

Chirp mass–distance distributions of the sources of gravitational waves

M. Ossowski

Institute of Theoretical Physics, Faculty of Physics, University of Warsaw, Pasteura 5, 02-093 Warsaw, Poland
e-mail: maciej.ossowski@fuw.edu.pl

Received 25 September 2019 / Accepted 26 January 2021

ABSTRACT

Context. The detection of gravitational waves emitted by binary black holes raises questions regarding the origin of binaries. There are several models in the literature involving binary evolution in both the field and clusters.

Aims. We compare the predictions of these models with observations and establish the reliability of this comparison.

Methods. We use the likelihood calculation to compare the models in the space spanned by the observed chirp mass and the luminosity distance of the source.

Results. We rank the models by their ability to explain all current gravitational wave detections. It is shown that the most probable models correspond to binary evolution with low metallicity. Several variants of such evolution have similar likelihoods. The globular cluster model, considered here, is disfavoured.

Conclusions. We present the usefulness of the method in distinguishing between models when new observations become available. We calculate the number of observations required to distinguish between each pair of models. We find that the number varies from ten to several thousand for some pairs of models, yet almost two-thirds of pairs are distinguishable with at most 100 observations.

Key words. gravitational waves – stars: black holes

1. Introduction

The discovery of gravitational waves (GWs; Abbott et al. 2016) marked the beginning of gravitational wave astronomy. The recent catalogue of gravitational wave sources (The LIGO Scientific Collaboration & the Virgo Collaboration 2019; Abbott et al. 2020) contains 50 candidates, 47 of which are binary black hole (BBH) mergers. The properties of the population were studied by Belczynski et al. (2017). In the mean time, several models of formation of BBHs have been proposed. The leading model is based on the assumption that all these sources form as a result of binary evolution. This idea has been explored by several authors in recent decades (Lipunov et al. 1997; Belczynski et al. 2002, 2008; Dominik et al. 2012). Recently a version of this scenario in which the stars undergo strong mixing and have uniform chemical composition was studied as an alternative option. Additionally, several authors discuss the possibility of enhanced formation of BBHs in globular clusters (Rodriguez et al. 2016; Askar et al. 2017).

The availability of data for 47 BBH mergers allows the model predictions to be compared with observations. We present these comparisons using Bayesian statistical inference. The paper is organised as follows: In Sect. 2 we present the principles of detection of BBHs, and Sect. 3 contains a description of the statistical method used. In Sect. 4 we present the models of BBH populations used for comparison and in Sect. 5 we discuss the properties of the populations that are important for this analysis. In Sect. 6 we briefly explain why signal-to-noise ratio distributions carry little information about the population that is useful for the Advanced Laser Interferometer Gravitational-Wave Observatory (LIGO), and finally in Sect. 7 we present results of the analysis of the chirp mass–luminosity distance distributions

in the context of current detections and the possibility of distinguishing between models when more data become available.

2. Detection in binary systems

Consider a BBH with masses m_1 and m_2 , which is moving in a circular orbit. Peters & Mathews (1963) showed that those systems will emit energy via GWs. The waveform of these GWs depends on the redshifted chirp mass $\mathcal{M}_z = \mathcal{M}(1+z)$, where chirp mass is defined as

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

The position of the BBH on the sky with respect to the detector is described in spherical coordinates by angles θ and ϕ . The orientation of the BBH is described by the angles ι and ψ in spherical coordinates, where the z -axis is parallel to the total angular momentum of the system (Finn 1996). Measuring distance to the system must take the geometry of spacetime into account. These cosmological effects are encompassed by the luminosity distance

$$D_L(z) = (1+z)D_H \int_0^z \frac{dz'}{E(z')},$$

where $E(z) = \sqrt{\Omega_m(1+z)^3 + \Omega_\Lambda}$ and $D_H = c/H_0$. We use the following parameters: $\Omega_m = 0.27$, $\Omega_\Lambda = 1 - \Omega_m = 0.73$, and $H_0 = 70.4 \text{ km s}^{-1} \text{ Mpc}^{-1}$. We assume that the BBH will be detected if its signal-to-noise ratio (S/N) ρ is greater than 8. The S/N here is defined as

$$\rho = \frac{A(\mathcal{M}_z)\Theta(\theta, \phi, \iota, \psi)}{D_L}, \quad (1)$$

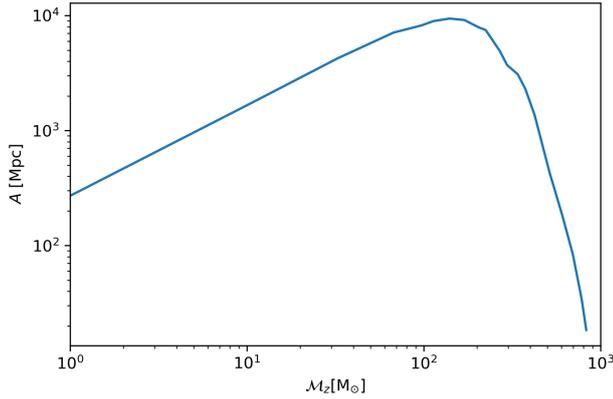


Fig. 1. Response of Advanced LIGO to the observed chirp mass.

where $A(M_z)$ is the luminosity distance at which an optimally oriented binary with redshifted chirp mass M_z is detected with an S/N of four and $\Theta(\theta, \phi, \iota, \psi)$ is the sensitivity of the detector to the orientation of the BBH. The orientation sensitivity $\Theta \in [0, 4]$ is

$$\Theta(\theta, \phi, \iota, \psi) = 2[F_+^2(1 + \cos^2 \iota)^2 + 4F_\times^2 \cos^2 \iota]^{1/2},$$

where F_+ and F_\times are the sensitivity of the detector to polarisations + and \times of the GW (Finn 1996) defined as

$$F_+ = \frac{1}{2}(1 + \cos^2 \theta) \cos 2\phi \cos 2\psi - \cos \theta \sin 2\phi \sin 2\psi, \quad (2)$$

$$F_\times = \frac{1}{2}(1 + \cos^2 \theta) \cos 2\phi \sin 2\psi + \cos \theta \sin 2\phi \cos 2\psi. \quad (3)$$

We consider the Advanced LIGO in configuration H1 (Martynov et al. 2016). The purpose of this choice is twofold: (i) to determine which model is preferred by up-to-date observations and (ii) to determine if this detector configuration is capable of differentiating between models. For $M_z \leq 100 M_\odot$, the dependence of the sensitivity of Advanced LIGO on chirp mass can be approximated by $A(M_z) \propto M_z^{5/6}$, which leads to the commonly used approximation of S/N $\rho \propto M_z^{5/6} \Theta / D_L$. The full range of $A(M_z)$, which is twice the horizon luminosity distance (i.e. distance for which $\rho = 8$ for a given detector), is presented in Fig. 1.

The dependence of sensitivity on chirp mass essentially encapsulates the ability of the interferometer to detect different frequencies which is a function of detector construction and corresponding noises. The main sources of the noise are seismic noises directly moving interferometer mirrors, thermal noise interacting with the beam, and quantum noise producing fluctuations in the number of photons in the beam. A full discussion on the noise sources can be found in Martynov et al. (2016).

3. Method of distinguishing models

To distinguish between the models we employ the method of statistical Bayesian inference. A broader description of this approach can be found in Gregory (2005). For the purpose of this analysis let us define:

$P(M_i | O)$ probability that model M_i is correct, given observation O

$P(O | M_i)$ probability of observing O , given that model M_i is correct

$P(M_i)$ prior probability that model M_i is correct.

We note that

$$P(O | M_i) = \mathcal{L}(O | M_i) := \prod_{k \in O} \frac{dP_i}{d\mathcal{B}}(\mathcal{B}_k), \quad (4)$$

where $\mathcal{L}(O | M_i)$ is the likelihood function for the model M_i , \mathcal{B} is the set of investigated observables, and k indexes the observations in O . Furthermore, we denote O as O_i to emphasise that it was generated using model M_i . The probability distribution of $\mathcal{B} = (D_L, \mathcal{M})$ for a given model is defined as follows:

$$\frac{dP_i}{d\mathcal{B}}(\mathcal{B}) = \sum_l \delta(D_L - D_{Ll}) \delta(\mathcal{M} - \mathcal{M}_l), \quad (5)$$

where for a given model M_i we estimate the above probability by calculating a set of 10^5 observations labelled by l and characterised by (D_{Ll}, \mathcal{M}_l) . The delta function is approximated using the value of the nearest bin on a 200 by 200 grid. Subsequently, for a model M_i we can calculate the likelihood of this model given up-to-date, real GW observations by

$$\mathcal{L}_{i,\text{Real}} = \prod_n \frac{dP_i}{d\mathcal{B}}(D_L = D_{Ln}, \mathcal{M} = \mathcal{M}_n), \quad (6)$$

where n covers the set of the real GW observations. Assuming that we are in the position of maximal ignorance with regard to the validity of the model, we let all $P(M_i)$ be equal. Then, following Gregory (2005), let us define the Bayes factor as

$$F_{ij} := \frac{P(M_i | O_i)}{P(M_j | O_i)} = \frac{P(O_i | M_i)}{P(O_i | M_j)}, \quad (7)$$

where the second equality is a result of the Bayes theorem,

$$P(M_i | O_i) = \frac{P(O_i | M_i)P(M_i)}{P(O_i)},$$

where $P(O_i)$ is the normalisation.

We would like to point out that $F_{ij} \gg 1$ means that observation O_i is in favour of model M_i over M_j and thus suggests that the former better represents reality than the latter. We analyse distributions of Bayes factors for different models to establish the threshold of distinguishability with p -value α . The threshold of distinguishability (TOD) with p -value α is defined as the minimal number of observations needed for $(1 - \alpha) \times 100\%$ of the Bayes factors F_{ij} to be greater than 1. To avoid numerical noise we also require that the same inequality holds for every number of the observations greater than the threshold. The process of calculating the TOD is as follows:

1. Generate an observation O_i using model M_i .
2. For each pair of models, calculate F_{ij} using the number of observations ranging logarithmically from 3 to 3000.
3. Repeat the previous point 10^4 times if the number of observations in O_i is smaller than 10 or 10^3 times otherwise, then analyse the distributions of the Bayes Factor based on this sample.

We note the following characteristics of this approach:

- There is no symmetry in reversing the models $F_{ij} \neq F_{ji}^{-1}$. This is because those Bayes factors use observations based on different models.

- The mean value of F_{ij} is expected to be greater than 1. Similarly to the previous point, this is because on average obtaining O_i is more probable with M_i than with M_j . This stresses the importance of using a high enough p -value.

- The TOD is dependent on the choice of the detector and the set of observables \mathcal{B} .

4. Investigated models

We surveyed a total of 34 models for compatibility with up-to-date observations. Additionally, 16 out of the above 34 models were surveyed for distinguishability based on their ability to explain observations. Of those 34 models, 32 are StarTrack models available online¹. To minimise numerical error it was additionally required that a model have at least 100 systems evolved to BBH. We now present an abridged description of the models; for the full description of the StarTrack models we refer to Dominik et al. (2012). The two non-StarTrack models are chemically homogeneous evolution and evolution in globular clusters.

4.1. StarTrack models

StarTrack is a population synthesis code focused on evolution of field stars and binary systems (Belczynski et al. 2002, 2008). It incorporates a twofold treatment of the Hertzsprung Gap donors in a common envelope (CE) phase. In the family of models *B*, the CE phase initiated by a Hertzsprung Gap donor always leads to a merger and formation of a Throne-Żytkow object. These models usually do not form a sufficient number of BBHs to explain observations.

In the family of models *A*, the energy balance is calculated following the Webbink (1984) formalism as

$$\left(\frac{GM_{\text{don},f}M_{\text{acc}}}{2A_f} - \frac{GM_{\text{don},i}M_{\text{acc}}}{2A_i} \right) = \frac{GM_{\text{don},i}M_{\text{don,env}}}{\lambda R_{\text{don,lob}}}, \quad (8)$$

where *i* and *f* stand for the initial and final, and M_{don} , M_{acc} , and $M_{\text{don,env}}$ are the masses of the donor, the accretor, and the ejected envelope. $R_{\text{don,lob}}$ is the Roche lobe radius of the donor at the beginning of the Roche Lobe Overflow (ROFL) and *A* is the binary separation. The energy balance determines whether the binary will shed its envelope, merge during CE, or coalesce. The λ parameter is used in energy balance of CE and is defined as

$$E_{\text{bind}} = - \frac{GM_{\text{don}}M_{\text{don,env}}}{\lambda R}, \quad (9)$$

where E_{bind} is the binding energy of the envelope and *R* is the radius of the donor.

The compact object merger can only occur if the binary survives the CE phase or phases. Therefore, we consider only the family of models *A*. StarTrack models can be categorised according to differences in the input physics: CE energy parameter λ , maximum neutron star (NS) mass, natal kick velocity, supernova engine, wind loss, and mass transfer treatment; for a more detailed discussion see Dominik et al. (2012). We used the model files published at the website syntheticuniverse.org and the description below refers to the models provided on that site.

4.1.1. Standard model

Standard model uses the λ parameter varying during evolution according to Xu & Li (2010), which is dubbed as *Nanjing* λ . Maximal NS mass is taken to be $2.5 M_{\odot}$. The mass transfer is half-conservative (half of the mass is ejected to infinity). This means that $f_a = 0.5$, defined as $\dot{M}_{\text{acc}} = f_a \dot{M}_{\text{don}}$, where \dot{M}_{acc} is ROFL mass transfer rate and \dot{M}_{don} is the ROFL mass transfer rate of the donor. Natal kick distribution is Maxwellian with $\sigma = 265 \text{ km s}^{-1}$ and the fall-back factor f_{fb} is according to Fryer et al. (2012). The fall-back factor f_{fb} is defined as $v_k = v_{\text{max}}(1 - f_{\text{fb}})$,

¹ syntheticuniverse.org

Table 1. Variants of StarTrack model.

Model	Changed parameter
V1	Fixed $\lambda = 0.01$
V2	Fixed $\lambda = 0.1$
V3	Fixed $\lambda = 1$
V4	Fixed $\lambda = 10$
V5	$M_{\text{NS,max}} = 3.0 M_{\odot}$
V6	$M_{\text{NS,max}} = 2.0 M_{\odot}$
V7	Natal kick distribution's $\sigma = 132.5 \text{ km s}^{-1}$
V8	Full BH natal kicks $f_{\text{fb}} = 1$
V9	No BH natal kicks $f_{\text{fb}} = 0$
V10	Delayed SN engine (Fryer et al. 2012)
V11	Wind mass loss reduced to 50%
V12	Fully conservative mass transfer $f_a = 1$
V13	Fully non-conservative mass transfer $f_a = 0$
V14	<i>Nanjing</i> λ increased by factor 5
V15	<i>Nanjing</i> λ decreased by factor 5

Notes. Except for the stated change, parameters are the same as in the Standard model.

where v_{max} is velocity drawn from the Maxwellian distribution and v_k is the final velocity. The Standard model uses a Rapid supernova engine, as described in Fryer et al. (2012).

4.1.2. Variants

In Table 1 we present variants of the StarTrack model. Parameters other than those mentioned under ‘Changed parameter’ remain the same as in the Standard model.

4.2. Other models

1. Chemically homogeneous evolution (CHE): One of the alternative channels of BBH evolution is CHE, described in de Mink & Mandel (2016). Here we study the model denoted ‘Default’ in Mandel & de Mink (2016). Owing to rapid rotation causing transport of matter between the core and the surface, CHE systems do not have a chemical gradient. Such stars gradually shrink, staying inside the Roche Lobe, and omit a transfer phase. Because of their non-violent evolution, the resulting BBHs may have a higher mass than is typical for these objects.

2. Globular Clusters (GCs): Morscher et al. (2013) suggest that the merger rate in GCs may be comparable to the rate in the field evolution. Because of this we include a GC model simulated with MOCCA code (Giersz et al. 2013) and broadly described in Askar et al. (2017). This model is a sum of approximately 2000 submodels with varying initial conditions: $Z \in [0.002-0.02]$ and distance from the galaxy centre between 1 and 50 kpc. All submodels share the Maxwellian natal velocity kick distribution with $\sigma = 265 \text{ km s}^{-1}$, the rotational velocity of a cluster of 220 km s^{-1} , and the IMF according to Kroupa (2001). The simulated GCs agree with observations of GCs in the Galaxy (Askar et al. 2017). Coalescing in a GC leads to faster BBH evolution, while those expelled and coalescing outside are subject to slower evolution; both scenarios are considered.

5. Characteristics of BBH population

We consider a space of three (not entirely independent) observables: $\{\rho, \mathcal{M}, D_L\}$. We also show in Sect. 6 that the luminosity distributions contain little information, and therefore we limit

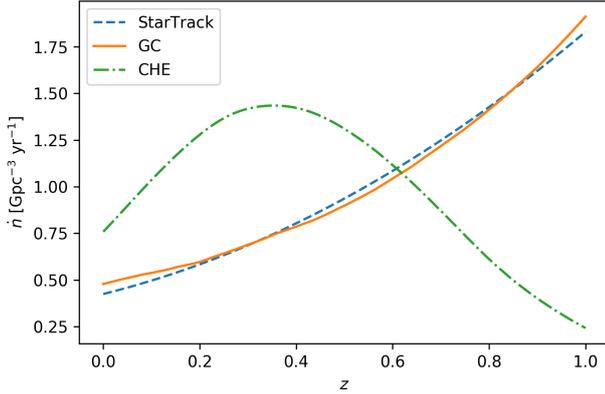


Fig. 2. Normalised merger rate distributions of the used models. The dashed, solid, and dot-dashed lines correspond to StarTrack, GC, and CHE models, respectively. StarTrack merger rate density is taken from model M10 from Belczynski et al. (2017).

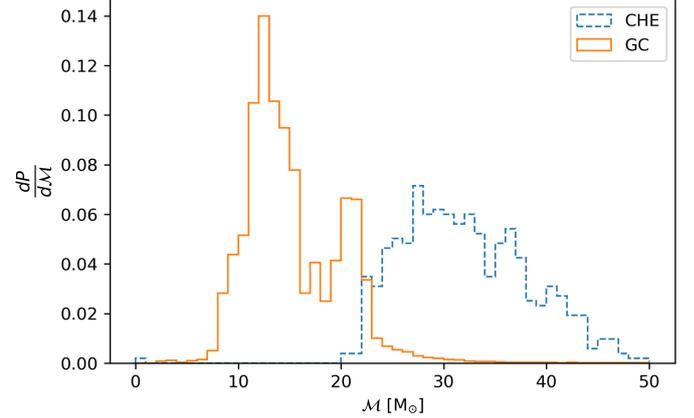


Fig. 3. Normalised intrinsic chirp mass distributions. The solid line corresponds to the GC model and the dashed line corresponds to the CHE model.

our analysis to two-dimensional chirp mass–luminosity distance distributions. For the purpose of this analysis, the BBH models have only two characteristic attributes: intrinsic chirp mass distribution and spatial distribution.

We have (Taylor et al. 2012)

$$\frac{d^4 N}{dt d\Theta dz d\mathcal{M}} = \frac{dV_c}{dz} \frac{\dot{n}(z)}{1+z} \mathcal{P}(\mathcal{M}) \mathcal{P}(\Theta), \quad (10)$$

where $\dot{n}(z)$ is the merger rate density containing information about the rate of merging of the BBH at a given redshift and $\mathcal{P}(\mathcal{M})$ is the observed chirp mass distribution (in contrast to the intrinsic chirp mass distribution used by Taylor et al. 2012). If $D_c(z) = D_H \int_0^z \frac{dz'}{E(z')}$ is comoving distance and V_c is comoving volume, then

$$\frac{dV_c}{dz} = \frac{4\pi D_c(z)^2 D_H}{E(z)}.$$

Integrating Eq. (10) over Θ , neglecting the time dependence, and substituting the variable z for D_L we get

$$\frac{d^2 N}{dD_L d\mathcal{M}} = \frac{4\pi D_c^2(z) D_H}{E(z) D_c(z) + D_H(1+z)} \frac{\dot{n}(z)}{1+z} \mathcal{P}(\mathcal{M}), \quad (11)$$

which we use as $P(O | M_i)$, setting \mathcal{B} to be $\{D_L, \mathcal{M}\}$.

Alternatively, integrating over \mathcal{M} and Θ we get

$$\frac{dN}{dz} = \frac{4\pi D_c(z)^2 D_H}{E(z)} \frac{\dot{n}(z)}{1+z}, \quad (12)$$

which is, when normalised, the probability of observing an event originating in the coalescence at the redshift z (regardless of its S/N). To choose the normalisation we limit redshift to z_{\max} , which is the solution of equation $\frac{A(M_c)\Theta_{\max}}{D_L(z)} = \rho_{\min}$ for the greatest mass in the model. Figure 2 presents the distributions of merger rate density. Because \dot{n} enters Eq. (12) linearly, and this equation itself is normalised, the absolute magnitude of merger rate density is unimportant.

In Fig. 3 we present the chirp mass distributions of CHE and GC models. The distributions for the StarTrack models can be found in Dominik et al. (2012); they typically average around $7 M_\odot$ for $Z = 1 Z_\odot$ and under $14 M_\odot$ for $Z = 0.1 Z_\odot$.

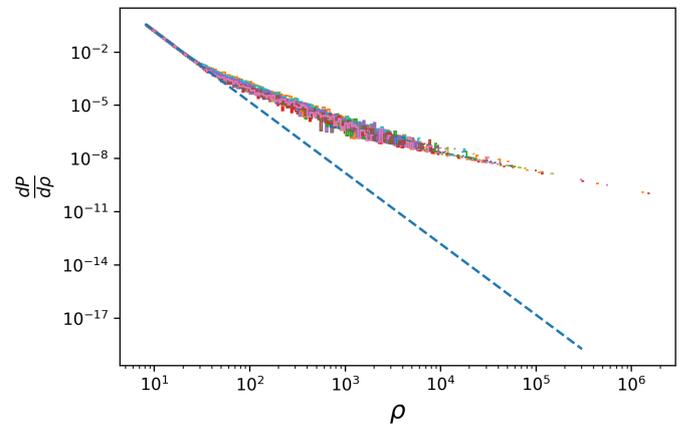


Fig. 4. Normalised S/N distribution. The dashed line represents the theoretical curve, with $\rho_{\min} = 8$. Other curves are the distributions of 16 simulated models (listed in Sect. 7.1) based on 10^5 binaries.

6. Luminosity distribution

Before investigating the mass–distance distribution, let us first consider a luminosity distribution. Because of the short range of Advanced LIGO (compared to other third-generation detectors, e.g., the *Einstein Telescope*) we can approximate the luminosity distance D_L with a physical distance R . In Eq. (1), we have $R \propto \rho^{-1}$. Then, assuming a uniform distribution of sources and a single value of mass, we have $N \propto V \propto R^3 \propto \rho^{-3}$ and $\frac{dN}{d\rho} \propto \rho^{-4}$. This raises a concern that, unless the BBH spatial distributions are sufficiently different, all S/N distributions may show little deviation from ρ^{-4} .

We demonstrate in Fig. 4 that for the majority of the events the distributions follow the theoretical curve, with only the most luminous events showing a difference. For this reason, we investigated the distributions of chirp mass–luminosity distance instead of S/N. The difference between the theoretical curve and calculations in the case of the most luminous events with $\rho > 10^2$ is mostly due to the mass distribution of merging BBHs. The heavy BBHs contribute to the tail with high ρ if the events are nearby, while for smaller values of ρ the S/N distribution is averaged over the large volume where events are detectable.

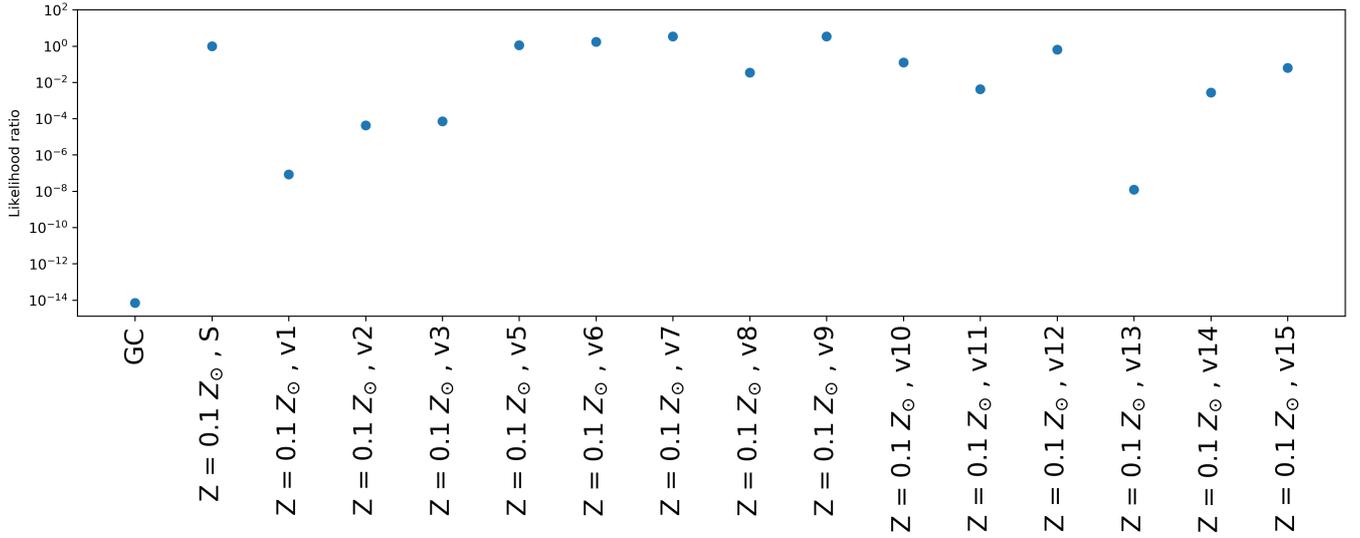


Fig. 5. Likelihoods for the 16 selected models that could explain all of the selected GW observations. The likelihoods are normalised by the likelihood of the Standard StarTrack model.

7. Results

7.1. Likelihoods

The updated list of GW observations can be found in the catalogues by [The LIGO Scientific Collaboration & the Virgo Collaboration \(2019\)](#) and [Abbott et al. \(2020\)](#). These observations are used in the calculation of the likelihood of explaining the observations by any single model (6). The likelihoods are calculated according to Eq. (4) with the chirp mass–luminosity distance taken as $P(D | M_i)$. In the calculation of probabilities, the errors are accounted for by weighting the probabilities with Gaussian distributions corresponding to the measurement uncertainty of the given observations.

Selected models with the highest likelihood are subsequently examined for distinguishability using Bayesian inference in Sect. 7.2. We present those likelihoods in Fig. 5. The demarcation point is the natural one; the likelihood of the next most likely model when compared to the likelihood of the Standard model is of the order of the numerical error.

All of the solar metallicity StarTrack models could explain only the closest observations with the smallest mass, with chirp mass of several solar masses and distance of less than 800 Mpc. On the contrary, the chirp mass distribution of the CHE model is too high to explain the mergers with chirp mass lower than $20 M_\odot$. The StarTrack models with $Z = 0.1 Z_\odot$ (with exception of the Variant 4 whose intrinsic mass distribution is too small) and GC are selected for calculations of thresholds. It is worth emphasising that none of the models are capable of explaining observations further than 3 Gpc.

Variant 9, characterised by no BH natal kick, is the most probable model with likelihood ≈ 3.5 times higher than the Standard model. Several models are slightly preferred to the Standard model: V5, V6, V7, and V9. Those models differ from the Standard model in that they have either a higher or lower NS mass and lower or no natal kick, respectively. Still none of these likelihoods are sufficiently high for the model to be a meaningful alternative to the Standard model. Any change to λ decreases the likelihood of the model by a factor of at least 100. An increase in wind mass loss (V11) reduces the likelihood by 8 orders of magnitude. Finally, the observations strongly disfavour the GC model, as it is less likely by 15 orders of magnitude.

7.2. Thresholds of distinguishability

For the calculation of the TOD, the data O_i are taken to have Gaussian errors with

$$\begin{aligned} \frac{\sigma_{D_L}}{D_L} &= \begin{matrix} +0.384 \\ -0.318 \end{matrix} \\ \frac{\sigma_M}{M} &= \begin{matrix} +0.112 \\ -0.077 \end{matrix}, \end{aligned}$$

which are average relative errors for all real observations. It should be noted that because the catalogues give a 90% symmetric interval error, the above errors were scaled to correspond to a standard deviation. Although the errors cited by the catalogues are not Gaussian, they differ from Gaussian mainly at the tails of the distribution and thus approximation is appropriate. The purpose of calculating the thresholds is to check whether likelihoods in the Fig. 5 are reliable with probability α , given the available number of observations. Table 2 suggests that two groups of parameters can be distinguished (i.e. thresholds for all pairs are under 47) with the current number of observations: wind mass reduction (Standard, V11) and $\lambda = \text{const}$ (V1, V2, V3, also for the purpose of the comparison Standard). Comparing this result with Fig. 5 we can conclude that changes to the Standard model parameters are unpreferable. Some groups of parameters are distinguishable, except for distinguishing with the Standard model. In this group there are: BH natal kick (V8, V9), mass transfer (V12, V13), and Nanjing λ (V14, V15). If for other reasons the Standard model would not be preferable, the results for those groups point to the likely correction of the parameters. The remaining groups: maximal NS mass (V5, V6), natal kick distribution (V7), and SN engine (V10) are not distinguishable. Finally, the GC, V11, and V13 models are (separately) distinguishable from the rest of the models. The overall situation may change if research based on other evidence limits the available model space.

An unexpected result is the strong anti-symmetry of the TOD for some models. We do not expect full symmetry in Tables 2 and 3 because the corresponding TODs are calculated using different observations. Still, in some cases, breaking of this symmetry is extreme enough that we propose an explanation based on the distributions for Standard StarTrack model and Variant

Table 2. Thresholds of distinguishability with p -value $\alpha = 0.1$.

–	GC	S	V1	V2	V3	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15
GC	–	4	3	4	10	4	3	4	4	4	4	3	3	5	5	4
S	4	–	13	35	22	n	n	n	22	n	155	5	28	6	58	28
V1	4	28	–	74	10	28	35	28	6	22	45	4	28	4	28	22
V2	6	35	17	–	28	35	45	35	8	35	58	4	35	5	533	17
V3	8	10	4	10	–	10	13	13	8	13	13	4	8	13	22	6
V5	4	121	17	45	22	–	n	n	22	n	121	5	28	6	58	28
V6	4	155	17	45	22	n	–	n	22	n	121	5	22	6	58	28
V7	4	121	13	28	22	255	255	–	28	n	74	5	13	6	45	22
V8	3	8	3	3	8	8	8	8	–	10	5	4	5	5	5	3
V9	4	95	10	22	22	121	121	199	35	–	45	5	17	6	45	17
V10	4	533	22	45	22	1832	874	417	17	326	–	5	155	6	74	35
V11	3	3	3	3	3	3	3	3	4	3	3	–	3	3	3	3
V12	4	n	22	45	22	n	n	n	17	n	417	5	–	6	74	45
V13	3	5	3	3	10	5	5	5	13	6	4	3	4	–	4	4
V14	6	22	8	35	121	22	22	22	8	22	22	4	17	8	–	10
V15	4	95	28	22	10	95	95	95	17	74	121	5	45	5	22	–

Notes. Displayed vertically are the tested models. Displayed horizontally are the models assumed to be correct, i.e. used to generate data O_i . Here, n signifies that the particular pair of models is indistinguishable even after 3000 observations.

Table 3. Thresholds of distinguishability with p -value $\alpha = 0.01$.

–	GC	S	V1	V2	V3	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15
GC	–	13	8	13	28	10	13	13	10	13	13	10	13	17	13	10
S	17	–	58	155	74	n	n	n	74	n	683	17	155	17	199	121
V1	13	95	–	199	28	95	74	74	22	74	95	13	95	10	95	58
V2	22	121	58	–	95	155	155	95	28	121	199	13	95	13	1832	58
V3	28	35	17	35	–	35	35	45	22	45	35	10	28	45	74	22
V5	17	417	58	155	95	–	n	n	58	n	417	17	95	17	199	121
V6	17	417	58	121	74	n	–	n	58	n	417	17	95	17	199	95
V7	13	417	45	95	74	874	874	–	95	n	199	17	74	17	155	74
V8	8	28	10	13	28	28	35	28	–	35	22	13	17	13	17	17
V9	13	255	45	74	74	417	417	533	95	–	199	17	58	17	121	58
V10	17	1832	58	199	74	n	3000	1431	45	1118	–	17	417	17	199	121
V11	13	3	6	5	3	3	3	3	13	3	3	–	5	3	3	5
V12	17	n	58	199	74	n	n	n	58	n	1431	13	–	17	199	155
V13	8	17	6	10	35	13	17	17	35	17	17	8	13	–	17	10
V14	22	74	28	95	533	74	74	74	28	74	74	13	45	22	–	28
V15	17	326	95	74	35	417	326	255	58	255	326	13	199	17	58	–

Notes. Displayed vertically are the tested model. Displayed horizontally are the models assumed to be correct, i.e. used to generate data O_i . Here, n signifies that the particular pair of models is indistinguishable even after 3000 observations.

12 with metallicity $Z = 0.1 Z_{\odot}$. If we assume that Variant 12 is correct, we can exclude the Standard model with only 28 observations (p -value = 0.9); yet the reverse is not true even for 3000 observations. The probability distribution of both models is concentrated in the area around $M \approx 28 M_{\odot}$. For Variant 12 the probability distribution is denser in this area than for the Standard model, while the latter distribution is overall more uniform. This fact is understated in Figs. 6 and 7 due to the logarithmic scale. When generating observations using Variant 12 one quickly exceeds the expected density for the Standard model, therefore the Variant 12 is distinguishable from the Standard model. On the other hand, more observations must be generated using the more uniform distribution of the Standard model in order to see that the density in the $M \approx 28 M_{\odot}$ area is not as high as expected for Variant 12.

8. Conclusions

Tables 2 and 3 show that given 100 observations we will likely be able to distinguish between almost 80% of the pairs of models. Given 1000 observations we will obtain p -value = 0.01 certainty in distinguishing between those models. Nevertheless, 16 out of the 240 pairs of models will remain indistinguishable, or rather we do not know how many observations we would need because of the imposed limit of 3000 observations in our analysis. The current number of observations allows us to distinguish between 173 out of 240 pairs of models with $\alpha = 0.1$ and between 102 pairs with $\alpha = 0.01$. This threshold of 100 observations will likely be reached once the full results of the next observational run O3 are published (Abadie et al. 2010). The most likely is Variant 9, followed closely by Variants 7, 6, and 5, and also the

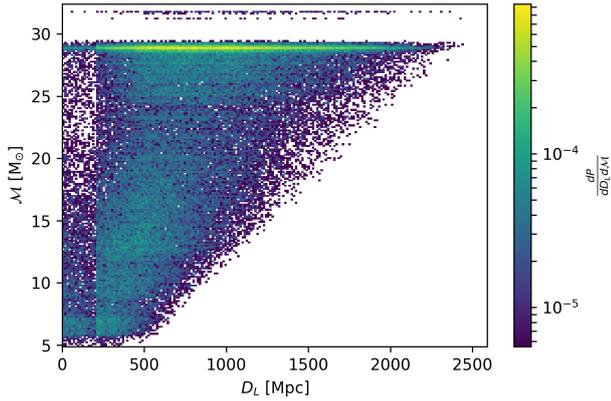


Fig. 6. Chirp mass–luminosity distance distribution for the Standard StarTrack model with $Z = 0.1 Z_{\odot}$.

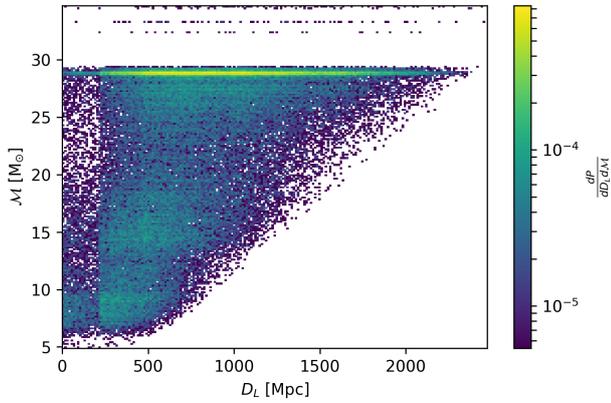


Fig. 7. Chirp mass–luminosity distance distribution for Variant 12 of the StarTrack model with $Z = 0.1 Z_{\odot}$.

Standard StarTrack model. Several StarTrack models with metallicity of $Z = 0.02$ are likely to explain observations from

O1 and O2, but no model is capable of explaining all observations from O3. The last runs contain mergers with chirp masses and distances that are greater than what the StarTrack models can account for. A possible solution is that the observed mergers followed several different channels of evolution and no single model can explain all of the observations.

Acknowledgements. Special thanks to Tomasz Bulik for indispensable discussion as well as the insights about stellar evolution physics. This work was supported by the TEAM/2016-3/19 grant from FNP.

References

- Abadie, J., Abbott, B. P., Abbott, R., et al. 2010, *Class. Quant. Grav.*, **27**, 173001
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016, *Phys. Rev. X*, **6**, 041014
- Abbott, R., Abbott, T. D., Abraham, S., et al. 2020, ArXiv e-prints [arXiv:2010.14527]
- Askar, A., Szkudlarek, M., Gondek-Rosińska, D., Giersz, M., & Bulik, T. 2017, *MNRAS*, **464**, L36
- Belczynski, K., Kalogera, V., & Bulik, T. 2002, *ApJ*, **572**, 407
- Belczynski, K., Kalogera, V., Rasio, F. A., et al. 2008, *ApJS*, **174**, 223
- Belczynski, K., Ryu, T., Perna, R., et al. 2017, *MNRAS*, **471**, 4702
- de Mink, S. E., & Mandel, I. 2016, *MNRAS*, **460**, 3545
- Dominik, M., Belczynski, K., Fryer, C., et al. 2012, *ApJ*, **759**, 52
- Finn, L. S. 1996, *Phys. Rev. D*, **53**, 2878
- Fryer, C. L., Belczynski, K., Wiktorowicz, G., et al. 2012, *ApJ*, **749**, 91
- Giersz, M., Heggie, D. C., Hurley, J. R., & Hypki, A. 2013, *MNRAS*, **431**, 2184
- Gregory, P. 2005, *Bayesian Logical Data Analysis for the Physical Sciences* (Cambridge: Cambridge University Press)
- Kroupa, P. 2001, *MNRAS*, **322**, 231
- Lipunov, V. M., Postnov, K. A., & Prokhorov, M. E. 1997, ArXiv e-prints [arXiv:astro-ph/9701134]
- Mandel, I., & de Mink, S. E. 2016, *MNRAS*, **458**, 2634
- Martynov, D. V., Hall, E. D., Abbott, B. P., et al. 2016, *Phys. Rev. D*, **93**, 112004
- Morscher, M., Umbreit, S., Farr, W. M., & Rasio, F. A. 2013, *ApJ*, **763**, L15
- Peters, P. C., & Mathews, J. 1963, *Phys. Rev.*, **131**, 435
- Rodriguez, C. L., Chatterjee, S., & Rasio, F. A. 2016, *Phys. Rev. D*, **93**, 084029
- Taylor, S. R., Gair, J. R., & Mandel, I. 2012, *Phys. Rev. D*, **85**, 023535
- The LIGO Scientific Collaboration, & the Virgo Collaboration 2019, *Phys. Rev. X*, **9**, 031040
- Webbink, R. F. 1984, *ApJ*, **277**, 355
- Xu, X.-J., & Li, X.-D. 2010, *ApJ*, **716**, 114