Current and future constraints on cosmology and modified gravitational wave friction from binary black holes

K. Leyde\textsuperscript{a,1}, S. Mastrogiovanni\textsuperscript{a,b}, D.A. Steer\textsuperscript{a}, E. Chassande-Mottin\textsuperscript{a}, C. Karathanasis\textsuperscript{c}

\textsuperscript{a}Université Paris Cité, CNRS, Astroparticule et Cosmologie, F-75006 Paris, France
\textsuperscript{b}Artemis, Université Côte d’Azur, Observatoire de la Côte d’Azur, CNRS, F-06304 Nice, France
\textsuperscript{c}Institut de Física d’Altes Energies (IFAE), Barcelona Institute of Science and Technology, Barcelona, Spain

E-mail: kleyde@apc.in2p3.fr, mastrosi@apc.in2p3.fr, steer@apc.in2p3.fr, ecm@apc.in2p3.fr, ckarathanasis@ifae.es

Abstract. Gravitational wave (GW) standard sirens are well-established probes with which one can measure cosmological parameters, and are complementary to other probes like the cosmic microwave background (CMB) or supernovae standard candles. Here we focus on dark GW sirens, specifically binary black holes (BBHs) for which there is only GW data. Our approach relies on the assumption of a source frame mass model for the BBH distribution, and we consider four models that are representative of the BBH population observed so far. In addition to inferring cosmological and mass model parameters, we use dark sirens to test modified gravity theories. These theories often predict different GW propagation equations on cosmological scales, leading to a different GW luminosity distance which in some cases can be parametrized by variables $\Xi_0$ and $n$. General relativity (GR) corresponds to $\Xi_0 = 1$. We perform a joint estimate of the population parameters governing mass, redshift, the variables characterizing the cosmology, and the modified GW luminosity distance. We use data from the third LIGO-Virgo-KAGRA observation run (O3) and find – for the four mass models and for three signal-to-noise ratio (SNR) cuts of 10, 11, 12 – that GR is consistently the preferred model to describe all observed BBH GW signals to date. Furthermore, all modified gravity parameters have posteriors that are compatible with the values predicted by GR at the 90% confidence interval (CI). We show that there are strong correlations between cosmological, astrophysical and modified gravity parameters. If GR is the correct theory of gravity, and assuming narrow priors on the cosmological parameters, we forecast an uncertainty of the modified gravity parameter $\Xi_0$ of 51% with $\sim$ 90 detections at O4-like sensitivities, and $\Xi_0$ of 20% with an additional $\sim$ 400 detections at O5-like sensitivity. We also consider how these forecasts depend on the current uncertainties of BBHs population distributions.

\textsuperscript{1}Corresponding author.
Introduction

The direct detection of gravitational waves (GWs) has produced a wealth of new results spanning a wide range of scientific areas. GW detections have clearly established the existence of a population of binary black holes (BBH) [1, 2], which otherwise eluded observations. Other types of GW sources including binary neutron stars (BNSs) [3] as well as composite systems...
with a neutron star and a black hole (NSBH) [4] have also been detected. The most recent GW transient catalog (GWTC-3) [2] observed with the LIGO-Virgo detectors [5–9] has reported nearly 100 compact binary sources. This has led to statistical studies of the population of these sources in terms of their distances, masses and spins [10–12], as well as to tests of general relativity (GR) (so far no deviations from GR have been observed) [13, 14].

GW observations can also constrain cosmological parameters in several different ways. If an electromagnetic (EM) signal is detected from the source, then its redshift \(z\) can be obtained from spectrometric measurements. On combining \(z\) with the luminosity distance obtained from gravitational wave data, it is possible to measure the Hubble constant \(H_0\), as first proposed in [15]. This ‘bright standard siren’ method was applied to the BNS merger GW170817 and its counterpart, resulting in an inferred value of \(H_0 = 70.0^{+12.0}_{-8.0}\) km s\(^{-1}\) Mpc\(^{-1}\) (notation for the 68\% confidence level) [16]. It is remarkable that a 20\% precision was achieved with this single event, showing that bright standard sirens are a complementary and independent method to measure \(H_0\) (see e.g. [17–19] for measurements using other methods). The accuracy of [16] can be further improved if the distance-inclination degeneracy can be broken, for example with astrophysical information about the gamma ray burst jet and its associated afterglow [20]. Using GW170817, it was also possible to measure the fractional difference, \(\mathcal{O}(10^{-16})\), between the propagation speed of GWs and EM-waves at LIGO-Virgo frequencies, which has led to constraints on different modified gravity theories (for a review, see [21] and papers within). Unfortunately, however, such joint observations of electromagnetic and gravitational waves are predicted to be rare in future science runs [22, 23].

Alternative approaches to determining \(z\) have been proposed that do not require a prompt electromagnetic signal, and can thus be applied to all types of GW sources, including BBH mergers. Following [15], a statistical estimate of the redshift can be obtained by identifying nearby galaxies in galaxy catalogs as potential hosts of the GW (compatible with its sky localization) [24–27]. This method has been applied to the BBH merger GW170814 [28] and to the event GW190814 [29], GWTC-2 [30] and to GWTC-3, resulting in \(H_0 = 68^{+13}_{-7}\) km s\(^{-1}\) Mpc\(^{-1}\) [31]. The GWTC-2 data in combination with galaxy catalog data has also been used to constrain cosmology and modified gravity theories in [32]. A similar idea is to cross-correlate the GW data with large scale structures from spectroscopic galaxy surveys (such as DESI or SPHEREx) [33–36]. Previous work also used this method to obtain forecasts to test modified gravity theories with third generation detectors [37, 38]. The degree of completeness of the electromagnetic survey is a key factor for the level of precision of the final redshift and hence cosmological parameter estimate.

In this paper we follow a different avenue, using solely compact binary source GW data, an approach often referred to as “dark GW sirens”. Indeed, the relation \(m_d = (1 + z)m_s\) between the source frame mass \(m_s\) and the detector frame mass \(m_d\) inferred from GW observations, together with the assumption that BBHs or BNSs each stem from a unique and universal mass distribution, can be used to infer the redshift. This approach has been used to fit both cosmological and source mass population parameters using BNSs (see [39] for 2nd generation (2G) detectors, [40] for 3G ones), and using BBHs (see [41] for 2G detectors, [42] for 3G ones, and [43] for both). Reference [41] estimates a potential precision for the Hubble parameter \(H(z = 0.7)\) of 6\% in an O4 like scenario.\(^1\) Note that this uncertainty depends on the assumed astrophysical priors such as the redshift distribution of sources.

\(^1\)Currently, the fourth observing run (O4) of the LIGO-Virgo-KAGRA collaboration is planned to start at the end of 2022. [https://www.ligo.org/scientists/GWEMalerts.php](https://www.ligo.org/scientists/GWEMalerts.php).
In the literature, some papers have started analyzing the potential of GW dark sirens to test for deviations from GR on cosmological scales. In this context, a standard approximation made in the literature is that only the dynamics of perturbations are modified relative to those of GR, so that the background evolution of the scale factor $a(t)$ is that of $\Lambda$CDM cosmology.

Focusing on tensor perturbations, the simplest setup is to consider modifications of gravity in which there are still 2 degrees of freedom propagating at the speed of light (at all energy scales), but whose energy dissipates differently relative to that in GR. In other words, the GW luminosity distance $d_{GW}^L(z)$ differs from the standard EM luminosity distance, see e.g. [44–46]. In [46], $d_{GW}^L(z)$ was parametrized in terms of two constants $\Xi_0$ and $n$; the corresponding form of $d_{GW}^L(z)$ was shown to be a good fit to certain modified gravity theories, including $f(R)$, non-local gravity, and others. In [47–49], focusing on models with extra spatial dimensions, $d_{GW}^L(z)$ was shown to depend on the comoving screening scale $R_c$ below which GR should be reproduced. A third, so called $c_M$-parametrization of $d_{GW}^L(z)$ has been proposed in [50–55], and models the effect of a time-dependent Planck mass in terms of dark energy content of the universe with a parameter $c_M$ (which vanishes in GR). Stability of this model has been studied numerically in [45].

In this context, [56] uses the GW dark siren method to constrain $c_M$: more explicitly, the assumption is that BBHs follow a broken power law mass distribution; then on fixing cosmological parameters such as the Hubble constant $H_0$, [56] uses the GWTC-2 catalog [57] together with upper-limits on the stochastic GW background to constrain $c_M$. Reference [58] also fixes cosmological parameters — in combination with a power law + peak mass model — and then uses GWTC-3 to constrain models with extra dimensions and a screening scale $R_c$. Reference [49] updated this work after the identification of a missing redshift-dependent factor in $d_{GW}^L(z)$. Finally, [59] does not fix cosmological parameters, and uses the dark standard siren method for the GWTC-3 events (with several signal-to-noise ratio (SNR) cuts), assuming a broken power law model only, and probes the parameters $\Xi_0$ and $n$. As we will stress below, see also [54], cosmological parameters are generally degenerate with modified gravity parameters, and so it is important to keep both free in order to avoid biased results.

In this paper we go significantly further in the analysis of GW dark sirens to test modified gravity and cosmology. Throughout we leave cosmological parameters free, and our aims are two-fold. First we want to probe the sensitivity of this method to the assumptions on the underlying mass distribution. We thus present a new analysis of the GWTC-3 catalog including, for the first time, a model selection of four different mass distributions and three modified gravity parametrisations. More precisely we study the broken power law, multi peak, power law + peak and truncated power law mass distributions, as well as the three parametrizations of $d_{GW}^L(z)$ outlined above. On calculating the associated Bayes factors, we show that the multi peak mass model is preferred over the other considered mass models, and find no evidence for modified gravity. We forecast which precision can be obtained for the measurement of the modified gravity parameter $\Xi_0$ with hundreds of GW events with future detector sensitivities [5–9, 60, 61]. We study how this measurement depends on the BBHs merger rate distribution uncertainties from current constraints [11] or on our ignorance of the value of the cosmological background parameters. Considering the credible intervals given by the current BBHs population uncertainties, we find that we will be able to constrain GR deviations with 20% precision with $\sim$ 500 BBHs detections. We find that our knowledge

\[2\] Where possible, we compare our results to those of [59], see sections 4 and 5. Note that we are also using independent codes.
of the cosmological background parameters does not strongly affect the precision on the GR deviation parameters. Moreover, we show that with the current distance reach, $\Xi_0$ is almost totally degenerate with the BBH redshift rate evolution, while this degeneracy will be broken with the extended reach enabled by future detector sensitivities.

The structure of the paper is as follows. Sec. 2 briefly reviews the different modified gravity luminosity distance parametrizations we consider. In Sec. 3 we introduce the Bayesian method and describe the population models that define the redshift and source frame mass distributions. In Sec. 4 the analysis is applied to GWTC-3 [2], and the results, including the posteriors of the modified gravity parameters are discussed. Sec. 5 describes forecasts with the future detectors and hundreds of GW events. We present our conclusions in Section 6.

2 Modified gravity and parametrization of the GW luminosity distance

Motivated by different problems of the standard cosmological model — such as the late-time accelerated expansion of the Universe — many theories have been proposed which modify GR on large scales (see [62–66] for reviews). In this paper, we assume that the background evolution is indistinguishable from a flat $\Lambda$CDM, described by the Hubble constant $H_0$, energy fraction in matter today $\Omega_m$ and in dark energy today $\Omega_\Lambda$, and consider theories in which there are still two tensor degrees of freedom as in GR. These generally satisfy a modified propagation equation, which in empty space is of the form (see e.g. [44])

$$h''_A + 2H(1 - \delta(\eta))h'_A + k^2 h_A = 0,$$

(2.1)

where $\eta$ is conformal time, $(.)' = \partial_\eta(.)$, $k$ is the comoving wavenumber, $A = (\times, \times)$ labels the two independent polarization components of the GWs, and the comoving Hubble parameter $H = a'/a$. Note that we have assumed that GWs propagate at the speed of light $c = 1$ at all frequencies. In GR, the function $\delta(\eta)$ vanishes; in modified gravity, it is non-zero, and its explicit form will depend on gravity theory under consideration.\footnote{In the specific case of Horndeski theories, $\delta(\eta)$ is one of the four functions of time which fully characterise linear perturbations around a cosmological background [53]. The stability of scalar and tensor perturbations imposes certain conditions on these 4 functions see e.g. [53, 67], which also must be satisfied by any parametrisation of those functions. See [68, 69] for further discussions and concrete examples.}

Eq. (2.1) can be solved using the WKB approximation to obtain the GW amplitude $h_A$ e.g. [45, 70, 71]. This scales as the inverse of the GW luminosity distance\footnote{For simplicity, we use $\delta(z)$ as a shorter replacement of $\delta(\eta(z))$.}

$$d_{GW}^L(z) = d_{EM}^L(z) e^{-\int_0^z dz' \frac{4(z')}{H(z')}}.$$  

(2.2)

where $d_{EM}^L(z)$ is the standard EM luminosity distance which, in a flat $\Lambda$CDM universe in the matter era (the GW events we consider are at $z \lesssim 1$), is given by

$$d_{EM}^L(z) = (1 + z) \int_0^z \frac{dz'}{H(z')}.$$  

(2.3)

with Hubble parameter

$$H(z) = H_0 \sqrt{\Omega_m(1 + z)^3 + \Omega_\Lambda}$$

and $\Omega_\Lambda = 1 - \Omega_m$. In this paper, following others (see [72] for a review), we focus on three parametrizations of $d_{GW}^L(z)$.
2.1 \( (\Xi_0, n) \) parametrization

The first parametrization of \( d_{GW}^L \) we consider was proposed in [71] and is given by

\[
d_{GW}^L = d_{EM}^L \left( \Xi_0 + \frac{1 - \Xi_0}{(1 + z)^n} \right),
\]

(2.4)

where both \( \Xi_0 \) and \( n \) are assumed positive. For \( z \ll 1 \), the two luminosity distances coincide, whereas for \( z \gg 1 \), \( d_{GW}^L \rightarrow \Xi_0 d_{EM}^L \). Notice that for \( \Xi_0 < 1 \), the GW luminosity distance is lower, which means that the GW source can be seen to larger (EM) distances than in GR. Thus, one expects to see a higher number of sources.

The parametrization of Eq. (2.4) has been shown to be a good fit to a number of different modified gravity theories. In the context of scalar-tensor theories, and in particular Horndeski theories [73, 74], when imposing that the speed of GWs is equal to the speed of light, \( \delta \) is related to one of the functions on the Horndeski Lagrangian (more exactly, \( G_4(\phi) \), see [72]). Indeed, physically this function \( G_4(\phi) \) leads to a time dependent Planck mass, and is the source of the modified friction term in Eq. (2.7). The two parameters \( \Xi_0 \) and \( n \) are then related to the change of this Planck mass at low and high redshifts [72]. Similar comments are true for DHOST theories [75, 76], scalar-tensor theories beyond Horndeski with second order equations of motion. The parametrization in Eq. (2.4) is also a good fit of the GW luminosity distance for a number of other theories, including \( f(R) \) gravity and non-local gravity: we refer the reader to Table 1 of [72] for a summary expressions for \( (\Xi_0, n) \) for these different theories (they depend on the parameters in the Lagrangian of the modified gravity theory, but potentially also on \( \Omega_m \)).

2.2 Extra dimensions

Many models of modified gravity have their origins in higher-dimensional spacetimes, for instance DGP models [47]. A characteristic of these theories is the existence of a new length scale, the comoving screening scale \( R_c \). Below this scale, the theory must pass the standard tests of general relativity, thus on scales \( d \ll R_c \), we expect to recover \( d_{GW}^L \sim d_{EM}^L \). On scales \( d \gg R_c \), modifications of gravity can become large and all the extra dimensions are probed by the gravitational field whose potential is therefore scales as \( \propto d^{-(D-2)/2} \), where \( D \) is the total number of spacetime dimensions. As a result, the relationship between the GW and EM luminosity distances can be parametrised by [48, 49]

\[
d_{GW}^L = d_{EM}^L \left[ 1 + \left( \frac{d_{EM}^L}{(1 + z)R_c} \right)^{\frac{D-4}{2}} \right],
\]

(2.5)

where \( n \) characterises the stiffness of the transition. Note that we define the parameter \( n \) differently with respect to [48]. In the following \( D, R_c \) and \( n \) will be taken as a free parameters, to be constrained with GW dark siren data.

2.3 \( c_M \) parametrization

The last parametrisation of \( d_{GW}^L \) we consider has been extensively used in the literature, see e.g. [50–53, 70]. It assumes that the modified friction term \( \delta(z) \) in Eq. (2.1) is proportional to the fractional dark energy density \( \Omega_{\Lambda}(z) \), namely

\[
\delta(z) = -\frac{c_M}{2} \frac{\Omega_{\Lambda}(z)}{\Omega_{\Lambda}} = -\frac{c_M}{2} \frac{1}{(1 + z)^3 \Omega_m + \Omega_{\Lambda}}.
\]

(2.6)
Gravity is thus modified at late times when dark energy dominates. This parametrization is not a good description of \( f(R) \) models \cite{77}. However, the advantage of (2.6) is its simple parametrization in terms of a single constant \( c_M \). It then follows from Eq. (2.2), see \cite{55}, that

\[
d_{\text{GW}}^L = d_{\text{EM}}^L \exp \left[ \frac{c_M}{2\Omega_\Lambda} \ln \frac{1+z}{[\Omega_m(1+z)^3 + \Omega_\Lambda]^{1/3}} \right]. \tag{2.7}
\]

In the context of bright GW standard sirens with EM counterparts, this parametrization of the luminosity distance has been investigated in \cite{54,55}. We will use it for dark sirens below.

3 Analysis framework

Starting from the observed set of \( N_{\text{obs}} \) GW detections associated with the data \( \{x\} = (x_1, ..., x_{N_{\text{obs}}}) \), such as the measured component masses and luminosity distance, we wish to infer hyperparameters \( \Lambda \) that describe the properties of the source population as a whole.

A hierarchical Bayesian analysis scheme can be used to calculate the posterior distribution of \( \Lambda \) \cite{78–80}, namely

\[
p(\Lambda|x, N_{\text{obs}}) \propto p(\Lambda) e^{-N_{\text{exp}}(\Lambda)} \left[ N_{\text{exp}}(\Lambda) \right]^{N_{\text{obs}}} \prod_{i=1}^{N_{\text{obs}}} \int \frac{p(x_i|\theta, \Lambda)p_{\text{pop}}(\theta|\Lambda)d\theta}{p_{\text{det}}(\theta, \Lambda)p_{\text{pop}}(\theta|\Lambda)d\theta}, \tag{3.1}
\]

where \( p(\Lambda) \) is a prior on the population parameters, \( \theta \) denotes the set of parameters intrinsic to each GW event, such as component spins, masses, luminosity distance, sky position, polarization angle, inclination, orbital angle at coalescence and the time of coalescence, \( N_{\text{exp}}(\Lambda) \) is the expected number of GW detections for a given \( \Lambda \) and a given observing time \( T_{\text{obs}} \), while \( N_{\text{obs}} \) is the number of detected events during an observation time \( T_{\text{obs}} \). The GW likelihood and probability of detecting a GW event with parameters \( \theta \) are denoted by \( p(x_i|\theta, \Lambda) \) and \( p_{\text{det}}(\theta, \Lambda) \), respectively. These expressions will depend on the sensitivity of the detector network. Furthermore, \( p_{\text{pop}}(\theta|\Lambda) \) represents the population-modeled prior. While the numerator in Eq. (3.1) accounts for the uncertainty on the measurement of the binary properties, the denominator correctly normalizes the posterior and includes selection effects \cite{78}.

The most probable values of \( \Lambda \) correspond to the population parameters that best fit the observed distribution of binaries, both in terms of the intrinsic parameters and of the number of events detectable in a given observation time. The population-modeled prior \( p_{\text{pop}}(\theta|\Lambda) \) is central for the hierarchical Bayesian analysis. When linking the redshift of the GW events with their luminosity distance (as measured from the data), the distribution of the component source frame masses and redshift is particularly important \cite{81}.

The hyperparameters \( \Lambda \) include: a set of cosmological background parameters \( H_0 \) and the matter energy density \( \Omega_m \), the parameters related to the GW propagation \( \Lambda_\alpha \) (see Sec. 2) and parameters used to describe the population of BBHs in source masses \( \Lambda_m \) and in redshift \( \Lambda_z \) (see Sec. 3.1-3.2). In this work, following \cite{31}, we assume that the source frame mass distributions of BBHs masses are independent from their redshift distribution, namely

\[
p_{\text{pop}}(m_{1,s}, m_{2,s}, z|\Lambda) = p(m_{1,s}, m_{2,s}|\Lambda_m) p(z|\Lambda_z, H_0, \Omega_m). \tag{3.2}
\]

We use phenomenological models for the source mass distribution \( p(m_{1,s}, m_{2,s}|\Lambda_m) \) and the (dimensionless) source spatial distribution \( p(z|\Lambda_z, H_0, \Omega_m) \). Once a population prior is pro-
vided, the number of expected detections can be calculated as

\[
N_{\text{exp}}(\Lambda) = R_0 T_{\text{obs}} \int p_{\text{det}}(m_{1,s}, m_{2,s}, z, \Lambda) \frac{p(m_{1,s}, m_{2,s}, \Lambda_m) f(z|\Lambda_z) \frac{dV_c}{dz}(H_0, \Omega_m) dz \, dm_{1,s} \, dm_{2,s}}{1 + z},
\]  

(3.3)

where \(R_0\) is the merger rate density today in units of Gpc\(^{-3}\)yr\(^{-1}\) (it defines the number of events per comoving volume per unit source frame proper time), and \(V_c\) is the comoving volume. The dimensionless function \(f(z|\Lambda_z)\) describes the evolution of the BBH merger rate with redshift. It is related to the merger rate of the sources as function of the redshift, as we now discuss.

### 3.1 Population-modeled priors: redshift

We assume the redshift prior is of the form

\[
p(z|\Lambda_z, H_0, \Omega_m) \propto f(z|\Lambda_z) \frac{dV_c}{dz}(H_0, \Omega_m).
\]

(3.4)

If \(f(z|\Lambda_z)\) is constant, all sources are distributed constantly in comoving volume. This describes all merging binaries including the ones not observable due to current sensitivities. The observed population can be obtained by accounting for the sensitivity of the detector network, namely by weighing each source with the probability \(p_{\text{det}}(\theta)\), see for example [78] for the explicit procedure. We model the rate evolution function heuristically as

\[
f(z|\gamma, \kappa, z_p) = \left(1 + \frac{1}{(1 + z_p)^\gamma + \kappa}\right) \frac{(1 + z)^\gamma}{1 + \left(\frac{1+z}{1+z_p}\right)^{\gamma+\kappa}},
\]

(3.5)

with three parameters that simply describe an initially increasing rate with an exponent \(\gamma\), followed by a decay with an exponent \(-\kappa\) for redshifts larger than the peak redshift \(z_p\). The redshift rate happens to be of the same form as the Madau-Dickinson star formation rate [82]\(^5\). However, we want to stress that Eq. (3.5) is not intended to model the star formation rate, nor the binary formation rate, but it is a simple model for the binary merger rate. The wide priors on the merger rate evolution we consider are given in App. A and result in a generic redshift distribution that can significantly differ from the one of the star formation rate.

### 3.2 Population-modeled priors: masses

The mass distributions are based on the four phenomenological models used in [11] to describe the primary mass source frame distribution. In essence these models describe the so called primary (heavier) source frame mass distribution as a power law between a minimum and maximum mass with a slope parameter. In order to account for possible accumulation points, certain models add overdensities in the mass spectrum governed by additional parameters, as we will now elaborate.

These models are designed to capture the current state of knowledge about the formation of stellar-mass BH. The pair instability supernovae process [84–98] foresees that BHs of masses larger than 40 – 50 M\(_\odot\) cannot be formed. Stars with higher masses lose either part of their mass (which motivates the inclusion of an accumulation point) or are entirely disrupted (which motivates the inclusion of a maximum mass).

\(^5\)Typical parameter values for the star evolution are \(\gamma \sim 1.9, \kappa \sim 3.4, z_p \sim 2.4\) [83].
The models of the study represent different attempts to fit the currently available data: the Truncated Power Law model, the Power law + peak: a Truncated Power Law model supplemented by a Gaussian distribution, or by two Gaussian peaks that is referred to as the Multi Peak model, and the Broken Power Law model. The secondary mass is described by a Truncated Power Law ranging from the same minimum mass as the primary mass distribution, to a maximum value given by the primary mass. The simplest model (Truncated Power Law) has sharp cutoffs, while the other three remaining models have a tapering window applied to the low-end boundary of their distribution. We provide more details about these mass models in App. A.

Overall, the global model including source population, cosmology and modified gravity aspects includes 11 to 20 parameters that are reviewed in App. A and Tables 3 and 4. The following sections present analyses that are based on a combination of the models presented above: Sec. 4 compares the four mass models, the three modified gravity models and GR, yielding a total of 16 combinations. Concerning the forecast of Sec. 5, the analysis is restricted to the power law + peak mass model in combination with the $\Xi_0$ modified gravity model.

4 Application to GWTC-3

This section presents the results of the joint parameter estimation of the mass distribution, redshift evolution and modified gravity parameters applied to the events of the GWTC-3 catalog [2]. A total of 48 runs are performed using the four mass models discussed in Sec. 3.2 combined with four different models for gravity: the baseline model is referred to as “GR” based on $\Lambda$CDM cosmology (no modification of GW propagation), and the three modified gravity models, introduced in Sec. 2 with a flat $\Lambda$CDM background, are referred to as $\Xi_0$, $D$ and $c_M$. The robustness of the results is evaluated by reproducing the analysis with three different SNR cuts.

The choices of the prior distributions for the mass models are summarized in App. A. The three parameters $\gamma$, $\kappa$ and $z_p$ of Eq. (3.5) used to model the merger rate evolution with redshift are generated from uniform priors as $\gamma \in U(0, 12)$, $z_p \in U(0, 4)$ and $\kappa \in U(0, 6)$. We use a uniform prior for $H_0$ compatible with values from CMB [99] and standard candle supernovae measurements [18], while we fix $\Omega_m = 0.3065$ [99]. As shown in [54] $H_0$ is expected to be correlated with the GR modification parameters. If $H_0$ and $\Omega_m$ are fixed to incorrect values this could potentially bias the estimation of the GR modification parameters. Thus, if a deviation from GR is concluded from the analysis, the assumed cosmological priors should be questioned. This motivates the additional use of “agnostic” or wide priors on the cosmological parameters. For the GW propagation parameters, we use a uniform prior in $\Xi_0 \in [0.3, 20]$ and $n \in [1, 100]$, for $c_M$ a uniform prior $c_M \in [-10, 50]$ and for the number of spacetime dimensions a uniform prior in the interval $D \in [3.8]$ is applied. We intentionally exclude negative values of $c_M < -10$ since otherwise $d_L^{GW}$ could decrease with redshift (for $z$ large). We have verified that this prior choice along with $z < 20$ used in the analysis produces a strictly increasing $d_L^{GW}$ in redshift.

We will see below that our results favor the baseline GR model and obtain posteriors for the modified gravity parameters that overlap with their “null” values in GR.

4.1 Description of the data set

Following the selection adopted in the last measurement of cosmological parameters performed by the LIGO-Virgo-Kagra collaboration (see [31] and associated public data release), we
select from the third GW transient catalog (GWTC-3) 42 GWs events consistent with a BBH source type with a network SNR > 11 and an Inverse False Alarm Rate (IFAR) of > 4 yr (taking the maximum value over the detection pipelines) [2]. These selection criteria exclude the asymmetric mass binary GW190814 [100] and the events associated with possible BNSs and NSBHs: GW170817, GW190425, GW200105, GW200115, GW190426 and GW190917 [12]. The event GW190521 [101] challenges our understanding of black hole formation from massive stars, and in particular of the pulsational pair instability supernova (PPISN) theory [91, 96–98]. Despite its high mass, it is compatible with the observed population of BBHs and its inclusion does not change the estimation of the cosmological and modified gravity parameters significantly [31, 59]. We therefore include it in the analysis. Given that some events such as GW200129_065458 show different mass ratio estimates depending on the waveform approximant, we use posterior distributions combined from the IMRPhenom [102, 103] and SEOBNR [104, 105] waveform families.

In order to check for possible systematics in the computation of selection biases, we also consider different choices for the SNR cut. While keeping the IFAR threshold fixed, we use a lower SNR cut of 10, leading to 60 selected events and a higher SNR cut of 12, leading to 35 selected events.

4.2 Results

4.2.1 Model selection

We will begin by discussing our findings in terms of model selection. Table 1 compares the Bayes factors between the different models and the GR+MULTI PEAK model for a given SNR cut. Fig. 1 and Tab. 2, give the marginal posteriors of the propagation parameters for the four source mass distributions and the three SNR cuts.

For all SNR cuts, we find that the preferred model is GR with a MULTI PEAK source mass model. The preference for the MULTI PEAK model is consistent with [12] and [31]. This is indicative of the two accumulation points around 10 M⊙ and 35 M⊙ in the BBH mass spectrum. The high Bayes factor for GR with current GW observations (given the priors on the cosmological parameters) shows that the introduction of additional modified gravity parameters to fit the observed BBH distribution is unnecessary. This is confirmed by the fact that the marginalized posteriors for the modified gravity parameters (see Fig. 1) are (within the 90% CI) consistent with their predicted GR values in all cases. If the observed mass distribution cannot be captured accurately by the mass model, the posteriors of the cosmological and modified gravity parameters can be considerably biased [81]. We find that at the current sensitivities and the quoted prior on the redshift distribution, the impact of the specific source frame mass model on the Ξ0 measurement is within the 90% uncertainty. Note that the TRUNCATED model is the least preferred model in terms of Bayes factors. This is a consequence of the TRUNCATED model not describing the observed mass spectrum of BBHs. The analysis shows furthermore that the MULTI PEAK mass model is preferred by a factor of 10 with respect to the simple BROKEN POWER LAW model. See App. C for further discussion on the mass model selection and related effect on the estimation of the modified gravity parameters.

With the nominal SNR cut at 11 and a source frame MULTI PEAK mass model, the number of spacetime dimensions is constrained to D = 4.6^{+2.6}_{-0.8}, the phenomenological Ξ0 = 2.1^{+3.6}_{-1.4} and the running Planck mass to cM = 1.2^{+4.4}_{-4.8}. Our results concerning Ξ0 with the BROKEN POWER LAW mass model are also consistent with the results shown in [59] at the 1σ level, for an SNR cut of 10, 11 and 12. However, our uncertainties are larger, mostly due
60 BBH events, SNR > 10, IFAR > 4 yr

| Model                        | GR  | Multi Peak | Power Law + Peak | Truncated |
|------------------------------|-----|------------|------------------|-----------|
| Broken Power Law             | -2.4| 0.0        | -1.2             | -6.3      |
| Multi Peak                  | -2.0| -0.2       | -1.7             | -6.4      |
| Power Law + Peak            | -3.2| -0.9       | -2.1             | -6.8      |
| Truncated                   | -3.0| -1.0       | -2.1             | -6.5      |

42 BBH events, SNR > 11, IFAR > 4 yr

| Model                        | GR  | Multi Peak | Power Law + Peak | Truncated |
|------------------------------|-----|------------|------------------|-----------|
| Broken Power Law             | -1.5| 0.0        | -0.8             | -3.2      |
| Multi Peak                  | -1.5| -0.0       | -0.9             | -3.4      |
| Power Law + Peak            | -1.9| -0.6       | -1.4             | -3.9      |
| Truncated                   | -1.9| -0.9       | -1.7             | -3.4      |

35 BBH events, SNR > 12, IFAR > 4 yr

| Model                        | GR  | Multi Peak | Power Law + Peak | Truncated |
|------------------------------|-----|------------|------------------|-----------|
| Broken Power Law             | -1.2| 0.0        | -1.1             | -2.6      |
| Multi Peak                  | -1.1| -0.4       | -1.2             | -2.8      |
| Power Law + Peak            | -2.1| -1.0       | -1.9             | -3.3      |
| Truncated                   | -1.9| -1.2       | -1.9             | -3.1      |

**Table 1**: Logarithm of the Bayes factor normalized to the preferred model (GR + Multi Peak) \( \log_{10} \left( \frac{p(\text{data}|\text{gravity model}, \text{mass model})}{p(\text{data}|\text{GR, multi peak})} \right) \) for different SNR cuts assuming a narrow prior for the cosmological parameters (cf. Table 3 of App. A). The preferred BBH mass and modified gravity model are highlighted in bold. The smaller the Bayes factor, the less likely the data can be explained by the given model.

the larger assumed prior on the Hubble constant. Excluding the prior of the rate of events \( R_0 \) and the priors on the cosmological parameters (both of which we assume to be more broad), our variable ranges are almost identical to the ones of [59]. Furthermore, different selection criteria using the maximum of SNR among the detection pipelines, lead to a slight variation of chosen events for a SNR cut of 11 when compared to [59]. The \( c_M \) model in [56] is constrained as \( c_M = -3.2^{+3.4}_{-2.0} \) for GWTC-2. The analysis presented here obtains a compatible value with a similar or larger uncertainty (depending on the SNR cut) for GWTC-3. Finally, [58] finds that the number of extra spacetime dimensions is \( D = 3.95^{+0.09}_{-0.07} \) with GWTC-3, when fixing the screening scale to 1 Mpc. However, the model of [58] lacks the correction of the redshift factor \((1+z)\) in the GW luminosity distance discussed in [49]. Nevertheless, our constraints
60 BBH events, SNR > 10, IFAR > 4 yr

|          | Broken Power Law | Multi Peak | Power Law + Peak | Truncated |
|----------|------------------|------------|------------------|-----------|
| $D$      | $5.8^{+2.0}_{-2.0}$ | $4.9^{+2.7}_{-1.2}$ | $4.8^{+2.8}_{-1.0}$ | $4.5^{+3.1}_{-0.8}$ |
| $\Xi_0$ | $1.6^{+2.3}_{-0.8}$ | $1.4^{+1.1}_{-1.1}$ | $1.3^{+1.2}_{-0.8}$ | $0.6^{+1.4}_{-0.2}$ |
| $c_M$    | $1.0^{+2.3}_{-2.6}$ | $0.5^{+2.5}_{-2.4}$ | $0.1^{+2.7}_{-2.1}$ | $-2.4^{+3.3}_{-1.4}$ |

42 BBH events, SNR > 11, IFAR > 4 yr

|          | Broken Power Law | Multi Peak | Power Law + Peak | Truncated |
|----------|------------------|------------|------------------|-----------|
| $D$      | $4.7^{+2.9}_{-0.9}$ | $4.6^{+2.6}_{-0.8}$ | $4.7^{+2.7}_{-0.9}$ | $4.8^{+2.8}_{-1.1}$ |
| $\Xi_0$ | $1.8^{+2.6}_{-1.2}$ | $2.1^{+3.6}_{-1.4}$ | $2.0^{+3.5}_{-1.3}$ | $0.7^{+3.0}_{-0.4}$ |
| $c_M$    | $0.5^{+4.1}_{-4.2}$ | $1.2^{+4.4}_{-4.8}$ | $1.1^{+4.3}_{-4.3}$ | $-2.5^{+5.2}_{-2.2}$ |

35 BBH events, SNR > 12, IFAR > 4 yr

|          | Broken Power Law | Multi Peak | Power Law + Peak | Truncated |
|----------|------------------|------------|------------------|-----------|
| $D$      | $4.8^{+2.8}_{-1.1}$ | $4.6^{+2.9}_{-0.9}$ | $4.8^{+2.9}_{-1.0}$ | $4.8^{+2.8}_{-1.1}$ |
| $\Xi_0$ | $1.2^{+1.4}_{-0.7}$ | $1.4^{+1.8}_{-0.8}$ | $1.4^{+1.8}_{-0.8}$ | $0.8^{+2.0}_{-0.5}$ |
| $c_M$    | $-0.1^{+2.8}_{-3.0}$ | $0.3^{+3.2}_{-3.3}$ | $0.4^{+3.2}_{-3.0}$ | $-1.8^{+4.6}_{-2.6}$ |

Table 2: Median and symmetric 90% confidence intervals of the modified gravity parameters and the selected mass models. The results are shown for different SNR cuts.

on the number of extra dimensions $D$ are compatible with [58].

Fig. 3 shows the 90% CI for the reconstructed GW luminosity distance-redshift relation (using the nominal SNR cut). Although the uncertainty levels for the models differ, all predicted GW luminosity distances overlap. Indeed, this is a good validity check of the analysis, since all modified gravity models use identical data.

One should also note, as shown in Fig. 3, that conversions of constraints between different modified gravity models are not trivial: When converting variables a prior is implicitly applied; a flat prior on $\Xi_0$ does not correspond to a flat prior on $D$ or $c_M$; one has to include the Jacobian for the change of variable (that will depend on the redshift value at hand). This complicates the comparison between different propagation models – for instance the rate of events today $R_0$ differs significantly for the $D$ modified gravity model with respect to the $c_M$ and $\Xi_0$ models. Appendix F of [59] derives an approximate relationship between the $c_M$ and $\Xi_0$ variables.

4.2.2 Correlation between model parameters

Let us now discuss the parameters of the analysis that most strongly correlate with the modified gravity parameters. In Fig. 2 we show the joint posterior distribution for the $\Xi_0$ model with the Multi Peak mass model and a SNR cut at 11. The $\Xi_0$ parameter correlates
Figure 1: The marginal distributions for the modified gravity parameters $D$ and $\Xi_0$ for all source mass models and for all three SNR cuts. Blue solid line: result obtained with 42 events with a SNR cut of 11. Orange dashed line: result obtained with a SNR cut of 12. Green dotted line: result obtained with an SNR cut of 10. The vertical black dashed lines indicate the value of the parameter in GR.

Significantly with the two parameters $\mu_{g,\text{high}}$ and $\mu_{g,\text{low}}$ which govern the mass features around $10 \, M_\odot$ and $34 \, M_\odot$, and with the rate evolution parameter $\gamma$. These correlations are analogous to the correlations between $H_0$ and $\gamma$, $\mu_{g,\text{high}}$ and $\mu_{g,\text{low}}$ observed in [31, 58, 59, 81].

Degeneracy between $\gamma$ and $\Xi_0$ — The posterior distribution in Fig. 2 shows a strong degeneracy between $\gamma$ and $\Xi_0$. Those two parameters appear to be approximately linearly related. The $\gamma$ vs $\Xi_0$ distribution exhibits a “ridge” of about constant height that corresponds to points with a comparable hierarchical likelihood that fit the data equally well. The projection of this ridge onto the $\gamma$ axis shown in the marginal distribution results in the rather high preferred value for $\gamma$.

Consequently, the data appear to be only informative on the ratio $\gamma/\Xi_0$: neither $\gamma$ nor $\Xi_0$ can be robustly measured individually. The constraint on $\Xi_0$ obtained with GWTC-3 is sensitive to the prior set on $\gamma$. As shown with Fig. 12 in App. C, larger priors on $\gamma$ correspond to a weaker constraint on $\Xi_0$. However, as detailed in Sec. 5, future detectors will observe much further. This will lead to the breakdown of the $\gamma - \Xi_0$ degeneracy, allowing more robust constraints on the individual parameters (cf. Sec. 5.2).

The same type of degeneracy is observed between $\gamma$ and $c_M$, and thus the same conclusions apply to the marginal distribution obtained with this other gravity model.
Figure 2: Joint posterior distribution obtained for the Multi Peak mass and $\Xi_0$ models using a GW event catalog produced with a SNR cut at 11. The blue line indicates the $\Xi_0 = 1$ value corresponding to GR and the contours indicate the 1 $\sigma$ and 2 $\sigma$ CI.

Degeneracy between $R_0$ and $\Xi_0$ — As opposed to [31, 81] (which are standard cosmological analyses and measure exclusively $H_0$), an extra correlation between $\Xi_0$ and the BBH merger rate density $R_0$ is also observed, as previously noted in [59]. The estimation of $R_0$ is related to the expected number of detected events $N_{\text{exp}}$. In fact, in the evaluation of the expected number of events, $H_0$ not only modifies the comoving volume as $1/H_0^3$ but also the redshift at which GW events will be detectable (since the SNR depends on the luminosity distance). These two effects roughly balance out such that the number of expected detections in a given time is weakly dependent on $H_0$. However, this is not the case when considering
modified GW propagation. The modified propagation leaves the comoving volume untouched (as it is defined with respect to the EM distance measure) but affects the average redshift at which it is possible to observe GW events. As a consequence, the number of expected detections per year strongly depends on the modified gravity parameters.

The extent of the correlation between $R_0$ and the modified propagation parameters depends on the assumed propagation model. In the case of the $\Xi_0$ model, $R_0$ can be constrained to a value similar to the one provided in [4]. This can be compared to Fig. 4, where the joint posterior of $R_0$ and the value of spacetime dimensions is shown. For this parametrization the correlation between $R_0$ and $D$ is high, affecting the obtained value of $R_0$: The long tail of large $D$ values allows for a tail of high rates — much larger values than in the GR case.

5 Forecasts with future GW detections

In this section the same analysis is repeated with simulated data sets representative of the upcoming O4 and O5 science runs. For each observation run we assume a 100% duty cycle for each detector. The LIGO detectors Hanford and Livingston (HL), are taken at the ‘Late High’ (O4) and ‘Design’ (O5) of the aLIGO noise levels [106]. For the Virgo detector (V) we simulate a ‘Late High’ (O4) and ‘Design’ (O5) noise curve [106].

We draw BBH masses from a fiducial POWER LAW + PEAK source frame mass population model (cf. Eq. A.7), compatible with studies from GWTC-3 [12]. The parameters of the model are $\alpha = 2.63$, $\beta = 1.26$, $m_{\text{min}} = 4.59 M_\odot$, $m_{\text{max}} = 86.22 M_\odot$, $\mu_g = 33.07 M_\odot$, $\sigma_g = 5.69 M_\odot$, $\lambda_{\text{peak}} = 0.1$, $\delta_m = 4.82 M_\odot$. This model includes only one sharp feature in the mass spectrum, thus representing a “pessimistic” scenario to constrain redshift with mass functions with respect to the MULTI PEAK model. Finally, the redshift evolution is assumed to follow the distribution of Eq. (3.5) with parameters $\gamma = 1.9$, $z_p = 2.4$ and $\kappa = 3.4$, similar to the ones considered in [83]. In particular, the power law exponent for low redshift events $\gamma$ is consistent with the most recent population study of LIGO-Virgo-KAGRA in [12].

\footnote{The values correspond to the median values obtained from GWTC-2 [11].}
Figure 4: The joint posterior distribution for the Multi Peak mass model and a SNR cut of 11. We show a subset of metaparameters, namely $D$, the GW propagation parameter (extra spacetime dimension), $R_0$, the rate of events today, $R_c$, the comoving screening scale, and $\gamma, z_p$ and $\kappa$, the parameters governing the redshift distribution of sources.
Two cases (both with $H_0 = 67.27$ km s$^{-1}$ Mpc$^{-1}$, $\Omega_m = 0.3166$ [17]) are analyzed: (i) The GR case with no GW propagation modification and (ii) A modified gravity model as proposed in [107], where $\Xi_0 = 1.8$, $n = 1.91$. Given the assumed population in redshift and source mass, a BBH merger rate density today of $R_0 = 23.9$ yr$^{-1}$ Gpc$^{-3}$ (compatible with current studies) and an observation period of 1 year, the GR case study predicts a total of 87 detections for O4 and 423 detections for O4+O5 with SNR $> 12$. In the modified gravity scenario, we take the deliberate choice of using the same mass and redshift BBH distributions and fix the same number of the GW detections. This imposes a rescaling of $R_0$ (or equivalently of the observing time) to the larger value of 61.4 yr$^{-1}$ Gpc$^{-3}$ since GW signals decay faster when $\Xi_0 > 1$. This choice of fixing the same number of GW data detections has been made to compare the constraints that we would be able to set on $\Xi_0$ in a GR and non-GR case, with the same amount of information from the data.

This section is organized as follows. In Sec. 5.1 the technical details of the analysis are provided, in Sec. 5.2 we forecast the precision of the modified GW propagation measurement for a GR universe and in Sec. 5.3 for a universe with $\Xi_0 = 1.8$ and $n = 1.91$. Finally, Sec. 5.4 discusses how current population uncertainties on the BBH distribution impact the forecasts.

5.1 Priors and other technical details of the analysis

For both cases (GR and modified gravity), two analyses are performed: (a) A full analysis of joint cosmological, redshift evolution, mass population and modified gravity parameters, using wide priors on the cosmological values. (b) A joint analysis of all parameters, but fixing the cosmology to measurement uncertainties from other probes such as the CMB [17]. The priors used are slightly different to the ones applied in Sec. 4, as reported in Tab. 3 and Tab. 4.

In order to evaluate the high-dimensional function $p(\Lambda | \{x\}, N_{\text{obs}})$ efficiently, the inference library bilby [108] and its Ensemble Monte Carlo Markov Chain implementation are used. The degree of convergence can be verified by studying the auto-correlation times of the sampling chains. We use 32 walkers with 50,000 to 100,000 convergence steps and we discard between 1 to 3 times the integrated auto correlation time steps as burn in.

Another key component is the generation of a proxy for the posterior samples associated with the intrinsic parameters for individual sources in the catalog. This proxy generation is based on a fit of the typical error made for the detector frame masses and luminosity distance. We use a new GW likelihood model calibrated using posteriors obtained from bilby. More details about the new GW likelihood model can be found in App. B.

In the following, all the results are reported at 68% symmetric credible intervals around the median. Relative uncertainties are computed from the average of the upper and lower sigma interval divided by the median.

5.2 Forecasts for a GR Universe

The Bayesian population inference scheme described in Sec. 3 is applied to the simulated data set described above. First, the results of the analysis for wide priors on the cosmological parameters are presented.

The Hubble constant is constrained at the 42% and 26% precision for O4 and O4+O5, respectively. The matter content $\Omega_m$ is essentially unconstrained for all runs. The complete results are shown in Fig. 13 of App. D where the posterior distributions obtained with agnostic priors for $H_0$ and $\Omega_m$ are shown for all parameters.

---

7 Though not essential for the present study, it is interesting to note that this value is compatible with the 2$\sigma$ CI obtained assuming GR in [12].
Figure 5: Mass and redshift distributions obtained for a source population simulated in the case of O4 and assuming GR. The population parameters are $\alpha = 2.63$, $\beta = 1.26$, $m_{\text{min}} = 4.59 \, M_\odot$, $m_{\text{max}} = 86.22 \, M_\odot$, $\mu_g = 33.07 \, M_\odot$, $\sigma_g = 5.69 \, M_\odot$, $\lambda_{\text{peak}} = 0.1$, $\delta_m = 4.82 \, M_\odot$. This event catalog contains 87 events for the O4 run.

Figure 6: Mass and redshift distributions obtained for a source population simulated in the case of O5 and assuming GR. We use the same parameters of the population as before. The O5 catalog contains 423 events. Compared to the O4 run, the detected events extend much further in redshift.

Assuming agnostic priors on the cosmological parameters, we obtain $\Xi_0 = 1.40^{+1.03}_{-0.58}$ and $\Xi_0 = 1.27^{+0.41}_{-0.33}$ for O4 and O4+O5 combined, respectively. The secondary parameter $n$ that governs modified gravity remains unconstrained, even in the best case with 510 events for O4+O5. This can be anticipated as, from the construction of the modified gravity model and for the GR case with $\Xi_0 = 1$, the model is perfectly degenerate under a change of $n$ (cf. Eq. (2.4)). Thus, notwithstanding a very large number of events, $n$ will be unconstrained if $\Xi_0 = 1$.

Of the redshift evolution parameters $\gamma$, $\kappa$ and $z_p$, the exponent $\gamma$ for the late redshift
evolution is the most constrained, not surprisingly as this parameter relates to the rate evolution at low redshifts. Compared to the previous case using the GWTC-3 events, the joint posterior on $\Xi_0$ and $\gamma$ in Fig. 13 appears much more localized around the true value. The event redshift distribution shown in Figure 6b extends further into redshift than the GWTC-3 data, allowing the $\Xi_0$ vs $\gamma$ degeneracy to be broken.

The $z_p$ posterior is flat for large values and vanishes at lower values of $z$ (the simulated BBH merger rate is strictly incompatible with $z_p = 0$ as it increases at low redshifts). The early evolution parameter $\kappa$ has a posterior which is almost flat, mostly because few events are detected at redshifts $z > z_p$. For O4, the maximum mass of the population $m_{\text{max}}$ has a broad distribution, ranging from $81 M_\odot$ to $102 M_\odot$ at the 1 sigma confidence level. This interval is reduced to $[80, 92] M_\odot$ for O4+O5 combined. Since the mass distributions follow a steeply decreasing power law (with $\alpha = 2.63$), only few sources are informative on the location of the upper mass cutoff — most observed sources have low mass.

The power law exponent for the redshift evolution $\gamma$ and $R_0$ are strongly degenerate. This is expected as a low rate of events today can be compensated (at first order) by an increase of the number of sources at higher redshifts. Furthermore, we find strong correlations between $H_0$ and the characteristic mass scales of the GW population as elaborated previously in [81]. Since the Hubble constant relates the source’s luminosity distance to the its redshift, it shifts the mass distribution to lower or higher values: A lower Hubble constant places sources generally at lower redshift and thus, source mass and detector frame mass differ less. Conversely, given the measured distribution of detector frame masses, a shift of $H_0$ to larger values can be compensated for by shifting the mass scales to lower values.

We also provide results using restricted priors on $H_0$ and $\Omega_m$. As discussed previously, this results in decreased error bars for $\Xi_0$. Fig. 7 compares the marginal posterior of the modified gravity parameter $\Xi_0$ (after O4+O5) applying agnostic priors with applying Planck priors to the cosmological parameters. The uncertainty of $\Xi_0$ is reduced by a factor of 1.5 when Planck priors on the cosmological parameters are assumed. The uncertainties on the modified gravity parametrization are reduced to $\Xi_0 = 1.47^{+0.92}_{-0.57}$ for O4 and $\Xi_0 = 1.08^{+0.27}_{-0.16}$ for O4+O5.
Again, \( n \) is unconstrained in both observation scenarios. All the correlations mentioned above apply here as well. Even so, since \( H_0 \) is much better constrained, the correlation \( H_0 - \Xi_0 \) has a weaker impact – slightly changed cosmological values do not affect the final uncertainties of the variables that parameterize the modified GW luminosity distance.

Compared to [59], both in the case of narrow\(^8\) and wide priors on the cosmological values, we obtain \( \Xi_0 \) constraints after O4+O5 twice as worse. For wide (narrow) priors, we obtain a relative error of \( \sigma_{\Xi_0} = 29\% \) (20\%), whereas the aforementioned source obtains 14\% (10\%). There are several differences in these analyses. With respect to [59] we assume a year time span of observation instead of five years, a power law + peak mass model instead of a broken power law model, and a SNR cut of 12 instead of 8. This leads to a large discrepancy in the number of observed events, \( \sim 4700 \) in the aforementioned case, \( \sim 500 \) here. Additionally, we apply a uniform in logarithm prior to the rate \( R_0 \), whereas [59] uses a uniform prior. Due to the those different assumptions, it is difficult to trace the origin of this discrepancy.

When the modified gravity parameters are set to their corresponding GR values (\( \Xi_0 = 1 \)), the Hubble constant is constrained to 39\% and 24\% (for the 1\( \sigma \) uncertainty) for O4 and O4+O5, respectively. This can be compared to the uncertainty of the Hubble constant of 67\% recently reported in [31] from 47 GW dark sirens. The full results with no modified gravity parametrization (and a comparison to previous works) are provided in Appendix E.

5.3 Assessing deviations from general relativity with a \( \Xi_0 = 1.8, n = 1.91 \) universe

We will now depart from the GR scenario and present the results for a universe in which the GW luminosity distance follows the modified relation of Eq.(2.4) with \( \Xi_0 = 1.8, n = 1.91 \). We simulate a mass and redshift distribution identical to the distribution before (cf. the parameters in the caption of Fig. 5) and we consider again 87 and 423 GW detections as in Sec. 5.2. We choose to retain the same number of events as this allows to see how \( \Xi_0 \) affects the measurement precision while keeping the same underlying population. Assuming the observing time (or \( R_0 \)) is kept fixed, changing \( \Xi_0 \) would, in principle, result in a different number of events as \( \Xi_0 \) changes the observable volume. Figure 14 shows the results obtained for O4+O5 sensitivities with large priors on the cosmological parameters, leading to \( \Xi_0 = 2.98^{+2.06}_{-1.18} \) for O4 and \( \Xi_0 = 2.31^{+1.07}_{-0.67} \) for O4+O5.

For the cosmological parameters in a narrow range consistent with Planck uncertainties in [17] the results are as follows: \( \Xi_0 = 2.68^{+2.03}_{-1.01} \) for O4 and \( \Xi_0 = 2.06^{+0.92}_{-0.47} \) for O4+O5. With respect to the previous agnostic priors, these are improvements, of about 6\% (O4) and 20\% (O4+O5) on the full width of the \( \Xi_0 \) uncertainty. After the observation run O5, GR could be excluded at the 2.3 sigma level, if the Universe follows a \( \Xi_0 = 1.8 \) model.

Our precision for \( \Xi_0 \) from O4+O5 corresponds to an increase in uncertainty of 45\% and 69\%, respectively for the case of agnostic and narrow priors on the cosmological parameters, when compared to [59]. While [59] assumes five years of observation time, the number of observed events is three times as high, since a lower rate of events today is simulated. Thus, our increased error bars are compatible with the theoretically expected improvement of \( 1/\sqrt{3} \approx 58\% \) from the larger number of observed events. This reasoning is only true when the hierarchical posterior has gaussianized, (see [81]). Additionally, as we show in Sec. 5.3, the forecast depends on the underlying population and the observed events thereof.

\(^8\)Note that we compare the analysis with narrow priors on the cosmological values with the result of [59] when the cosmological parameters are fixed.
We end this section by investigating whether these O4+O5 constraints can give interesting information on a specific underlying theory. In particular we consider the class of quadratic degenerate higher-order scalar tensor (DHOST) theories with action given by \cite{76, 109, 110}

\[ S[g_{\mu\nu}, \phi] = \int d^4x \sqrt{-g} \left[ F_0(\phi, X) + F_1(\phi, X)\Box \phi + F_2(\phi, X)R + \sum_{I=1}^5 A_I(\phi, X)L^{(2)}_I \right], \] (5.1)

where \( X = g^{\mu\nu} \partial_\mu \phi \partial_\mu \phi \) with the five possible Lagrangians \( L^{(2)}_A \)

\[ L^{(2)}_1 = \phi^\mu \phi_{\mu\nu}, \quad L^{(2)}_2 = (\phi^\nu)^2, \quad L^{(2)}_3 = \phi^\nu \phi^\rho \phi_{\rho\sigma}, \]
\[ L^{(2)}_4 = \phi^\mu \phi^\nu \phi^\rho \phi^\sigma, \quad L^{(2)}_5 = (\phi^\rho \phi_{\rho\sigma})^2, \] (5.2)

with \( \phi^\mu \phi^\nu = \nabla_\nu \nabla_\mu \phi \), and \( \phi_\mu = \nabla_\mu \phi \). For the functions appearing in the action Eq. (5.1) we take the same form as in \cite{111}

\[ F_0 = c_2 X, \quad F_1 = 0, \quad F_2 = \frac{M_0^2}{2} + c_4 X^2, \quad A_3 = -8c_4 - \beta. \] (5.3)

For stability, the constant \( c_2, c_4 \) and \( \beta \) must satisfy \cite{111}

\[ c_2 > 0, \quad c_4 > 0, \quad \beta < -\frac{32}{3} c_4. \] (5.4)

We chose \( M_0 \) to be the Planck mass. Given these constants, one can calculate the friction term \( \delta(\eta) \) \cite{111}, and hence an effective \( \Xi_0 \) and \( n \). We have considered the parameter ranges \( c_2 \in [0, 60], \beta \in [-40, 0] \) and \( c_4 \in [0, 2] \), and found that \( |\Xi_0 - 1| \lesssim 0.3 \). Therefore, with O4+O5 sensitivities, we will not be able to constrain the theory any further.

### 5.4 Impact of underlying BBH population on the \( \Xi_0 \) determination

The precision with which one can infer GW propagation parameters such as \( \Xi_0 \) is also determined by the true underlying population of BBH mergers, which is currently uncertain.

To investigate the dependency of the \( \Xi_0 \) precision on the true underlying BBH population we further simulate 75 different populations with different minimum mass, position of the Gaussian peak, power law slope etc.. These 75 populations are taken at random from the population samples in \cite{9} of \cite{11}, and a full analysis, where we marginalize over the total rate of events \( R_0 \), is performed for each. The analysis relies on the same detectors and same sensitivities as in the O4+O5 scenario mentioned earlier. The event rate \( R_0 \) (or the observing time) is tuned to keep the number of detected events constant and equal to 87 for O4 and 423 for O5. This choice allows to single out the impact of the underlying BBH distribution on the measurement precision on \( \Xi_0 \). The cosmological parameters are fixed to the Planck values as before, and we assume \( \Xi_0 = 1 \) and \( n = 0 \). For each run, a synthetic population is generated based on a randomly chosen sample for the metaparameters (mass and redshift evolution) taken from the posterior samples in \cite{11}. To speed up the convergence of the Markov Chain, we neglect the uncertainties of detector frame masses and luminosity distance. Therefore, these simulations represent a lower limit on the precision that can be reached for \( \Xi_0 \). This provides us with the variability and robustness of the precision obtained for \( \Xi_0 \) with the current population uncertainties.

\[ \text{https://dcc.ligo.org/public/0171/P2000434/003/Population_Samples.tar.gz} \]
Figure 8: The set of $\Xi_0$ posteriors for various population realizations with different metaparameters, drawn according to the uncertainties in [11]. We have marked the best constrained run (in terms of 1 sigma) in red. We find that half of the runs have an uncertainty of $\sigma_{\Xi_0} \leq 12\%$ (1 sigma)). However, of these 75 runs, we find 4 outliers that have significantly larger uncertainties on $\Xi_0$, with $\sigma_{\Xi_0} \geq 43\%$.

Fig. 8 presents the distribution of the $\Xi_0$ posterior for the different underlying populations. When averaged over all 75 runs, an uncertainty of $\sigma_{\Xi_0} = 0.16$ is obtained, where $\sigma_{\Xi_0}$ denotes the 1 $\sigma$ error. If we include 90% of the runs (discarding the extreme 5% of all cases), we have a minimum uncertainty of $\sigma_{\Xi_0,\text{min}} = 0.05$ and a maximum uncertainty of $\sigma_{\Xi_0,\text{max}} = 0.43$. We find that 50% of runs, have an uncertainty of $\sigma_{\Xi_0} \leq 0.12$, while less than 20% of runs have an uncertainty $\sigma_{\Xi_0} \geq 0.20$.

As shown in App. F, there is a significant variability for different realizations of the same population. The uncertainties given here should not be taken as absolute estimates of the expected errors, since we assume no uncertainties of the intrinsic GW signal parameters. Indeed, when we compare the average 1 $\sigma$ interval to the 1 $\sigma$ interval from Sec. 5.2 we find a factor of $\mathcal{O}(2)$ difference. Thus, the scatter of 1 $\sigma$ confidence levels described above, should be multiplied by a factor of that order.

6 Conclusions

This work presented the current and future constraints on cosmology and modified gravity arising from dark GW sirens. We estimated the population distribution of BBHs jointly with cosmological background parameters and parameters governing the modified GW luminosity distance.
The analysis of the most recent publicly released data in \cite{12} shows no evidence for deviations from GR at cosmological scales. Current data set constraints on the three theories of gravity considered. Based on an event selection cut at a SNR of 11, in combination of a multi peak source frame mass model, the phenomenological parameter for the GW friction model is constrained to $\Xi_0 = 2.1^{+3.6}_{-1.4}$, the number of spacetime dimensions to $D = 4.6^{+2.6}_{-0.8}$, and the running Planck mass to $c_M = 1.2^{+4.4}_{-4.8}$. This study evidences the interdependence of these variables and the parameters that govern the population of BBH sources. Compared to existing works, for example \cite{59}, we have considered a larger number of modified gravity models, and a broader set of mass distributions. For those, we have explicitly calculated Bayes factors and find no evidence for modified gravity.

The results obtained with the GWTC-3 catalog agree with the current literature reviewed in Sec. 4, which is indicative of their robustness.

We also discussed the constraints on GR deviations that one can set with the next science runs O4 and O5. For future observation scenarios we find that the phenomenological parameter $\Xi_0$ can be constrained to 29\% (assuming wide priors on the cosmological parameters) and 20\% (assuming Planck priors on the cosmology) after O4 and O5 combined. In the case of a $\Xi_0 = 1.8$ and $n = 1.91$ universe, we find 38\% (wide cosmological priors) and 34\% (narrow cosmological priors) after O4+O5 combined. The value of these constraints depend on the simulated population, that is: the true metaparameters, and the specific events observed. To understand this effect, we have generated synthetic populations according to current population knowledge. Following O4+O5, the $\Xi_0$ posterior has an average uncertainty of 16\%, if $\Xi_0 = 1$ and if the intrinsic parameter uncertainties are neglected.

Due to different assumptions on future observation runs (such as different number of observed events, underlying mass models and selection criteria), the forecast of the $\Xi_0$ uncertainties in Sec. 5 are difficult to compare to existing results. However, our results are generically compatible with the result of \cite{59} in the forecast for modified gravity, see Sec. 5.3.

Further steps could include the effect of spins, which are expected to be correlated with the mass ratio of the two component masses \cite{112}. It was shown that lensing is likely to become non-negligible for events of O4 and O5 \cite{113, 114}. At the latest for detectors as LISA or third generation observatories such as Einstein Telescope and Cosmic Explorer \cite{115, 116} we expect these effects to become very important due to the longer light paths through the lensing distribution (see \cite{117} for LISA). Furthermore, one expects the metallicity of stars to change over cosmic history. From simulations of the resulting BHs mass distribution \cite{118}, this evolution carries over to the position of the pair instability mass accumulation point. We have also neglected the time delay of the star formation rate and BBH coalescence. How these vectors impact the cosmological estimate was recently investigated in \cite{119}. We would thus like to extend the present analysis to a mass population that varies with redshift. Since we do not use any galaxy catalogs, this analysis may be more robust against calibration uncertainties from galaxy catalogs such as the estimation of its completeness or the estimation of its luminosity function. However, waveform systematics and higher order modes are certain to have an impact of the cosmological parameter uncertainties and we leave it to future work to quantify the expected discrepancies.
Appendices

A Priors and models

The four phenomenological mass models that we employ (see App. B of [11]) can be written as a linear combination of two types of statistical distributions. The first is a truncated power law \( P(x|x_{\text{min}}, x_{\text{max}}, \alpha) \) described by a slope \( \alpha \), and lower and upper bounds \( x_{\text{min}}, x_{\text{max}} \) at which there is a hard cutoff

\[
P(x|x_{\text{min}}, x_{\text{max}}, \alpha) \propto \begin{cases} x^\alpha & (x_{\text{min}} \leq x \leq x_{\text{max}}) \\ 0 & \text{Otherwise.} \end{cases} \tag{A.1}
\]

The second is a Gaussian distribution with mean \( \mu \) and standard deviation \( \sigma \),

\[
G(x|\mu, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left[ -\frac{(x-\mu)^2}{2\sigma^2} \right]. \tag{A.2}
\]

The source mass priors for the BBH populations that we consider are factorized as

\[
\pi(m_1, s, m_2, s | \Phi_m) = \pi(m_1, s | \Phi_m) \pi(m_2, s | m_1, s, \Phi_m), \tag{A.3}
\]

where \( \pi(m_1, s | \Phi_m) \) is the distribution of the primary mass component while \( \pi(m_2, s | m_1, s, \Phi_m) \) is the distribution of the secondary mass component given the primary. For all of the mass models, the secondary mass component \( m_2, s \) is described with a truncated power law (PL) with slope \( \beta \) between a minimum mass \( m_{\text{min}} \) and a maximum mass \( m_{\text{max}} \)

\[
\pi(m_2, s | m_{\text{min}}, m_{\text{max}}, \alpha) = P(m_2, s | m_{\text{min}}, m_{\text{max}}, \beta), \tag{A.4}
\]

while the primary mass is described with the several models discussed in the following paragraphs. For some of the phenomenological models, we also apply a smoothing factor to the lower end of the mass distribution

\[
\pi(m_1, s, m_2, s | \Phi_m) = [\pi(m_1, s | \Phi_m) \pi(m_2, s | m_1, s, \Phi_m)] S(m_1, s | \delta_m, m_{\text{min}}) S(m_2, s | \delta_m, m_{\text{min}}), \tag{A.5}
\]

where \( S \) is a sigmoid-like window function that adds a tapering of the lower end of the mass distribution. See Eq. (B6) and Eq. (B7) of [11] for the explicit expression of the window function.

The four phenomenological mass models are summarized in the following list. In Table 4, we report the prior ranges used for the population hyper-parameters.

- **Truncated Power Law**: The distribution of the primary mass \( m_{1,s} \) is described with a truncated power law with slope \( -\alpha \) between a minimum mass \( m_{\text{min}} \) and a maximum mass \( m_{\text{max}} \).

\[
\pi(m_{1,s} | m_{\text{min}}, m_{\text{max}}, \alpha) = P(m_{1,s} | m_{\text{min}}, m_{\text{max}}, -\alpha). \tag{A.6}
\]

- **Power Law + Peak**: The primary mass component is modeled as a superposition of a truncated PL, with slope \( -\alpha \) between a minimum mass \( m_{\text{min}} \) and a maximum mass \( m_{\text{max}} \), plus a Gaussian component with mean \( \mu_g \) and standard deviation \( \sigma_g \),

\[
\pi(m_{1,s} | m_{\text{min}}, m_{\text{max}}, \alpha, \lambda_g, \mu_g, \sigma_g) = (1 - \lambda_g) P(m_{1,s} | m_{\text{min}}, m_{\text{max}}, -\alpha) + \lambda_g G(m_{1,s} | \mu_g, \sigma_g). \tag{A.7}
\]
• **Broken Power Law**: The distribution of $m_{1,s}$ follows a PL between a minimum mass $m_{\text{min}}$ and a maximum mass $m_{\text{max}}$. The broken power law model is characterized by two PL slopes $\alpha_1$ and $\alpha_2$ and by a breaking point between the two regimes at $m_{\text{break}} = m_{\text{min}} + b(m_{\text{max}} - m_{\text{min}})$, where $b$ is a number $\in [0, 1]$. The broken PL model is

$$
\pi(m_{1,s}|m_{\text{min}}, m_{\text{max}}, m_{\text{break}}, \alpha_1, \alpha_2) = P(m_{1,s}|m_{\text{min}}, m_{\text{break}}, -\alpha_1) + \frac{P(m_{\text{break}}|m_{\text{min}}, m_{\text{break}}, -\alpha_1)P(m_{1,s}|m_{\text{break}}, m_{\text{max}}, -\alpha_2)}{P(m_{\text{break}}|m_{\text{break}}, m_{\text{max}}, -\alpha_2)}.
$$

(A.8)

• **Multi Peak**: The distribution of $m_{1,s}$ is described as a PL between a minimum mass $m_{\text{min}}$ and a maximum mass $m_{\text{max}}$ with two additional Gaussian components with means $\mu_{g,\text{low}}, \mu_{g,\text{high}}$ and standard deviations $\sigma_{g,\text{low}}, \sigma_{g,\text{high}}$. The parameter $\lambda_g$ is the fraction of events in the two Gaussian components while the $\lambda_{g,\text{low}}$ is the fraction of events in the lower Gaussian component. We denote the set of parameters that describe the MULTI PEAK as $\Lambda_m = \{m_{\text{min}}, m_{\text{max}}, \alpha, \lambda_g, \mu_{g,\text{high}}, \sigma_{g,\text{high}}, \lambda_{g,\text{low}}, \mu_{g,\text{low}}, \sigma_{g,\text{low}}\}$.

$$
\pi(m_{1,s}|\Lambda_m) = (1 - \lambda_g)P(m_{1,s}|m_{\text{min}}, m_{\text{break}}, -\alpha) + \lambda_g \lambda_{g,\text{low}} G(m_{1,s}|\mu_{g,\text{low}}, \sigma_{g,\text{low}}) + \lambda_g (1 - \lambda_{g,\text{low}}) G(m_{1,s}|\mu_{g,\text{high}}, \sigma_{g,\text{high}}).
$$

(A.9)
| Parameter | Units | Prior GWTC–3 (Sec. 4) | Prior O4+O5 (Sec. 5) | Description |
|-----------|-------|------------------------|-----------------------|-------------|
| $\gamma$ | -     | $U(0, 12)$              | $U(0, 8)$             | Late redshift evolution in Madau-Dickinson |
| $z_p$    | -     | $U(0, 4)$              | $U(0, 6)$             | Characteristic redshift in Madau-Dickinson |
| $\kappa$ | -     | $U(0, 6)$              | $U(0, 8)$             | Early redshift evolution in Madau-Dickinson |
| $R_0$    | Gpc$^{-3}$yr$^{-1}$ | $U(0, 1000)$ | Log Uniform(0.1, 100) Log Uniform(10, 300) | Event Rate Today |
| $H_0$    | kms$^{-1}$Mpc$^{-1}$ | $U(65, 77)$ | $U(30, 130)$ Log Uniform(0.05, 0.4) Log Uniform(0.3082, 0.3250) | Hubble constant |
| $\Omega_m$ | -     | 0.3065                | $U(0.05, 0.4)$ Log Uniform(0.3082, 0.3250) | Matter content Universe today |
| $\Xi_0$  | -     | $U(0.3, 20)$          | $U(0.3, 10)$          | Modified gravity parameters, see Eq. (2.4) |
| $n$      | -     | $U(1, 5)$             | $U(1, 5)$             | Modified gravity parameters, see Eq. (2.7) |
| $D$      | -     | $U(3.8, 8)$           | -                     | Modified gravity parameters, see Eq. (2.5) |
| $R_e$    | Mpc   | Log Uniform(10, 10$^5$) | -                     | Modified gravity parameters, see Eq. (2.7) |
| $n$      | -     | Log Uniform(0.1, 100) | -                     | Modified gravity parameters, see Eq. (2.7) |
| $c_M$    | -     | $U(-10, 50)$          | -                     | Modified gravity parameters, see Eq. (2.7) |

**Table 3:** Overview of the model hyper-parameters along with their associated prior distribution that we assume for the full joint analysis with icarogw. We apply two priors to the cosmological parameters: agnostic priors spanning a wide range of values, as well as informative priors obtained from current Planck measurements. When the O4+O5 analysis targets the modified gravity model, the prior distribution on $R_0$ is wider. For the priors of the four different mass models, consider Table 4.
| Parameter  | Units | Prior GWTC–3 (Sec. 4) | Prior O4+O5 (Sec. 5) | Description |
|------------|-------|-----------------------|----------------------|-------------|
|            |       | Truncated             | Broken PL            | Multi Peak  |
| α          | -     | $\mathcal{U}(1.5, 12)$| n.a.                 | $\mathcal{U}(1.5, 12)$ | (Negative) PL slope of $m_{1,d}$ |
| $\alpha_1$ | -     | n.a.                  | $\mathcal{U}(1.5, 12)$| n.a.      | (Negative) PL slope for masses below $m_{\text{break}}$ |
| $\alpha_2$ | -     | n.a.                  | $\mathcal{U}(1.5, 12)$| n.a.      | (Negative) PL slope for masses above $m_{\text{break}}$ |
| β          | -     | $\mathcal{U}(-4, 12)$ | $\mathcal{U}(-4, 12)$| $\mathcal{U}(-4, 12)$ | PL slope of $m_{2,d}$ |
| $m_{\text{min}}$ | M$_\odot$ | $\mathcal{U}(2, 10)$ | $\mathcal{U}(2, 10)$ | $\mathcal{U}(2, 10)$ | Minimum mass |
| $m_{\text{max}}$ | M$_\odot$ | $\mathcal{U}(50, 200)$ | $\mathcal{U}(50, 200)$ | $\mathcal{U}(50, 200)$ | Maximum mass |
| $m_{\text{break}}$ | M$_\odot$ | n.a.             | $\mathcal{U}(2, 200)$ | n.a.      | Mass scale where PL slope changes |
| $\lambda_g$ | -     | n.a.                  | $\mathcal{U}(0, 1)$ | $\mathcal{U}(0, 1)$ | $\mathcal{U}(0.01, 0.5)$ Fraction of events in Gaussian |
| $\mu_g$    | M$_\odot$ | n.a.                  | $\mathcal{U}(20, 50)$ | $\mathcal{U}(20, 50)$ | $\mathcal{U}(10, 50)$ Mean of Gaussian distribution |
| $\sigma_g$ | M$_\odot$ | n.a.                  | $\mathcal{U}(0.4, 10)$ | $\mathcal{U}(0.4, 10)$ | $\mathcal{U}(0, 40)$ Standard deviation of Gaussian |
| $\lambda_{g,\text{low}}$ | -     | n.a.                  | n.a. | $\mathcal{U}(0, 1)$ | n.a. Fraction of events in 2nd Gaussian |
| $\mu_{g,\text{low}}$ | M$_\odot$ | n.a.                  | n.a. | $\mathcal{U}(7, 15)$ | n.a. Mean of the 2nd Gaussian component |
| $\sigma_{g,\text{low}}$ | M$_\odot$ | n.a.                  | n.a. | $\mathcal{U}(0.4, 5)$ | n.a. Standard deviation of the 2nd Gaussian component |
| $\delta_m$ | M$_\odot$ | n.a.                  | $\mathcal{U}(0, 10)$ | $\mathcal{U}(0, 10)$ | $\mathcal{U}(0, 20)$ Smoothing factor at minimum mass |

Table 4: Priors of the four different mass models for the O3 real data analysis and the forecast of O4+O5. The abbreviation n.a. stands for “non applicable”. 


B Likelihood model

For a forecast on the estimation of cosmological and the GW luminosity friction parameters under the assumptions of a source mass distribution, we will need to produce extensive GW data. In order to reduce computation time, we want to avoid to run a full parameter estimation for the order of a $\sim 90$ GW events for O4, and $\sim 400$ events for O5. Therefore, in this appendix we construct a proxy that mimics the expected measurement uncertainties. Since the analysis outlined in Section 3 takes into account the detector frame masses, as well as the GW luminosity distance, we provide samples of these three variables.

An approximant should satisfy the following conditions: (i) Capture the uncertainties we expect for a realistic analysis. (ii) Be computationally cheap. (iii) Be compatible with the set of detected events used to calculate the selection effect. The first point will be verified with the simulation of a full MCMC analysis using bilby, performing a parameter estimation for several GW events. The reader will see in the following that point (ii) is fulfilled by construction of the likelihood. We will only consider the amplitude shaping parameters here, namely the component masses, the GW luminosity distance, and $f$, a function of the inclination, sky position and polarization. We assume that the SNR $\rho$ takes the following scaling

$$\rho = N \chi(M) \sqrt{f},$$  \hspace{1cm} (B.1)

where $\chi$ is a fit, the parameters of which we obtain from simulated data with bilby [108]. This function quantifies the chirp mass dependency of the SNR. We assume a fit of the form

$$\chi(M) = n \frac{M^{\alpha_1}}{M^{\alpha_1 - \alpha_2} + M^{\alpha_1 - \alpha_2}},$$  \hspace{1cm} (B.2)

where the free parameters are $\alpha_1, \alpha_2$ and $M_c$. The normalization $n$ is redundant of the prefactor $N$ in Eq. (B.1). This fit interpolates between two PLs above and below a characteristic chirp mass $M_c$ with exponents of $\alpha_i$. The fit assumed here is plotted in Figure 9b. Finally, the factor $f$ can be seen as a geometrical factor, that takes into account the inclination, polarization angle and sky position of our source

$$f = \sum_{a \in \text{det}} \left[ \left( F_{+}(a) \frac{1 + \cos(\iota)}{2} \right)^2 + (F_{\times}(a) \cos(\iota))^2 \right],$$  \hspace{1cm} (B.3)

where the index $a$ runs over all detectors included in the analysis, $\iota$ is the inclination and $F$ are the antenna response functions. Again, we fit the $f$ distribution from a set of simulations. The fit of the $f$ distribution can be found in Figure 9a for the HLV network at O4 sensitivity. Finally, the normalization factor $N$ is constrained from calculating the number of detected sources with bilby.

For the likelihood model or the distribution of measured values we make the following assumptions:

- The square of the measured network SNR follows a non-central chi square distribution with the number of degrees of freedom corresponding to the number of detectors included in the analysis.
- We assume Gaussian distributions on the chirp mass $M$ and the symmetric mass ratio $\eta$, where the standard deviation is inversely proportional to the network SNR.
Figure 9: Consistency of the likelihood approximant for the HLV network at O4 sensitivity

- For the likelihood distribution of the amplitude factor $f$, we fitted the $f$ distribution for a full (aligned spins) parameter analysis. As a first order approximation, we take this likelihood to be independent of SNR. It is only dependent on the measured value of $f$.
Figure 10: Distribution obtained for the observed populations after application of the selection criterion. This figure compares the distribution obtained with bilby (orange/real) to the one based on the SNR approximation in Eq. (B.1) (blue/approx). The distribution obtained through the two methods are compatible overall with minor differences in their tails: the distribution based on the SNR approximation has fewer events at lower masses (a) and (b), and goes to further distances (c). All these populations are for O4 sensitivity.

C Comparison of results obtained with the PL and PLG source frame mass models

Here we elaborate further on the impact on the source frame mass model by comparing the results obtained with the TRUNCATED (PL) and POWER LAW + PEAK (PLG) models, see Figure 12. As indicated by its lower Bayes factor the PL model is too simple to fit all the features in the observed population. To fit the overdensity in the GWTC-3 event distribution at \(\sim\)40 solar masses the PL model results in a shallower slope of the primary mass distribution in comparison with the PLG model. PL thus predicts more sources at higher masses than the PLG mass model. The observed numbers of events at high masses being fixed, PL compensates with a lower maximum mass. Because of strong correlation between the maximum mass and
Figure 11: Comparison between a full parameter estimation using *bilby* (red) and the posterior proxy (blue) for four different GW events. The posterior distributions obtained for Events 2 in (c) and 3 in (d) slightly disagree in their position (value around which the likelihood is peaked) but are consistent in terms of scatter.

$\Xi_0$ (see Sec. 4.2.2), the lower $m_{\text{max}}$ thus results in a lower $\Xi_0$ for the PL when compared to the PLG. This appendix thus evidences the strong impact that the choice of mass model can have on the measurement of $\Xi_0$. To obtain a reliable measurement of modified GR parameters it is indispensable to test a range of different mass models and evaluate their goodness-of-fit by comparing their Bayes factor.
Figure 12: Results obtained with GWTC-3 data using wide priors on $H_0$ and on $\gamma$. The figure compares the posterior distribution based on the PL mass model (red) to the PLG mass model (blue). A SNR cut of 11 and an IFAR of 4 years is applied to the events in the GWTC-3 catalog. Level curves at 1, 2 and 3 sigmas are indicated.

D Full results for the O4 & O5 science runs assuming agnostic cosmological priors

In this Appendix, we show some selected corner plots for the iCarogw analysis of the posterior for the metaparameters, the analysis which was outlined in Section 3. In Figure 13 we present the corner plot for the analysis of O4 and O5 produced with a simulated data set based on a flat $\Lambda$CDM universe with no friction term (GR case) and using agnostic priors on the cosmology. Figure 14 is obtained with synthetic data that models a universe with a modified GW propagation. This figure shows the analysis for O4 and O5, with broad priors on the
cosmological parameters. The results of Sections 5.2 and 5.3 are based on the posteriors presented in these two figures.

E  Analysis for the O4 & O5 science runs with the GR model

Here we study the expected precision with which we expect to recover cosmological parameters when both the simulated dataset and the analysis model is based on GR (i.e., the modified gravity parameters are fixed to their default GR values).

Fig. 15 and 16 show the results obtained with simulated O4 and O5 science runs, leading to 87 and 423 detected events respectively. In both cases, the matter content of the universe

Figure 13: Full results for O4+O5 assuming GR (no friction term) and using wide priors on the cosmology. We indicate the level curves of the posterior at 1, 2 and 3 sigmas. The true values are shown by orange lines.
Figure 14: Full results for O4+O5 assuming modified gravity and using wide priors on the cosmology. We indicate the level curves of the posterior at 1, 2, and 3 sigmas. The true values are shown by orange lines.

$\Omega_m$ is essentially unconstrained. We find that $H_0 = 76^{+32}_{-26}$ km s$^{-1}$ Mpc$^{-1}$ for O4 and $H_0 = 85^{+21}_{-20}$ km s$^{-1}$ Mpc$^{-1}$ (i.e., 24% precision) for O5. This is only a marginal improvement to the Hubble constant measurement (26% precision) when $\Xi_0$ and $n$ were left to vary. When comparing our results obtained for O5 with those in [41] (12% after one year of observation$^{10}$), we find larger uncertainties on the Hubble constant by a factor of $\sim 2$. However, it is difficult to make clear conclusions as those comparisons are not strictly "apple-to-apple". The analysis presented here assumes a SNR cutoff of 12, while [41] assumes a cutoff of 8, resulting in a factor of 2.5 difference in the number of observed events. With 5 years of observation time,

$^{10}$This uncertainty was obtained by adapting and running the published code used by [41].
Figure 15: Results obtained for a simulated O4 science run (87 detected events) with the GR model. (Modified gravity parameters are fixed in this analysis to their default GR value).

the Hubble constant measurement in [41] only marginally improves in precision to 11%, an error decay slower that the standard asymptotic law $\propto 1/\sqrt{N_{\text{obs}}}$. For a pivotal redshift of $z \approx 0.7$ we find an uncertainty of $H(z)$ of 23% which is much larger than the value of 6% in [41] after one year of observation. As [59] points out, [41] uses a redshift evolution model that only depends on one parameter, which can decrease the $H_0$ error bars by a factor of 2. Compared to [59], we find a difference of 20% in the predicted error bars for $H_0$. The population in [59] is assumed to be a BROKEN POWER LAW, whereas we assume a POWER LAW + PEAK mass population.

F Posterior distribution of $\Xi_0$ for different realizations of the event catalog

In order to understand the fluctuations for fixed population parameters, we perform several population analyses, where the underlying population is fixed to the parameters of Section 5. We assume a GR universe and a HLV network with a sensitivity and observation times
Figure 16: Results obtained for a simulated O5 science run (423 detected events) with the GR model. (Modified gravity parameters are fixed in this analysis to their default GR value).

as given in Section 5. Fig. 17 presents the posteriors on $\Xi_0$ for 10 identical runs, where we assume 510 detected events. For each new run, we generate a new catalog of events. The $\Xi_0$ posterior displays a scatter depending on the realizations of the population.

Acknowledgments

The authors are grateful for computational resources provided by the LIGO Laboratory and supported by the National Science Foundation Grants PHY-0757058 and PHY-0823459. This research has made use of data or software obtained from the Gravitational Wave Open Science Center (gw-openscience.org), a service of LIGO Laboratory, the LIGO Scientific Collaboration, the Virgo Collaboration, and KAGRA. LIGO Laboratory and Advanced LIGO are funded by the United States National Science Foundation (NSF) as well as the Science and Technology Facilities Council (STFC) of the United Kingdom, the Max-Planck-Society (MPS), and the State of Niedersachsen/Germany for support of the construction of Advanced
Figure 17: $\Xi_0$ posterior obtained for 10 different realizations of the BBH population with the same metaparameters. To make this check feasible computationally, the uncertainties on the intrinsic GW parameters (masses and GW luminosity distance) were neglected.

LIGO and construction and operation of the GEO600 detector. Additional support for Advanced LIGO was provided by the Australian Research Council. Virgo is funded, through the European Gravitational Observatory (EGO), by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale di Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions by institutions from Belgium, Germany, Greece, Hungary, Ireland, Japan, Monaco, Poland, Portugal, Spain. The construction and operation of KAGRA are funded by Ministry of Education, Culture, Sports, Science and Technology (MEXT), and Japan Society for the Promotion of Science (JSPS), National Research Foundation (NRF) and Ministry of Science and ICT (MSIT) in Korea, Academia Sinica (AS) and the Ministry of Science and Technology (MoST) in Taiwan. Numerical computations were performed on the DANTE platform, APC, France. We would like to thank Marco Crisostomi for helpful discussion on the computation of modified luminosity distances in DHOST models. The Foundation CFM pour la Recherche in Paris has generously supported KL during his doctorate. SM is supported by the ANR COSMERGE project, grant ANR-20-CE31-001 of the French Agence Nationale de la Recherche. SM acknowledges the support of the LabEx UnivEarthS (ANR-10-LABX-0023 and ANR-18-IDEX0001), of the European Gravitational Observatory and of the Paris Center for Cosmological Physics.

References

[1] LIGO Scientific, Virgo collaboration, Observation of Gravitational Waves from a Binary Black Hole Merger, *Phys. Rev. Lett.* **116** (2016) 061102 [1602.03837].

[2] LIGO Scientific, Virgo, KAGRA collaboration, GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run, 2111.03606.

[3] LIGO Scientific, Virgo collaboration, GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral, *Phys. Rev. Lett.* **119** (2017) 161101 [1710.05832].
LIGO Scientific, KAGRA, VIRGO collaboration, *Observation of Gravitational Waves from Two Neutron Star–Black Hole Coalescences*, *Astrophys. J. Lett.* **915** (2021) L5 [2106.15163].

and J Aasi, B.P. Abbott, R. Abbott, T. Abbott, M.R. Abernathy, K. Ackley et al., *Advanced LIGO*, *Classical and Quantum Gravity* **32** (2015) 074001.

ALIGO collaboration, *Sensitivity and performance of the Advanced LIGO detectors in the third observing run*, *Phys. Rev. D* **102** (2020) 062003 [2008.01301].

M. Tse et al., *Quantum-Enhanced Advanced LIGO Detectors in the Era of Gravitational-Wave Astronomy*, *Phys. Rev. Lett.* **123** (2019) 231107.

[8] VIRGO collaboration, *Advanced Virgo: a second-generation interferometric gravitational wave detector*, *Class. Quant. Grav.* **32** (2015) 074001.

[9] LIGO Scientific, *Sensitivity and performance of the Advanced LIGO detectors in the third observing run*, *Phys. Rev. D* **102** (2020) 062003 [2008.01301].

[10] M. Tse et al., *Quantum-Enhanced Advanced LIGO Detectors in the Era of Gravitational-Wave Astronomy*, *Phys. Rev. Lett.* **915** (2021) L5 [2106.15163].

[11] LIGO Scientific, VIRGO, KAGRA Scientific collaboration, *The population of merging compact binaries inferred using gravitational waves through GWTC-3*, 2112.06861.

[12] B.F. Schutz, *Determining the Hubble Constant from Gravitational Wave Observations*, *Nature* **323** (1986) 310.

[13] A.G. Riess, S. Casertano, W. Yuan, L.M. Macri and D. Scolnic, *Large Magellanic Cloud Cepheid Standards Provide a 1% Foundation for the Determination of the Hubble Constant and Stronger Evidence for Physics beyond ΛCDM*, *Astrophys. J.* **876** (2019) 85 [1903.07603].

[14] K.C. Wong et al., *H0LiCOW – XIII. A 2.4 per cent measurement of H0 from lensed quasars: 5.3σ tension between early- and late-Universe probes*, *Mon. Not. Roy. Astron. Soc.* **498** (2020) 1420 [1907.04869].

[15] K. Hotokezaka, E. Nakar, O. Gottlieb, S. Nissanke, K. Masuda, G. Hallinan et al., *A Hubble constant measurement from superluminal motion of the jet in GW170817*, *Nature Astron.* **3** (2019) 940 [1807.04869].

[16] J.M. Ezquiaga and M. Zumalacárregui, *Dark Energy in light of Multi-Messenger Gravitational-Wave astronomy*, *Front. Astron. Space Sci.* **5** (2018) 44 [1807.09241].

[17] S. Mastrogiovanni, R. Duque, E. Chassande-Mottin, F. Daigne and R. Mochkovitch, *What role will binary neutron star merger afterglows play in multimessenger cosmology?*, 2012.12836.
[41] W.M. Farr, M. Fishbach, J. Ye and D. Holz, A Future Percent-Level Measurement of the Hubble Expansion at Redshift 0.8 With Advanced LIGO, *Astrophys. J. Lett.* **883** (2019) L42 [1908.09084].

[42] Z.-Q. You, X.-J. Zhu, G. Ashton, E. Thrane and Z.-H. Zhu, Standard-siren cosmology using gravitational waves from binary black holes, *Astrophys. J.* **908** (2021) 215 [2004.00036].

[43] J.M. Ezquiaga and D.E. Holz, Jumping the Gap: Searching for LIGO’s Biggest Black Holes, *Astrophys. J. Lett.* **909** (2021) L23 [2006.02211].

[44] I.D. Saltas, I. Sawicki, L. Amendola and M. Kunz, Anisotropic Stress as a Signature of Nonstandard Propagation of Gravitational Waves, *Phys. Rev. Lett.* **113** (2014) 191101 [1406.7139].

[45] A. Nishizawa, Generalized framework for testing gravity with gravitational-wave propagation. I. Formulation, *Phys. Rev. D* **97** (2018) 104037 [1710.04825].

[46] E. Belgacem, Y. Dirian, S. Foffa and M. Maggiore, Gravitational-wave luminosity distance in modified gravity theories, *Phys. Rev. D* **97** (2018) 104066 [1712.08108].

[47] G. Dvali, G. Gabadadze and M. Porrati, 4D gravity on a brane in 5D Minkowski space, *Physics Letters B* **485** (2000) 208 [hep-th/0005016].

[48] I.D. Saltas, I. Sawicki, L. Amendola and M. Kunz, Anisotropic Stress as a Signature of Nonstandard Propagation of Gravitational Waves, *Phys. Rev. Lett.* **113** (2014) 191101 [1406.7139].

[49] A. Nishizawa, Generalized framework for testing gravity with gravitational-wave propagation. I. Formulation, *Phys. Rev. D* **97** (2018) 104037 [1710.04825].

[50] E. Belgacem, Y. Dirian, S. Foffa and M. Maggiore, Gravitational-wave luminosity distance in modified gravity theories, *Phys. Rev. D* **97** (2018) 104066 [1712.08108].

[51] G. Dvali, G. Gabadadze and M. Porrati, 4D gravity on a brane in 5D Minkowski space, *Physics Letters B* **485** (2000) 208 [hep-th/0005016].

[52] C. Deffayet and K. Menou, Probing Gravity with Spacetime Sirens, *Astrophys. J. Lett.* **668** (2007) L143 [0709.0003].

[53] M. Corman, A. Ghosh, C. Escamilla-Rivera, M.A. Hendry, S. Marsat and N. Tamanini, Constraining cosmological extra dimensions with gravitational wave standard sirens: from theory to current and future multi-messenger observations, arXiv e-prints (2021) arXiv:2109.08748 [2109.08748].

[54] E. Bellini, A.J. Cuesta, R. Jimenez and L. Verde, Constraints on deviations from ΛCDM within Horndeski gravity, *JCAP* **02** (2016) 053 [1509.07816].

[55] D. Alonso, E. Bellini, P.G. Ferreira and M. Zumalacárregui, Observational future of cosmological scalar-tensor theories, *Phys. Rev. D* **95** (2017) 063502 [1610.09290].

[56] E. Bellini, R. Jimenez and L. Verde, Signatures of Horndeski gravity on the Dark Matter Bispectrum, *JCAP* **05** (2015) 057 [1504.04341].

[57] E. Bellini and I. Sawicki, Maximal freedom at minimum cost: linear large-scale structure in general modifications of gravity, *JCAP* **07** (2014) 050 [1404.3713].

[58] S. Mastrogiacomi, L. Haegel, C. Karathanasis, I. Magaña Hernandez and D.A. Steer, Gravitational wave friction in light of GW170817 and GW190521, *J. Cosmology Astropart. Phys.* **2021** (2021) 043 [2010.04047].

[59] M. Lagos, M. Fishbach, P. Landry and D.E. Holz, Standard sirens with a running Planck mass, *Phys. Rev. D* **99** (2019) 083504 [1901.03321].

[60] J.M. Ezquiaga, Hearing gravity from the cosmos: GWTC-2 probes general relativity at cosmological scales, *Phys. Lett. B* **822** (2021) 136665 [2104.05139].

[61] LIGO Scientific, Virgo collaboration, GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run, *Phys. Rev. X* **11** (2021) 021053 [2010.14527].

[62] I. Magana Hernandez, Constraining the number of spacetime dimensions from GWTC-3 binary black hole mergers, arXiv e-prints (2021) arXiv:2112.07650 [2112.07650].

[63] M. Mancarella, E. Genoud-Prachex and M. Maggiore, Cosmology and modified gravitational wave propagation from binary black hole population models, *PhRvD* **105** (2022) 064030 [2112.05728].
[60] KAGRA collaboration, *Overview of KAGRA: Detector design and construction history*, *PTEP* 2021 (2021) 05A101 [2005.05574].

[61] KAGRA collaboration, *Interferometer design of the KAGRA gravitational wave detector*, *Phys. Rev. D* 88 (2013) 043007 [1306.6747].

[62] T. Clifton, P.G. Ferreira, A. Padilla and C. Skordis, *Modified Gravity and Cosmology*, *Phys. Rept.* 513 (2012) 1 [1106.2476].

[63] C. Deffayet and D.A. Steer, *A formal introduction to Horndeski and Galileon theories and their generalizations*, *Class. Quant. Grav.* 30 (2013) 214006 [1307.2450].

[64] A. Joyce, B. Jain, J. Khoury and M. Trodden, *Beyond the Cosmological Standard Model*, *Phys. Rept.* 568 (2015) 1 [1407.0059].

[65] K. Koyama, *Cosmological Tests of Modified Gravity*, *Rept. Prog. Phys.* 79 (2016) 046902 [1504.04623].

[66] L. Heisenberg, *A systematic approach to generalisations of General Relativity and their cosmological implications*, *Phys. Rept.* 796 (2019) 1 [1807.01725].

[67] X. Gao and D.A. Steer, *Inflation and primordial non-Gaussianities of ‘generalized Galileons’*, *JCAP* 12 (2011) 019 [1107.2642].

[68] M. Denissenya and E.V. Linder, *Gravity’s Islands: Parametrizing Horndeski Stability*, *JCAP* 11 (2018) 010 [1808.00013].

[69] E. Belgacem, Y. Dirian, S. Foffa and M. Maggiore, *Modified gravitational-wave propagation and standard sirens*, *Phys. Rev. D* 98 (2018) 023510 [1805.08731].

[70] E. Belgacem et al., *Testing modified gravity at cosmological distances with LISA standard sirens*, *JCAP* 07 (2019) 024 [1906.01593].

[71] G.W. Horndeski, *Second-order scalar-tensor field equations in a four-dimensional space*, *Int. J. Theor. Phys.* 10 (1974) 363.

[72] D. Langlois and K. Noui, *Degenerate higher derivative theories beyond Horndeski: evading the Ostrogradski instability*, *JCAP* 02 (2016) 034 [1510.06930].

[73] E.V. Linder, *Challenges in connecting modified gravity theory and observations*, *Phys. Rev. D* 95 (2017) 023518 [1607.03113].

[74] I. Mandel, W.M. Farr and J.R. Gair, *Extracting distribution parameters from multiple uncertain observations with selection biases*, *Mon. Not. Roy. Astron. Soc.* 486 (2019) 1086 [1809.02063].

[75] E. Thrane and C. Talbot, *An introduction to Bayesian inference in gravitational-wave astronomy: Parameter estimation, model selection, and hierarchical models*, *PASA* 36 (2019) e010 [1809.02293].

[76] S. Vitale, D. Gerosa, W.M. Farr and S.R. Taylor, *Inferring the properties of a population of compact binaries in presence of selection effects*, 2007.05579.
[81] S. Mastrogiovanni, K. Leyde, C. Karathanasis, E. Chassande-Mottin, D.A. Steer, J. Gair et al., Cosmology in the dark: On the importance of source population models for gravitational-wave cosmology, 2103.14663.

[82] P. Madau and M. Dickinson, Cosmic Star Formation History, Ann. Rev. Astron. Astrophys. 52 (2014) 415 [1403.0007].

[83] T. Callister, M. Fishbach, D. Holz and W. Farr, Shouts and Murrns: Combining Individual Gravitational-Wave Sources with the Stochastic Background to Measure the History of Binary Black Hole Mergers, Astrophys. J. Lett. 896 (2020) L32 [2003.12152].

[84] W.W. Ober, M.F. El Eid and K.J. Fricke, Evolution of massive pregalactic stars-part two-nucleosynthesis in pair creation supernovae and pregalactic enrichment, Astronomy and Astrophysics 119 (1983) 61.

[85] J.R. Bond, W.D. Arnett and B.J. Carr, The evolution and fate of Very Massive Objects, Astrophysical Journal 280 (1984) 825.

[86] W. Glatzel, K.J. Fricke and M.F. El Eid, The fate of rotating pair-unstable carbon-oxygen cores, A&A 149 (1985) 413.

[87] A. Heger and S.E. Woosley, The Nucleosynthetic Signature of Population III, ApJ 567 (2002) 532 [astro-ph/0107037].

[88] H. Umeda and K. Nomoto, Pop III hypernova nucleosynthesis and abundances in very metal-poor halo stars, astro-ph/0205365.

[89] E. Scannapieco, P. Madau, S. Woosley, A. Heger and A. Ferrara, The Detectability of pair-production supernovae at z<~6, Astrophys. J. 633 (2005) 1031 [astro-ph/0507182].

[90] D. Kasen, S.E. Woosley and A. Heger, Pair Instability Supernovae: Light Curves, Spectra, and Shock Breakout, Astrophys. J. 734 (2011) 102 [1101.3336].

[91] S.E. Woosley, Pulsational Pair-Instability Supernovae, Astrophys. J. 836 (2017) 244 [1608.08939].

[92] Z. Barkat, G. Rakavy and N. Sack, Dynamics of Supernova Explosion Resulting from Pair Formation, Phys. Rev. Lett. 18 (1967) 379.

[93] A. Heger and S.E. Woosley, The nucleosynthetic signature of population III, Astrophys. J. 567 (2002) 532 [astro-ph/0107037].

[94] E. Chatzopoulos and J.C. Wheeler, Effects of Rotation on the Minimum Mass of Primordial Progenitors of Pair Instability Supernovae, Astrophys. J. 748 (2012) 42 [1201.1328].

[95] K.-J. Chen, S. Woosley, A. Heger, A. Almgren and D. Whalen, Two-Dimensional Simulations of Pulsational Pair-Instability Supernovae, Astrophys. J. 792 (2014) 28 [1402.4134].

[96] M. Spera and M. Mapelli, Very massive stars, pair-instability supernovae and intermediate-mass black holes with the SEVN code, Mon. Not. Roy. Astron. Soc. 474 (2017) 4739 [1706.06109].

[97] N. Giacobbo, M. Mapelli and M. Spera, Merging black hole binaries: the effects of progenitor’s metallicity, mass-loss rate and Eddington factor, Mon. Not. Roy. Astron. Soc. 474 (2018) 2959 [1711.03556].

[98] K. Belczynski et al., The Effect of Pair-Instability Mass Loss on Black Hole Mergers, Astron. Astrophys. 594 (2016) A97 [1607.03116].

[99] PLANCK collaboration, Planck 2015 results. XIII. Cosmological parameters, Astron. Astrophys. 594 (2016) A13 [1502.01589].

[100] LIGO SCIENTIFIC, VIRGO collaboration, GW190814: Gravitational Waves from the Coalescence of a 23 Solar Mass Black Hole with a 2.6 Solar Mass Compact Object, Astrophys. J. Lett. 896 (2020) L44 [2006.12611].
LIGO Scientific, Virgo collaboration, GW190521: A Binary Black Hole Merger with a Total Mass of 150\,M\odot, Phys. Rev. Lett. 125 (2020) 101102 [2009.01075].

J.E. Thompson, E. Fauchon-Jones, S. Khan, E. Nitoglia, F. Pannarale, T. Dietrich et al., Modeling the gravitational wave signature of neutron star black hole coalescences, Phys. Rev. D 101 (2020) 124059 [2002.08383].

G. Pratten et al., Computationally efficient models for the dominant and subdominant harmonic modes of precessing binary black holes, Phys. Rev. D 103 (2021) 104056 [2004.06503].

J.E. Thompson, E. Fauchon-Jones, S. Khan, E. Nitoglia, F. Pannarale, T. Dietrich et al., Modeling the gravitational wave signature of neutron star black hole coalescences, Phys. Rev. D 101 (2020) 124059 [2002.08383].

S. Ossokine et al., Multipolar Effective-One-Body Waveforms for Precessing Binary Black Holes: Construction and Validation, Phys. Rev. D 101 (2020) 124059 [2002.08383].

G. Pratten et al., Computationally efficient models for the dominant and subdominant harmonic modes of precessing binary black holes, Phys. Rev. D 103 (2021) 104056 [2004.06503].

A. Matas et al., Aligned-spin neutron-star–black-hole waveform model based on the effective-one-body approach and numerical-relativity simulations, Phys. Rev. D 102 (2020) 043023 [2004.10001].

KAGRA, LIGO Scientific, Virgo, VIRGO collaboration, Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA, Living Rev. Rel. 21 (2018) 3 [1304.0670].

E. Belgacem, Y. Dirian, A. Finke, S. Foffa and M. Maggiore, Gravity in the infrared and effective nonlocal models, JCAP 04 (2020) 010 [2001.07619].

G. Ashton et al., BILBY: A user-friendly Bayesian inference library for gravitational-wave astronomy, Astrophys. J. Suppl. 241 (2019) 27 [1811.02042].

J. Ben Achour, M. Crisostomi, K. Koyama, D. Langlois, K. Noui and G. Tasinato, Degenerate higher order scalar-tensor theories beyond Horndeski up to cubic order, JHEP 12 (2016) 100 [1608.08135].

M. Crisostomi, K. Koyama and G. Tasinato, Extended Scalar-Tensor Theories of Gravity, JCAP 04 (2016) 044 [1602.03119].

M. Crisostomi, K. Koyama, D. Langlois, K. Noui and D.A. Steer, Cosmological evolution in DHOST theories, JCAP 01 (2019) 030 [1810.12070].

T.A. Callister, C.-J. Haster, K.K.Y. Ng, S. Vitale and W.M. Farr, Who Ordered That? Unequal-mass Binary Black Hole Mergers Have Larger Effective Spins, Astrophys. J. Lett. 922 (2021) L5 [2106.00521].

K.K.Y. Ng, K.W.K. Wong, T. Broadhurst and T.G.F. Li, Precise LIGO Lensing Rate Predictions for Binary Black Holes, Phys. Rev. D 97 (2018) 023012 [1703.06319].

S.-S. Li, S. Mao, Y. Zhao and Y. Lu, Gravitational lensing of gravitational waves: A statistical perspective, Mon. Not. Roy. Astron. Soc. 476 (2018) 2220 [1802.05089].

M. Punturo et al., The third generation of gravitational wave observatories and their science reach, Class. Quant. Grav. 27 (2010) 084007.

E.D. Hall and M. Evans, Metrics for next-generation gravitational-wave detectors, Class. Quant. Grav. 36 (2019) 225002 [1902.09485].

G. Cusin and N. Tamanini, Characterization of lensing selection effects for LISA massive black hole binary mergers, Mon. Not. Roy. Astron. Soc. 504 (2021) 3610 [2011.15109].

R. Farmer, M. Renzo, S.E. de Mink, P. Marchant and S. Justham, Mind the gap: The location of the lower edge of the pair instability supernovae black hole mass gap, 1910.12874.

S. Mukherjee, The redshift dependence of black hole mass distribution: Is it reliable for standard sirens cosmology?, 2112.10256.

S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. Jiménez Forteza et al., Frequency-domain gravitational waves from nonprecessing black-hole binaries. I. New
numerical waveforms and anatomy of the signal, Phys. Rev. D 93 (2016) 044006 [1508.07250].

[121] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. Jiménez Forteza et al., Frequency-domain gravitational waves from nonprecessing black-hole binaries. II. A phenomenological model for the advanced detector era, Phys. Rev. D 93 (2016) 044007 [1508.07253].