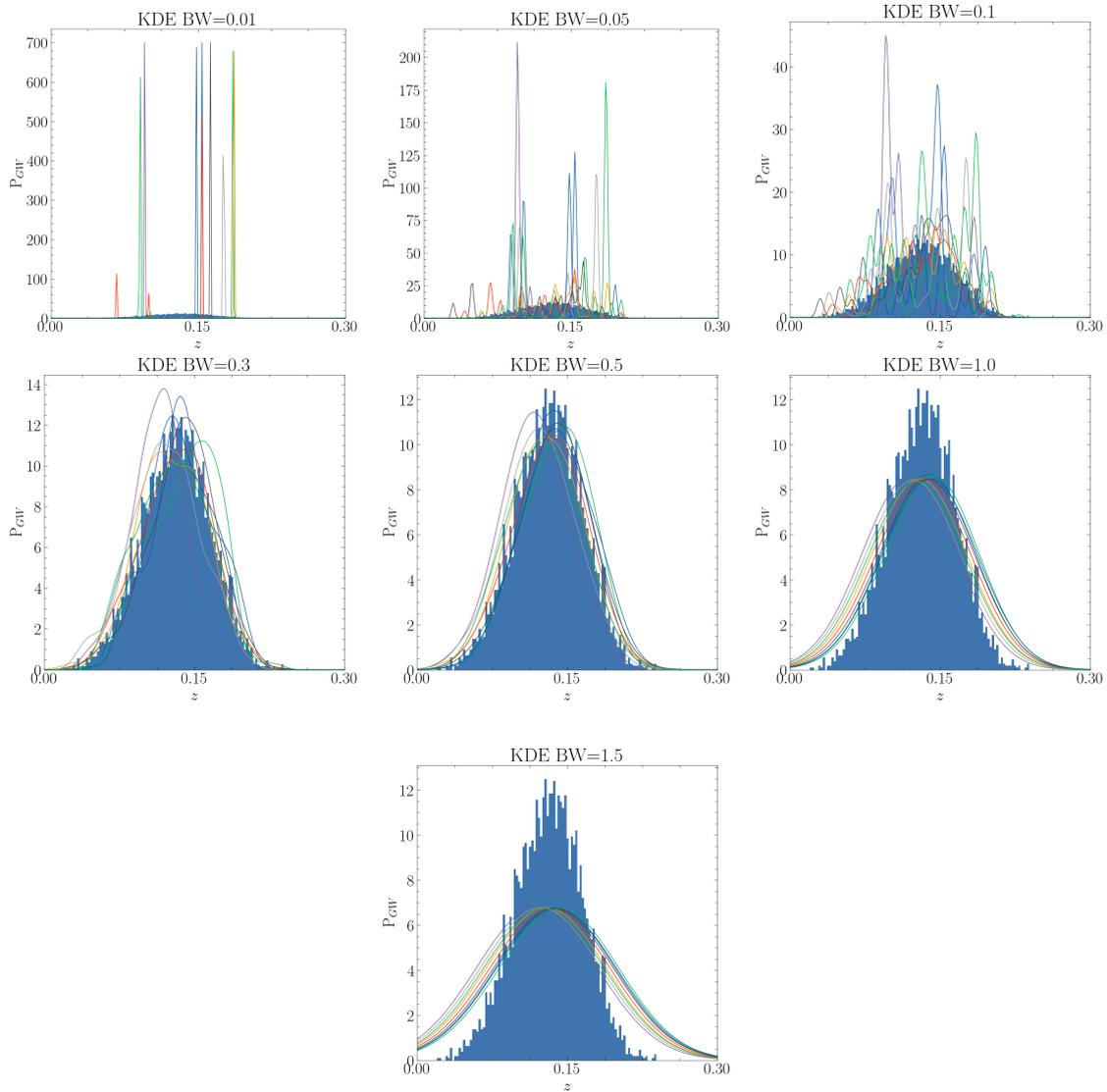


Figure 4.11: KDE probability distributions overlapped to the samples distribution of an O4 GW event. Each panel corresponds to the result obtained with a different KDE bandwidth

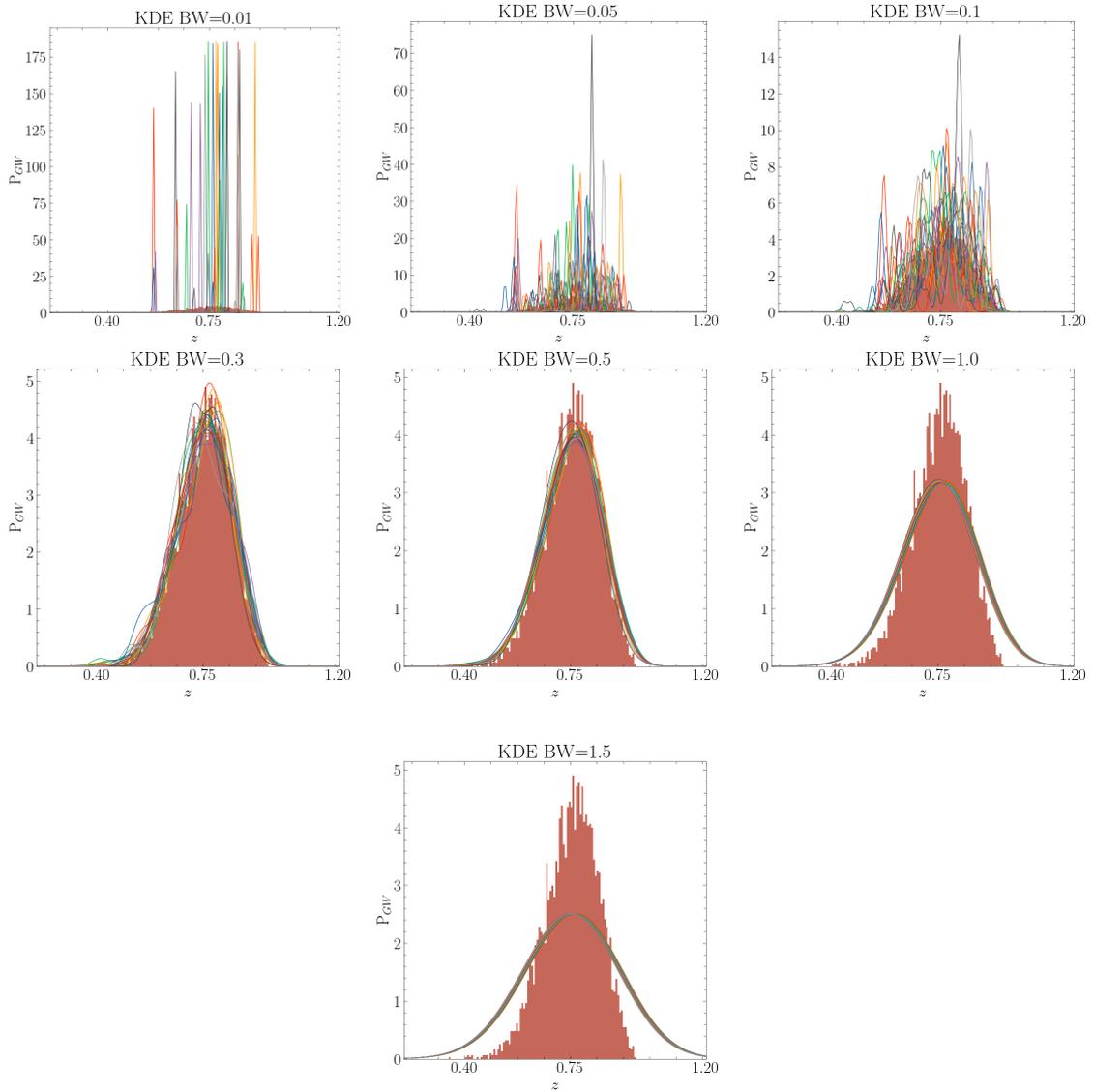


Figure 4.12: KDE probability distributions overplotted with the samples distribution of an O5 GW event. Each panel corresponds to the result obtained with a different KDE bandwidth.

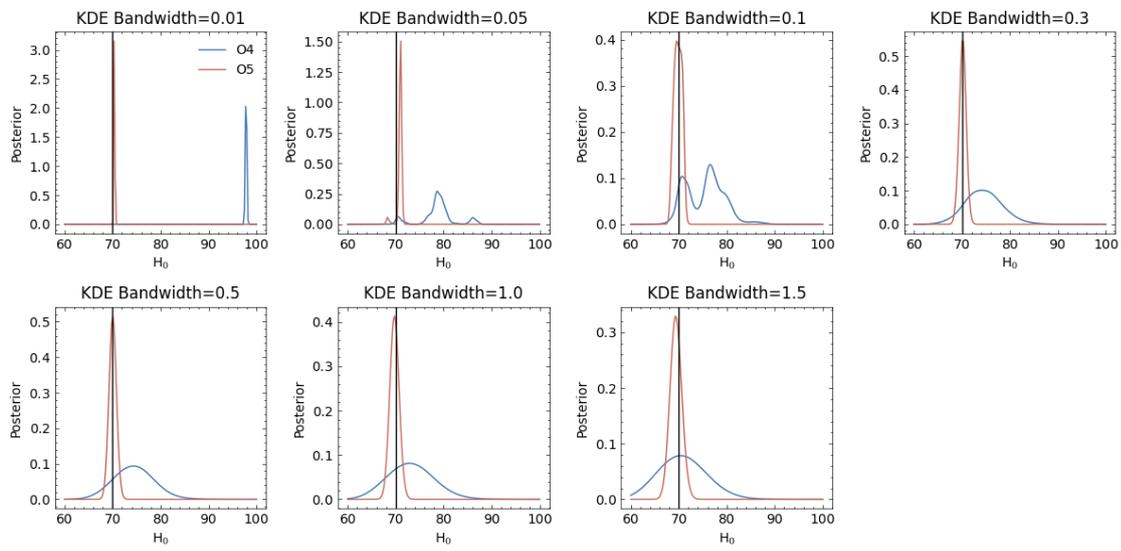


Figure 4.13: H_0 posterior distributions. Every panel shows the result obtained with both O4 and O5 using different KDE bandwidth. The black vertical lines indicate the true value, i.e. $H_0 = 70$ [km/s/Mpc].

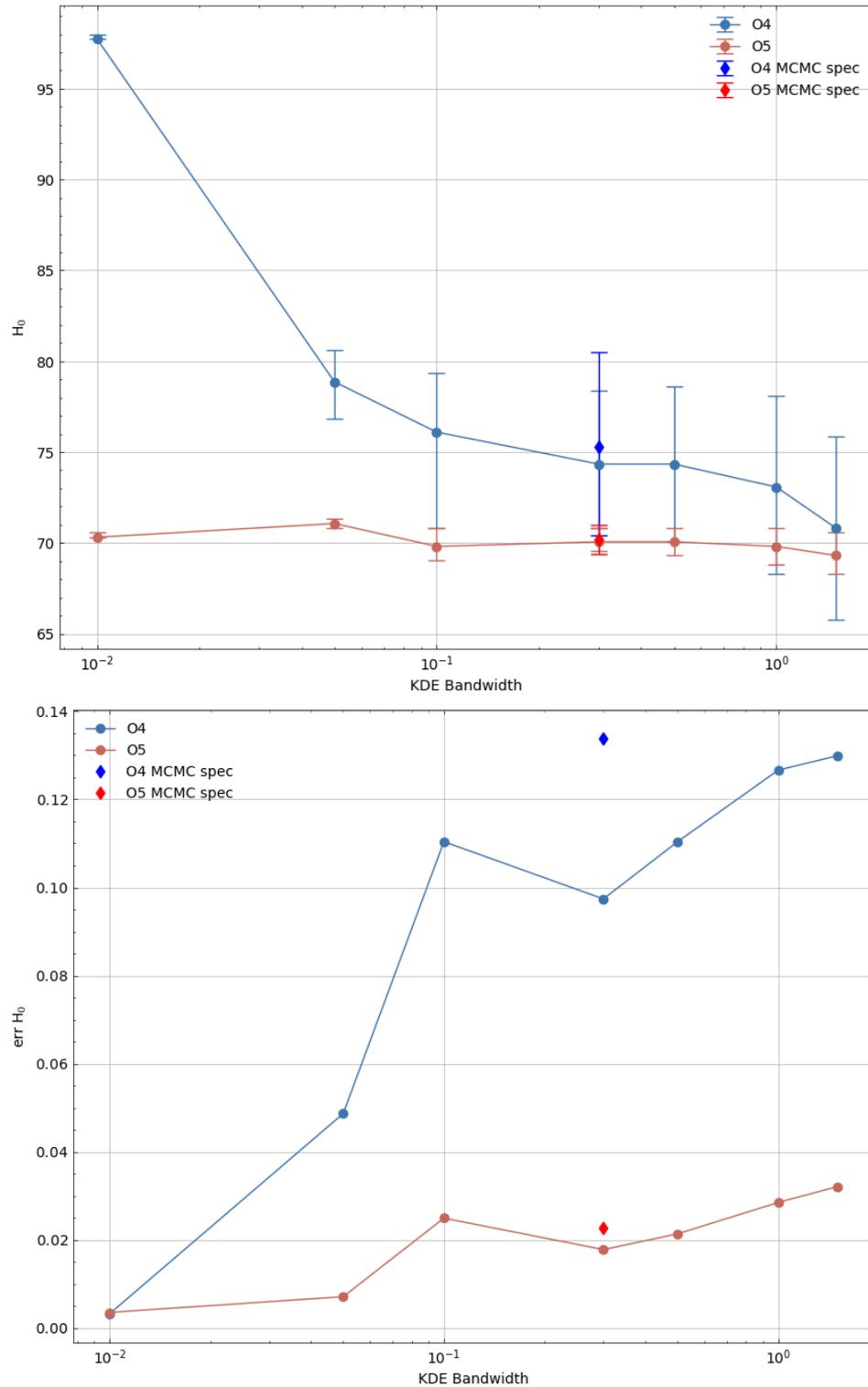


Figure 4.14: In both panels the blue points are obtained from the O4 dataset while the red ones from the O5 dataset. In the top panel, the median values of H_0 are represented with their error bars obtained with different KDE bandwidth. The single points in dark blue and red represent the values obtained with complete MCMC analyses of Borghi et al. (2024). In the bottom panel, the percentage errors of the values in the top panel are represented.

4.2. DEPENDENCE OF THE RESULTS ON THE ASSUMED CATALOGS 97

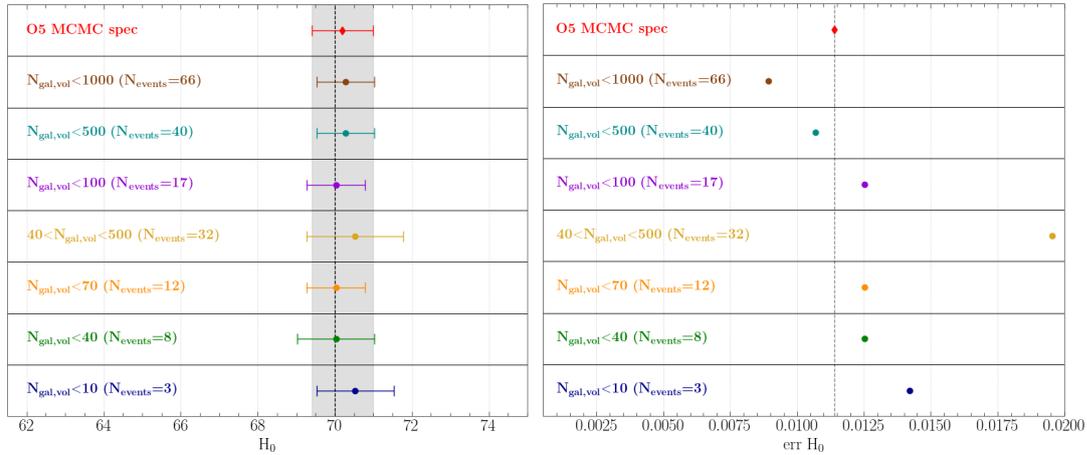


Figure 4.15: On the left panel the H_0 results obtained with the different samples using $z_{\text{err}} = 0.001$ are shown. The grey band represents the 1σ CI of the full MCMC analysis. On the right panel, the percentage errors of the corresponding values shown in the left panel are displayed. The dashed black vertical line represents the value corresponding to the full MCMC analysis. In both panels, different colours represent different subsamples.

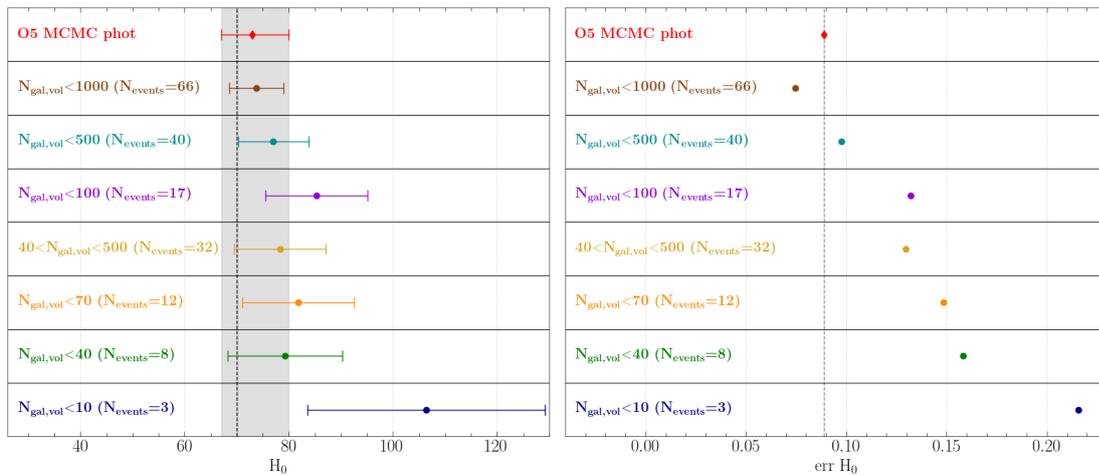


Figure 4.16: In the left panel are represented the H_0 results obtained with the different samples using $z_{\text{err}} = 0.05$. The grey band represents the 1σ CI of the full MCMC analysis. In the right panel, the percentage errors of the corresponding values shown in the left panel are displayed. The dashed black vertical line represents the value corresponding to the full MCMC analysis. In both panels, different colours represent different subsamples.

Chapter 5

Conclusions and Future Prospects

Since the first detection of GW150914 (Abbott et al., 2016), BBHs become a hot topic in both the fields of Astrophysics and Cosmology. Thus, always new methods have been developed in order to extract information from their detected signals and to the improvement of the technologies for the instruments, i.e. interferometers, that allow their detection. The subsequent detection of GW170817 (Abbott et al., 2017b) opened the doors to multi-messenger astronomy, further increasing the interest in these signals. With the LIGO-Virgo-KAGRA detector network, an increasing number of events have been detected and collected in catalogues. From the cosmological point of view, GWs have become an interesting model-independent new cosmological probe, as the loudness of these sources is directly from assuming the validity of general relativity, therefore, they provide a direct measurement of the luminosity distance d_L . The issue is that this measurement is degenerate with the redshift (a more massive binary at higher redshift produces the same signal as a lower massive one at lower redshift), and hence to derive cosmological constraints it is crucial to break this degeneracy by including in the likelihood the information about the redshift distribution of the sources, either directly detecting the electromagnetic counterpart (e.g. for BNS) or by statistically associating it with the help of a galaxy catalogue (as for BBH). This makes them standard sirens (bright and dark sirens, respectively, Moresco et al. 2022).

To maximise their scientific return, new methods to use GW events as cosmological probes have been developed. In order to obtain more accurate results on

cosmological parameters, such as H_0 , it is fundamental not only to rely only on the accuracy of the data but also to study and understand better the underlying astrophysics. Therefore, it is fundamental to learn as much as possible about the astrophysical population generating GWs. Thus, since the majority of the current detected signals come from the merging of binary black holes (up to the Observing Run 3, we have roughly one hundred BBHs and only one BNS with detected electromagnetic counterpart), it is of primary importance to study their population and, in particular, their mass distribution as a function of redshift. Many probability mass functions have been proposed to reproduce the available data; thanks to the increasing number of detected events and to the improved measurement accuracy, it has been possible to better characterise the distribution and obtain better cosmological constraints (Abbott et al., 2023a).

Current measurements have made huge leaps forward, yet our ability to discern features in the BBH mass distribution remains limited. It is therefore crucial to assess whether future measurement will allow us to better disentangle between mass function models, understand more about the distribution of BBH masses, and hence derive physical models about their origin. In parallel, it is also crucial to develop robust model selection algorithms to distinguish between the various proposed models.

With this aim, on one hand using mock catalogues of GW events that simulate the next O4 and O5 catalogues, I have studied four proposed mass function models Abbott et al. (2021) using the CHIMERA code in combination with a nested sampler of the posterior (*dynesty*) to obtain parameter constraints and the models' Bayesian evidence. On the other hand, the field of GWs is a relatively novel topic; thus, it is important not only to study if and how results improve with the newest and more accurate data, but even to assess how these results are impacted by systematics. Therefore, I have studied how the posterior distribution of cosmological parameters (in particular the Hubble constant H_0) changes when different assumptions are made in considering the galaxy and GW catalogues. In particular, I explored using different values of the galaxies' redshift error, different KDE bandwidths for the events posterior distributions evaluation, and different subsamples of the data mock catalogues.

5.1 Main results

In this section, the main results obtained within this work are summarised.

1. Samplers profiling

The first part of my work is devoted to the study of different techniques to sample the likelihood. As presented in sec. 2.3, there are many different methods to sample the parameter space and many more codes that apply these methods, therefore, it is important to test these samplers to understand their strengths and weaknesses and assess which is the best one for your own analysis. In particular, I tested three different samplers `emcee` (Foreman-Mackey et al., 2013), `zeus` (Karamanis et al., 2021) affine samplers and `dynesty` (Speagle, 2020) nested sampler, focusing on their sampling efficiency from a computational time point of view and accuracy in constraining parameters.

- As a preliminary test, I have checked the samplers' performances by analysing (as a test bed) the measured values of the Hubble parameters $H(z)$ with a cosmological model where the dark energy equation of state can vary with time (i.e. the $w_0 - w_a$ parameterisation). The results are summarised in Tab. 2.2. I found that the three samplers return consistent parameter constraints. Actually, a slight difference is observed between the 1σ confidence levels obtained with the two affine samplers and those obtained with the nested sampler, where the last ones are found to be a bit larger than the others. This can be explained by taking into account the fact that `dynesty` is an evidence-orientated sampler, therefore, the assessments of parameters constraints its slightly worse than what can be obtained through affine sampling.
- Using the same dataset, I have performed two different scaling tests: in the first one, I changed the number of walkers of the two affine samplers and checked how their running time changed; in the second one, I changed the number of iterations of all the three samplers and checked again how their running time changes. The results are shown in Fig. 2.10. I have seen that, at least for this specific model, despite what

is written in Karamanis et al. (2021), `emcee` results to be approximately 5 times faster than `zeus` in both tests. In addition, I have seen that, in this case of study, the sampling time of `dynesty` remains quite constant when increasing the number of iterations. Thus, even if it is always slower than `emcee`, with few iterations it is slower than `zeus` but with more iterations it becomes faster.

2. Codes development

In addition to the actual data analysis, I needed to implement new codes to perform it, integrating them with the actual CHIMERA pipeline.

- Building on the code structure of the tests described before, I implemented the `dynesty` sampler in the analysis code of CHIMERA (Borghini et al., 2024). This implied in particular the redefinition of the parameter prior probability function because, differently from the affine samplers, `dynesty` needs to have the priors scaled to the D-dimensional unit cube, where D indicates the number of parameters. In parallel, I have implemented a code structure in the CHIMERA analysis code that allows the user to select the preferred sampler.
- The CHIMERA's mass function library was lacking the broken power law model. Therefore, before starting the analysis, I needed to implement it. To do it, I followed the structure and the notation of the already present mass functions. I split the total probability distribution in the primary mass function, its normalisation, the secondary mass function, and its normalisation. Actually, these last two parts were already present as they are common to other mass distributions. For the primary mass probability distribution I followed the definition of eq. 2.16 adding in the regime $m_{break} < m_1 < m_h$ the term $m_{break}^{(-\alpha_1 + \alpha_2)}$. This addition was necessary to ensure a smooth transition between the two power laws.
- For the model comparison part of my analysis, I needed to encode the procedure for the production of the Posterior Predictive Checks. I followed the steps described in Edelman et al. (2022) implementing the functions of CHIMERA.

3. **Exploring the constraints on the BBH mass functions with O4 and O5 data** The first objective of this work was to study the different BBH mass functions using simulations of the future O4 and O5 datasets.

- The studied mass functions are those mentioned in sec. 2.2, namely: Truncate Power Law, Broken Power Law, Power Law plus one Gaussian Peak, Power Law plus two Gaussian Peak. To derive constraints for the population parameters, a nested sampling code was run for each mass distribution and each catalogue. It is important to remember that the two catalogues are made of mock data that have been generated using the PLP model as fiducial distribution. Therefore, it is possible to directly compare the common parameter, as shown Tab. 5.1. I noticed that the PLP model, as expected, is the one that better retrieves the fiducial parameters. It can be noticed that, in general, β , that is the slope of the secondary mass distribution, is not well reconstructed. This does not much affect the global shape of the mass functions that are well reconstructed. Furthermore, it is possible to see that the accuracy of the results obtained with the O5 data is indeed greater than in the case of O4. Last but not least, the PL2P model actually is a competitive model, as in this case it identifies the second peak as the first one hence returning to be basically a PLP, which is the real distribution of the analysed data.
- I kept track of the CPU time needed for the different analyses to converge. As expected, as can be seen from Tab. 4.6, the higher the number of parameters, the higher the CPU time. In addition, it can be noticed that the O5 analyses took more time than their O4 equivalent, despite what one may think a priori. Actually, at the time of writing some of the analyses are still running.
- To analyse the results more quantitatively, the different models were compared considering three implemented model selection criteria: Bayesian evidence, Deviance Information Criterion (DIC), and the Posterior Predictive Checks. In the case of the analysis performed on the O4 data, this criterion produces conflicting results. Indeed, the Bayesian evidence

Model (O4)	α	β	m_5	m_{95}
Fiducial	3.4	1.1	6.93	35.36
TPL	$3.05^{+0.14}_{-0.16}$	$3.19^{+0.58}_{-0.97}$	$6.28^{+0.21}_{-0.30}$	$24.8^{+2.1}_{-2.2}$
BPL	$2.21^{+0.25}_{-0.27}$	$2.63^{+0.78}_{-0.73}$	$6.14^{+0.47}_{-0.62}$	$32.5^{+1.9}_{-2.3}$
PLP	$3.90^{+0.32}_{-0.29}$	$2.16^{+0.79}_{-0.68}$	$6.51^{+0.37}_{-0.42}$	$34.1^{+1.6}_{-2.0}$
PL2P	3.77 ± 0.51	$2.35^{+1.19}_{-0.89}$	$6.39^{+0.58}_{-0.68}$	$33.7^{+2.3}_{-3.0}$
Model (O5)	α	β	m_5	m_{95}
Fiducial	3.4	1.1	6.93	35.36
TPL	$2.9^{+0.11}_{-0.19}$	$1.97^{+0.63}_{-0.41}$	$6.96^{+0.21}_{-0.27}$	$29.6^{+4.0}_{-1.8}$
BPL	$2.07^{+0.30}_{-0.29}$	$1.62^{+0.68}_{-0.23}$	$7.02^{+0.42}_{-0.58}$	$37.0^{+2.5}_{-4.5}$
PLP	$3.42^{+0.13}_{-0.18}$	$1.81^{+0.59}_{-0.18}$	6.88 ± 0.52	$36.31^{+0.64}_{-0.91}$
PL2P	$3.29^{+0.40}_{-0.39}$	$1.77^{+0.87}_{-0.67}$	$6.91^{+0.47}_{-0.67}$	$36.7^{+2.0}_{-2.2}$

Table 5.1: Common mass functions parameters obtained from the nested sampling analysis for both the O4 and O5 catalogue.

indicates that the best model is the BPL while the DIC indicates as favoured model the PL2P. From the PPCs, instead, we can see that the PLP and the PL2P predicted curves reproduce well the observed one, while the BPL is slightly in contrast with the observations. Therefore, we can conclude that with O4 the TPL model can be safely discarded while the other ones remain less distinguishable. On the other hand, in the O5 case, all the criteria agree with each other on the fact that the best-fitting model is the PLP. Nevertheless, the PL2P model has to be kept in consideration as the comparison criteria do not provide decisive evidence that it can be discarded, in agreement with what has been concluded by looking at the parameter constraints.

4. Assessing the dependence of the results on the catalogs assumptions

In parallel with the study of the mass functions, I have performed 1-dimensional analyses to investigate how the resulting posterior distributions of hyperparameters are affected by: the selection of different values of the error on the galaxies redshift z_{err} , the selection of different KDE bandwidths, and the choice of different cuts of the number of galaxies per localisation volume

$N_{gal,vol}$. The posteriors of H_0 are obtained on a fixed grid of hyperparameters, keeping each time all fixed except the one of interest (to minimise the CPU time). Here in particular I focused on the constraints on the Hubble constant H_0 .

- Using a series of z_{err} , that goes from the best one $z_{err} = 0.001$ to the worst one $z_{err} = 0.1$, it is possible to observe that the probability distributions shift toward values of H_0 higher than the fiducial one when increasing the galaxies redshift errors. The cause of this effect could be traced to how the redshift distribution is treated in the code when it is convolved with an expected "uniform in comoving volume" distribution (see eq. 2.6), but this needs further investigations. The O5 results are less sensitive to the worsening of the redshift accuracy, the above-described effect does not appear until $z_{err} < z_{phot} = 0.05$ as the values remain constant around $H_0 \simeq 71$ [km/s/Mpc] then increasing rapidly with the worst accuracies. Indeed, with low z_{err} they have percentage errors lower than $\sim 5 - 4\%$ than the O4 results, while with high z_{err} they are consistent with each other. Comparing the 1-dimensional results with those obtained with the full MCMC analyses of Borghi et al. (2024), it can be seen that, in the case of $z_{err} = z_{phot}$, the 1-dimensional results have lower percentage errors, while with $z_{err} = z_{spec} = 0.001$ they are consistent with each other. This can be due to the fact that in my analyses all the parameters but H_0 are fixed to their fiducial values, removing some degeneracies such as the one with μ_g .
- Different values of the KDE bandwidth estimate differently the probability distribution of the GW events. This affects the resulting H_0 posterior distribution. It can be seen that the O5 results are not affected much by the choice of the bandwidth, even though, as can be seen from Fig. 4.12, not all the values can reproduce well the event samples distribution. On the other hand, the O4 results depend heavily on the KDE bandwidth choice; in particular, low values of it create artifacts in the H_0 probability distribution.
- I have studied different subsamples of the O5 catalogue using different

$N_{gal,vol}$ cuts, i.e. those displayed in Tab. 4.9, using both $z_{err} = z_{phot}$ and $z_{err} = z_{spec}$. In the case of z_{spec} the results obtained from the different subsamples agree with the fiducial and the best full MCMC value. Analyses that consider only the best three events are not as constraining as those that include events with a higher $N_{gal,vol}$. On the other hand, excluding the worst localised events improves the accuracy of the results. Thus, for future analyses with galaxy catalogues that provide redshift with $z_{err} = z_{spec}$, events with very few $N_{gal,vol}$ are not mandatory to obtain competitive constraints and events with a high number of $N_{gal,vol}$ can be discarded to reduce the computational load. Studying also the complementary subsamples, it is possible to see that considering only events with $N_{gal,vol} > 40$ we can obtain results consistent with the results of the entire MCMC analyses. On the other hand, in the case of $z_{err} = z_{phot}$ it is more relevant also to have a sufficient number of events instead of a few well-localised events, indeed the best constraints are obtained using the subsample $N_{gal,vol} < 1000$, this is due to the fact that with this kind of redshift accuracy the information is mainly given by the BBH events instead of by the galaxies. Therefore, for future analyses with galaxy catalogues that provide this redshift accuracy, it is important to have a high number of events, thus including even the worst localised, in order to improve the constraints' accuracy. This study paves the way to define an optimal sample of GW BBH data to be used as standard dark sirens to derive cosmological constraints.

5.2 Future prospects

Given all the results obtained, it is finally possible to discuss some of the possible future developments of this work.

First of all, as the analyses' running times of this work are not satisfactory, it is necessary to integrate more efficiently the nested sampling algorithm. This could be done by optimising its parallelisation and integrating it with other public codes, e.g. Bilby (Ashton et al., 2019).

Furthermore, the 1-dimensional analyses on the posterior distributions of H_0 expose a bias on H_0 toward high values when increasing the error on the redshift. This bias has to be explored in more detail to be accounted for in future analyses. One possible path to follow is to test if a different function describing the redshift galaxy distribution could help in removing this issue.

Moreover, in the literature, BBH mass function models different from those used in this work can be found. Therefore, it would be interesting to finalise MCMC analyses with them also applying the presented model selection criteria in order to compare them with those already studied. In particular, it would be useful to implement more general and less parametric mass functions, as those presented in Abbott et al. (2023b).

Moreover, considering that my analyses have been performed on GW mock data, that have been generated assuming as fiducial BBH mass distribution the PLP model, I already knew which was the mass function that has to be retrieved. Therefore, in the future it could be explored a new analysis by generating new simulations, possibly within a blind challenge, to properly assess if a generic BBH probability mass distribution can be retrieved.

Bibliography

- Abbott, B. P. et al. (2016). Observation of Gravitational Waves from a Binary Black Hole Merger. *Physical Review Letters*, 116:061102.
- Abbott, B. P. et al. (2017a). A gravitational-wave standard siren measurement of the hubble constant. *Nature*, 551:85–88.
- Abbott, B. P. et al. (2017b). GW170817: Observation of gravitational waves from a binary neutron star inspiral. *Physical Review Letters*, 119:161101.
- Abbott, B. P. et al. (2019). Properties of the binary neutron star merger GW170817. *Physical Review X*, 9:011001.
- Abbott, R. et al. (2021). GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run. *arXiv e-prints*, page arXiv:2111.03606.
- Abbott, R. et al. (2021). Observation of gravitational waves from two neutron star–black hole coalescences. *The Astrophysical Journal Letters*, 915:L5.
- Abbott, R. et al. (2021). Population Properties of Compact Objects from the Second LIGO-Virgo Gravitational-Wave Transient Catalog. *The Astrophysical Journal Letters*, 913:L7.
- Abbott, R. et al. (2023a). Constraints on the Cosmic Expansion History from GWTC-3. *The Astrophysical Journal*, 949:76.
- Abbott, R. et al. (2023b). Population of Merging Compact Binaries Inferred Using Gravitational Waves through GWTC-3. *Physical Review X*, 13:011048.

- Abdalla, E. et al. (2022). Cosmology intertwined: A review of the particle physics, astrophysics, and cosmology associated with the cosmological tensions and anomalies. *Journal of High Energy Astrophysics*, 34:49–211.
- Adams, M. R., Cornish, N. J., and Littenberg, T. B. (2012). Astrophysical model selection in gravitational wave astronomy. *Physical Review D*, 86:124032.
- Aghanim, N. et al. (2020). Planck 2018 results. VI. Cosmological parameters. *Astronomy & Astrophysics*, 641:A6.
- Amaro-Seoane, P. et al. (2017). Laser Interferometer Space Antenna. *arXiv e-prints*, page arXiv:1702.00786.
- Ashton, G. et al. (2019). Bilby: A user-friendly bayesian inference library for gravitational-wave astronomy. *The Astrophysical Journal Supplement Series*, 241:27.
- Ashton, G. et al. (2022). Nested sampling for physical scientists. *Nature Reviews Methods Primers*, 2(1).
- Belczynski, K. et al. (2012). Missing Black Holes Unveil the Supernova Explosion Mechanism. *The Astrophysical Journal*, 757:91.
- Borghi, N., Mancarella, M., Moresco, M., Tagliazucchi, M., Iacovelli, F., Cimatti, A., and Maggiore, M. (2024). Cosmology and astrophysics with standard sirens and galaxy catalogs in view of future gravitational wave observations. *arXiv e-prints*.
- Borghi, N., Moresco, M., and Cimatti, A. (2022). Toward a Better Understanding of Cosmic Chronometers: A New Measurement of $H(z)$ at $z \sim 0.7$. *The Astrophysical Journal Letters*, 928:L4.
- Branchesi, M. et al. (2023). Science with the Einstein Telescope: a comparison of different designs. *Journal of Cosmology and Astroparticle Physics*, 2023:068.
- Carretero, J., Castander, F. J., Gaztañaga, E., Crocce, M., and Fosalba, P. (2014). An algorithm to build mock galaxy catalogues using MICE simulations. *Monthly Notices of the Royal Astronomical Society*, 447:646–670.

- Coles, P. and Lucchin, P. (2003). *Cosmology: The Origin and Evolution of Cosmic Structure*. Wiley.
- Collaboration, D. et al. (2016). The desi experiment part i: Science, targeting, and survey design.
- Crocce, M., Castander, F. J., Gaztañaga, E., Fosalba, P., and Carretero, J. (2015). The MICE Grand Challenge lightcone simulation – II. Halo and galaxy catalogues. *Monthly Notices of the Royal Astronomical Society*, 453:1513–1530.
- Dalal, N., Holz, D. E., Hughes, S. A., and Jain, B. (2006). Short GRB and binary black hole standard sirens as a probe of dark energy. *Physical Review D*, 74:063006.
- Del Pozzo, W. (2012). Inference of cosmological parameters from gravitational waves: Applications to second generation interferometers. *Physical Review D*, 86:043011.
- Dodelson, S. and Schmidt, F. (2003). *Modern Cosmology*. Academic Press, Elsevier Science.
- Edelman, B., Doctor, Z., Godfrey, J., and Farr, B. (2022). Ain’t No Mountain High Enough: Semiparametric Modeling of LIGO-Virgo’s Binary Black Hole Mass Distribution. *The Astrophysical Journal*, 924:101.
- Evans, M. et al. (2021). A horizon study for cosmic explorer: Science, observatories, and community.
- Ezquiaga, J. M. and Holz, D. E. (2022). Spectral sirens: Cosmology from the full mass distribution of compact binaries. *Physical Review Letters*, 129:061102.
- Foreman-Mackey, D., Hogg, D. W., Lang, D., and Goodman, J. (2013). emcee: The MCMC Hammer. *Publications of the Astronomical Society of the Pacific*, 125:306.
- Fosalba, P., Crocce, M., Gaztañaga, E., and Castander, F. J. (2015). The MICE grand challenge lightcone simulation – I. Dark matter clustering. *Monthly Notices of the Royal Astronomical Society*, 448:2987–3000.

- Fosalba, P., Gaztañaga, E., Castander, F. J., and Crocce, M. (2014). The MICE Grand Challenge light-cone simulation – III. Galaxy lensing mocks from all-sky lensing maps. *Monthly Notices of the Royal Astronomical Society*, 447:1319–1332.
- Gair, J. R. et al. (2023). The Hitchhiker’s Guide to the Galaxy Catalog Approach for Dark Siren Gravitational-wave Cosmology. *The Astrophysical Journal*, 166:22.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis, Third edition*. Chapman and Hall/CRC.
- Goodman, J. and Weare, J. (2010). Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science*, 5:65–80.
- Handley, W. (2022). Next generation cosmological analysis with nested sampling. In *Tensions in Cosmology*, Corfù.
- Hjorth, J. et al. (2017). The Distance to NGC 4993: The Host Galaxy of the Gravitational-wave Event GW170817. *The Astrophysical Journal Letters*, 848:L31.
- Hobson, M. P., Efstathiou, G. P., and Lasenby, A. N. (2006). *General Relativity: An Introduction for Physicists*. Cambridge University Press.
- Holz, D. E. and Hughes, S. A. (2005). Using gravitational-wave standard sirens. *The Astrophysical Journal*, 629:15–22.
- Hulse, R. A. and Taylor, J. H. (1975). Discovery of a pulsar in a binary system. *The Astrophysical Journal Letters*, 195:L51–L53.
- Iacovelli, F., Mancarella, M., Foffa, S., and Maggiore, M. (2022a). Forecasting the detection capabilities of third-generation gravitational-wave detectors using gwfast. *The Astrophysical Journal*, 941:208.
- Iacovelli, F., Mancarella, M., Foffa, S., and Maggiore, M. (2022b). Gwfast: A fisher information matrix python code for third-generation gravitational-wave detectors. *The Astrophysical Journal Supplement Series*, 263:2.

- Karamanis, M. and Beutler, F. (2021). Ensemble slice sampling: Parallel, black-box and gradient-free inference for correlated & multimodal distributions. *Statistics and Computing*, 31.
- Karamanis, M., Beutler, F., and Peacock, J. A. (2021). zeus: a python implementation of ensemble slice sampling for efficient Bayesian parameter inference. *Monthly Notices of the Royal Astronomical Society*, 508:3589–3603.
- Koposov, S., Speagle, J., et al. (2023). joshspeagle/dynesty: v2.1.3.
- Kunz, M., Trotta, R., and Parkinson, D. R. (2006). Measuring the effective complexity of cosmological models. *Physical Review D*, 74:023503.
- Laureijs, R. et al. (2011). Euclid definition study report.
- Liddle, A., Mukherjee, P., and Parkinson, D. (2006). Model selection in cosmology. *Astronomy and Geophysics*, 47:4.30–4.33.
- Liddle, A. R. (2007). Information criteria for astrophysical model selection. *Monthly Notices of the Royal Astronomical Society*, 377:L74–L78.
- Linder, E. V. (2003). Exploring the Expansion History of the Universe. *Physical Review Letters*, 90:091301.
- Luo, Z., Guo, Z., Jin, G., Wu, Y., and Hu, W. (2020). A brief analysis to taiji: Science and technology. *Results in Physics*, 16:102918.
- Madau, P. and Dickinson, M. (2014). Cosmic Star-Formation History. *Annual Review of Astronomy and Astrophysics*, 52:415–486.
- Maggiore, M. (2007). *Gravitational Waves. Vol. 1: Theory and Experiments*. Oxford University Press.
- Maggiore, M. (2018). *Gravitational Waves. Vol. 2: Astrophysics and Cosmology*. Oxford University Press.
- Mancarella, M., Borghi, N., Foffa, S., Genoud-Prachex, E., Iacovelli, F., Maggiore, M., Moresco, M., and Schulz, M. (2022a). Gravitational-wave cosmology with

- dark sirens: state of the art and perspectives for 3g detectors. In *Proceedings of 41st International Conference on High Energy physics — PoS(ICHEP2022)*, ICHEP2022. Sissa Medialab.
- Mancarella, M. and Genoud-Prachex, E. (2022). Cosmostatgw/mgcosmopop: v1.0.0.
- Mancarella, M., Genoud-Prachex, E., and Maggiore, M. (2022b). Cosmology and modified gravitational wave propagation from binary black hole population models. *Physical Review D*, 105:064030.
- Mandel, I., Farr, W. M., and Gair, J. R. (2019). Extracting distribution parameters from multiple uncertain observations with selection biases. *Monthly Notices of the Royal Astronomical Society*, 486:1086–1093.
- Moresco, M. (2015). Raising the bar: new constraints on the Hubble parameter with cosmic chronometers at $z \sim 2$. *Monthly Notices of the Royal Astronomical Society*, 450:L16–L20.
- Moresco, M. et al. (2012). Improved constraints on the expansion rate of the Universe up to $z \sim 1.1$ from the spectroscopic evolution of cosmic chronometers. *Journal of Cosmology and Astroparticle Physics*, 2012:006.
- Moresco, M. et al. (2016). A 6% measurement of the Hubble parameter at $z \sim 0.45$: direct evidence of the epoch of cosmic re-acceleration. *Journal of Cosmology and Astroparticle Physics*, 2016:014.
- Moresco, M. et al. (2022). Unveiling the Universe with emerging cosmological probes. *Living Reviews in Relativity*, 25:6.
- Punturo, M. et al. (2010). The Einstein Telescope: a third-generation gravitational wave observatory. *Classical and Quantum Gravity*, 27:194002.
- Ratsimbazafy, A. L., Loubser, S. I., Crawford, S. M., Cress, C. M., Bassett, B. A., Nichol, R. C., and Väisänen, P. (2017). Age-dating luminous red galaxies observed with the Southern African Large Telescope. *Monthly Notices of the Royal Astronomical Society*, 467:3239–3254.

- Riess, A. G. et al. (2022). A comprehensive measurement of the local value of the hubble constant with 1/s/mpc. Uncertainty from the hubble space telescope and the SH0es team. *The Astrophysical Journal Letters*, 934:L7.
- Schutz, B. F. (1986). Determining the hubble constant from gravitational wave observations. *Nature*, 323:310–311.
- Simon, J., Verde, L., and Jimenez, R. (2005). Constraints on the redshift dependence of the dark energy potential. *Physical Review D*, 71:123001.
- Skilling, J. (2004). Nested Sampling. *AIP Conference Proceedings*, 735:395–405.
- Skilling, J. (2006). Nested sampling for general Bayesian computation. *Bayesian Analysis*, 1:833–859.
- Speagle, J. S. (2020). DYNesty: a dynamic nested sampling package for estimating Bayesian posteriors and evidences. *Monthly Notices of the Royal Astronomical Society*, 493:3132–3158.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Linde, A. (2014). The Deviance Information Criterion: 12 Years on. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 76:485–493.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 64:583–639.
- Stern, D., Jimenez, R., Verde, L., Kamionkowski, M., and Stanford, S. A. (2010). Cosmic chronometers: constraining the equation of state of dark energy. I: $H(z)$ measurements. *Journal of Cosmology and Astroparticle Physics*, 2010:008.
- Talbot, C. and Thrane, E. (2018). Measuring the Binary Black Hole Mass Spectrum with an Astrophysically Motivated Parameterization. *The Astrophysical Journal*, 856:173.
- Thrane, E. and Talbot, C. (2019). An introduction to bayesian inference in gravitational-wave astronomy: Parameter estimation, model selection, and hierarchical models. *Publications of the Astronomical Society of Australia*, 36.

- Trotta, R. (2008). Bayes in the sky: Bayesian inference and model selection in cosmology. *Contemporary Physics*, 49:71–104.
- Weber, J. (1960). Detection and generation of gravitational waves. *Physical Review*, 117:306–313.
- Weinberg, S. (1972). *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*. John Wiley & Sons, Inc.
- Zevin, M. et al. (2020). Exploring the Lower Mass Gap and Unequal Mass Regime in Compact Binary Evolution. *The Astrophysical Journal Letters*, 899:L1.
- Zhang, C., Zhang, H., Yuan, S., Liu, S., Zhang, T.-J., and Sun, Y.-C. (2014). Four new observational $H(z)$ data from luminous red galaxies in the Sloan Digital Sky Survey data release seven. *Research in Astronomy and Astrophysics*, 14:1221–1233.

Appendix A

Corner plots

This Appendix includes full corner plots for the O4 and O5 analyses assuming four different BBH mass functions: truncated power law (TPL; fig. A.1), broken power law (BPL; fig. A.2), power law + peak (PLP; fig. A.3), and power law + 2 peaks (PL2P; fig. A.4). Then, it includes corner plots comparing the results using nested sampling (`dynesty`) and affine-invariant sampling (`emcee`) for the O4 (fig. A.5) and O5 (fig. A.6) runs.

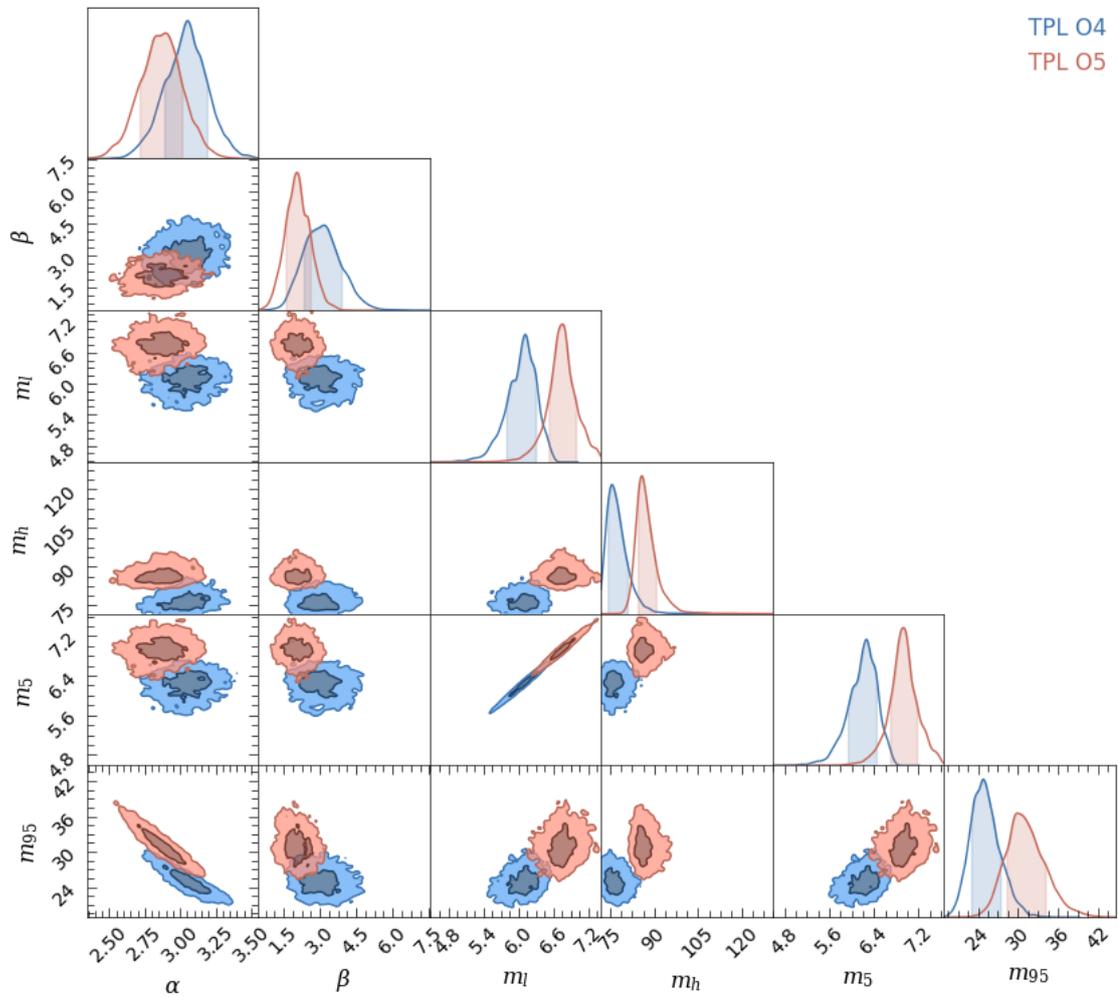


Figure A.1: Corner plot of the TPL parameters distributions. In blue the distributions obtained with the O4 data, in red the distributions obtained with the O5 data

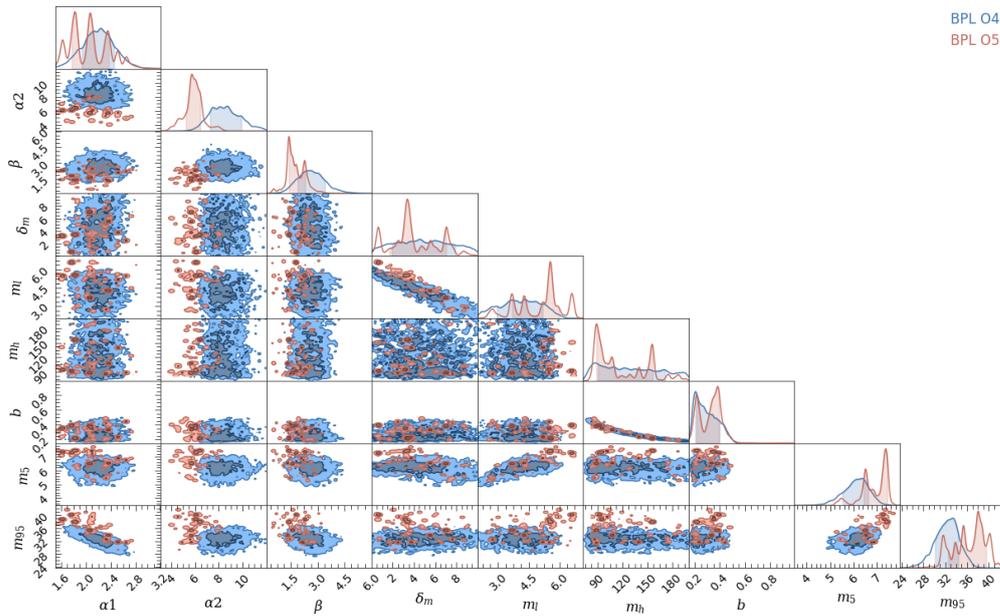


Figure A.2: Corner plot of the BPL parameters distributions. In blue the distributions obtained with the O4 data, in red the distributions obtained with the O5 data

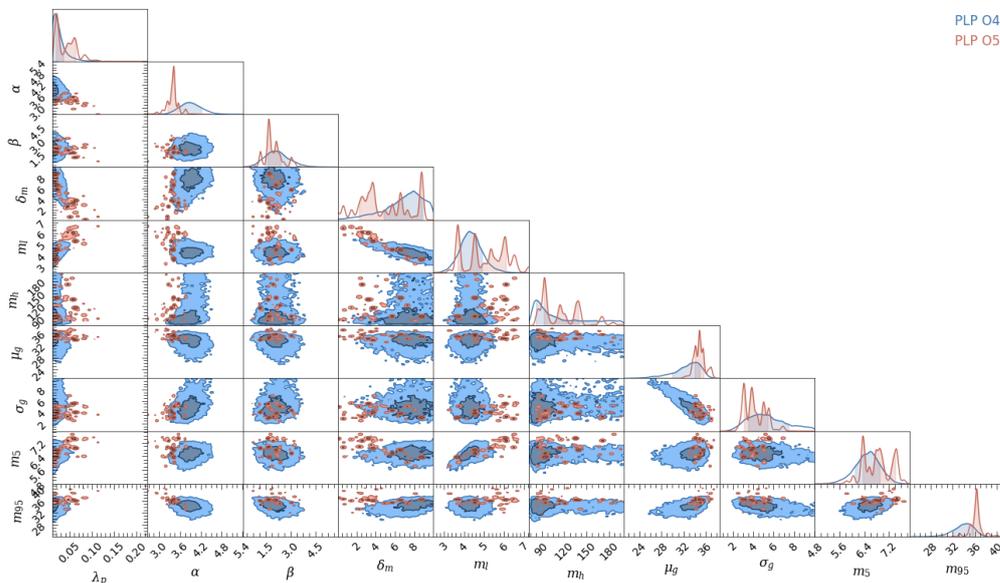


Figure A.3: Corner plot of the PLP parameters distributions. In blue the distributions obtained with the O4 data, in red the distributions obtained with the O5 data

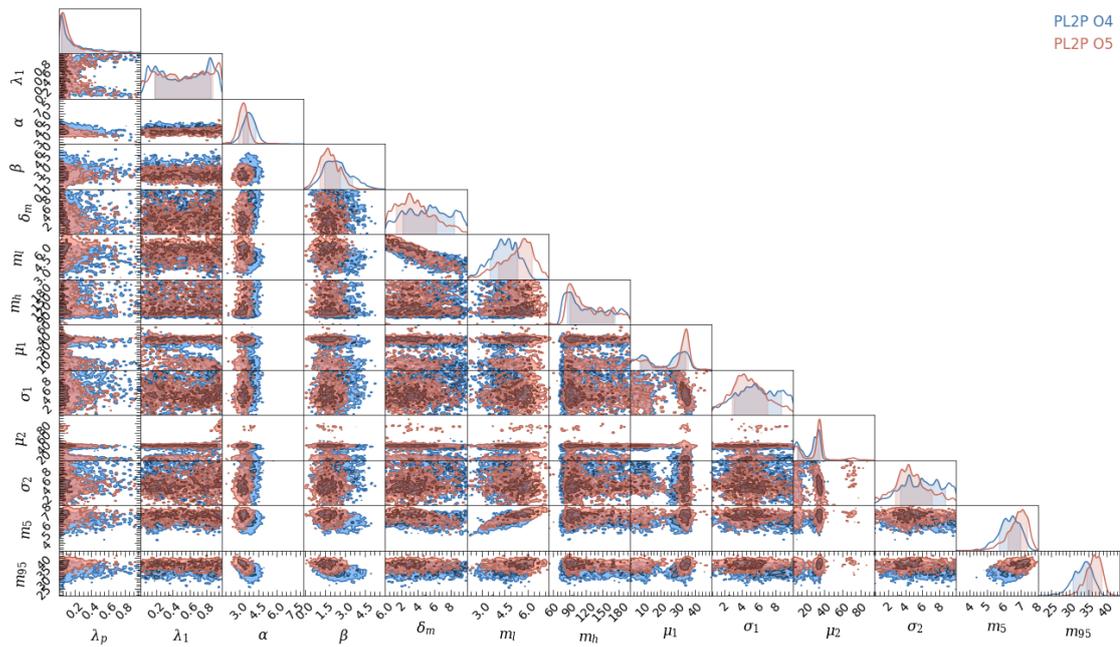


Figure A.4: Corner plot of the PL2P parameters distributions. In blue the distributions obtained with the O4 data, in red the distributions obtained with the O5 data

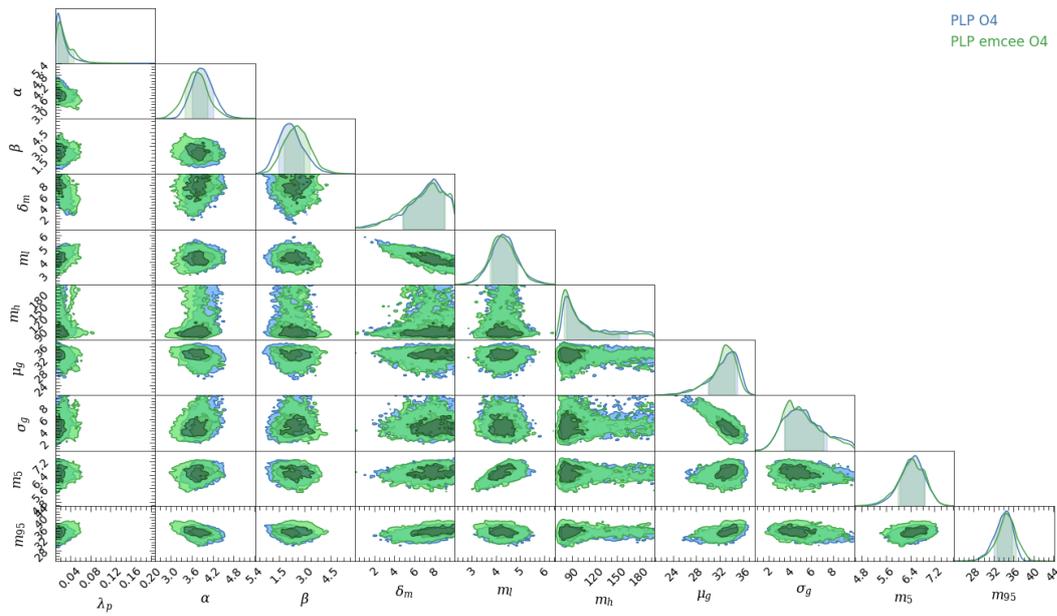


Figure A.5: Corner plot of the PLP parameters distributions obtained with the O4 data with the `dynesty` and `emcee` samplers. In blue the distributions obtained with `dynesty`, in green the distributions obtained with `emcee`

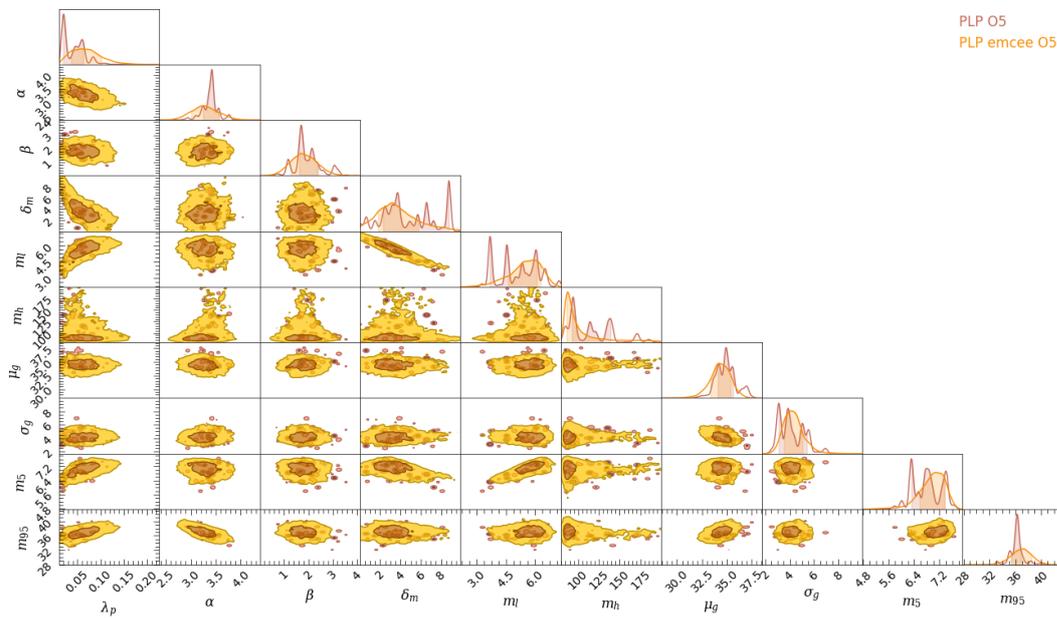


Figure A.6: Corner plot of the PLP parameters distributions obtained with the O5 data with the `dynesty` and `emcee` samplers. In red the distributions obtained with `dynesty`, in yellow the distributions obtained with `emcee`