Mapping the inhomogeneous Universe with Standard Sirens: Degeneracy between inhomogeneity and modified gravity theories

Marios Kalomenopoulos, Sadegh Khochfar, Jonathan Gair, Shun Arai

1 Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh EH9 3HJ, UK
2 Max Planck Institute for Gravitational Physics, Potsdam Science Park, Am Muhlenberg 1, D-14476 Potsdam, Germany
3 Center for Gravitational Physics, Yukawa Institute for Theoretical Physics, Kyoto University, Kyoto 606-8502, Japan

ABSTRACT

The detection of gravitational waves (GWs) and an accompanying electromagnetic (E/M) counterpart have been suggested as a future probe for cosmology and theories of gravity. In this paper, we present calculations of the luminosity distance of sources taking into account inhomogeneities in the matter distribution that are predicted in numerical simulations of structure formation. We assume mock GWs sources, with known redshift, based on binary population synthesis models, between redshifts \( z = 0 \) and \( z = 5 \). We show that present systematic limits of observations allow for a wide range of effective inhomogeneous models to be consistent with a homogeneous and isotropic Friedman-Lemaître-Robertson-Walker background model. Increasing the number of standard sirens from 150 to 350 and up to \( z = 5 \) helps shrink the confidence contours (68% C.L.) by \( \sim 35\% \), but still leaving the inhomogeneity parameters loosely constrained. In addition, we show that inhomogeneities resulting from clustering of matter can mimic certain classes of modified gravity theories, or other effects that dampen GW amplitudes, and deviations larger than \( \delta \nu \sim O(0.1) \) (99% C.L.) to the extra friction term \( \nu \), from zero, would be necessary to distinguish them. This limit is determined by observational errors (signal-to-noise ratio) and to improve on this constraint by an order of magnitude, one would need to reduce systematic observational errors to about one-fifth of their current values.

Key words: methods: theory – cosmology: gravitational waves - inhomogeneous universe - modified gravity

1 INTRODUCTION

The standard model of cosmology Lambda-Cold-Dark-Matter (ΛCDM) is based on the assumption that the universe is homogeneous and isotropic, which is supported on large-scales by precise CMB measurements (Akrami & et al 2018). These symmetry hypotheses lead to the well-known Friedmann-Lemaître-Robertson-Walker (FLRW) metric to describe the Universe's geometry. Although the standard model has passed successfully many tests, various “tensions” exist (Verde et al. 2019), so independent confirmations of its basic assumptions are important.

Tests checking the homogeneity and isotropy of the Universe, based on electromagnetic (E/M) observations, have been proposed and performed in the literature (for example Clifton et al. (2008); Busti et al. (2012); Dhawan et al. (2018)), showing consistency of the standard model, but require more data (so far simple inhomogeneous models are consistent with homogeneity at \( < 2 \sigma \) level). At the same time, the robustness of some basic ΛCDM assumptions, like the isotropy or the spatial curvature of the Universe, have recently been under debate (see Nielsen et al. (2016); Rubin & Hayden (2016); Colin et al. (2019); Rubin & Heitlauf (2019); Di Valentino et al. (2020); Efstathiou & Gratton (2020); Heinesen & Buchert (2020); Migkas et al. (2020)). Most of the above analyses require good, model-independent distance measurements, which in general, are difficult to obtain.

The first direct detection of GWs with an E/M counterpart (“standard siren”) (Abbott et al. 2017c) presents an

1 Historically, “Standard Sirens” refer to every compact binary - BBH, BNS, BH-NS - with or without an E/M counterpart (Schutz 1986; Holz & Hughes 2005). Here we are going to use the term

Historically, “Standard Sirens” refer to every compact binary - BBH, BNS, BH-NS - with or without an E/M counterpart (Schutz 1986; Holz & Hughes 2005). Here we are going to use the term
alternative way to study fundamental physics and provide an independent probe for assessing our basic assumptions for the Universe. The simultaneous determination of the luminosity distance, based solely on general relativity (GR), and the localisation of the source from the E/M observation, which was demonstrated for example with the first observation of a binary Neutron Star (Abbott et al. 2017b), has opened new possibilities to test our cosmological theories, in a model-independent way, by exploiting the observed distances.

There are already various proposals of how the luminosity distances inferred from GWs standard sirens observations can be used for cosmology. Examples include the determination of the Hubble parameter (Schutz 1986; Abbott et al. 2017d). Moreover, (Seto et al. 2001) and (Yagi et al. 2012) suggest a direct measurement of cosmic acceleration with GW sirens by observing a phase shift in the signal (a deviation from the value that is expected in an FLRW spacetime, can be a sign for deviations from the Cosmological Principle’s homogeneity). Further applications of standard sirens include constraints on cosmic anisotropies, e.g. the dipole anisotropy (Cai et al. 2018; Lin et al. 2018), and the spatial curvature, e.g. (Wei 2018).

At the same time, various classes of modified theories of gravity predict that distances inferred from GWs deviate from those of their E/M counterpart. These theories include a number of new degrees of freedom to dynamically describe the late-time acceleration of the Universe (e.g. Joyce et al. 2016; Ferreira 2019). The additions to GR can introduce several new effects on the propagation of GWs, like a different propagation speed or an amplitude decay, compared to photons (Saltas et al. 2014, 2018; Nishizawa 2018; Amendola et al. 2018; María Ezquiaga & Zumalacárregui 2018). However, so far tests with current GWs observations (Abbott et al. 2016, 2019a,c), that investigate a massive graviton and phenomenological models of extra-dimensions, show consistency with GR.

As laid out above, several different physical processes can lead to a reduction in the amplitude of GWs as they propagate through the Universe. In this paper, we investigate the measurement of luminosity distances of GWs propagating through inhomogeneities and under the effects of modified gravity theories. To model the effects of inhomogeneities, different approaches have been taken. In the literature (see e.g. Fleury 2015; Helbig 2019, for an overview), but here we will concentrate on the Dyer-Roeder (DR) relation (Dyer & Roeder 1972) and a modification of it, the modified DR relation (mDR) (Clarkson et al. 2012). In a previous study (Yoo et al. 2007) investigated the impact of inhomogeneities on GWs, concentrating only on an extreme inhomogeneous (empty) DR universe, with a uniform distribution of same mass objects. The focus of that study was, however, on the information that could be obtained about the lenses, using strong lensing of GW events. In this paper we investigate the possible constraints on arbitrary inhomogeneous models and on inhomogeneous models derived from cosmological N-body simulations. We also examine possible degeneracies of these effects with modified gravity theories.

The paper is structured as follows: In Section 2 we study the effects on the GW luminosity distance produced by modified gravity theories and an inhomogeneous background and present the various models we are interested in constraining, in Section 3 we describe the cosmological simulations and the numerical techniques employed, in Section 4 we discuss our results and in Section 5 we summarise our findings and discuss possible future directions.

## 2 DISTANCES FROM GWS

A GW detection with an E/M counterpart leads to two observable quantities for cosmology: the distance to the source $d_{\text{gw}}^L$ and its redshift $z$. The dominant, quadrupole contribution (Maggiore 2008) gives:

$$h = \frac{4}{d_{\text{gw}}^L} \left( \frac{GM_c}{c^2} \right)^{5/3} \left( \frac{\pi f_{\text{gw}}}{c} \right)^{2/3},$$

where $h$ is the amplitude of the GW, $c$ is the speed of light, $G$ Newton’s constant, $f_{\text{gw}}(t)$ the frequency of the GW at the observer, $d_{\text{gw}}^L$ is the luminosity distance and $M_c = (1+z)M_c$ is the “redshifted chirp mass” and $M_c = (m_1m_2)^{3/5}/(m_1+m_2)^{1/5}$ is the chirp mass, with $m_1$ and $m_2$ the individual masses of the compact objects. For binary black hole systems, the information of redshift is completely degenerate with the chirp mass, and thus a cosmological model, which makes a correspondence between $z$ and $d_{\text{gw}}^L$, cannot be constrained by a GW observation alone. Redshift is inferred by an E/M counterpart, or by a galaxy that is identified as the host of a GW event (Nishizawa 2017). Hereafter, we consider GW events where the redshift $z$ is known. The redshift and luminosity distance have been measured by a number of E/M observations. The measured luminosity distance, $d_{\text{L}}^{E/M}(z)$ is consistent with that of a flat ΛCDM universe (ignoring radiation):

$$d_{\text{L}}^{E/M}(z) = (1+z) \frac{c}{H_0} \int_0^1 \frac{dx}{\sqrt{\Omega_m x + \Omega_\Lambda}}.$$  

where $\Omega_m$, $\Omega_\Lambda$ are the dimensionless densities for matter and the cosmological constant respectively and $H_0$ the Hubble parameter. This fact allows us to assume that the luminosity distance is unique with the underlying cosmological model, i.e., $d_{\text{L}}^{E/M}(z) = d_{\text{E/M}}(z)$. However, this assumption should be carefully inspected with GW observations. Any deviations between luminosity distances derived from the GW and the E/M counterpart are potential signs of differences caused by systematic deviations from the underlying model assumptions. In the following sections we review the effects caused by modified gravity theories and inhomogeneities in the Universe, and the possibility to constrain them.

---

2 There is a degeneracy with orbit inclination which can be lifted with precise observations of the two polarisations or E/M observations (Baker et al. 2019).

3 For binary neutron stars, the tidal effects can permit a direct redshift measurement.
2.1 Modified Gravity

We start by reviewing the effects of modified gravity that can lead to a decrease on the observed amplitude of a GW\(^4\). We are interested in these, due to their possible degeneracy with the presence of inhomogeneities, that also result in a drop of the observed amplitude (Section 2.2). The new additions to GR can introduce several new effects on the propagation of GWs (Saltas et al. 2014, 2018; Nishizawa 2018; Amendola et al. 2018; María Ezquiaga & Zumalacárregui 2018). These are model dependent, but effectively they are summarised in the equation below (primes denote derivatives w.r.t. conformal time):

\[
h''_{ij} + [2 + \nu(\tau)]H h'_{ij} + [c_\tau(\tau)^2 k^2 + a(\tau)^2 \mu^2]h_{ij} = a(\tau)^2 \Gamma_{ij}. \tag{3}
\]

where the standard evolution of tensor perturbations is modified by \(\nu\) an additional friction term, \(c_\tau\) the GW propagation speed, \(\mu\) the graviton mass and \(\Gamma_{ij}\) a source term. The case: \(\nu = \mu = 0\) and \(c_\tau = 1\) corresponds to standard GR. Focusing only on the terms affecting the amplitude, and neglecting any source terms, we have:

\[
h''_{ij} + [2 + \nu(\tau)]H h'_{ij} + k^2 h_{ij} = 0. \tag{4}
\]

The general class of theories that predict a friction term \(\nu\) is the class of scalar-tensor theories (Saltas et al. 2018) and theories that break some fundamental assumptions, introducing for example non-locality (Belgacem et al. 2018a) or extra dimensions (Corman et al. 2020). Here, because of the friction term in front of the Hubble drag, the distances measured from GWs and E/M signals can differ. Usually the GWs distance is compared to the standard \(\Lambda\) CDM one - see eq. (2) - which is calculated using the redshift obtained from the E/M observation. Then, the relation between the GW and E/M luminosity distance is given by (Belgacem et al. 2018a, 2020):

\[
d_L^{GW}(z) = d_L^{E/M}(z) \exp \left\{ - \int_0^z \frac{\nu(\zeta')}{2} \frac{d\zeta'}{1 + \zeta'} \right\}. \tag{5}
\]

For constant \(\nu\), it is easy to solve this relation analytically:

\[
d_L^{GW}(z) = d_L^{E/M}(z) (1 + z)^{-\nu/2}. \tag{6}
\]

This case was investigated in (Nishizawa & Arai 2019), where the authors found that future constraints on \(\nu\) can reach 1%. The GW170817 event provided a very loose constraint on \(\nu\), in the range \(-75.3 \leq \nu \leq 78.4\), when considered a constant. The constraints are weaker when \(\nu\) is Taylor-expanded as \(\nu = \nu_0 - \nu_1 H_0 t_{LB}\), with \(t_{LB}\) the look-back time, and \(\nu_0, \nu_1\) are fitted (Arai & Nishizawa 2018). Throughout this paper, we assume that \(\nu\) is constant.

2.2 Inhomogeneous models

Inhomogeneous models affect both GWs and E/M distances on the same way. This is because they assume that both signals travel in paths that deviate from the smooth background, although without altering GR. A critical investigation of these effects is necessary, since comparing the GWs distance with the one based on \(\Lambda\) CDM can lead to the false assumption of a deviation from GR, while the source of the discrepancy would be a simplified approximation of the E/M distance, when inhomogeneities should be taken into account.

Below we review how different models can affect the angular diameter distance, and hence the luminosity distance that we would infer.

2.2.1 The FLRW distance

As a reference, we recall that for a smooth, homogeneous universe, the angular diameter distance \(D\), is given from the differential equation (Clarkson et al. 2012):

\[
d^2D \frac{dD}{dz^2} + \left( \frac{d \ln H}{dz} + \frac{2}{1 + z} \right) \frac{dD}{dz} = -\frac{3}{2} \frac{\Omega_m H_0^2}{H^2(1 + z)} D, \tag{7}
\]

where \(H(z)^2 = H_0^2 [\Omega_m (1 + z)^3 + \Omega_\Lambda + \Omega_k (1 + z)^2] \).

2.2.2 The DR distance

The Dyer-Roeder (DR) distance models a bundle of rays travelling through voids ("empty beams"), while assuming that the general background follows the standard FLRW geometry. Its usefulness is due to the fact that it yields the largest possible distance (for a given redshift) for light bundles which have not passed through a caustic (Schneider et al. 1992).

On the other hand, if the universe is not very clumpy, the Dyer-Roeder distances and those of the smooth FLRW universe are not too different (Fukugita et al. 1992; Schneider et al. 1992).

The DR distance is given by the solution to the following differential equation:

\[
d^2D \frac{dD}{dz^2} + \left( \frac{d \ln H}{dz} + \frac{2}{1 + z} \right) \frac{dD}{dz} = -\frac{3}{2} \frac{\Omega_m H_0^2}{H^2(1 + z)} \alpha(z) D, \tag{8}
\]

where \(H(z)^2 = H_0^2 [\Omega_m (1 + z)^3 + \Omega_\Lambda + \Omega_k (1 + z)^2]\) and only an extra factor \(\alpha(z)\) - the inhomogeneity factor - has been added to the RHS compared to the FLRW case. For \(\alpha = 1\), we return to the standard FLRW result, while \(\alpha = 0\) describes an "empty" beam, with the intermediate values denoting different levels of under-dense regions. Values of \(\alpha > 1\) correspond to over-densities.

2.2.3 The modified DR distance

As we mentioned above, the DR approximation assumes that we can disentangle the inhomogeneities encountered by photons from the background density, which assumes a smooth FLRW background. However, photons only experience the local curvature and expansion, so a simple modification would be to take into account the different expansion dynamics caused by the local anisotropies. This leads to a
modified version of the DR formula:\(^5\):
\[
\frac{d^2D}{dz^2} + \left( \frac{1 + z}{H(z)^2} \right) \left[ 3a(z)\Omega_m(1 + z) + 2\Omega_k \right] \frac{dD}{dz} = \frac{3}{2} \Omega_m \frac{H_0^2}{H(z)^2} (1 + z)\alpha(z)D,
\]
with \(H(z)^2 = H_0^2 [a(z)\Omega_m(1 + z)^3 + \Omega_L + \Omega_k (1 + z)^2]\), where the inhomogeneities parameter \(\alpha(z)\), has now been included in all the density terms (\(\rho_m \rightarrow a\rho_m\)), but its derivatives have been neglected, as in (Clarkson et al. 2012).

As can be seen from Figure 1, the distance of the DR and mDR approximations (in the extreme case of totally empty beams, i.e. \(\alpha = 0\)) vs the FLRW case can become quite important for high-z observations, and a careful modelling is needed if we want to extract accurate parameters.

Although these redshifts (\(\geq 1\)) are large enough to reach homogeneity, in most cases we observe focused light “beams”, so we usually do not see an average over the whole sky (Weinberg 1976). Hence, inhomogeneities at relatively large redshifts still need to be accounted for (Fleury et al. 2017; Dhawan et al. 2018).

2.2.4 The inhomogeneity parameter

The distances above are not mathematically exact solutions of GR, but are effective models trying to capture the impact of inhomogeneities on observed distances (Schneider et al. 1992; Mattsson 2010). Different parameterisations have been suggested in the literature for such effective models. Here we will focus on the following two (Linder 1988; Bolejko 2011):

(i) \(\alpha(z) = a_0 + a_1z\),
(ii) \(\alpha(z) = 1 + f(z)\langle\delta\rangle_D\),

where \(a_0, a_1\) are arbitrary constants, the function \(f(z)\) is chosen as \(f(z) = (1 + z)^{-5/3}\) to be consistent with the weak lensing approximation\(^6\) (Bonvin et al. 2006) and \(\langle\delta\rangle_D\) denotes the average present-time density contrast along a ray, with \(\delta = \delta\rho/\rho\). The two parameterisations are connected at small redshifts via \(a_0 = 1 + \langle\delta\rangle_D\) and \(a_1 = -5\langle\delta\rangle_D/4\) (see Appendix B for details). We obtain the values of \(\delta\) from cosmological simulations described in the next section.

2.3 Constraining the inhomogeneity parameter

The differential equations above give the angular diameter distance in each specific case. We solve them, using as initial conditions \(D(z = 0) = 0\) and \(D'(z = 0) = c/H_0\), and assuming, for consistency with the later sections, the cosmological parameters of WMAP9 (Hinshaw et al. 2013), with \((\Omega_m, \Omega_L, \Omega_k, \ H_0) = (0.285, 0.715, 0, 69.5)\).

To transform to luminosity distances we use the reciprocity distance relation (Etherington 1933; Ellis 2007), which holds for an arbitrary spacetime geometry, as long as photons are conserved, and connects luminosity and angular diameter distances via:

\[
d_L = (1 + z)^2D\tag{10}
\]

Concerning the errors in distance determination, we follow the analysis of (Cai & Yang 2017). We neglect the errors of the spectroscopic redshift determination, since they are negligible compared to the errors in the luminosity distance.

For the errors in the GW luminosity distance we take into account the uncertainty from instrumental errors (Li 2015):

\[
\sigma_{\text{inst}} = \frac{2d_L}{\rho},
\]

where \(\rho\) is the Signal-to-Noise Ratio, and the factor of 2 comes from marginalising over the uncertainty of the inclination’s determination. Also, there is an additional error, connected to weak lensing effects (Sathyaprakash et al. 2010; Zhao et al. 2011):

\[
\sigma_{\text{lens}} = 0.05zd_L.
\]

Hence, the total error (for a single event) is:

\[
\sigma_{d_L} = \sqrt{\left(\frac{2d_L}{\rho}\right)^2 + (0.05zd_L)^2}
\]

We add these errors in quadrature, assuming they originate from independent sources, as one part is systematic and the other statistical. In the following, when calculating the \(\chi^2\) fit, in order to estimate the errors we are going to consider the luminosity distance of the “fiducial” case as their source. We assume that all sources are independent, and include a scatter in their distances, allowing them to follow a normal distribution \(N(d_L, \sigma)\), with \(d_L\) the mean and \(\sigma\) the standard deviation, for each redshift, as shown in Figure 2. A constant value of \(\rho = 8\), the minimum value usually required for detection, is also used as a conservative limit for the estimation of the luminosity distance error. Later on, we also investigate the consequences of choosing a more optimistic value, that would be standard for future GWs observatories.

2.4 Merger Rates

The redshift distribution of potentially observable GW sources (Figure 3) is given by (Zhao et al. 2011):

\[
P(z) \sim \frac{4\pi\Omega_{gw}^2(z)R(z)}{H(z)(1 + z)},
\]

Figure 1. An extreme example (“empty beam” case, \(\alpha = 0\)) of the differences between the different distant measures. The three models start deviating significantly for \(z \geq 1\).
the secondary star are: (i) the mass of the primary star assumptions used for the calculation in Schneider et al. (Portegies Zwart & Yungelson 1998), where the main SeBa The rates are based on the binary population synthesis code with E/M counterparts (BH-NS & NS-NS) given by (Schneider et al. 2001; Zhao et al. 2011):

\[
R(z) = \begin{cases} 
  1 + 2z, & z \leq 1 \\
  \frac{3}{2}(5 - z), & 1 < z < 5 \\
  0, & z \geq 5. 
\end{cases}
\]  

The rates are based on the binary population synthesis code SeBa (Portegies Zwart & Yungelson 1998), where the main assumptions used for the calculation in Schneider et al. (2001) are: (i) the mass of the primary star \( m_1 \) is determined using the mass function described by (Scalo 1986) between 0.1 and 100 \( M_\odot \). (ii) For a given \( m_1 \), the mass of the secondary star \( m_2 \) is randomly selected from a uniform distribution between a minimum of 0.1 \( M_\odot \) and the mass of the primary star, (iii) The semi-major axis distribution is taken flat in \( \log(a_{\text{ini}}) \) ranging from Roche-lobe contact up to \( 10^9 R_\odot \) and (iv) The initial eccentricity distribution is independent of the other orbital parameters. The results are constrained by our Galaxy, but, using the observed star formation rate, they are extended until redshift \( z = 5 \).

In the rate equation - eq.(14) - we can assume, without affecting our results, the estimates for a smooth FLRW cosmology, since we are interested in the deviations of distances between some observed sources. This should not be affected by how we populate the sources (see also Appendix A). The evolution of merger rate as a function of redshift has also been constrained from the up-to-date detections (Abbott et al. 2019b), but with poor results so far due to the small statistics. Only the presence of a significant number of lensed events could change considerably the distribution, however this does not seem to be the case (Hannuksela et al. 2019). Nevertheless, the method proposed in this paper is not invalidated by the distribution of sources, however the latter is important for the estimated error contours.

3 SIMULATIONS

To calculate realistic density anisotropies, we rely on cosmological dark matter only simulations run with Gadget-3 (Springel 2005) for the LEGACY project.

The latter is composed by two primary volumes of 1600 Mpc/h and 100 Mpc/h box sizes run down to \( z = 0 \) with 2048³ resolution elements, as well as a set of zoom-in simulations on the larger box, with size 83 Mpc/h, and an effective resolution of 32768³ (1700³). These simulations have been designed to sample \(-2, -1, 0, +1, \) and \(+2 \) \( \sigma \) of the mean density value as well as extremely high (cluster) and very low (void) density regions and are therefore ideal to study different environmental effects.

For our analysis, we use the data from the big 1600 Mpc/h run, at \( z = 0 \), with an effective resolution of 2048³ particles and \( 5.43 \times 10^{10} M_\odot \) mass resolution which will suffice for the purpose of this investigation.

3.1 Cloud-in-cells implementation

For our simulation box, we perform a Cloud-in-Cell (CiC) interpolation scheme to deposit the particles to specific grid points (Hockney & Eastwood 1988). The weighting function is:

\[
W(x - x_p) = \begin{cases} 
  1 - \frac{|x - x_p|}{L}, & \text{if } |x - x_p| \leq L \\
  0, & \text{otherwise,} 
\end{cases}
\]  

where \( L \) is the resolution of the CiC box and \( x_p \) the centres of the grid cells.

3.1.1 Density calculation

We investigate a number of resolutions, from a grid with 128³ cells to a grid with 1024³ cells. From these we calculate the density anisotropies along straight trajectories and calculate the mean value for each ray\(^7\) to find \( \langle \delta \rangle_{1D} \) (Figure 4). For all resolutions the mean 3D density contrast \( \langle \delta \rangle \) is zero.

As expected, the lowest resolution (128³) returns a distribution closer to the mean contrast, since we average out on larger volumes. The highest resolution run (1024³) produces a distribution with tails that describe the under/over-densities at small scales. In the following we exploit the data from this run.

\[\text{See Appendix C for details.}\]
First of all, in both cases we recover the input models at the $1\sigma$-level. Secondly, we observe a similar trend as (Clarkson et al. 2012), in that different inhomogeneous models can lead to quite distinct results as shown by the dashed-line contours compared to the solid line ones. These uncertainties in the modelling, limit precision cosmology in terms of being able to constrain physical properties (e.g. the optical depth), modified gravity theories or parameterisations of the dark energy equation of state to high accuracy (Chevallier & Polarski 2001; Linder 2003).

Thirdly, as expected, we observe that both parameterisations lead to more stringent constraints when more events are detected. Also, high-z observations are required to better distinguish between the models, being consistent with Figure 1. The two parameters ($a_0, a_1$) are almost unconstrained in the DR case, while $a_0$ can reach $\sim 50\percnt$ accuracy if a mDR model is considered. However, the latter seems almost unaffected by either $z_{\text{max}}$ or the number of sources.

A similar conclusion results from the analysis of 350 GW sources, but for different maximum redshifts $z_{\text{max}} = 1.5, 3$. The relevant figures demonstrate clearly that the maximum redshift is a much more important parameter than the number of sources, since the different distance measures start to deviate significantly at larger redshifts. This would allow a direct probe of the scales where the Universe reaches homogeneity, especially if the errors can be reduced by de-lensing techniques (Lewis & Challinor 2006).

4 RESULTS

4.1 Inhomogeneity Constraints

To fit our model with GW standard sirens observations we use a $\chi^2$ goodness of fit defined as:

$$\chi^2 = \sum \frac{[d_L(z) - d^\text{GW}_L(z, \alpha, \Omega)]^2}{\sigma_{d_L}(z)^2},$$

(17)

where the summation includes all distinct observations (we use here 350 mock GWs events, unless stated otherwise. This corresponds to a conservative limit for future detectors (Maggiore et al. 2020; Abbott et al. 2017a) for redshifts $z \geq 2$ per year cycle, but seems to be a quite optimistic limit for a 5-year LISA mission (Klein et al. 2010) and the next observing run of the current GWs detectors (Abbott et al. 2018). The $\sigma$ term summarises the different possible parameterisations. For each “fiducial” distance, the $\alpha$ parameters take all the values between 0 to 1, where for each one a different value for $\chi^2$ is calculated. The redshift of the sources are randomly drawn from the distribution in Figure 3. We denote as $d_L(z)$ the “fiducial” distance, that would be specified in each case (see below), and with this we calculate the errors using eq. (13). $d^\text{GW}_L$ denotes the distance based on the inhomogeneous models we study. We finally calculate the constraints that can be put on the inhomogeneity/modified gravity parameters. In our contour fits we plot the $1\sigma, 2\sigma, 3\sigma$ confidence intervals. The filled ones correspond to the DR model and the dashed ones to the mDR model.

4.1.1 Scaling with sources and redshift

We begin by investigating how the constraints scale with number of sources and with maximum value of redshift. For this we simulate 150, 350 and 500 sources and investigate redshifts $z = 1.5, 3, 5$. The filled contours correspond to constraints based on the DR distance and the dashed ones on the mDR distance, both using the first parameterisation of section 2.2.4 ($a(z) = a_0 + a_1 z$).

Our results are summarised in Figure 5. We start by testing how well we recover an inhomogenous DR model with input parameters ($a_0, a_1$) = (0.5, 0.5) and a FLRW model. First of all, in both cases we recover the input models at the $1\sigma$-level.

4.1.2 Constraints from numerical simulations

We now follow the same procedure exploiting the second parameterisation in 2.2.4. This has a clearer physical interpretation, related directly with the density anisotropies along the line-of-sight. Here we keep the maximum redshift and the number of sources fixed at $z_{\text{max}} = 5$, $N = 350$, and consider as “fiducial” distances the ones based on the $\delta$ parameterisation ($a(z) = 1 + f(z)/(\delta 1D)$).

To calculate them, we follow the following procedure: For each mock source, we pick randomly a 1D density contrast from the high resolution distribution of Figure 4 and a random redshift as before. We calculate the “fiducial” distance by numerically solving the DR equation, since in this model this would be equivalent to the weak lensing approximation. With them, we try to constrain the DR and mDR models, based on the first parameterisation of section 2.2.4 ($a(z) = a_0 + a_1 z$).

An example of the possible deviation is shown in Figure 6, where we have used the two limiting cases ($\pm 3\sigma$) based on our simulations. Since these correspond to small perturbations in the metric, the effects are small. More extreme cases are shown in Figure 7. The constraints on ($a_0, a_1$) are shown in Figure 8. We see that the presence of small inhomogeneities along the ray results in distance estimates consistent with FLRW, confirming some previous semi-analytical and numerical studies (Mortsell 2002; Kaiser & Peacock 2016; Adamek et al. 2019). However, we note that the result
Figure 5. The left column shows how the constraining power of standard sirens changes with number of sources. The two upper figures in the right column show the effects of maximum redshift, when total number of sources is held fixed. The input distances in these cases follow a DR model with \((a_0, a_1) = (0.5, 0.5)\). The bottom right plot shows a “sanity check” fit to an FLRW input. Filled contours show constraints on the DR models, while dashed lines on the mDR ones. S/N ratio is \(\rho = 8\).
Figure 6. An example of the most extreme expected deviations from FLRW when using the $\delta$ parameterisation, based on simulation data. Values higher (lower) than the FLRW one, correspond to mean under (over) density.

Figure 7. Ratio of modified distances over the FLRW one for two cases of big under-densities: “empty” case, $\langle \delta \rangle_{1d} = -1$ (dot-dashed lines) and “semi-empty” case, $\langle \delta \rangle_{1d} = -0.5$ (solid lines).

Figure 8. Constraints on the inhomogeneity parameters $(a_0, a_1)$ (as above) based on a realistic density distribution from numerical simulations. For both cases there is consistency with a FLRW background.

only holds for weak inhomogeneities$^8$, since non-linear effects could potentially have an important contribution (e.g. Bolejko 2018).

4.1.3 Limitations by observational errors

Finally, we note that the improvement of the constraints is going to be limited by the observational errors, more specifically $\rho$, especially if the deviations from FLRW are small. We summarise this in Figure 9, where we show that the addition of a huge number of extra observations is needed for a significant shrinkage of the confidence contours.

This is also true for the constraints on modified gravity models (see section 4.2). There we see, in Figures 15 and 17, that the constraints can be greatly improved if the observational errors are significantly decreased (around 5 times smaller). This demonstrates, as shown above, that small inhomogeneities are statistically equivalent to the FLRW approximation.

4.2 Modified gravity effects

As we have seen, the measurement of accurate distances from GWs can be in itself a independent probe of the inhomogeneity of our Universe. Inhomogeneous models with under-densities, in general predict larger distances than the FLRW ones and this effect could have more general implications, since every physical mechanism that modifies the amplitude of the wave, could lead to a similar deviation (María Ezquiaga & Zumalacárregui 2018; Belgacem et al. 2018a, 2019; Wei 2019; Zhou et al. 2019). Figure 11 provides a simple example of this correspondence.

Of course, an inhomogeneous universe will lead to a different propagation than FLRW, for both photons and GWs, however this strengthens our point that an accurate determination of the underlying geometry is necessary before investigating deviations of modified gravity models, which are usually compared with the expectation from FLRW. This could lead to important implications if the underlying geometry is not FLRW, since in this case modified gravity effects are degenerate with respect to inhomogeneous models.

However, as can been seen in the example of Figure 11, precise distance measurements, that would significantly reduce the observed errors, could disentangle the two effects in simple models, since they have a different redshift dependence (concave vs convex curves). Also it is worth noticing that inhomogeneous models could possibly only mimic gravity modifications that lead to increased distance, so with $\nu < 0$.

For a more detailed comparison, we repeat our $\chi^2$ fit, where in this case we use as “fiducial” input a modified gravity model with effective parameter $\nu = -0.5$. We then constrain the values of $(a_0, a_1)$ for the two inhomogeneous models, using 350 mock observations that reach $z_{\text{max}} = 5$. This leads to Figure 13. We see that even a small number of sources ($N = 30$) results in quite strong constraints, indicating that only extreme inhomogeneities can lead to equivalent results. Of course, this is an arbitrary example, but it

$^8$ Note that the maximum mean under-density we find in our simulations is of the order of $\delta \rho / \rho \sim -0.4$. 

MNRAS 000, 1–15 (2020)
Figure 9. Evolution of constraints (as above) with increasing number of sources. Here we assume that weak lensing errors are random and independent. S/N ratio is $\rho = 8$.

Figure 10. Evolution of constraints (as above) with increasing S/N ratio. From $\rho = 8$ to $\rho = 30$ and $\rho = 50$, consistent with current and future observatories’ capabilities.

Figure 11. A comparison of the effect on the GW distances between inhomogeneous models and a general modified gravity parameterisation. Constant parameters are $(a_0, a_1) = (0, 0), (a_0, a_1) = (0.2, 0.1)$ for DR and mDR models respectively, and $\nu = -0.5$ for modified gravity. To be compared with Figure 12.

Figure 12. A comparison of the effect on the GW distances between inhomogeneous models and a general modified gravity parameterisation. Constant parameters are $(a_0, a_1) = (0, 0), (a_0, a_1) = (0.2, 0.1)$ for DR and mDR models respectively, and $\nu = -0.5$ for modified gravity.

demonstrates our previous point that “realistic” inhomogeneous models are not able to mimic all values of the modification parameters. The best fit parameters lead to distances as shown in Figure 12, which deviate significantly from the FLRW ones and are probably unphysical. However, modified gravity models that result in less extreme values of $(a_0, a_1)$ consistent with the bounds shown in Figure 13 would not be distinguishable from inhomogeneous models given present observational facilities.

Finally, we invert the procedure and exploiting the $\delta$ parameterisation we try to fit the best modified gravity model and put limits to the values of $\nu$ that could be disentangled from small inhomogeneity effects. As can be seen in Figure 14, deviations bigger than $\sim 0.1$ from the FLRW value
Figure 13. Contour fits for the two inhomogeneous distance models, when the “fiducial” distances follow a modified gravity model with $\nu = -0.5$. S/N ratio is $\rho = 8$.

Figure 14. $\chi^2$ fit of a modified gravity model to realistic inhomogeneous distances from a cosmological simulation. The vertical lines correspond to the 99% confidence limits. Even a large number of observations requires deviations from $\nu = 0$ of the order of $O(0.1)$ to disentangle modified gravity effects. S/N ratio is $\rho = 8$.

($\nu = 0$) are needed, in order to be possible to disentangle the modified gravity effect from inhomogeneities, even for a large number of observations. Assuming very small observational errors (around 1/5 of the reported ones), deviations bigger than $\sim 0.01$ from the GR value ($\nu = 0$) are enough to disentangle modified gravity effects (Figure 15). This shows that the intrinsic scatter of small inhomogeneities is not significant enough to manage to mimic large deviations from the GR case. We investigate these effects further in section 4.3.

At the same time, a small number of observations can lead to serious misidentifications of inhomogeneity effects with deviations from GR. We demonstrate this effect and quantify the number of events needed for better convergence in Figure 16, where we draw the likelihood of the $\nu$ parameter, for different numbers of events. At least 100 standard sirens would be needed for convergence to the “real” value with about $\delta \nu \sim \pm 0.1$ accuracy. The same order of accuracy is obtained when we triple the number of sources (left-panel in Figure 14), which demonstrates that already at these sample numbers intrinsic observational errors become the limiting factor, as we have commented above.

4.3 Varying S/N

So far we have kept the signal-to-noise ratio constant, and more specifically we have set $\rho = 8$. This corresponds to the standard detection limit that is used in current observations. Moreover it allowed us to investigate the more pessimistic scenario possible.

However, as we have mentioned above, future observations are expected to have a higher S/N ratio, which would
automatically reduce the errors and improve our constraints.

In Figure 17, we show how the improved accuracy is going to shrink the uncertainty on the calculation of the friction parameter \( \nu \) (we refer here to the width of the distribution in Figure 16). The improvement can be as much as 100% for a small number of observations. For a large number of events \((N \geq 350)\), the enhancement is smaller, but it seems that it has a more important contribution in increasing accuracy, compared to the number of observations.

Finally, we extend our study to a larger range for \( \rho \) in Figure 18. Here we study how the S/N ratio is affecting our ability to constrain the inhomogeneity parameters \((a_0, a_1)\). We assume as baseline the worst case scenario, \( \rho = 8 \), \( N = 10 \) and calculate the improvement expected when increasing either \( \rho \), or \( N \) or both. We quantify this by comparing the areas of the 68% contours with the baseline case, for example a 5 times improvement, would lead to a 5 times smaller area. Again, as above, we demonstrate that the upgrade of our detectors (higher \( \rho \)) is much more important than the number of events.

Future runs (Abbott et al. 2018) and detectors (Klein et al. 2016; Abbott et al. 2017a; Maggiore et al. 2020) are expected to achieve this enhancement both in terms of \( \rho \) and number of detections.

5 CONCLUSIONS

In this work, we propose the use of GWs standard sirens as a quite clean and model-independent probe for studying inhomogeneities in the universe. Modelling the inhomogeneities with two effective distance formulae, the DR and mDR, we put constraints on the underlying geometry of our Universe. Furthermore, we investigate the degree of degeneracy between the impact of inhomogeneities and of modified gravity theories on the luminosity distance of GWs. We claim that a possible confusion on a test of gravity occurs due to the indistinguishability of gravity effects from inhomogeneity.

More specifically, we show that:

- Constraints to effective inhomogeneity parameters are possible from future standard sirens observations, though in most cases these will be weak. These are more dependent on the horizon redshift of the observations, than on the number of sources observed. We see that low redshift observations lead to very weak constraints on inhomogeneous models, making it difficult to disentangle them from other effects. Hence, high redshift observations \((z \geq 1.5)\), are needed to provide a clearer probe for inhomogeneities. For these, future detectors (Klein et al. 2016; Abbott et al. 2017a; Maggiore et al. 2020) are necessary.

- Realistic inhomogeneities, based on numerical simulations of cosmological structure formation, lead to constraints consistent with an FLRW geometry, but still the constraints are weak, unless a significant number of sources are detected \((N \geq 350)\). These are expected for the next generation of ground based detectors, but they constitute a very optimistic limit for the near future space detectors, like LISA (Amaro-Seoane et al. 2013). At some point, our constraining power will be severely limited by observational errors (low S/N ratio), unless deviations from FLRW are significant.

- A modified propagation due to an inhomogeneous background can lead to constraints on the geometry of the universe itself. We have neglected any angular dependence, but since we are considering narrow beams, a possible presence of anisotropies could be directly constrained by future observations.

- An inhomogeneous background can “mimic” modified gravity models in the amplitude decay of a GW. This should be taken into account, when trying to constrain parameters in these models. Most extreme cases can be easily disentangled, but we have shown that modifications in the \( \nu \) parameter, of the order of \( \mathcal{O}(0.1) \), would be needed to disentangle these effects from inhomogeneities. At the same time, a significant number of standard sirens \((N \geq 100)\) is necessary to avoid misidentifications. Similar care should be taken, when constraining other physical phenomena with similar effects on observed distances, like the average opacity of the intergalactic medium (Wei 2019; Zhou et al. 2019).

- The improvement of our observing accuracy, quantified with the S/N ratio, will play the more significant role in
Figure 17. Improvement of accuracy in the determination of the friction term $\nu$ with number of events and S/N ratio. The effect of $\rho$ is more important than increasing the number of observations.

Figure 18. Effect of different signal-to-noise ratio $\rho$ and number of sources $N_{\text{sources}}$ on the constraints on the inhomogeneity parameters ($a_0$, $a_1$). Improvement is quantified on how much the 68% contour area is shrunk compared to the worst case scenario of $\rho = 8$, $N = 10$.

our constraining power for both inhomogeneity and modified gravity parameters, compared to an increased number of detections.

ACKNOWLEDGEMENTS

We want to thank Tjonnie Li for useful clarifications. SA thanks for the hospitality at Royal Observatory in Edinburgh during the Nagoya University and The University of Edinburgh Joint-Degree Program. For the analysis we used Mathematica (Wolfram Research 2019), Matplotlib (Hunter 2007) and NumPy (Oliphant 2006; Van Der Walt et al. 2011).
Rubin D., Hayden B., 2016, (https://arxiv.org/abs/1610.08972)
Rubin D., Heitlauf J., 2019, arXiv e-prints, (https://arxiv.org/abs/1912.02191)
Saltas I. D., Sawicki I., Amendola L., Kunz M., 2014, Phys. Rev. Lett., (https://arxiv.org/abs/1406.7139)
Saltas I. D., Amendola L., Kunz M., Sawicki I., 2018, preprint, (https://arxiv.org/abs/1812.03969)
Sathyaprakash B. S., Schutz B. F., Van Den Broeck C., 2010, Classical and Quantum Gravity, (https://arxiv.org/abs/0906.4151)
Scalo J. M., 1986, Fundamentals Cosmic Phys.
Schneider P., Ehlers J., Falco E. E., 1992, Gravitational Lenses. Springer-Verlag, doi:10.1007/978-3-662-03758-4
Schneider R., Ferrari V., Matarrese S., Portegies Zwart S. F., Schneider P., Ehlers J., Falco E. E., 1992, Gravitational Lenses.
Scalo J. M., 1986, Fundamentals Cosmic Phys.
Wei J.-J., 2019, (https://arxiv.org/abs/1902.00223)
Van Der Walt S., Colbert S. C., Varoquaux G., 2011, Computing in Science & Engineering, 13, 22
Verde L., Treu T., Riess A. G., 2019, Nature Astronomy, (https://arxiv.org/abs/1812.02327)
Van Der Walt S., Colbert S. C., Varoquaux G., 2011, Computing in Science & Engineering, 13, 22
Wei J.-J., 2018, (https://arxiv.org/abs/1806.09781)
Wei J.-J., 2019, (https://arxiv.org/abs/1902.00223)
Weinberg S., 1976, ApJ, (https://arxiv.org/abs/astro-ph/0002055)
Wei J.-J., 2019, (https://arxiv.org/abs/1902.00223)
Verde L., Treu T., Riess A. G., 2019, Nature Astronomy, (https://arxiv.org/abs/1907.10625)
Wei J.-J., 2019, (https://arxiv.org/abs/1902.00223)
Weinberg S., 1976, ApJ, (http://articles.adsabs.harvard.edu/pdf/1976ApJ...208L...1W)
Wolfram Research I., 2019, (https://www.wolfram.com/mathematica)
Yagi K., Nishizawa A., Yoo C.-M., 2012, (https://arxiv.org/abs/1204.1670)
Yoo C.-M., Nakao K.-i., Kozaki H., Takahashi R., 2007, ApJ, 655, 691
Zhao W., van den Broeck C., Baskaran D., Li T. G. F., 2011, Phys. Rev. D, (https://arxiv.org/abs/1009.0206v4)
Zhou L., Fu X., Peng Z., Chen J., 2010, (https://arxiv.org/abs/1912.02327)

APPENDIX A: THE EFFECT OF AN INHOMOGENEOUS UNIVERSE TO MERGER RATES

Using a modified “Hubble expansion”, given by:

\[
\dot{H}(z) = H_0[a(z)\Omega_m(z)(1+z)^3 + \Omega_{\Lambda,0} + \Omega_k,0(1+z)^2]^{1/2}, \tag{A1}
\]

and the modified comoving distance,

\[
\tilde{D}(z) = c \int_0^z \frac{1}{\dot{H}(z')} dz', \tag{A2}
\]

the merger rates are:

\[
P(z) \sim \frac{4\pi \tilde{D}(z) R(z)}{\dot{H}(z)(1+z)}. \tag{A3}
\]

Although this may lead to some differences (Figure A1), the effect is small. Also, we want to emphasize a caveat of this analysis: the uncertainties of the stellar population and evolution models (summarised in R(z)) should be more important than the effects of inhomogeneities, so the distribution of merger rates per redshift isn’t a clear probe (except maybe for some very extreme cases that are not plausible). Hence this reinforces our arguments that it’s safe to assume a homogeneous background when estimating the mergers’ distribution.

APPENDIX B: COMPARISON OF PARAMETERISATIONS

In the main text we study two parameterisations proposed by (Linder 1988; Babichev 2011):

(i) \(a(z) = a_0 + a_1 z\),
(ii) \(a(z) = 1 + D(z)(\delta)_{1D}\),

where we chose the function \(D(z)\) as \(D(z) = (1+z)^{-5/4}\) in order to be consistent with the weak lensing approximation (Bonvin et al. 2006). The \(\langle \delta \rangle_{1D}\) denotes the average present-time density contrast along a ray.

Although the first one is more general, the two parameterisations are connected at small redshifts. A Taylor expansion of (ii) gives:

\[
a(z) = 1 + (1+z)^{-5/4}(\delta)_{1D} = 1 + (\delta)_{1D} - \frac{5}{4} z(\delta)_{1D}. \tag{B1}
\]

With the following identifications: \(a_0 = 1 + (\delta)_{1D}\) and \(a_1 = -5(\delta)_{1D}/4\), we see how the \(a\) parameters are connected to inhomogeneities along the line of sight.

APPENDIX C: DENSITY DISTRIBUTION AND RAY-TRACING

In the main text we calculated the distribution of mean densities along straight trajectories (see Figure 4). Although this approximation is valid, when anisotropies are small, as a check we performed the same exercise using a ray-tracing code. The latter propagates the rays along the potential...
GWs as inhomogeneity probes

Figure C1. Distribution of densities along 1D rays ($\langle \delta \rangle_{1D}$ + 1) on simulations of different resolution, at $z = 0$. The higher resolution run, which corresponds to a density averaging in $\sim 8$ Mpc$^3$ cubes gives the more interesting tails, being able to resolve better the small-scale structure.

The results are shown in Figure C1 and are consistent with the ones we described in the main text, validating our analysis.

This paper has been typeset from a TeX/LaTeX file prepared by the author.

---

9 A detailed examination of our ray-tracing method will be described in a future work. Here we use it as a sanity check.