

PAPER • OPEN ACCESS

Testing the nature of gravitational wave propagation using dark sirens and galaxy catalogues

To cite this article: Anson Chen *et al* JCAP02(2024)035

View the [article online](#) for updates and enhancements.

You may also like

- [Joint cosmological and gravitational-wave population inference using dark sirens and galaxy catalogues](#)
Rachel Gray, Freija Beirnaert, Christos Karathanasis et al.
- [Constraints on the Cosmic Expansion History from GWTC-3](#)
R. Abbott, H. Abe, F. Acernese et al.
- [Things That Might Go Bump in the Night: Assessing Structure in the Binary Black Hole Mass Spectrum](#)
Amanda M. Farah, Bruce Edelman, Michael Zevin et al.

Testing the nature of gravitational wave propagation using dark sirens and galaxy catalogues

Anson Chen ^a, Rachel Gray ^{a,b} and Tessa Baker ^{a,c}

^a*Department of Physics and Astronomy, Queen Mary University of London,
Mile End Road, London E1 4NS, United Kingdom*

^b*SUPA, University of Glasgow,
Glasgow G12 8QQ, United Kingdom*

^c*Institute of Cosmology and Gravitation, University of Portsmouth,
Burnaby Road, Portsmouth PO1 3FX, United Kingdom*

E-mail: anson.chen@ligo.org, rachel.gray@ligo.org, tessa.baker@ligo.org

ABSTRACT: The dark sirens method enables us to use gravitational wave events *without* electromagnetic counterparts as tools for cosmology and tests of gravity. Furthermore, the dark sirens analysis code GWCOSMO can now robustly account for information coming from both galaxy catalogues and the compact object mass distribution. We present here an extension of the GWCOSMO code and methodology to constrain parameterized deviations from General Relativity that affect the propagation of gravitational waves under the Λ CDM expansion history of the universe. We show results of our analysis using data from the GWTC-3 gravitational wave catalogues, in preparation for application to the O4 observing run. After testing our pipelines using the First Two Years mock data set, we reanalyse 46 events from GWTC-3, and combine the posterior for BBH and NSBH sampling results for the first time. We obtain joint constraints on H_0 and parameterized deviations from General Relativity in the Power Law + Peak BBH population model. With increased galaxy catalogue support in the future, our work sets the stage for dark sirens to become a powerful tool for testing gravity.

KEYWORDS: Bayesian reasoning, Gravitational waves in GR and beyond: theory , modified gravity

ARXIV EPRINT: [2309.03833](https://arxiv.org/abs/2309.03833)

Contents

1	Introduction	1
2	GW propagation beyond GR	3
2.1	(Ξ_0, n) parameterization	4
2.2	Horndeski class parameterization	5
2.3	Extra dimensions	7
3	Modifications to GWCOSMO	7
3.1	Bayesian framework of GWCOSMO	7
3.2	Prior on source masses and merger rate evolution	12
4	Tests on mock data	12
4.1	Simulation setup	12
4.2	Results	13
5	Reanalysing the GWTC-3 data	14
5.1	Parameter ranges	14
5.2	Dark sirens results	15
5.2.1	1D measurement of Ξ_0	15
5.2.2	BBH joint analysis	17
5.2.3	NSBH joint analysis	20
5.3	Combined BBH and NSBH joint analysis	20
6	Conclusions	23
A	Corner plots for GWTC-3 reanalysis	26
B	Posterior samples for GWTC-3 reanalysis	27

1 Introduction

As the imprint of a propagating gravitational field perturbation, gravitational waves (GWs) are the most natural tool with which to test the nature of gravity. The nature of their generation and subsequent propagation means that they uniquely access both the strong-field regime of gravity near compact objects, as well as the weak-field regime of gravity on cosmological scales [1, 2]. Furthermore, unlike many electromagnetic (EM) observables, they carry direct information about the luminosity distance of their source. When combined with redshift information, we can therefore use gravitational wave data to probe the luminosity distance-redshift relation — and hence key cosmological parameters like the Hubble constant — in a manner distinct to traditional methods such as the Cosmic Microwave Background or the local distance ladder [3, 4]. Some departures from General Relativity (GR) will also impact the luminosity distance-redshift relation, and hence can be tested in this way.

For both tests of gravity and measurements of H_0 , the presence of an EM counterpart to a GW source — a “bright siren” event first introduced by [5] — provides the most direct

and (usually) powerful constraints. This was seen most spectacularly with binary neutron star (BNS) merger GW170817 and its EM counterpart [6, 7]. Here, the localization of the accompanying kilonova allowed the galaxy NGC4993 to be uniquely identified as the host galaxy of GW170817 [8]. With an additional correction for the peculiar velocity of NGC4993, from this single event alone the Hubble constant was estimated to be $H_0 = 70.0_{-8.0}^{+12.0} \text{ km s}^{-1} \text{ Mpc}^{-1}$ [9]. The implications of GW170817 for tests of gravity were even more impactful. The arrival of the GRB counterpart just $\sim 1.7 \text{ s}$ after the GW merger signal constrained the propagation speed of GWs (c_T) in the frequency band of terrestrial detectors¹ at the level $|1 - c_T/c| \leq 3 \times 10^{-15}$ [7]. This led to stringent constraints on various modified gravity models, such as scalar-tensor and vector-tensor theories [14–16], Born-Infeld gravity [17] and parity-violating gravity [18].

However, so far bright sirens have proved relatively elusive. With current broad uncertainties on the underlying event rates, it remains unclear to what extent we can rely on them as cosmological tools. This has led to the development of an alternative set of techniques for GW cosmology that do *not* rely on EM counterparts, known as dark siren methods. These techniques instead employ galaxy catalogues, together with features in the mass spectrum of compact objects, to supply redshift information about the possible hosts of GW sources. Under these methods, the redshift of a GW host galaxy cannot be uniquely identified; instead the galaxy catalogue is used to construct a prior distribution for the redshift along specific lines of sight (in a pixellated fashion [19]). This prior is combined with GW data in a hierarchical Bayesian formalism [20–23] to infer cosmological parameters. The population model of compact objects also informs this result because, assuming a known distribution of compact object masses in the source frame, this is related to the distribution that GW detectors actually measure (in ‘detector frame’) by factors of $(1 + z)$. Hence for any distinct features in the mass distribution we can effectively compare their assumed and measured locations to infer information about source redshifts, again in a statistical sense only.

To date, the predominant focus of dark sirens development has been to constrain the Hubble constant (but see also [24, 25]). The primary goal of this paper is to introduce an extension of a key dark sirens tool, described below, which allows the method to test extensions of GR impacting the luminosity distance-redshift relation, i.e. one of the most generic effects of cosmologically-motivated modified gravity theories. With the fourth observing run of the LIGO-Virgo-KAGRA (LVK) collaboration underway, we ready our tools for application to the increasing stack of GW detections, to yield (we anticipate) increasingly precise results.

At present, two sophisticated code packages have been independently developed to measure cosmological parameters with the bright siren and/or the dark siren methods: GWCOSMO [19, 26] and ICAROGW [27, 28]. Both pipelines are now capable of drawing on information from both the statistical host galaxy distribution from galaxy catalogues, and also via constraining compact object population models and merger rate redshift evolution models. The ability to use information from both galaxy catalogues and the population

¹A caveat is that an effective field approach to modified gravity still allows a deviation of c_T from c at an energy scale lower than the terrestrial detection band [10]. This scenario can be probed with massive BBHs by the Laser Interferometer Space Antenna (LISA) alone [11], or with stellar origin BBHs by joint detection of LVK and LISA [12, 13], or possibly by pulsar timing arrays.

model simultaneously is an important recent development, detailed in [26], which ensures a robust H_0 result.²

After the O3 observation run [30], 46 dark sirens in GWTC-3 with signal-to-noise ratios (SNRs) larger than 11 were selected for statistical host galaxy identification with the GLADE+ galaxy catalogue [19] using GWCOSMO v1.0.0 [31]. These yield a measurement of $H_0 = 67_{-12}^{+13} \text{ km s}^{-1} \text{ Mpc}^{-1}$. Combining the dark siren result with the bright siren GW170817 yields $H_0 = 68_{-6}^{+8} \text{ km s}^{-1} \text{ Mpc}^{-1}$. Meanwhile the analysis with 42 BBHs using ICAROGW 2.0 gives $H_0 = 71_{-30}^{+35} \text{ km s}^{-1} \text{ Mpc}^{-1}$ with the assumption of the compact object binary merger rate proportional to the galaxy luminosity, and $H_0 = 43_{-18}^{+48} \text{ km s}^{-1} \text{ Mpc}^{-1}$ without the assumption [27]. Although these measurements are not more precise than those from EM probes, they serve as an early step on a pathway towards an independent and competitive probe of H_0 . We note here that there exists a third approach to constraining H_0 with GW data, by fully cross-correlating GW events and galaxy clustering [32–34].

We wish to capitalise on these early successes of the dark sirens methodology, and likewise apply them to test the laws of gravity. The purpose of this paper is to describe the extension of the GWCOSMO software to achieve precisely this. We test our implementation in the local universe using the First Two Years mock data, and then apply the methodology to real data from GWTC-3 events. Our results show good agreement with those from other groups where similar modified gravity effects are included, for example in the ICAROGW code described in [24, 27], and also an independent code pipeline appeared in [25]. While we can't currently include the expansion history of the universe beyond the Λ CDM model due to computational cost (see more discussion in section 3), this lays the groundwork for future tests of gravity at a cosmic scale in O4 that have not been possible in previous observing runs.

The paper is constructed in five sections. In section 2 we review possible modifications to the GW luminosity distance outside of GR, and introduce three classes of parameterization for this phenomenon. In section 3 we revisit the theoretical framework of GWCOSMO and develop its extension to features beyond GR. Section 4 and section 5 show the reanalysis of mock bright siren data and GWTC-3 events respectively, now including modified gravity parameters with the latest version of GWCOSMO. We conclude and discuss further lines of development in section 6.

2 GW propagation beyond GR

A common feature of gravity theories beyond GR is a modification to the effective friction term that impacts GW propagation. Although in principle some alternative theories of gravity can change the expansion history of the universe, in this work we focus on theories that create unique effects in GW propagation whilst remaining close to a Λ CDM expansion history (as this is well supported by current observational data). In theories with this feature, in the Jordan frame (where matter is minimally coupled to the metric), the modified GW propagation equation can be written as [25, 35–39]

$$\tilde{h}''_A + 2\mathcal{H}[1 - \delta(\eta)]\tilde{h}'_A + c^2 k^2 \tilde{h}_A = 0, \quad (2.1)$$

²Previously, the first version of ICAROGW [29] constrained H_0 , population and merger rate models without utilising galaxy catalogue information; meanwhile GWCOSMO v1.0.0 [19] measured H_0 using galaxy catalogue information, but fixing the population and merger rate model.

where $\tilde{h}(\eta, k)$ is the GW amplitude in Fourier space, subscript A denotes the $+$ or \times polarization of GWs, η is conformal time, $'$ denotes the derivative with respect to η , and $\mathcal{H} = a'/a$ is the conformal Hubble parameter. Here $\delta(\eta)$ is the modification to the friction term introduced by modified gravity theories, where GR is recovered when $\delta(\eta) = 0$. It is shown in section 4.1.4 of [40] that the solution to eq. (2.1) in GR gives $\tilde{h}(\eta, k) \propto 1/d_L^{\text{EM}}(z)$, where $d_L^{\text{EM}}(z)$ is the EM luminosity distance defined by

$$d_L^{\text{EM}}(z) = (1+z) \int_0^z \frac{c d\tilde{z}}{H_0 \sqrt{\Omega_{m,0}(1+\tilde{z})^3 + \Omega_{\Lambda,0}}} \quad (2.2)$$

in the late-time universe. $\Omega_{m,0}$ and $\Omega_{\Lambda,0}$ are the matter density today and the dark energy density today respectively. When $\delta(\eta) \neq 0$, it is shown in [35] that one can absorb the changes to the GW propagation equation by defining a new effective conformal Hubble factor. To do this, we introduce a new scale factor \tilde{a} such that

$$\frac{\tilde{a}'}{\tilde{a}} = \mathcal{H}[1 - \delta(\eta)]. \quad (2.3)$$

When expressed in terms of the new scale factor, the friction term now has the same form as in GR. From the solution to the transformed equation we obtain a new quantity called the GW luminosity distance $d_L^{\text{GW}}(z)$, which is related to $d_L^{\text{EM}}(z)$ by

$$d_L^{\text{GW}}(z) = \frac{a(z)}{\tilde{a}(z)} d_L^{\text{EM}}(z). \quad (2.4)$$

In a non-GR theory, we now have $\tilde{h}(\eta, k) \propto 1/d_L^{\text{GW}}(z)$. Eq. (2.3) can be written as

$$\left(\log \frac{a}{\tilde{a}} \right)' = \delta(\eta) \mathcal{H}(\eta). \quad (2.5)$$

Integrating eq. (2.5) allows us to re-express the relation between $d_L^{\text{GW}}(z)$ and $d_L^{\text{EM}}(z)$ as:

$$d_L^{\text{GW}}(z) = d_L^{\text{EM}}(z) \exp \left\{ - \int_0^z \frac{dz'}{1+z'} \delta(z') \right\}. \quad (2.6)$$

The exact form of $\delta(z)$ depends on the specific gravity model under consideration (some subtleties related to this parameterisation are discussed in [41]). We can also choose to parameterise the entire term in curly brackets above in some sensible manner, which we demonstrate below. In the next section we present the analysis of three models/parameterizations implemented in GWCOSMO.

2.1 (Ξ_0, n) parameterization

As mentioned above, there exist many gravity theories that can introduce a non-standard friction term in the GW propagation equation. Under these circumstances, rather than work with a large number of individual models, it can be efficient to instead work with a general parameterized form of the ratio $d_L^{\text{GW}}(z)/d_L^{\text{EM}}(z)$. Constraints on parameters belonging to specific gravity models are then obtainable via a mapping from the parameterised constraint.

A widely-adopted and simple parameterization for the ratio of $d_L^{\text{GW}}(z)/d_L^{\text{EM}}(z)$ is given by [36]

$$\frac{d_L^{\text{GW}}(z)}{d_L^{\text{EM}}(z)} = \Xi_0 + \frac{1 - \Xi_0}{(1+z)^n}, \quad (2.7)$$

where Ξ_0 and n are the parameters controlling the variation from GR. This parameterization was originally obtained by fitting to a class of non-local gravity models in [35, 36, 42], however, its form can be motivated on more general grounds as we describe below momentarily. As a result, this parameterization has been widely adopted in testing GR with dark sirens, for example in [24, 25, 27, 43, 44], and with other techniques such as using BNS mass distribution [45] and strongly lensed GW events [46]. Some mappings of Ξ_0 and n to parameters in exact gravity models are listed in table 1 of [38].

The GR limit of eq. (2.7) corresponds to $\Xi_0 = 1$, so that the GW distance is the same as the EM luminosity distance. This limit is also recovered for $z \sim 0$, largely irrespective of Ξ_0 . Conversely, when $z \gg 1$ the ratio $d_L^{\text{GW}}(z)/d_L^{\text{EM}}(z)$ approaches a constant set by the value of Ξ_0 . This is motivated by the hypothesis that any deviation from GR predominantly affects the late universe where dark energy dominates, so $d_L^{\text{GW}}(z)/d_L^{\text{EM}}(z)$ deviates from 1 as redshift increases under effects beyond GR, and approaches constant at high redshift since the deviation from GR is suppressed in early universe.

The value of n determines how sharply the transition of the ratio from 1 to Ξ_0 take place as redshift increases. In the low redshift limit $z \ll 1$, the ratio can be approximated at the linear order in z :

$$\frac{d_L^{\text{GW}}(z)}{d_L^{\text{EM}}(z)} = 1 - zn(1 - \Xi_0) + O(z^2), \quad (2.8)$$

which is often useful in the regime of the local universe. When n is fixed as a parameter, its fiducial value is often taken to be $n = 1.91$, the value predicted by the RT non-local gravity model when measuring Ξ_0 [25]. This may seem a somewhat arbitrary choice, but it will not significantly impact the results of this work (typically we will allow n to vary as a free parameter with broad priors anyway). Note that GR is also recovered as n approaches 0, meaning we can potentially expect to see some degeneracy between constraints on Ξ_0 and n , if the data are consistent with GR.

In figure 1 we plot the ratio of $d_L^{\text{GW}}(z)/d_L^{\text{EM}}(z)$ for some different values of Ξ_0 . The ratio is greater than 1 when $\Xi_0 > 1$ and lower than 1 when $\Xi_0 < 1$, for $n \neq 0$.

2.2 Horndeski class parameterization

In addition to the generic (Ξ_0, n) parameterization, we also implement a specific parameterization originating from the Horndeski family of extensions to GR. The Horndeski action describes the most generic scalar-tensor theory of gravity with a second-order equation of motion [47, 48]. The observation of the luminal speed of gravitational waves by GW170817 and its EM counterpart [6] strongly constrains particular terms in the original Horndeski action, though see the caveats discussed in [10]. Setting these terms to zero, the remaining reduced Horndeski action is [14–16, 49]

$$S = \int d^4x \sqrt{-g} \left[\frac{1}{2} M_{\text{eff}}^2(\phi) R + K(\phi, X) - G_3(\phi, X) \square \phi \right] + S_m(g_{\mu\nu}, \psi_m), \quad (2.9)$$

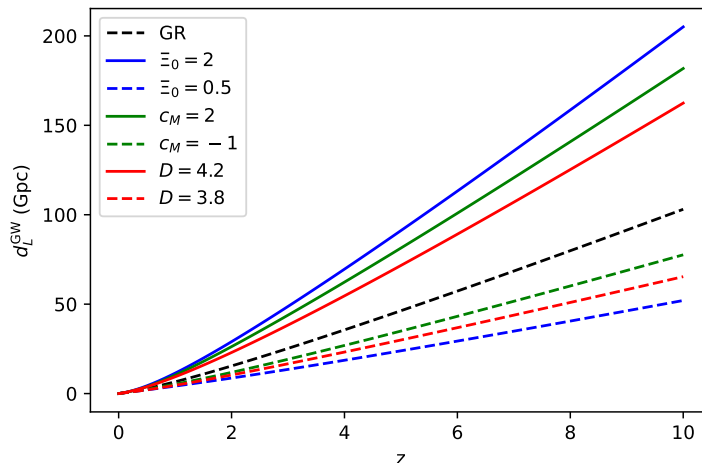


Figure 1. GW luminosity distances in the three modified gravity models analysed in this paper compared to GR. We fix $n = 1.91$ in the (Ξ_0, n) parameterization, and $R_c = 100$ Mpc in the extra-dimension model.

where ϕ is the scalar degree of freedom, M_{eff} is the effective Planck mass which evolves with ϕ , X is a kinetic term defined by $X \equiv -\nabla^\mu \phi \nabla_\mu \phi / 2$, S_m is the matter action, and ψ_m are matter fields minimally coupled to the metric $g_{\mu\nu}$. When M_{eff} equals the Planck mass M_{P} , and $K = G_3 = 0$, the action reduces to GR. It has been shown in the literature (e.g. [39, 50, 51]) that the GW propagation equation derived from the reduced Horndeski action is given by

$$\tilde{h}'' + \mathcal{H}[2 + \alpha_M(z)]\tilde{h}' + c^2 k^2 \tilde{h} = 0, \quad (2.10)$$

where α_M is the running rate of the effective Planck mass defined by

$$\alpha_M(z) \equiv \frac{d \ln(M_{\text{eff}}/M_{\text{P}})^2}{d \ln a}. \quad (2.11)$$

We see from eq. (2.1) that the GW friction term $\delta(z)$ is related to $\alpha_M(z)$ by $\delta(z) = -\alpha_M(z)/2$. A widely-used and simple parameterization for $\alpha_M(z)$ is that it is proportional to the fractional dark energy density [52–55] as

$$\alpha_M(z) = c_M \frac{\Omega_\Lambda(z)}{\Omega_{\Lambda,0}} = c_M \frac{1}{\Omega_{m,0}(1+z)^3 + \Omega_{\Lambda,0}}, \quad (2.12)$$

where c_M is the free parameter to be constrained, and $\Omega_\Lambda(z)$ is the fractional dark energy density. This assumption associates the deviation from GR to the growth of the dark energy density parameter. On one hand, this seems a plausible assumption if a cosmological modified gravity theory fulfils its principle motivation to replace dark energy. On the other, it has been argued that eq. (2.12) is not representative of the true evolution of mainstream gravity theories [56]. We will use eq. (2.12) as a general agnostic expression; a different parameterization can be straightforwardly replaced in our calculations. For instance another way of parameterization can be found in [57].

With the relation $\delta(z) = -\alpha_M(z)/2$, one obtains by substituting the above parameterization into eq. (2.6) that

$$d_L^{\text{GW}}(z) = d_L^{\text{EM}}(z) \exp \left\{ \frac{c_M}{2\Omega_{\Lambda,0}} \ln \frac{1+z}{[\Omega_{m,0}(1+z)^3 + \Omega_{\Lambda,0}]^{1/3}} \right\}. \quad (2.13)$$

This parameterization only depends on one free parameter, c_M . GR is recovered when $c_M = 0$. We have $d_L^{\text{GW}}(z)/d_L^{\text{EM}}(z) > 1$ when $c_M > 0$, and $d_L^{\text{GW}}(z)/d_L^{\text{EM}}(z) < 1$ when $c_M < 0$, as shown in figure 1.

2.3 Extra dimensions

Apart from scalar-tensor theories, additional spacetime dimensions are introduced in some extended theories of gravity, which can also leave an imprint on GW propagation. For instance, in Dvali-Gabadadze-Porrati (DGP) theory [58, 59], gravitational forces propagate in higher dimensions beyond 4-dimension spacetime, while ordinary matter is confined on the 3-dimension spatial brane. This leads to an effective ‘leakage’ of the perceived gravitational forces over cosmological distance scales, which transforms into an extra damping term in the GW propagation equation. It is shown in [60] that the GW and EM luminosity distances in some theories with extra dimensions are related by

$$d_L^{\text{GW}}(z) = d_L^{\text{EM}}(z) \left\{ 1 + \left[\frac{d_L^{\text{EM}}(z)}{(1+z)R_c} \right]^{n_D} \right\}^{\frac{D-4}{2n_D}}, \quad (2.14)$$

where D is the number of spacetime dimensions, R_c is the comoving scale which controls the transitions from the GR 4-dimensional regime at small scales to the higher-dimensional regime on large scales, and n_D controls the sharpness of the transition around R_c . On small scales where $d_L^{\text{EM}} \ll R_c$, GR is recovered and hence d_L^{GW} is indistinguishable from d_L^{EM} . On large scales where $d_L^{\text{EM}} \gg R_c$, the ratio of $d_L^{\text{GW}}/d_L^{\text{EM}}$ grows with the comoving distance to the power of $(D-4)/2$. Constraints on D with bright and dark sirens have been studied in [24, 60].

The comparison between d_L^{GW} and d_L^{EM} for the three modified gravity parameterizations with a selection of typical values is shown in figure 1. We can see that in all of the three models, $d_L^{\text{GW}}(z)/d_L^{\text{EM}}(z)$ can either be larger or smaller than 1, depending on the parameter value. It starts from 1 at low redshift and deviates from 1 as redshift increases. Therefore the deviation from GR is more significant at high redshift, so at the simplest level, GW events further away from may be more helpful in testing these gravity models. As we will see in sections 4 and 5, the reality of this statement will be complicated by the sky localisations and distance errors of each event.

3 Modifications to GWCOSMO

3.1 Bayesian framework of GWCOSMO

The detailed Bayesian statistical framework of the GWCOSMO code is presented in [19, 23, 26], which we will briefly review in this section. Then we will show how the posterior probabilities of modified gravity parameters are built. They are included in a set of hyper-parameters Λ

to be measured with the bright siren and the dark siren methods. These hyper-parameters also include cosmological parameters like H_0 , and parameters describing both the compact object population distribution and the merger rate redshift evolution, which will be discussed in section 3.2.

Given N_{det} detected GW events with observation data $\{x_{\text{GW}}\}$, the posterior for a set of hyper-parameters Λ is given by [61, 62]

$$p(\Lambda|\{x_{\text{GW}}\}, \{D_{\text{GW}}\}, I) \propto p(\Lambda|I)p(N_{\text{det}}|\Lambda, I) \prod_i^{N_{\text{det}}} p(x_{\text{GW}i}|D_{\text{GW}i}, \Lambda, I), \quad (3.1)$$

where D_{GW} is a parameter representing whether a GW event is detected or not, which takes the value of 1 or 0. I represents any other information or assumptions not explicitly contained in the parameter set Λ . The detection criteria of an event is that its SNR surpasses the SNR threshold we choose. The selected SNR threshold in our analysis will be discussed in the next two sections. The prior $p(\Lambda|I)$ is uniform in the analysis for all parameters in Λ . $p(N_{\text{det}}|\Lambda, I)$ is the probability of detecting N_{det} events. It depends on the intrinsic astrophysical rate of GW events R . As discussed in [61, 63], by choosing a flat prior on R in a log scale, we obtain the combined prior $p(\Lambda, R|I) \propto 1/R$, so that the dependence on R is dropped out when marginalizing $p(N_{\text{det}}|\Lambda, I)$ over R . Beyond this $p(N_{\text{det}}|\Lambda, I)$ has no further dependence on hyper-parameters. With Bayes' theorem, the likelihood for obtaining a set of GW observations x_{GW} can be expanded into

$$\begin{aligned} \prod_i^{N_{\text{det}}} p(x_{\text{GW}i}|D_{\text{GW}i}, \Lambda, I) &\propto \prod_i^{N_{\text{det}}} \frac{\int p(x_{\text{GW}i}|\theta, \Lambda, I)p(\theta|\Lambda, I)d\theta}{\int p(D_{\text{GW}i}|\theta, \Lambda, I)p(\theta|\Lambda, I)d\theta} \\ &\propto \left[\prod_i^{N_{\text{det}}} \int p(x_{\text{GW}i}|\theta, \Lambda, I)p(\theta|\Lambda, I)d\theta \right] \left[\int p(D_{\text{GW}}|\theta, \Lambda, I)p(\theta|\Lambda, I)d\theta \right]^{-N_{\text{det}}}, \end{aligned} \quad (3.2)$$

where θ represents a parameter set describing a detected GW signal, which must be marginalised over in order to find posterior for Λ . The parameters in θ of interest to us in this work include the detector frame masses m_1^{det} and m_2^{det} , redshift z , and sky location Ω . $p(x_{\text{GW}i}|\theta, \Lambda, I)$ in the numerator is obtained from the posterior samples of GW signal parameters estimated for each detected event, assuming a cosmological model. Here only cosmological hyper-parameters are relevant, while population and merger rate hyper-parameters enter in $p(\theta|\Lambda, I)$. $p(D_{\text{GW}i}|\theta, \Lambda, I)$ in the denominator is called the detection probability of GW events given a cosmology model. In GWCOSMO it is computed with a large number of injections of GW events. The injections are simulated detections of GW signals that are generated with the GW waveform injected with noises. The injections we used are generated with the IMRPhenomPv2 waveform over the sensitivity curves of specific detector network configurations such as O2, O3 or O4 set ups. Provided the selected SNR threshold, injections meeting the SNR threshold are used for computing the detection probability.

The posterior computed above is affected by the selection effects, which are the effects caused by the sensitivity of GW detectors, our choices of the population and cosmology model, prior ranges of θ and the SNR threshold in computing detection probability. If the prior range of parameters cannot cover the space that may be detected from GW events,

the posterior will be biased. However, it is not always more beneficial for the prior range to be larger, because some parameters may be degenerate, so that one of the unconstrained parameters may bias constraints on another one (we will see an example of this in section 5). In addition, the SNR threshold determines whether an event is counted as detected. It must be consistent with the SNR threshold of LVK detection, so it can't be too low. On the other hand, if the threshold is too high, we will have too few events to obtain a tighter constraint. Therefore the prior range and the SNR threshold need to be selected carefully.

The probabilistic distributions of the redshift and sky location of the GW event are marginalised over in different ways for the bright siren and the dark siren methods. For a bright siren, a prior on the redshift and the sky location can be retrieved from the host galaxy identified from an EM counterpart with high accuracy. In GWCOSMO the redshift prior of a bright siren $p(z|\Lambda, I)$ is approximated by a Gaussian distribution with the mean and the standard deviation being the same as the measured redshift of the host galaxy and its redshift uncertainty. (This in principle could be generalised to non-Gaussian redshift measurements in future.)

On the other hand, the prior on redshift and the sky location in the dark siren method is obtained with galaxy catalogue information. Without an EM counterpart, a GW event is localized within a patch of the sky that is too large to identify its host galaxy. However, galaxies recorded in catalogues within this patch of the sky are potential host galaxies that can provide redshift information of the GW source. The redshift prior of the GW source is then obtained by combining the redshift distribution of galaxies in the catalogues within the sky localization area of the event upon the prior of the comoving redshift volume.

In GWCOSMO the redshift prior is computed for each pixel on the sky [19]. The resolution in dividing the sky is determined by a parameter called `nside` introduced by `healpy` [64, 65]. For larger `nside`, there are more pixels in the whole sky and correspondingly fewer galaxies in each pixel; this allows for more finely-grained features in across different patches in the sky, but takes longer to evaluate numerically. For a typical value of `nside` = 64, the sky is divided into 49,152 pixels. The pixelation of the line-of-sight (LOS) redshift prior reduces the bias caused by variations in galaxy catalogue incompleteness — the completeness is different for sky areas explored by different galaxy surveys that make up the GLADE galaxy catalogue [66]. Assuming a uniform completeness across the sky will overestimate or underestimate the completeness of some areas, creating bias in GW host identification. Therefore using galaxy catalogue with specific completeness in each pixel, which is defined according to the apparent magnitude threshold of the survey, will reduce such bias.

The final form of the posterior with the dark siren approach is given by

$$\begin{aligned}
 p(\Lambda|\{x_{\text{GW}}\}, \{D_{\text{GW}}\}, I) &\propto p(\Lambda|I)p(N_{\text{det}}|\Lambda, I) \left[\iint p(D_{\text{GW}}|z, \theta, \Lambda, I)p(\theta|\Lambda, I) \sum_j^{N_{\text{pix}}} p(z|\Omega_j, \Lambda, I)d\theta dz \right]^{-N_{\text{det}}} \\
 &\times \prod_i^{N_{\text{det}}} \left[\iint \sum_j^{N_{\text{pix}}} p(x_{\text{GW}_i}|z, \Omega_j, \theta, \Lambda, I)p(\theta|\Lambda, I)p(z|\Omega_j, \Lambda, I)d\theta dz \right],
 \end{aligned}
 \tag{3.3}$$

where $p(z|\Omega_j, \Lambda, I)$ is the LOS redshift prior for each pixel at a sky location Ω_j , which can be pre-computed before estimating the posterior, of which the details are presented in [26].

Because of this, we do not include modified gravity models that change the universe expansion history, in which the LOS redshift prior needs to be recomputed while varying modified gravity parameters, and it is very expensive at the moment. In this paper, the GLADE+ galaxy catalogue in the K band is used for the analysis. Although the K band is less complete than the B_J band, its incompleteness model is better understood than the B_J band, and hence gives more reliable results. Unlike the K band, the B_J band suffers from an issue that its incompleteness is not well-captured by a simple magnitude threshold.

In the GR version of GWCOSMO, H_0 is the only cosmological parameter in Λ . Here we extend GWCOSMO to the estimation of sets of modified gravity parameters such as Ξ_0 , n in eq. (2.7), c_M in eq. (2.13), or D in eq. (2.14). Since these parameters modify $d_L^{\text{GW}}(z)$, but leave $d_L^{\text{EM}}(z)$ unchanged, the LOS redshift prior from the galaxy catalogue is not affected. However, modification is required when marginalising over the source frame masses of compact object binaries, because z inferred from $d_L^{\text{GW}}(z)$ now differs from that from $d_L^{\text{EM}}(z)$ under different values of modified gravity parameters. Inside a pixel, the numerator integrand of the likelihood is given by

$$p(x_{\text{GW}i}|z, \Lambda, I) = \int_{m_{\text{min}}}^{m_{\text{max}}} \int_{m_{\text{min}}}^{m_1^{\text{s}}} p(x_{\text{GW}i}|z, m_1^{\text{s}}, m_2^{\text{s}}, \Lambda, I) p(m_1^{\text{s}}, m_2^{\text{s}}|\Lambda, I) dm_2^{\text{s}} dm_1^{\text{s}}, \quad (3.4)$$

where m_1^{s} and m_2^{s} are the component masses in the GW source frame, with the assumption $m_1^{\text{s}} > m_2^{\text{s}}$. m_{min} and m_{max} are the minimum and the maximum mass for a source of that type, given by the population model. The prior on source masses $p(m_1^{\text{s}}, m_2^{\text{s}}|I)$ is inferred from the compact object mass distribution models [67]. However, the posterior samples from GW observation data are generated in the detector frame. In addition, it is $d_L^{\text{GW}}(z)$ that is measured from the GW signal instead of z . Therefore we have

$$p(x_{\text{GW}i}|z, \Lambda, I) = \int_{m_{\text{min}}}^{m_{\text{max}}} \int_{m_{\text{min}}}^{m_1^{\text{s}}} p(x_{\text{GW}i}|d_L^{\text{GW}}(z, \Lambda), m_1^{\text{det}}[m_1^{\text{s}}, z], m_2^{\text{det}}[m_2^{\text{s}}, z], I) p(m_1^{\text{s}}, m_2^{\text{s}}|I) dm_2^{\text{s}} dm_1^{\text{s}}, \quad (3.5)$$

Applying Bayes' theorem we obtain

$$p(x_{\text{GW}i}|d_L^{\text{GW}}, m_1^{\text{det}}, m_2^{\text{det}}, I) = \frac{p(d_L^{\text{GW}}, m_1^{\text{det}}, m_2^{\text{det}}|x_{\text{GW}i}, I) p(x_{\text{GW}i}|I)}{\pi(d_L^{\text{GW}}, m_1^{\text{det}}, m_2^{\text{det}}|I)} \propto \frac{p(d_L^{\text{GW}}, m_1^{\text{det}}, m_2^{\text{det}}|x_{\text{GW}i}, I)}{\pi(d_L^{\text{GW}}, m_1^{\text{det}}, m_2^{\text{det}}|I)}, \quad (3.6)$$

where $\pi(d_L^{\text{GW}}, m_1^{\text{det}}, m_2^{\text{det}}|I)$ is the prior applied to d_L^{GW} , m_1^{det} and m_2^{det} when computing the posterior samples of a GW event. The posterior samples are effectively the numerator of eq. (3.6). For the GWTC-3 data set, the posterior samples are generated with two options: $\pi(d_L^{\text{GW}}) \propto (d_L^{\text{GW}})^2$ or $\pi(d_L^{\text{GW}})$ uniform in comoving volume prior. In this paper we used the data set with $\pi(d_L^{\text{GW}}) \propto (d_L^{\text{GW}})^2$. Meanwhile $\pi(m_1^{\text{det}}, m_2^{\text{det}})$ is uniform. Then we convert the posterior samples of d_L^{GW} into those of z by

$$p(d_L^{\text{GW}}(z, \Lambda), m_1^{\text{det}}, m_2^{\text{det}}|x_{\text{GW}i}, I) = p(z, m_1^{\text{det}}, m_2^{\text{det}}|x_{\text{GW}i}, \Lambda, I) \left| \frac{\partial z}{\partial d_L^{\text{GW}}} \right|. \quad (3.7)$$

The conversion from d_L^{GW} to z , and the Jacobian $|\partial z / \partial d_L^{\text{GW}}|$ for posterior samples needs to be modified under different parameterizations, which is the main modification compared to

the GR version of GWCOSMO. Finally, the transformation from source masses into detected masses gives

$$dm_1^s dm_2^s = \left| \frac{\partial(m_1^s, m_2^s)}{\partial(m_1^{\text{det}}, m_2^{\text{det}})} \right| dm_1^{\text{det}} dm_2^{\text{det}} = \frac{1}{(1+z)^2} dm_1^{\text{det}} dm_2^{\text{det}}. \quad (3.8)$$

Following the steps above, the expression for the integrand in eq. (3.5) becomes (remember this is part of the numerator in eq. (3.3)):

$$p(x_{\text{GW}i}|z, \Lambda, I) \propto \int_{m^{\text{det}}(m_{\text{min}}, z)}^{m^{\text{det}}(m_{\text{max}}, z)} \int_{m^{\text{det}}(m_{\text{min}}, z)}^{m^{\text{det}}(m_1^s, z)} \frac{p(z, m_1^{\text{det}}, m_2^{\text{det}}|x_{\text{GW}i}, \Lambda, I)}{(d_L^{\text{GW}})^2} \left| \frac{\partial z}{\partial d_L^{\text{GW}}} \right| \times \frac{p(m_1^s[m_1^{\text{det}}, z], m_2^s[m_2^{\text{det}}, z]|I)}{(1+z)^2} dm_2^{\text{det}} dm_1^{\text{det}}. \quad (3.9)$$

The posterior samples on z computed from d_L^{GW} given Λ , m_1^{det} and m_2^{det} for a GW event can be approximated as a sum over delta-functions:

$$p(z, m_1^{\text{det}}, m_2^{\text{det}}|x_{\text{GW}i}, \Lambda, I) \approx \frac{1}{N_{\text{pos}}} \sum_{k=1}^{N_{\text{pos}}} \delta(z - z_k(d_{L,k}^{\text{GW}}, \Lambda)) \delta(m_1^{\text{det}} - m_{1,k}^{\text{det}}) \delta(m_2^{\text{det}} - m_{2,k}^{\text{det}}), \quad (3.10)$$

where k denotes the k^{th} posterior sample and N_{pos} denotes the total number of posterior samples. Inserting this back to eq. (3.9) we obtain the result as

$$p(x_{\text{GW}i}|z, \Lambda, I) \propto \frac{1}{N_{\text{pos}}} \sum_{k=1}^{N_{\text{pos}}} p(m_1^s(m_{1,k}^{\text{det}}, z), m_2^s(m_{2,k}^{\text{det}}, z)|I) \left| \frac{\partial z}{\partial d_L^{\text{GW}}} \right| \frac{1}{(d_L^{\text{GW}})^2 (1+z)^2} \delta(z - z_k(d_{L,k}^{\text{GW}}, \Lambda)). \quad (3.11)$$

The above quantity is then multiplied by the LOS redshift prior³ $p(z|\Omega_j, \Lambda, I)$, and integrated over redshift. Summing the results over all pixels one obtains the numerator of the likelihood as indicated in eq. (3.3). The final result of the numerator is given by

$$p(x_{\text{GW}i}|\Lambda, I) \propto \frac{1}{N_{\text{pos}}} \sum_{k=1}^{N_{\text{pos}}} p(m_1^s(m_{1,k}^{\text{det}}, z_k), m_2^s(m_{2,k}^{\text{det}}, z_k)|I) \left| \frac{\partial z}{\partial d_L^{\text{GW}}} \right|_{z=z_k} \frac{1}{(d_{L,k}^{\text{GW}})^2 (1+z_k)^2} \times \sum_j^{N_{\text{pix}}} p(z_k|\Omega_j, \Lambda, I). \quad (3.12)$$

The details of computing the LOS redshift prior $p(z_k|\Omega_j, \Lambda, I)$ are described in [26]. Similarly, injections in the detector frame need to be converted into the source frame when marginalising the detection probability in the denominator of the likelihood over redshift and component masses. The detection probability is given in the same form by

$$p(D_{\text{GW}}|\Lambda, I) \propto \frac{1}{N_{\text{inj}}} \sum_{k=1}^{N_{\text{inj}}} p(m_1^s(m_{1,\text{inj},k}^{\text{det}}, z_k), m_2^s(m_{2,\text{inj},k}^{\text{det}}, z_k)|I) \left| \frac{\partial z}{\partial d_L^{\text{GW}}} \right|_{z=z_{\text{inj},k}} \frac{1}{(d_{L,\text{inj},k}^{\text{GW}})^2 (1+z_{\text{inj},k})^2} \times \sum_j^{N_{\text{pix}}} p(z_{\text{inj},k}|\Omega_j, \Lambda, I). \quad (3.13)$$

³In reality, a kernel density estimate (KDE) of $p(x_{\text{GW}i}|z, \Lambda, I)$ is created first, and multiplied with the redshift prior.

By inserting the results of the numerator and the denominator back into eq. (3.3), and multiplying over all the GW events, we can thus obtain the posterior on the parameter set Λ .

3.2 Prior on source masses and merger rate evolution

The source mass prior $p(m_1^s, m_2^s|I)$ is obtained from the mass distribution model. There are several different BBH mass distribution models adopted in the LVK analysis, such as the Broken Power Law model and the Power Law + Peak model [31, 67]. In our analysis, we use the Power Law + Peak model for black hole distribution, because it is preferred over other models from the population analysis of O3 data [68], and is adopted by a number of similar works [24, 26, 27]. We use a uniform distribution for neutron stars as in the LVK GWTC-3 cosmology paper [31]. There are 8 parameters in the description of the Power Law + Peak model. In our analysis of GWTC-3 events in section 5, we will study cases where these parameters are fixed or varied. In the cases where they are varied, the priors on them are uniformly distributed, and are listed in table 1 in section 5.

The explicit redshift prior actually depends on the presence of GW sources, which we denote by s . By Bayes' theorem it is given that

$$p(z|s, \Omega_j, \Lambda, I) = \frac{p(s|z, \Omega_j, \Lambda, I)p(z|\Omega_j, \Lambda, I)}{p(s|\Omega_j, \Lambda, I)}. \quad (3.14)$$

The term $p(s|\Omega_j, \Lambda, I)$ is cancelled for appearing in both the numerator and the denominator of eq. (3.3). Assuming that GW sources are distributed uniformly in the sky on the scale of our pixels, the dependence on Ω_j can be dropped. Then $p(s|z)$ depends on the redshift evolution of the merger rate of compact object binaries. Following the LVK cosmology paper [31], we adopt a parameterization motivated by connection of GW events to the star formation rate in [69], which has the form

$$p(s|z, \gamma, k, z_p) = [1 + (1 + z_p)^{-\gamma-k}] \frac{(1+z)^\gamma}{1 + [(1+z)/(1+z_p)]^{\gamma+k}}. \quad (3.15)$$

The free parameters γ , k and z_p control the merger rate redshift evolution, and are included in the hyper-parameters to be measured. The merger rate evolution can be scaled by an overall factor R_0 , but it is not included in the current GWCOSMO analysis.

4 Tests on mock data

4.1 Simulation setup

Because of the limited number of detected GW events at the present time, we first use mock GW events to test the measurement of modified gravity parameters by GWCOSMO. We use the First Two Years of Electromagnetic Follow-Up with Advanced LIGO and Virgo (F2Y) mock data set [70] to test our bright siren methodology. This catalogue was generated by injecting around 50,000 BNS mergers with the LIGO and Virgo sensitivity curves mimicking the O2 run. The criteria for a detected bright siren is having single detector SNR over 4, and a network SNR over 12. As a result only approximately 500 BNS events are marked as detected events, among which 250 of them are randomly selected for parameter estimation. We used these 250 bright sirens for our test of gravity with GWCOSMO; we note this is the same mock data challenge as employed for H_0 in [23].

We generated our own BNS injections for computing the detection probability. We used the same O2 LIGO and Virgo sensitivity curves as in [70], which are the “early” and “mid” sensitivity curves for L1 and H1 in LIGO technical document [71], and the O2 noise curve for V1 from [72]. Each of the three detectors has a duty factor of 76.8% as indicated in figure 9 of [70]. We generated 50,000 injections with the SNR threshold of 12, GW luminosity distance between 0 and 500 Mpc, and the BNS component mass between 1.2 and $1.6 M_{\odot}$, assuming $m_1 > m_2$.

A significant feature of the F2Y dataset is that it is generated with a local universe approximation, given that the redshifts of detectable BNS events are expected to be very low. The linear cosmology approximation $d_L^{\text{EM}} \approx cz/H_0$ is applied for the EM counterparts of BNS injections. The BNS d_L^{GW} is then parameterized with the linearized d_L^{EM} in the three modified gravity models introduced in section 2. We note that in the case of the Hubble constant analysis, the detection probability under a low redshift approximation simply grows as H_0^3 [21]; in the case of modified gravity it acquires additional corrections. On the other hand, the detected BNS masses have no redshift correction since this is negligible in the nearby universe. The mass prior we used in the analysis is uniform in $[1.2, 1.6] M_{\odot}$, since the F2Y BNSs were generated with component masses between 1.2 and $1.6 M_{\odot}$. Such a mass prior is much narrower than the range of $[1.0, 3.0] M_{\odot}$ used in later LVK analyses: neutron stars were generally expected to have masses strongly peaked around $1.4 M_{\odot}$ before much heavier neutron stars were detected in events from the O3 observing run.

4.2 Results

We first measure Ξ_0 in the (Ξ_0, n) parameterization shown in eq. (2.7) with a uniform prior in $[0.05, 10]$, while fixing other cosmological parameters as $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and $n = 1.91$.⁴ The normalized posterior of Ξ_0 for each bright siren is plotted in black, and the combined posterior is plotted in red in figure 2. The measured value with 1σ bound is $\Xi_0 = 0.96_{-0.19}^{+0.19}$, which is consistent with GR. Our results show that GWCOSMO is capable of constraining Ξ_0 well in the simplified case of a one-parameter analysis with a number of bright sirens. The number of bright sirens needed to obtain this bound (250) may seem quite large; this is because, as discussed in section 2.1, Ξ_0 is most relevant to high-redshift GW events, and the mock bright sirens used here are all at low redshifts.

However, the 1D measurement of Ξ_0 can be misleading, because it has assumed H_0 to be a fixed value. Since H_0 and Ξ_0 both determine the conversion from GW distances to redshifts, they are expected to be degenerate. Ideally it is more sensible to perform a 2D H_0 - Ξ_0 joint analysis. However, the number of posterior samples for the F2Y mock BNS is ~ 1000 , which is very low compared to the real events that have tens of thousands of posterior samples. Therefore the estimation of redshift for F2Y data is less accurate. Our joint analysis shows that the GR value ($H_0 = 70, \Xi_0 = 1$) is on the edge of the 3σ bound of the 2D joint posterior. Due to the low number of posterior samples for the F2Y data, the 2D result is unreliable.

We further perform the same analysis on c_M in the Horndeski class parameterization shown in eq. (2.13) and D in the extra-dimension parameterization shown in eq. (2.14). We fix H_0 , mass distribution and merger rate redshift evolution parameters as in the Ξ_0 measurement.

⁴Note that other cosmological parameters such as Ω_{M0} will not be varied in any of our analyses: they are well-measured by EM probes already, so we adopt their central values.

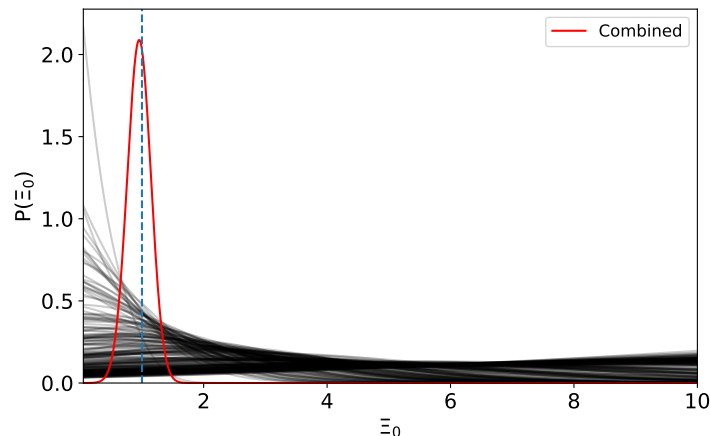


Figure 2. The normalized posterior probability distribution for Ξ_0 for individual events (black) and the combined one (red). The blue vertical line labels $\Xi_0 = 1$ recovery of GR.

In the measurement of c_M , the values of $\Omega_{m,0}$ and $\Omega_{\Lambda,0}$ are needed to reconstruct the evolution of the fraction of dark energy density as described in eq. (2.12). We adopt $\Omega_{m,0} = 0.3065$ from the Planck result [73] as in the LVK cosmology paper [31], and $\Omega_{\Lambda,0} = 1 - \Omega_{m,0}$ as other components of energy density are negligible today. We apply the uniform prior for c_M in $[-5, 10]$. The normalized posterior of c_M for individual events and the combined posterior are plotted in the left panel of figure 3. The 1σ bound measured value is $c_M = -0.18^{+0.71}_{-0.72}$, which is consistent with GR where $c_M = 0$.

In the measurement of D in the extra-dimension parameterization shown in eq. (2.14), we choose to fix $R_c = 100$ Mpc, which is the same magnitude as that of the luminosity distances for BNS events, so that the modified gravity effect can be probed at such distances. This value may not match the one motivated from fundamental physics, but it allows us to test the constraint on D at a scale of the BNS mock data we have available. In addition we also fix $n_D = 1$. The prior of D is uniform in $[3.5, 5.0]$. The effect of varying D with fixed R_c and n_D is shown in figure 1. The normalized posterior of D for individual events and the combined posterior are plotted in the right panel of figure 3. The 1σ bound measured value is $D = 4.015^{+0.027}_{-0.027}$, which is consistent with the GR limit of this class of theories.

In summary, we obtain good 1D constraints on modified gravity parameters in all of the three alternative gravity models using 250 mock bright sirens, and recover consistency with GR in all cases (as of course expected, as the mock data is generated under GR). However, because of the low number of posterior samples of the F2Y data, 2D joint constraints of H_0 and modified gravity parameters are not reliable. Still, our test shows that GWCOSMO is capable of measuring deviations beyond GR with the bright siren method. It will be used to test gravity with bright sirens that are expected to be detected in the O4 run.

5 Reanalysing the GWTC-3 data

5.1 Parameter ranges

In this section we show the results of reanalysing the GWTC-3 events with GWCOSMO, including modified gravity parameters. We use the same 46 dark sirens from the GWTC-3

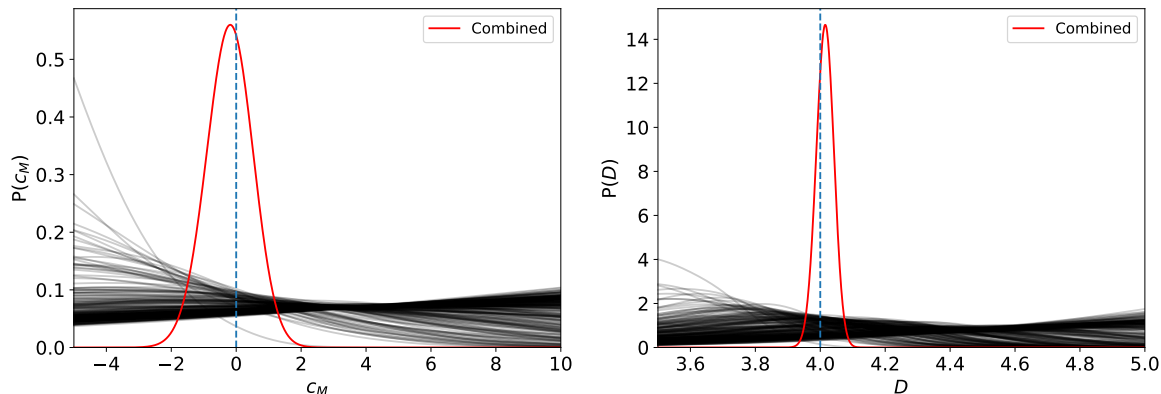


Figure 3. The normalized posterior probability distribution for c_M (left panel) and D (right panel) for individual events (dark) and the combined constraint (red). The blue vertical lines indicate the GR limits of $c_M = 0$ and $D = 4$.

catalogues as in the LVK GWTC-3 cosmology paper [31] with net SNR larger than 11, including 42 BBH mergers, 1 BNS merger GW190425, and 3 NSBH mergers GW190814, GW200105_162426 and GW200115_042309. The posterior samples of GWTC-2.1 [74] are used in computing the numerator of the likelihood for events from O1 to O3a, and GWTC-3 samples [30] for events in O3b. The denominator of the likelihood is computed using 2×10^6 GW injections with the SNR threshold set to 11. The injected detected masses range from 1 to $500 M_\odot$, and the injected GW distances are varied between 0.1 and $\sim 20,000$ Mpc. Such a high value for the maximum GW distance is needed to cover the GW distance for extreme values of the modified gravity parameters for $z \in (0, 10)$.

The LOS redshift prior we used is pre-computed using the GLADE+ galaxy catalogue in the K band, with resolution of $n_{\text{side}} = 64$. We will focus on the $\Xi_0 - n$ parameterization for the next few subsections, but the other modified gravity parameters will be treated analogously, and their final results will be discussed in section 5.3. Given the maximum redshift $z = 10$ of the LOS redshift prior, we need to choose joint prior bounds of H_0 and Ξ_0 such that the maximum injected GW distance corresponds to redshift not larger than 10. As a result, we vary Ξ_0 in a range of $[0.3, 10]$ when fixing $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$. A combination of the maximum H_0 and the minimum Ξ_0 gives the largest redshift for the maximum injected GW distance, so we need to reduce the higher bound of H_0 or raise the lower bound of Ξ_0 when varying both of them. We chose a range of $[20, 140] \text{ km s}^{-1} \text{ Mpc}^{-1}$ for H_0 and $[0.35, 10]$ for Ξ_0 for the joint analysis.

5.2 Dark sirens results

5.2.1 1D measurement of Ξ_0

We first measure the posterior of Ξ_0 in eq. (2.7) with a flat prior on $[0.3, 10]$, while fixing the other two cosmological parameters to be $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and $n = 1.91$. We also fix the parameters in the Power Law + Peak mass prior as in the LVK cosmology paper [31], which are $\alpha = 3.78$, $\beta = 0.81$, $m_{\text{min}}^{\text{BH}} = 4.98 M_\odot$, $m_{\text{max}}^{\text{BH}} = 112.5 M_\odot$, $\delta_m = 4.8 M_\odot$, $\mu_g = 32.27 M_\odot$, $\sigma_g = 3.88 M_\odot$ and $\lambda_p = 0.03$. A brief description of the physical significance

Parameter	Definition	Prior
H_0	Hubble constant	$\mathcal{U}(20.0, 140.0)$
Ξ_0	Modified gravity parameter controlling high- z limit of distance ratio in eq. (2.7)	$\mathcal{U}(0.35, 10.0)$
n	Modified gravity parameter controlling steepness of distance ratio in eq. (2.7)	$\mathcal{U}(0.1, 10.0)$
α	The power of the power law component in the primary mass distribution	$\mathcal{U}(1.5, 8.0)$
β	The power of the power law component in the mass ratio distribution	$\mathcal{U}(-4.0, 6.0)$
$m_{\min}^{\text{BH}} [M_\odot]$	The minimum mass of the mass distribution	$\mathcal{U}(2.0, 10.0)$
$m_{\max}^{\text{BH}} [M_\odot]$	The maximum mass of the mass distribution	$\mathcal{U}(50.0, 200.0)$
λ_p	Fraction of the model in the Gaussian component	$\mathcal{U}(0.0, 0.5)$
μ_g	Mean of the Gaussian component in the primary mass distribution	$\mathcal{U}(10.0, 50.0)$
σ_g	Width of the Gaussian component in the primary mass distribution	$\mathcal{U}(0.1, 20.0)$
δ_m	Range of mass tapering at the lower end of the mass distribution	$\mathcal{U}(0.0, 15.0)$
γ	The power of the power law distribution of the rate evolution before redshift z_p	$\mathcal{U}(0.0, 12.0)$
k	The power of the power law distribution of the rate evolution after redshift z_p	$\mathcal{U}(0.0, 10.0)$
z_p	The redshift turning point between two power law distributions	$\mathcal{U}(0.0, 10.0)$

Table 1. Prior of the parameters in the Power Law + Peak BBH mass distribution.

of these parameters is given in table 1. In the cases of NSBH and BNS, we fix $m_{\min}^{\text{NS}} = 1.0 M_\odot$ and $m_{\max}^{\text{NS}} = 3.0 M_\odot$. The merger rate evolution parameters are also fixed with the same values $\gamma = 4.59$, $k = 2.86$ and $z_p = 2.47$. The normalized posterior of Ξ_0 for each individual event is plotted in figure 4. We can see that there are three typical patterns in the distribution of $p(\Xi_0)$ for different events: favoring low Ξ_0 such as GW190910_112807, favoring high Ξ_0 such as GW191216_213338, and peaking at a certain value of Ξ_0 such as GW170809_082821. The reasons for these patterns can be found by looking at the posterior sample distribution in z converted from d_L^{GW} for different Ξ_0 for each event, which we will elaborate on in appendix B. The measured value from the combined posterior with 1σ bound is $\Xi_0 = 1.06_{-0.18}^{+0.25}$, which is consistent with GR.

For most dark siren events in GWTC-3, the redshift support from the galaxy catalogue does not have a significant effect on improving the precision of the Ξ_0 measurement. The event with the most galaxy support is GW190814, which is shown in figure 5; here we compare the analysis with the galaxy catalogue and an empty catalogue (no galaxies). $p(\Xi_0)$ with galaxy catalogue support peaks at lower Ξ_0 compared to the flat distribution without the support, which shows that galaxy catalogues have the potential to increase the precision of

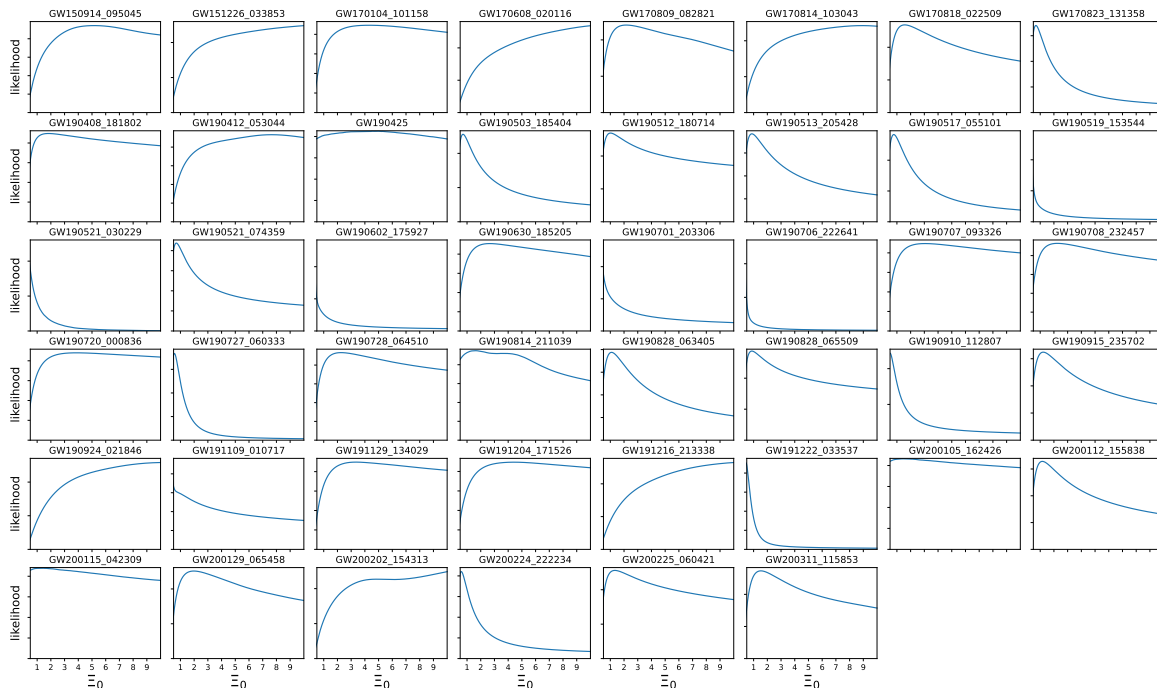


Figure 4. The one-dimensional likelihood of Ξ_0 from individual selected events used in our analysis. Note that all other parameters are held fixed. These 1D curves demonstrate the range of possible curve shapes obtained when galaxy catalogue support is moderately weak; see appendix B for a detailed explanation.

modified gravity constraints with dark sirens. The galaxy catalogue information has only a weak effect on the current dark siren analysis; this is because the present localization of GW events has a relatively low accuracy, so that the information from clustering of galaxies gets washed out as it is averaged over too many pixels. If the localization is small enough, our method is sensitive to clustering of galaxies so that the results will favour some values of H_0 and Ξ_0 . In addition, the galaxy catalogue is incomplete, especially at high redshift, so it provides limited support to GW events further away.

In future, with an increase in the sensitivity of the KAGRA detector, and the addition of LIGO India to the terrestrial network, the localization of GW events will be improved. Moreover, galaxy surveys by new detectors such as the Dark Energy Spectroscopic Instrument (DESI) [75] and the Euclid space telescope [76] in the future will provide millions of further spectroscopic galaxy redshifts, particularly at higher redshifts where catalogue support is currently lacking. This will further accelerate the precision of dark sirens constraints.

5.2.2 BBH joint analysis

Next we perform a joint posterior measurement for *all* hyper-parameters under the (Ξ_0, n) MG parameterization with nested sampling in GWCOSMO. We adopt uniform priors for all hyper-parameters including the mass distribution model and the merger rate redshift evolution model as listed in table 1. We present the corner plot of the joint posteriors for 42 BBH events from nested samplings for 6 of the most interesting parameters in figure 6, which

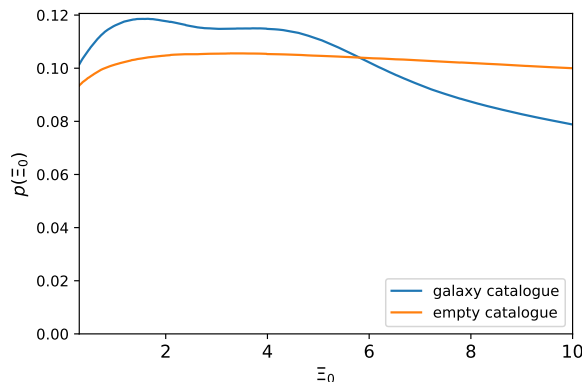


Figure 5. The 1D posterior of Ξ_0 for GW190814, with both the GLADE+ galaxy catalogue and an empty catalogue analysis (not galaxies). We see that for this event the clustering in the galaxy catalogue (dis)favors certain values of Ξ_0 , creating a bumpy structure in the likelihood. Comparable features can be seen in the H_0 likelihood shown in figure 9 of [26].

are γ , $m_{\text{max}}^{\text{BH}}$, μ_g , H_0 , Ξ_0 and n . The full corner plot with all parameters is shown in figure 11 in appendix A. We can see that some BBH mass distribution parameters are moderately well-constrained; however, k and z_p in the merger rate redshift evolution model are poorly constrained, which is similar to results in the LVK cosmology paper [31]. In addition, the estimated value of H_0 is $55.93^{+38.35}_{-26.29} \text{ km s}^{-1} \text{ Mpc}^{-1}$, again an uncertainty similar to previous results. The estimated value for Ξ_0 is $1.29^{+1.22}_{-0.67}$, which is consistent with GR; as expected, the result is less constrained than that in the 1D measurement because of co-variance with other parameters. Similar results have been obtained using the ICAROGW software [24, 27]. Meanwhile n is measured to be $0.79^{+3.86}_{-0.65}$, which is lower than the fiducial value of 1.91, but with quite broad uncertainties. The reason that $p(n)$ favors low values of n is likely that the $\Xi(z)$ parameterization reduces to GR when $n = 0$, so a shift towards low values of n is an alternative fit to the data (instead of $\Xi_0 = 1$), if the data are consistent with GR. Moreover, as shown by the 2D Ξ_0 - n posterior contour, Ξ_0 is better constrained when n is larger. This is because the redshift-dependent term with the inverse power of n in eq. (2.7) becomes less significant when n is large, leaving the constant Ξ_0 as the more dominant term. Therefore the constraint is put more directly on Ξ_0 when n is large.

From the contour plot we can see that H_0 and Ξ_0 are partially degenerate with each other. This is because $d_L^{\text{GW}} \propto \Xi_0/H_0$ as shown in eq. (2.2) and (2.7), so that in the conversion of posterior samples from d_L^{GW} to z in computing the likelihood, varying H_0 and Ξ_0 while keeping the ratio Ξ_0/H_0 unchanged gives the same redshift. So the result that the estimated value of H_0 is higher than that in the LVK cosmology paper [31] is likely caused by Ξ_0 favoring a slightly higher value than GR. Similar degeneracy can also be found in figure 13 of [24].

Apart from H_0 , we can also see that Ξ_0 is strongly degenerate with γ . In the computation of the likelihood, the LOS redshift prior is multiplied by the merger rate redshift evolution, which is approximately proportional to $(1+z)^\gamma$ when $z \ll z_p$. Given that all GW events in GWTC-3 have redshift less than 1, and from star formation arguments z_p is typically considered to have a value around 2, the power of γ dominates the redshift evolution of the merger rate. This can also explain why k , which is the power of the redshift evolution

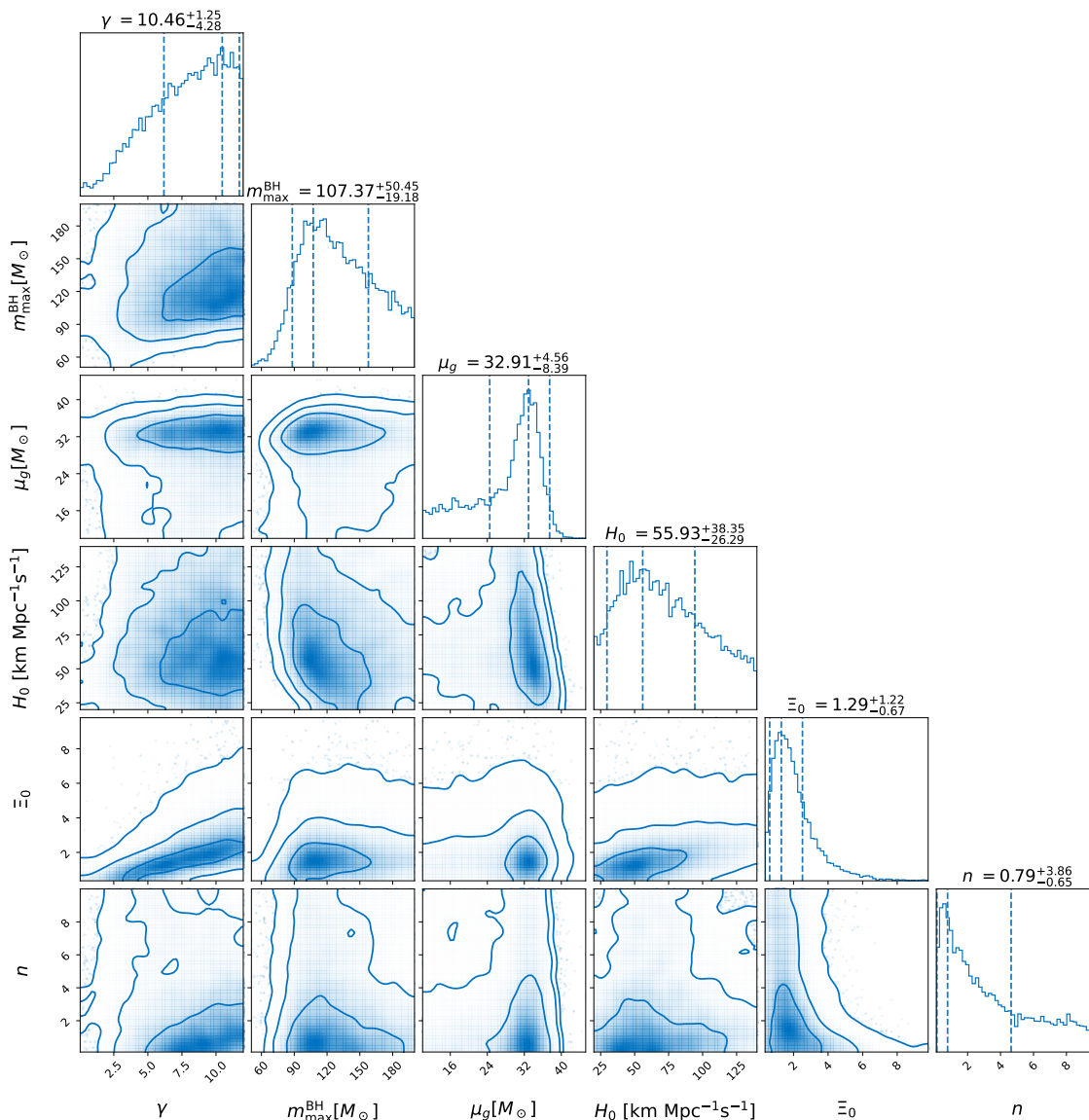


Figure 6. Selected corner plot of joint posteriors for 42 BBH events from the GWTC-3 catalogue computed by the GWCOSMO sampling run. The contours show 1, 2 and 3σ bound for the 2D posteriors. The vertical lines in the 1D histogram plot correspond to the peak values and 1σ bound of the parameters in the title. The full corner plot including all parameters is shown in figure 11. A brief description of the mass and merger rate hyper-parameters is given in table 1.

when $z > z_p$, and z_p are poorly constrained with GWTC-3 events. So the numerator of the likelihood is proportional to Ξ_0 and $(1+z)^\gamma$, leading to a strong degeneracy between Ξ_0 and γ . A similar feature is also found in figure 12 of the ICAROGW results [24], where the distribution of $p(\gamma)$ with the Power Law + Peak BBH population model is similar to ours. Given that the distribution of $p(\gamma)$ favors a higher value, if we allow the prior bound of γ to be larger, the distribution of $p(\Xi_0)$ will migrate to higher values because of the degeneracy. The prior ranges for γ , k and z_p are chosen to be sufficiently wide in consideration of the effect of a possible time delay between the formation and the merger of the binary.

Having tested our method with the F2Y mock data set at low redshift in section 4, we validate our results for BBHs at higher redshift by sanity check against independent results from other groups, for instance ICAROGW [27, 28], and they show good agreement. Although a realistic mock BBH data set applicable at high redshift is desirable for validating our method, generating it is computationally expensive and is beyond the scope of this work.

5.2.3 NSBH joint analysis

We then perform joint analysis on all hyper-parameters with the three NSBH mergers by nested samplings implemented in GWCOSMO. The NSBH events selected are GW190814, GW200105_162426 and GW200115_042309. Note that GW190814 is not a confirmed NSBH, but we choose to treat it as one for this analysis. In addition to all of the parameters in the analysis with BBHs, parameters m_{\min}^{NS} and m_{\max}^{NS} for the uniform neutron star population model are included in the estimation. We use the uniform prior of $m_{\min}^{\text{NS}} \in [0.5, 1.5]M_{\odot}$ and $m_{\max}^{\text{NS}} \in [1.501, 5]M_{\odot}$. Due to the small number of events, most of the parameters in the BBH population model and the merger rate evolution model are poorly constrained. Furthermore, the marginalized posteriors for some of the parameters show misleading results, as shown in the selected contours in figure 7. For example, the posteriors for μ_g and σ_g indicate a narrow Gaussian peak at $\sim 24 M_{\odot}$ for the black hole population model. This is caused alone by the black hole with mass $\sim 23 M_{\odot}$ in GW190814. The same appearance is probably also shown in the constraint of m_{\max}^{NS} , which is driven by the assumption that GW190814's secondary mass is a neutron star with mass $\sim 2.6 M_{\odot}$. Meanwhile the constraints on cosmological parameters are generally weaker than those with the BBHs due to the lower number of NSBH events.

5.3 Combined BBH and NSBH joint analysis

We can improve the constraints on cosmological parameters by combining the results of BBHs and NSBHs. We believe this is the first time a combined dark sirens analysis of BBHs and NSBHs has been applied to tests of gravity. Since the sampling analysis for NSBHs includes two more parameters, m_{\max}^{NS} and m_{\min}^{NS} , the sampling data from BBHs and NSBHs cannot be combined simply. We first compute a kernel density estimation (KDE) function for the multi-dimensional joint posterior of the cosmological parameters with the sampling data. Then we compute the joint posterior for each gridded value of the cosmological parameter space with the KDE for BBHs and NSBHs, and multiply the two posteriors over the space. Finally we obtain the posterior of each parameter by marginalizing the joint multi-dimensional posterior over other parameters. The combined posteriors for H_0 , Ξ_0 and n are shown in figure 8. The estimated values of H_0 , Ξ_0 and n for the combined posterior are between the BBH posterior and the NSBH posterior, with a slightly smaller 1σ uncertainty than the individual BBH and NSBH posteriors, which is as expected. Again, we see the partial degeneracies between Ξ_0 and H_0 , and Ξ_0 and n are maintained.

The results for the other two modified gravity models introduced in this work for the combined posterior of BBHs and NSBHs are shown in figure 9 and 10. For the Horndeski class model, we apply a uniform prior for c_M in $[-4, 10]$. The lower prior bound is restricted by the maximum injected GW distance. The combined posterior gives $c_M = 1.5_{-2.1}^{+2.2}$, which is consistent with the GR prediction of $c_M = 0$ within a 1σ bound. Principally due to the

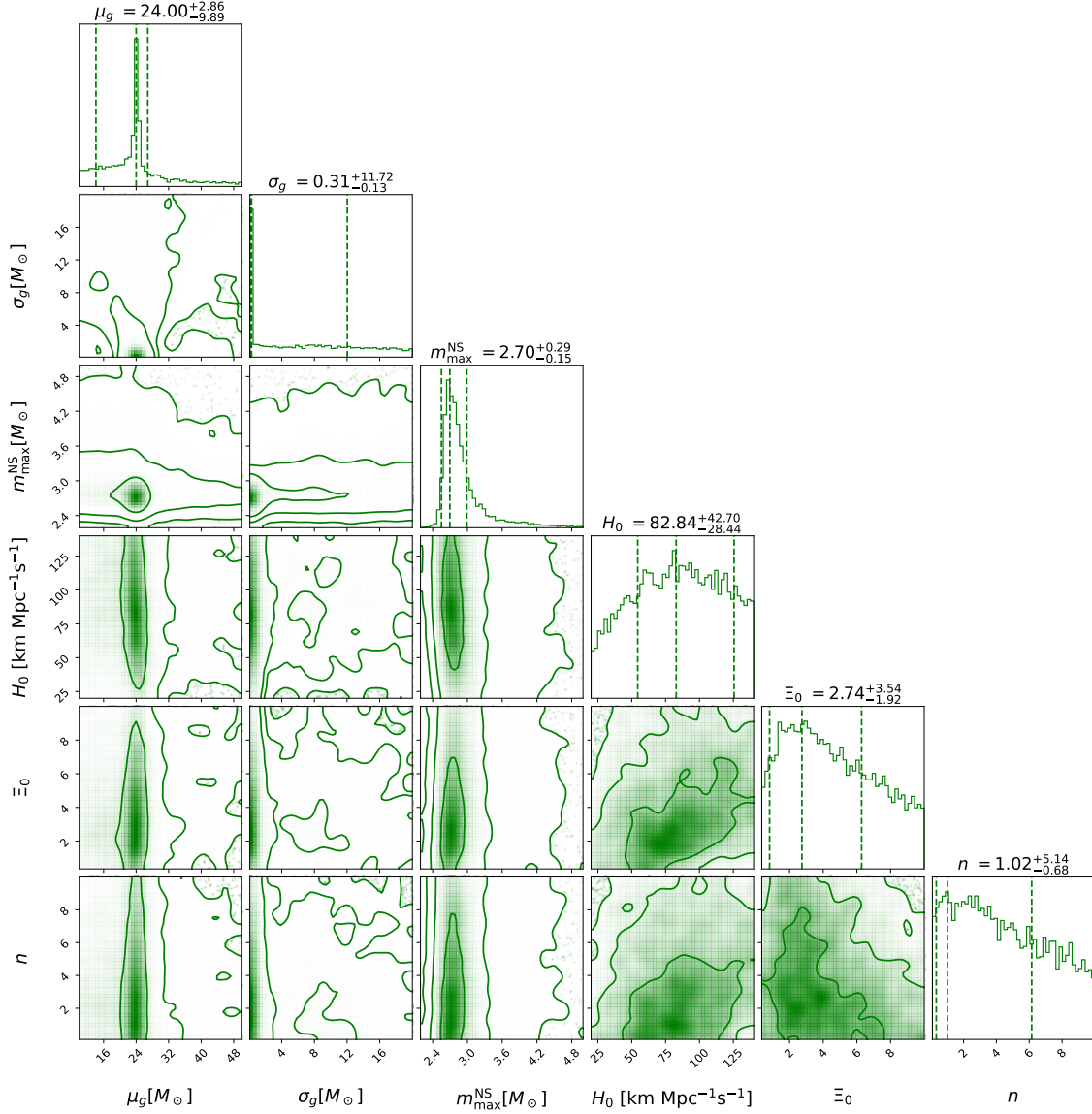


Figure 7. Selected corner plot of joint posteriors for 3 NSBH events from the GWTC-3 catalogue computed by the GWCOSMO sampling run. The contours show 1, 2 and 3 σ bound for the 2D posteriors. The vertical lines in the 1D histogram plot correspond to the peak values and 1 σ bound of the parameters in the title. A brief description of the mass and merger rate hyper-parameters is given in table 1.

higher redshifts of the dark siren events, this constraint is much narrower than that obtained by the bright siren GW170817 [54], and is competitive with that from spectral sirens [77]. We can also see slight degeneracy between H_0 and c_M in the contour.

On the other hand, for the extra-dimension model, we use a uniform prior for $D \in [3.7, 6]$, $\log(R_c/\text{Mpc}) \in [0.5, 6]$ and $n_D \in [0.3, 10]$. We obtain $D = 4.07^{+1.01}_{-0.23}$, which is consistent with 4-dimensional spacetime within 1 σ bound, but the distribution of $p(D)$ has a long tail at higher values. This constraint is slightly better than that obtained with the spectral siren method [78]. Meanwhile the constraint on R_c is $\log(R_c/\text{Mpc}) = 3.93^{+1.55}_{-0.55}$, corresponding

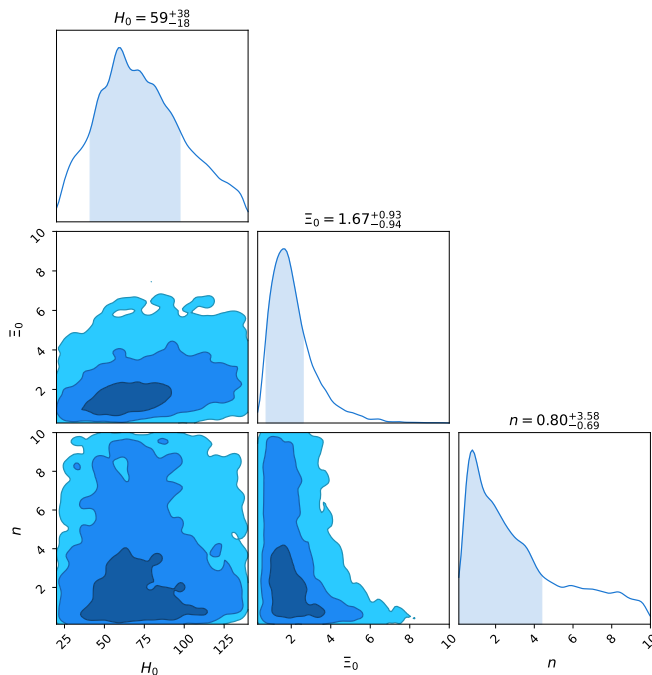


Figure 8. Posteriors of H_0 , Ξ_0 and n by combining results of BBHs and NSBHs from the GWCOSMO sampling run.

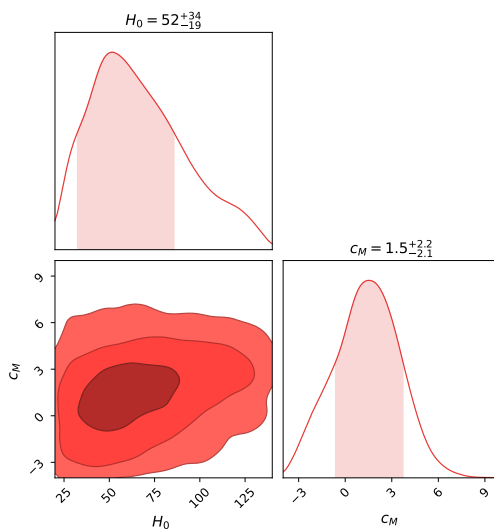


Figure 9. Posteriors of H_0 and c_M by combining results of BBHs and NSBHs from the GWCOSMO sampling run.

to $R_c \sim 10$ Gpc, which is around the same scale as the Hubble radius. The drop of the posterior of $\log(R_c/\text{Mpc})$ at the upper bound is artificial due to the over-smoothing of KDE; the distribution of samples in R_c is actually flat at the upper limit, as GR is recovered when R_c approaches infinity. In figure 10 we trim the range of R_c where the over-smoothing takes place. We have also included n_D in the joint posterior, but n_D is essentially unconstrained. Compared with the other two modified gravity models, there is little degeneracy among H_0 , D and $\log R_c$.

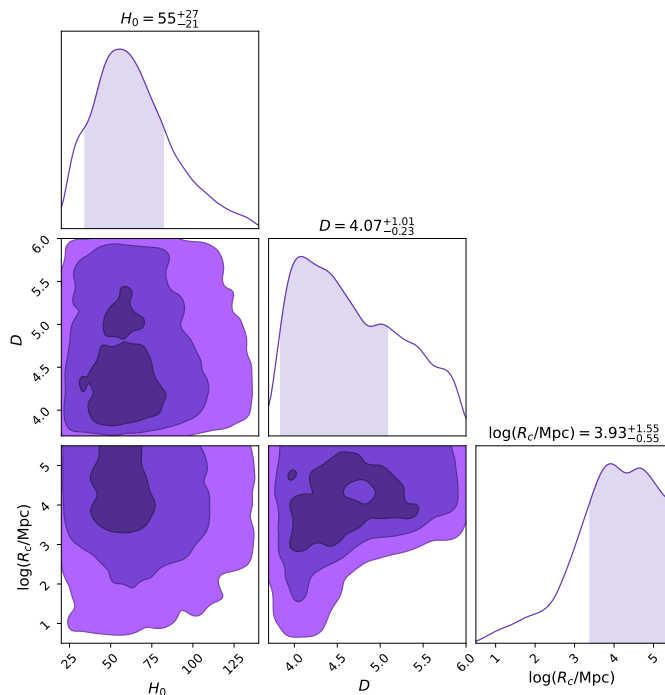


Figure 10. Posteriors of H_0 , D and $\log R_c$ by combining results of BBHs and NSBHs from the GWCOSMO sampling run.

6 Conclusions

The detection of gravitational waves offers us a powerful tool for testing theories of gravity beyond GR. The detection of bright siren GW170817 kicked off these investigations by providing a strong constraint on a particular aspect of modified GW propagation, namely the propagation speed. However, due to its relative proximity, the bounds obtained on other aspects of modified GW propagation remain relatively weak. We stress that an event ideal for constraining the Hubble parameter (e.g. one at low redshift) is not necessarily an ideal system for tests of gravity and fundamental physics. Therefore, high redshift events should not be always discarded in favour of apparently simpler analyses at low redshift.

Although more bright siren detections are expected going forward, their rate and redshift distribution remains currently rather unknown. The dark sirens technique, which statistically identifies host galaxies of GW events with galaxy catalogues, enables us to avoid pinning our hopes for GW cosmology and tests of fundamental physics on the uncertain bright siren landscape. Although dark sirens are individually less constraining than bright sirens as cosmological probes, their advantage lies in both their number and security (meaning that, they rely only on regular BBH and NSBH mergers, which we are confident will continue to be detected in increasing numbers). Any moderately well-localised GW event can be used for a dark sirens analysis, provided it meets a modest SNR threshold and has satisfactory parameter estimation results. As such, at least tens (or possibly ~ 100) of suitable events are expected during the recently-commenced O4 observation run. We can expect to see corresponding improvements in the parameter constraints presented here on the timescale of a year or two.

In this work we described the modification to the GWCOSMO software pipeline for constraining modified gravity parameters that affect GW luminosity distances, with both bright sirens and dark sirens. A crucial step forward is that these additional MG parameters can now be constrained concurrently with hyper-parameters describing the mass distribution and redshift evolution of compact objects, as well as the Hubble constant.

To validate our pipeline and explore the constraining power of future data sets, we first performed our tests of gravity on 250 mock bright sirens from the F2Y mock data catalogue. We computed the 1D posteriors on Ξ_0 , c_M and D in the three modified gravity parameterizations of section 2 respectively, recovering the expected consistency with GR. The joint posterior on H_0 and Ξ_0 with the F2Y mock bright sirens remained weak: it is difficult to constrain modified propagation effects with bright sirens located at low redshifts, where changes to the GW luminosity distances are small.

Turning our attention to the current data, we then computed the posterior probability on the modified gravity parameters in addition to mass distribution and rate evolution hyper-parameters with 46 dark sirens from the GWTC-3 catalogue. The constraint on Ξ_0 from the marginalized posterior with 42 BBHs is $1.29_{-0.67}^{+1.22}$, which is consistent with GR. We observed high degeneracy between Ξ_0 and γ (a parameter controlling the redshift evolution of the merger rate), which makes the choice of the prior bound of γ important for tests of gravity. This is a good example of how constraints on cosmology, astrophysics and fundamental physics are entangled, at least for the present. On the other hand, it is computationally expensive to generate the quantity of mock data for dark siren needed, and the current levels of uncertainty of the constraints are very large, so it is sufficient to have tested the method using one set of mock data, which is the bright siren case, and sanity checked the results against independent results of other groups using the same GW data from the realistic GWTC-3 catalogue for the dark siren case. We leave the generation of an accurate dark siren mock dataset and testing our method with it for future works.

Finally, we combined for the first time the results of 42 BBHs and 3 NSBHs, which yields $\Xi_0 = 1.67_{-0.94}^{+0.93}$. Furthermore, the combined posterior of BBHs and NSBHs gives the constraint of $c_M = 1.5_{-2.1}^{+2.2}$ in the Horndeski class model, and $D = 4.07_{-0.23}^{+1.01}$ in the extra-dimension model, all of which are consistent with GR.

At present, the effects of adding galaxy catalogue information are most pronounced for event GW190814, due to its good localisation and moderately low redshift; they are not highly significant for most other events. This is primarily due to the relatively low redshift range for which the GLADE+ catalogue employed our analysis has high completeness. Therefore, for most of the GW events to date, the bulk of their GW luminosity distance distribution lies beyond the catalogue. The good news is that other spectroscopic galaxy catalogues with higher completeness and redshift range are forthcoming, such as those from the DESI [75] and Euclid [76] experiments (see [79] for an initial dark siren analysis with the DESI catalogue). In some sense, this is the much better issue to have: we know that galaxy catalogue completeness will improve dramatically on a timescale of five years or less. We do not currently know what the corresponding improvements for bright siren counts will be. On longer timescales, the measurement of cosmological and modified gravity parameters with direct cross-correlation between GW events and galaxy surveys has been forecast in [80–84] (see also [85, 86]). This can serve as an independent probe in the 3G era.

A secondary effect in the current constraints is localization of GW events, which was particularly large for some events in the first two LVK observing runs. When ‘averaged’ over a large sky area we expect the galaxy clustering signal — which is key for (dis)favouring some parameter values — to be washed out. Given the incompleteness issue above, this is not a major stumbling block at present. With developments at the KAGRA and LIGO-India sites progressing, observations by three widely-separated detectors in future will lead to corresponding improvements in event localization, and should prevent this effect from ever becoming a limiting factor.

In this analysis we selected three specific parameterizations of modified GW luminosity distances to study. Whilst these functional forms are commonly used and motivated by general properties of current MG models, they are intended as representative examples only. The analysis pipelines developed in this work can be easily adapted to use alternative parameterizations or specific gravity models, if desired. On the other hand, the current version of GWCOSMO is unable to include the modified gravity effects in the expansion history of the universe, because it uses the pre-computed LOS redshift prior in computing the posterior. Although the dependence on H_0 and the modified gravity parameters that leads to friction in GW propagation is dropped out from the LOS redshift prior, the effects that changes the expansion history of the universe would change the LOS redshift prior, so it needs to be computed each time when varying the parameters in such models. Recomputing the LOS priors for each set of cosmological parameter values would be prohibitively expensive at present. In this work we focus on expanding the current version of GWCOSMO to the test of modified gravity without recomputing the LOS redshift prior. Moreover, in the perspective of theory motivation, although in principle the background of the universe is allowed to be different from the Λ CDM model, in practice other lines of evidence indicate that it is relatively close to the Λ CDM model. Hence we first study theories that stay close to Λ CDM expansion but modify luminosity distances. In future work we intend to broaden the range of options in GWCOSMO, including effects beyond modified GW luminosity distances.

As the fourth LVK observing run progresses, there may well be significant advances or new discoveries in some of the ingredients that make up the dark sirens recipe, e.g. the mass distribution of compact objects, its possible evolution with redshift, or the merger rates of different source types (as analysed with GWTC-3 [68]). With our current framework in hand, these are all straightforward to accommodate. Such improvements will help to pin down the astrophysical parts of the analysis, lifting degeneracies with the cosmological and beyond-GR parameters. We have set the stage for improved dark siren tests of gravity, and eagerly anticipate the forthcoming GW and galaxy data.

Acknowledgments

The authors would like to thank Benoit Revenu and Christos Karathanasis for assistance with GW injection computations. A.C. is supported by a PhD grant from the Chinese Scholarship Council (grant no. 202008060014), and the STSM GravNet grant during his visit to the NBI. The research of R.G. was supported by ERC Starting Grant *SHADE* 949572 and STFC grant ST/V005634/1. T.B. is supported by ERC Starting Grant *SHADE* (grant no. StG 949572) and a Royal Society University Research Fellowship (grant no. URF\R1\180009). This

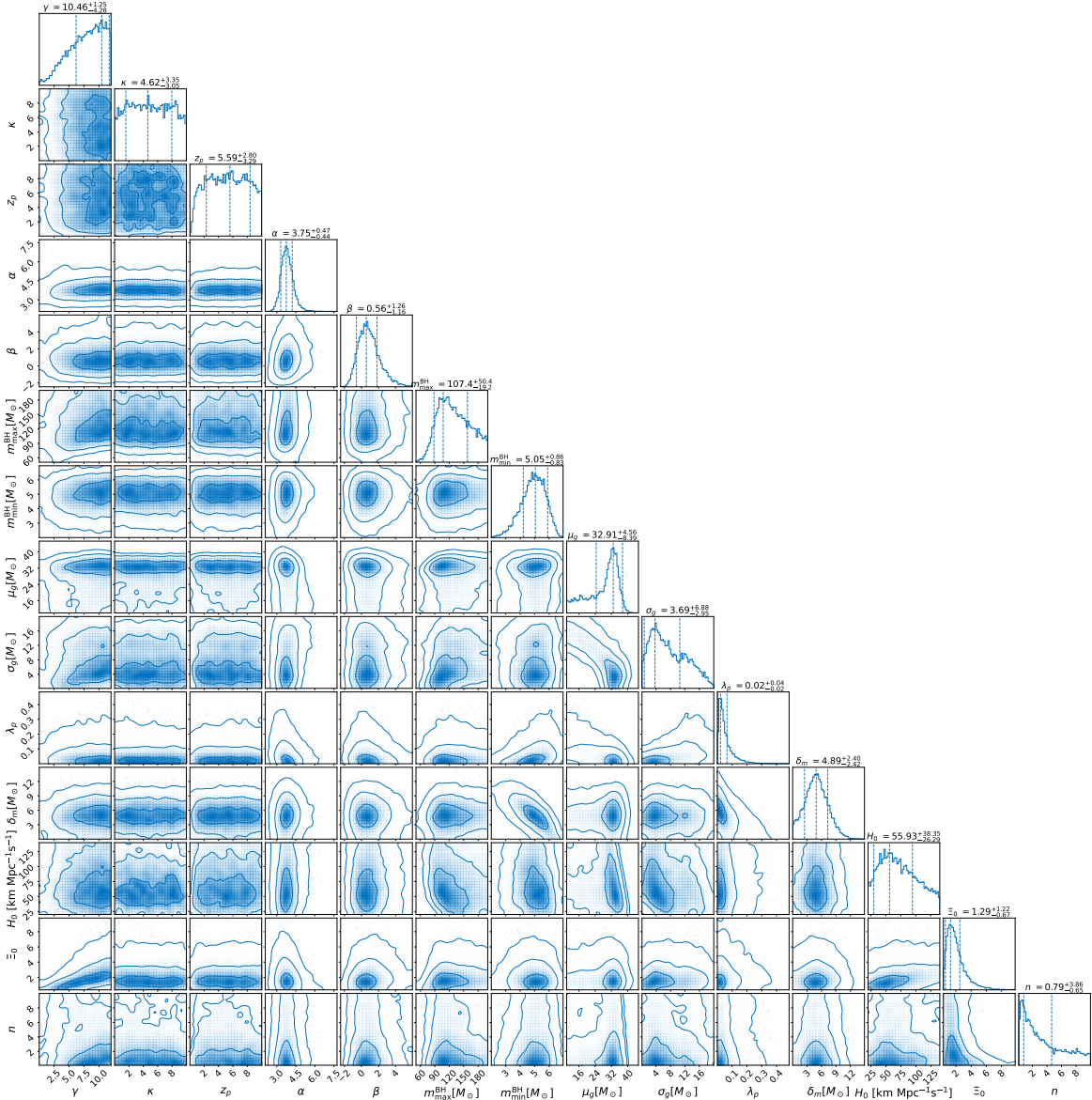


Figure 11. Full corner plot of joint posteriors for 42 BBH events from the GWTC-3 catalogue. The contours show 1, 2 and 3σ bound for the 2D posteriors. The vertical lines in the 1D histogram plot correspond to the peak values and 1σ bound of parameters in the title.

material is based upon work supported by NSF’s LIGO Laboratory, which is a major facility fully funded by the National Science Foundation. The authors are grateful for computational resources provided by the LIGO Laboratory and supported by National Science Foundation Grants PHY-0757058 and PHY-0823459.

A Corner plots for GWTC-3 reanalysis

We present the full corner plot for the joint posteriors obtained from the GWCOSMO sampling run with 42 BBHs in GWTC-3 in figure 11. γ , κ and z_p in the merger rate evolution model

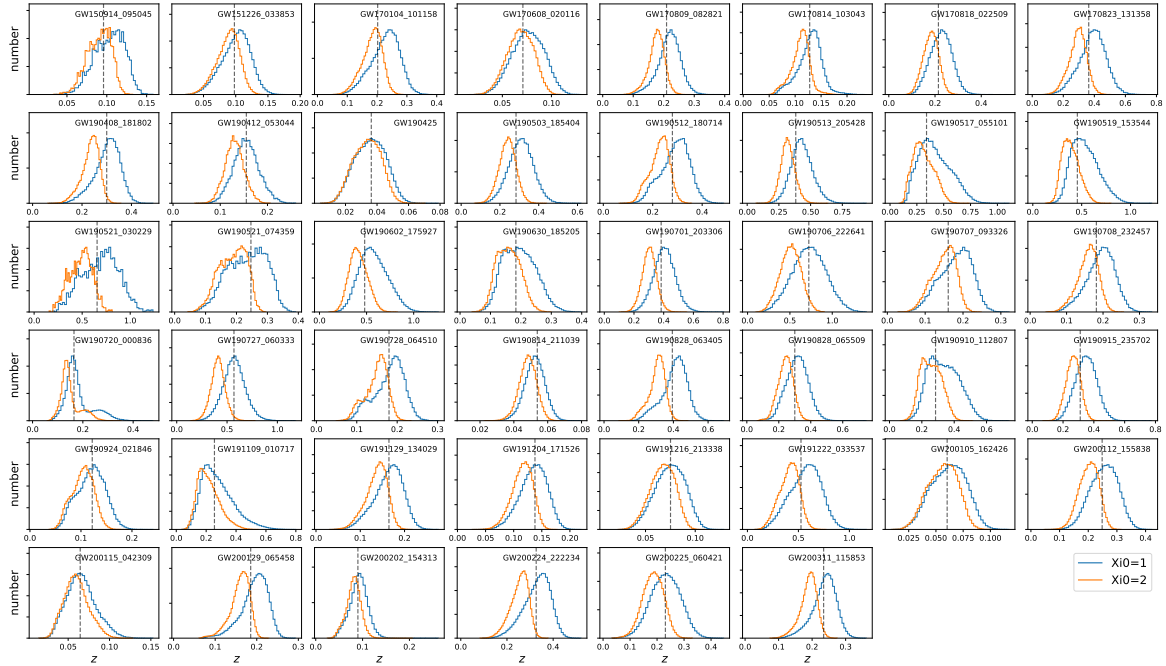


Figure 12. Posterior samples converted from GW luminosity distance to redshift for the 46 events from the GWTC-3 catalogue with different values of Ξ_0 . The grey dashed vertical line marks the estimated redshift in GR from GWTC-3.

are essentially unconstrained. On the other hand, parameters for BBH population model and cosmological parameters are constrained at a better level. There exists notable degeneracy between $H_0 - \Xi_0$, $\gamma - \Xi_0$, $\mu_g - \sigma_g$ and $m_{\min}^{\text{BH}} - \sigma_m$ (see discussion in the main text).

B Posterior samples for GWTC-3 reanalysis

Here we explain the three varieties of Ξ_0 posteriors observed from individual events in figure 4. In figure 12 we plot histograms of the redshift posterior samples for the 46 events selected from GWTC-2.1 and GWTC-3. The redshifts of posterior samples are converted from d_L^{GW} with two values, $\Xi_0 = 1$ (GR) and $\Xi_0 = 2$ with $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and $n = 1.91$. The grey dashed vertical line marks the estimated redshift in GR from GWTC-3. We can see that the posterior samples shift to lower redshift as Ξ_0 increases. For the events in which the peak of the samples is located at higher redshift than the estimated redshift, the peak crosses through to the other side of the estimated redshift when Ξ_0 increases. As a result, there is a peak in the likelihood of Ξ_0 over the prior range of Ξ_0 , see for example GW170809_082821 in figure 4. For the events in which the peaks of the samples for different Ξ_0 values are on the same side of the estimated redshift, the likelihood of Ξ_0 continuously decreases instead of having a peak, for example GW190910_112807 in figure 4. For the events in which the posterior samples shift little when Ξ_0 increases, the likelihood is dominated by the denominator that is continuously decreasing, so the likelihood is continuously increasing, for example GW191216_213338 in figure 4.

References

- [1] LIGO SCIENTIFIC and VIRGO collaborations, *Tests of general relativity with binary black holes from the second LIGO-Virgo gravitational-wave transient catalog*, *Phys. Rev. D* **103** (2021) 122002 [[arXiv:2010.14529](#)] [[INSPIRE](#)].
- [2] LIGO SCIENTIFIC, VIRGO and KAGRA collaborations, *Tests of General Relativity with GWTC-3*, [arXiv:2112.06861](#) [[INSPIRE](#)].
- [3] L. Verde, T. Treu and A.G. Riess, *Tensions between the Early and the Late Universe*, *Nat. Astron.* **3** (2019) 891 [[arXiv:1907.10625](#)] [[INSPIRE](#)].
- [4] E. Di Valentino et al., *Snowmass2021 — Letter of interest cosmology intertwined. Part II. The Hubble constant tension*, *Astropart. Phys.* **131** (2021) 102605 [[arXiv:2008.11284](#)] [[INSPIRE](#)].
- [5] B.F. Schutz, *Determining the Hubble Constant from Gravitational Wave Observations*, *Nature* **323** (1986) 310 [[INSPIRE](#)].
- [6] LIGO SCIENTIFIC and VIRGO collaborations, *GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral*, *Phys. Rev. Lett.* **119** (2017) 161101 [[arXiv:1710.05832](#)] [[INSPIRE](#)].
- [7] B.P. Abbott et al., *Gravitational Waves and Gamma-rays from a Binary Neutron Star Merger: GW170817 and GRB 170817A*, *Astrophys. J. Lett.* **848** (2017) L13 [[arXiv:1710.05834](#)] [[INSPIRE](#)].
- [8] LIGO SCIENTIFIC et al. collaborations, *Multi-messenger Observations of a Binary Neutron Star Merger*, *Astrophys. J. Lett.* **848** (2017) L12 [[arXiv:1710.05833](#)] [[INSPIRE](#)].
- [9] LIGO SCIENTIFIC and VIRGO collaborations, *A gravitational-wave standard siren measurement of the Hubble constant*, *Nature* **551** (2017) 85 [[arXiv:1710.05835](#)] [[INSPIRE](#)].
- [10] C. de Rham and S. Melville, *Gravitational Rainbows: LIGO and Dark Energy at its Cutoff*, *Phys. Rev. Lett.* **121** (2018) 221101 [[arXiv:1806.09417](#)] [[INSPIRE](#)].
- [11] LISA COSMOLOGY Working Group, *Measuring the propagation speed of gravitational waves with LISA*, *JCAP* **08** (2022) 031 [[arXiv:2203.00566](#)] [[INSPIRE](#)].
- [12] T. Baker, E. Barausse, A. Chen, C. de Rham, M. Pieroni and G. Tasinato, *Testing gravitational wave propagation with multiband detections*, *JCAP* **03** (2023) 044 [[arXiv:2209.14398](#)] [[INSPIRE](#)].
- [13] I. Harry and J. Noller, *Probing the speed of gravity with LVK, LISA, and joint observations*, *Gen. Rel. Grav.* **54** (2022) 133 [[arXiv:2207.10096](#)] [[INSPIRE](#)].
- [14] T. Baker, E. Bellini, P.G. Ferreira, M. Lagos, J. Noller and I. Sawicki, *Strong constraints on cosmological gravity from GW170817 and GRB 170817A*, *Phys. Rev. Lett.* **119** (2017) 251301 [[arXiv:1710.06394](#)] [[INSPIRE](#)].
- [15] J.M. Ezquiaga and M. Zumalacárregui, *Dark Energy After GW170817: Dead Ends and the Road Ahead*, *Phys. Rev. Lett.* **119** (2017) 251304 [[arXiv:1710.05901](#)] [[INSPIRE](#)].
- [16] P. Creminelli and F. Vernizzi, *Dark Energy after GW170817 and GRB170817A*, *Phys. Rev. Lett.* **119** (2017) 251302 [[arXiv:1710.05877](#)] [[INSPIRE](#)].
- [17] S. Jana, G.K. Chakravarty and S. Mohanty, *Constraints on Born-Infeld gravity from the speed of gravitational waves after GW170817 and GRB 170817A*, *Phys. Rev. D* **97** (2018) 084011 [[arXiv:1711.04137](#)] [[INSPIRE](#)].
- [18] A. Nishizawa and T. Kobayashi, *Parity-violating gravity and GW170817*, *Phys. Rev. D* **98** (2018) 124018 [[arXiv:1809.00815](#)] [[INSPIRE](#)].

- [19] R. Gray, C. Messenger and J. Veitch, *A pixelated approach to galaxy catalogue incompleteness: improving the dark siren measurement of the Hubble constant*, *Mon. Not. Roy. Astron. Soc.* **512** (2022) 1127 [[arXiv:2111.04629](#)] [[INSPIRE](#)].
- [20] W. Del Pozzo, *Inference of the cosmological parameters from gravitational waves: application to second generation interferometers*, *Phys. Rev. D* **86** (2012) 043011 [[arXiv:1108.1317](#)] [[INSPIRE](#)].
- [21] H.-Y. Chen, M. Fishbach and D.E. Holz, *A two per cent Hubble constant measurement from standard sirens within five years*, *Nature* **562** (2018) 545 [[arXiv:1712.06531](#)] [[INSPIRE](#)].
- [22] M. Fishbach et al., *A Standard Siren Measurement of the Hubble Constant from GW170817 without the Electromagnetic Counterpart*, *Astrophys. J. Lett.* **871** (2019) L13 [[arXiv:1807.05667](#)] [[INSPIRE](#)].
- [23] R. Gray et al., *Cosmological inference using gravitational wave standard sirens: A mock data analysis*, *Phys. Rev. D* **101** (2020) 122001 [[arXiv:1908.06050](#)] [[INSPIRE](#)].
- [24] K. Leyde, S. Mastrogiovanni, D. Steer, E. Chassande-Mottin and C. Karathanasis, *Current and future constraints on cosmology and modified gravitational wave friction from binary black holes*, *JCAP* **09** (2022) 012 [[arXiv:2202.00025](#)] [[INSPIRE](#)].
- [25] A. Finke, S. Foffa, F. Iacovelli, M. Maggiore and M. Mancarella, *Cosmology with LIGO/Virgo dark sirens: Hubble parameter and modified gravitational wave propagation*, *JCAP* **08** (2021) 026 [[arXiv:2101.12660](#)] [[INSPIRE](#)].
- [26] R. Gray et al., *Joint cosmological and gravitational-wave population inference using dark sirens and galaxy catalogues*, *JCAP* **12** (2023) 023 [[arXiv:2308.02281](#)] [[INSPIRE](#)].
- [27] S. Mastrogiovanni et al., *Joint population and cosmological properties inference with gravitational waves standard sirens and galaxy surveys*, *Phys. Rev. D* **108** (2023) 042002 [[arXiv:2305.10488](#)] [[INSPIRE](#)].
- [28] S. Mastrogiovanni et al., *ICAROGW: A python package for inference of astrophysical population properties of noisy, heterogeneous and incomplete observations*, [arXiv:2305.17973](#) [[INSPIRE](#)].
- [29] S. Mastrogiovanni et al., *On the importance of source population models for gravitational-wave cosmology*, *Phys. Rev. D* **104** (2021) 062009 [[arXiv:2103.14663](#)] [[INSPIRE](#)].
- [30] LIGO SCIENTIFIC, VIRGO and KAGRA collaborations, *GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo during the Second Part of the Third Observing Run*, *Phys. Rev. X* **13** (2023) 041039 [[arXiv:2111.03606](#)] [[INSPIRE](#)].
- [31] LIGO SCIENTIFIC et al. collaborations, *Constraints on the Cosmic Expansion History from GWTC-3*, *Astrophys. J.* **949** (2023) 76 [[arXiv:2111.03604](#)] [[INSPIRE](#)].
- [32] S. Mukherjee, A. Krolewski, B.D. Wandelt and J. Silk, *Cross-correlating dark sirens and galaxies: measurement of H_0 from GWTC-3 of LIGO-Virgo-KAGRA*, [arXiv:2203.03643](#) [[INSPIRE](#)].
- [33] S. Bera, D. Rana, S. More and S. Bose, *Incompleteness Matters Not: Inference of H_0 from Binary Black Hole-Galaxy Cross-correlations*, *Astrophys. J.* **902** (2020) 79 [[arXiv:2007.04271](#)] [[INSPIRE](#)].
- [34] M. Oguri, *Measuring the distance-redshift relation with the cross-correlation of gravitational wave standard sirens and galaxies*, *Phys. Rev. D* **93** (2016) 083511 [[arXiv:1603.02356](#)] [[INSPIRE](#)].
- [35] E. Belgacem, Y. Dirian, S. Foffa and M. Maggiore, *Gravitational-wave luminosity distance in modified gravity theories*, *Phys. Rev. D* **97** (2018) 104066 [[arXiv:1712.08108](#)] [[INSPIRE](#)].
- [36] E. Belgacem, Y. Dirian, S. Foffa and M. Maggiore, *Modified gravitational-wave propagation and standard sirens*, *Phys. Rev. D* **98** (2018) 023510 [[arXiv:1805.08731](#)] [[INSPIRE](#)].

- [37] A. Nishizawa, *Generalized framework for testing gravity with gravitational-wave propagation. Part I. Formulation*, *Phys. Rev. D* **97** (2018) 104037 [[arXiv:1710.04825](#)] [[INSPIRE](#)].
- [38] E. Belgacem et al., *Testing modified gravity at cosmological distances with LISA standard sirens*, *JCAP* **07** (2019) 024 [[arXiv:1906.01593](#)] [[INSPIRE](#)].
- [39] T. Baker and I. Harrison, *Constraining Scalar-Tensor Modified Gravity with Gravitational Waves and Large Scale Structure Surveys*, *JCAP* **01** (2021) 068 [[arXiv:2007.13791](#)] [[INSPIRE](#)].
- [40] M. Maggiore, *Gravitational Waves. Volume 1: Theory and Experiments*, Oxford University Press (2007) [[DOI:10.1093/acprof:oso/9780198570745.001.0001](#)] [[INSPIRE](#)].
- [41] A.E. Romano and M. Sakellariadou, *Mirage of Luminal Modified Gravitational-Wave Propagation*, *Phys. Rev. Lett.* **130** (2023) 231401 [[arXiv:2302.05413](#)] [[INSPIRE](#)].
- [42] E. Belgacem, Y. Dirian, S. Foffa and M. Maggiore, *Nonlocal gravity. Conceptual aspects and cosmological predictions*, *JCAP* **03** (2018) 002 [[arXiv:1712.07066](#)] [[INSPIRE](#)].
- [43] M. Mancarella, E. Genoud-Prachex and M. Maggiore, *Cosmology and modified gravitational wave propagation from binary black hole population models*, *Phys. Rev. D* **105** (2022) 064030 [[arXiv:2112.05728](#)] [[INSPIRE](#)].
- [44] M. Mancarella et al., *Gravitational-wave cosmology with dark sirens: state of the art and perspectives for 3G detectors*, *PoS ICHEP2022* (2022) 127 [[arXiv:2211.15512](#)] [[INSPIRE](#)].
- [45] A. Finke, S. Foffa, F. Iacovelli, M. Maggiore and M. Mancarella, *Modified gravitational wave propagation and the binary neutron star mass function*, *Phys. Dark Univ.* **36** (2022) 100994 [[arXiv:2108.04065](#)] [[INSPIRE](#)].
- [46] A. Finke, S. Foffa, F. Iacovelli, M. Maggiore and M. Mancarella, *Probing modified gravitational wave propagation with strongly lensed coalescing binaries*, *Phys. Rev. D* **104** (2021) 084057 [[arXiv:2107.05046](#)] [[INSPIRE](#)].
- [47] G.W. Horndeski, *Second-order scalar-tensor field equations in a four-dimensional space*, *Int. J. Theor. Phys.* **10** (1974) 363 [[INSPIRE](#)].
- [48] C. Deffayet, X. Gao, D.A. Steer and G. Zahariade, *From k-essence to generalised Galileons*, *Phys. Rev. D* **84** (2011) 064039 [[arXiv:1103.3260](#)] [[INSPIRE](#)].
- [49] J. Sakstein and B. Jain, *Implications of the Neutron Star Merger GW170817 for Cosmological Scalar-Tensor Theories*, *Phys. Rev. Lett.* **119** (2017) 251303 [[arXiv:1710.05893](#)] [[INSPIRE](#)].
- [50] E. Bellini and I. Sawicki, *Maximal freedom at minimum cost: linear large-scale structure in general modifications of gravity*, *JCAP* **07** (2014) 050 [[arXiv:1404.3713](#)] [[INSPIRE](#)].
- [51] M. Lagos, E. Bellini, J. Noller, P.G. Ferreira and T. Baker, *A general theory of linear cosmological perturbations: stability conditions, the quasistatic limit and dynamics*, *JCAP* **03** (2018) 021 [[arXiv:1711.09893](#)] [[INSPIRE](#)].
- [52] E. Bellini, A.J. Cuesta, R. Jimenez and L. Verde, *Constraints on deviations from Λ CDM within Horndeski gravity*, *JCAP* **02** (2016) 053 [Erratum *ibid.* **06** (2016) E01] [[arXiv:1509.07816](#)] [[INSPIRE](#)].
- [53] D. Alonso, E. Bellini, P.G. Ferreira and M. Zumalacárregui, *Observational future of cosmological scalar-tensor theories*, *Phys. Rev. D* **95** (2017) 063502 [[arXiv:1610.09290](#)] [[INSPIRE](#)].
- [54] M. Lagos, M. Fishbach, P. Landry and D.E. Holz, *Standard sirens with a running Planck mass*, *Phys. Rev. D* **99** (2019) 083504 [[arXiv:1901.03321](#)] [[INSPIRE](#)].

- [55] S. Mastrogiovanni, D. Steer and M. Barsuglia, *Probing modified gravity theories and cosmology using gravitational-waves and associated electromagnetic counterparts*, *Phys. Rev. D* **102** (2020) 044009 [[arXiv:2004.01632](#)] [[INSPIRE](#)].
- [56] E.V. Linder, *Challenges in connecting modified gravity theory and observations*, *Phys. Rev. D* **95** (2017) 023518 [[arXiv:1607.03113](#)] [[INSPIRE](#)].
- [57] R. D’Agostino and R.C. Nunes, *Probing observational bounds on scalar-tensor theories from standard sirens*, *Phys. Rev. D* **100** (2019) 044041 [[arXiv:1907.05516](#)] [[INSPIRE](#)].
- [58] G.R. Dvali, G. Gabadadze and M. Porrati, *4D gravity on a brane in 5D Minkowski space*, *Phys. Lett. B* **485** (2000) 208 [[hep-th/0005016](#)] [[INSPIRE](#)].
- [59] A. Padilla, *Ghost free brane world bigravity*, *Class. Quant. Grav.* **21** (2004) 2899 [[hep-th/0402079](#)] [[INSPIRE](#)].
- [60] M. Corman, C. Escamilla-Rivera and M.A. Hendry, *Constraining extra dimensions on cosmological scales with LISA future gravitational wave siren data*, *JCAP* **02** (2021) 005 [[arXiv:2004.04009](#)] [[INSPIRE](#)].
- [61] 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 [[arXiv:1809.02063](#)] [[INSPIRE](#)].
- [62] 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*, in *Handbook of Gravitational Wave Astronomy*, C. Bambi, S. Katsanevas and K.D. Kokkotas eds., Springer, Singapore (2020), pp. 1–60 [[arXiv:2007.05579](#)] [[DOI:10.1007/978-981-15-4702-7_45-1](#)] [[INSPIRE](#)].
- [63] M. Fishbach, D.E. Holz and W.M. Farr, *Does the Black Hole Merger Rate Evolve with Redshift?*, *Astrophys. J. Lett.* **863** (2018) L41 [[arXiv:1805.10270](#)] [[INSPIRE](#)].
- [64] A. Zonca et al., *healpy: equal area pixelization and spherical harmonics transforms for data on the sphere in Python*, *J. Open Source Softw.* **4** (2019) 1298 [[INSPIRE](#)].
- [65] K.M. Górski et al., *HEALPix — A Framework for high resolution discretization, and fast analysis of data distributed on the sphere*, *Astrophys. J.* **622** (2005) 759 [[astro-ph/0409513](#)] [[INSPIRE](#)].
- [66] G. D’Álya et al., *GLADE+ : an extended galaxy catalogue for multimessenger searches with advanced gravitational-wave detectors*, *Mon. Not. Roy. Astron. Soc.* **514** (2022) 1403 [[arXiv:2110.06184](#)] [[INSPIRE](#)].
- [67] R. Abbott et al., *Population Properties of Compact Objects from the Second LIGO-Virgo Gravitational-Wave Transient Catalog*, *Astrophys. J. Lett.* **913** (2021) L7 [[arXiv:2010.14533](#)] [[INSPIRE](#)].
- [68] KAGRA, VIRGO and LIGO SCIENTIFIC collaborations, *Population of Merging Compact Binaries Inferred Using Gravitational Waves through GWTC-3*, *Phys. Rev. X* **13** (2023) 011048 [[arXiv:2111.03634](#)] [[INSPIRE](#)].
- [69] P. Madau and M. Dickinson, *Cosmic Star Formation History*, *Ann. Rev. Astron. Astrophys.* **52** (2014) 415 [[arXiv:1403.0007](#)] [[INSPIRE](#)].
- [70] L.P. Singer et al., *The First Two Years of Electromagnetic Follow-Up with Advanced LIGO and Virgo*, *Astrophys. J.* **795** (2014) 105 [[arXiv:1404.5623](#)] [[INSPIRE](#)].
- [71] L. Barsotti and P. Fritschel, *Early aligo configurations: example scenarios toward design sensitivity*, LIGO-T1200307-v4 (2012).

- [72] KAGRA, LIGO SCIENTIFIC, VIRGO and VIRGO collaborations, *Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA*, *Living Rev. Rel.* **21** (2018) 3 [[arXiv:1304.0670](#)] [[INSPIRE](#)].
- [73] PLANCK collaboration, *Planck 2015 results. Part XIV. Dark energy and modified gravity*, *Astron. Astrophys.* **594** (2016) A14 [[arXiv:1502.01590](#)] [[INSPIRE](#)].
- [74] LIGO SCIENTIFIC and VIRGO collaborations, *GWTC-2.1: Deep extended catalog of compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run*, *Phys. Rev. D* **109** (2024) 022001 [[arXiv:2108.01045](#)] [[INSPIRE](#)].
- [75] A. Dey et al., *Overview of the DESI Legacy Imaging Surveys*, *Astron. J.* **157** (2019) 168 [[arXiv:1804.08657](#)] [[INSPIRE](#)].
- [76] G.D. Racca et al., *The Euclid mission design*, in *Space Telescopes and Instrumentation 2016: Optical, Infrared, and Millimeter Wave*, proceedings of the *SPIE Astronomical Telescopes + Instrumentation*, Edinburgh, U.K., 26 June–1 July 2016, H.A. MacEwen, G.G. Fazio, M. Lystrup, N. Batalha, N. Siegler and E.C. Tong eds., SPIE (2016) [*Proc. SPIE* **9904** (2016) 99040O] [[arXiv:1610.05508](#)] [[INSPIRE](#)].
- [77] J.M. Ezquiaga, *Hearing gravity from the cosmos: GWTC-2 probes general relativity at cosmological scales*, *Phys. Lett. B* **822** (2021) 136665 [[arXiv:2104.05139](#)] [[INSPIRE](#)].
- [78] I. Magana Hernandez, *Constraining the number of spacetime dimensions from GWTC-3 binary black hole mergers*, *Phys. Rev. D* **107** (2023) 084033 [[arXiv:2112.07650](#)] [[INSPIRE](#)].
- [79] A. Palmese, C.R. Bom, S. Mucesh and W.G. Hartley, *A Standard Siren Measurement of the Hubble Constant Using Gravitational-wave Events from the First Three LIGO/Virgo Observing Runs and the DESI Legacy Survey*, *Astrophys. J.* **943** (2023) 56 [[arXiv:2111.06445](#)] [[INSPIRE](#)].
- [80] S. Mukherjee and B.D. Wandelt, *Beyond the classical distance-redshift test: cross-correlating redshift-free standard candles and sirens with redshift surveys*, [arXiv:1808.06615](#) [[INSPIRE](#)].
- [81] S. Mukherjee, B.D. Wandelt and J. Silk, *Probing the theory of gravity with gravitational lensing of gravitational waves and galaxy surveys*, *Mon. Not. Roy. Astron. Soc.* **494** (2020) 1956 [[arXiv:1908.08951](#)] [[INSPIRE](#)].
- [82] S. Mukherjee, B.D. Wandelt and J. Silk, *Testing the general theory of relativity using gravitational wave propagation from dark standard sirens*, *Mon. Not. Roy. Astron. Soc.* **502** (2021) 1136 [[arXiv:2012.15316](#)] [[INSPIRE](#)].
- [83] S. Mukherjee, B.D. Wandelt, S.M. Nissanke and A. Silvestri, *Accurate precision Cosmology with redshift unknown gravitational wave sources*, *Phys. Rev. D* **103** (2021) 043520 [[arXiv:2007.02943](#)] [[INSPIRE](#)].
- [84] C.C. Diaz and S. Mukherjee, *Mapping the cosmic expansion history from LIGO-Virgo-KAGRA in synergy with DESI and SPHEREx*, *Mon. Not. Roy. Astron. Soc.* **511** (2022) 2782 [[arXiv:2107.12787](#)] [[INSPIRE](#)].
- [85] G. Cañas-Herrera, O. Contigiani and V. Vardanyan, *Learning How to Surf: Reconstructing the Propagation and Origin of Gravitational Waves with Gaussian Processes*, *Astrophys. J.* **918** (2021) 20 [[arXiv:2105.04262](#)] [[INSPIRE](#)].
- [86] J. Fonseca, S. Zazzera, T. Baker and C. Clarkson, *The observed number counts in luminosity distance space*, *JCAP* **08** (2023) 050 [[arXiv:2304.14253](#)] [[INSPIRE](#)].