

Gravitational wave cosmology with extreme mass-ratio inspirals

Danny Laghi^{1*}, Nicola Tamanini^{2,3}, Walter Del Pozzo¹,

Alberto Sesana⁴, Jonathan Gair², Stanislav Babak⁵

¹ *Dipartimento di Fisica “Enrico Fermi”, Università di Pisa, and INFN Sezione di Pisa, Pisa I-56127, Italy*

² *Max-Planck-Institut für Gravitationsphysik, Albert-Einstein-Institut, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany*

³ *Laboratoire des 2 Infinis - Toulouse (L2IT-IN2P3), Université de Toulouse, CNRS, UPS, F-31062 Toulouse Cedex 9, France*

⁴ *Dipartimento di Fisica “G. Occhialini”, Università di Milano - Bicocca, Piazza della Scienza 3, 20126 Milano, Italy*

⁵ *Laboratoire Astroparticule et Cosmologie, CNRS, Université de Paris, 10 rue Alice Domon et Léonie Duquet, 75013 Paris, France*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

The Laser Interferometer Space Antenna (LISA) will open the mHz frequency window of the gravitational wave (GW) landscape. Among all the new GW sources expected to emit in this frequency band, extreme mass-ratio inspirals (EMRIs) constitute a unique laboratory for astrophysics and fundamental physics. Here we show that EMRIs can also be used to extract relevant cosmological information, complementary to both electromagnetic (EM) and other GW observations. By using the loudest EMRIs (SNR>100) detected by LISA as dark standard sirens, statistically matching their sky localisation region with mock galaxy catalogs, we find that constraints on H_0 can reach $\sim 1.1\%$ ($\sim 3.6\%$) accuracy, at the 90% credible level, in our best (worst) case scenario. By considering a dynamical dark energy (DE) cosmological model, with Λ CDM parameters fixed by other observations, we further show that in our best (worst) case scenario $\sim 5.9\%$ ($\sim 12.3\%$) relative uncertainties at the 90% credible level can be obtained on w_0 , the DE equation of state parameter. Besides being relevant in their own right, EMRI measurements will be affected by different systematics compared to both EM and ground-based GW observations. Cross validation with complementary cosmological measurements will therefore be of paramount importance, especially if convincing evidence of physics beyond Λ CDM emerges from future observations.

Key words: Gravitational waves astrophysics – Cosmology – Extreme mass ratio inspirals

1 INTRODUCTION

The first direct detection of gravitational waves (GWs) in 2015 (Abbott et al. 2016a) ended a long experimental quest and opened a new observational window onto the Universe. Since then, the three ground-based interferometers currently operated by the LIGO-Virgo Collaboration, have announced a total of 50 candidate GW events emitted from binary black hole (BBH) mergers, binary neutron star (BNS) mergers, and possible binary neutron star-black hole mergers (Abbott et al. 2019a, 2020a,b). These observations have shed new light on the physics of compact objects (Abbott et al. 2019a,c) and have allowed tests of General Relativity in the dynamical strong field regime for the first time (Abbott et al. 2016b, 2019b,d, 2020c). With observational runs currently ongoing and improvements in the sensitivity of the interferometers planned for the forthcoming years, many more detections are expected with a progressive increase in the parameter estimation accuracy. These new observations are quickly consolidating GWs into a new observational science, namely *GW*

astronomy (Barack et al. 2019). Although the direct scope of GW observations consists in recovering the intrinsic and astrophysical properties of GW sources, interesting cosmological information can be extracted from the data as well, especially if associated electromagnetic (EM) counterparts can be identified. Cosmological inference can thus be considered a subpart of GW astronomy, which we can call *GW cosmology*.

1.1 Gravitational-wave cosmology

The poster example of how GW observations can be used to infer cosmological constraints is given by GW170817, the first BNS merger ever detected (Abbott et al. 2017b). The observation of an EM counterpart to the GW signal (Abbott et al. 2017d) allowed the identification of the host galaxy of the source, and thus the use of the event as a cosmic distance indicator. In this context such GW sources are usually referred to as *standard sirens* (Schutz 1986; Holz & Hughes 2005; Nissanke et al. 2010). The simultaneous measurement of both the luminosity distance (from the GW waveform) and the redshift (from EM observations) of a GW source provides data points to fit the so-called *distance-redshift relation*

* e-mail:danny.laghi@df.unipi.it

(Peebles 1993; Weinberg 2008), which links the luminosity distance to the redshift of each point in the Universe and is a function of the cosmological parameters characterizing the cosmic background expansion. At low redshift this relation becomes the *Hubble law*, that only depends on the Hubble constant H_0 . Given the low redshift ($z \simeq 0.01$) of GW170817, this event provided constraints on H_0 only. The results are in general agreement, though not competitive, with previous measurements (Abbott et al. 2017c). Future observations of similar events will reduce the uncertainty on H_0 (Dalal et al. 2006; Nissanke et al. 2013; Chen et al. 2018), and possibly will help solving the tension on its measured value between local and CMB observations (e.g. Ade et al. (2016); Aghanim et al. (2018); Riess et al. (2016, 2019); Mörtzell & Dhawan (2018); Feeney et al. (2019)).

An EM counterpart is not strictly necessary to use compact binary mergers such as BBHs and BNSs as standard sirens. By matching the sky localisation region of GW sources – which can be inferred from the GW measurements – with galaxy catalogs, one might in fact be able to extract complementary information on the redshift of the sources, without the need of spotting an EM counterpart. The idea was originally proposed by Schutz (1986) and it has subsequently been used and developed in different analyses (Holz & Hughes 2005; MacLeod & Hogan 2008; Del Pozzo 2012; Petiteau et al. 2011; Chen et al. 2018; Gray et al. 2020). It has already been tested with real data collected by the LIGO and Virgo detectors (Fishbach et al. 2019; Soares-Santos et al. 2019; Abbott et al. 2020d), though the constraints obtained so far with this “statistical” method are not competitive with the ones derived from GW170817 and its EM counterpart, mainly because of the poor spatial resolution of the current network of ground-based interferometers. Future observations, taken with an enlarged network of ground-based GW detectors, will allow for better cosmological measurements (Chen et al. 2018), mainly thanks to the improved sky localisation accuracy. Other complementary methods, which analogously do not require the identification of an EM counterpart, might yield interesting results as well (Taylor & Gair 2012; Del Pozzo et al. 2017; Oguri 2016; Mukherjee & Wandelt 2018; Mukherjee et al. 2020c,d,a,b; Farr et al. 2019; Ezquiaga & Holz 2020). The era of precise cosmological measurements with GWs will however start only with next generation interferometers, such as the *Einstein Telescope* (ET) (Punturo et al. 2010; Maggiore et al. 2020; Sathyaprakash et al. 2010; Belgacem et al. 2019a) and the *Cosmic Explorer* (Abbott et al. 2017a; Reitze et al. 2019a,b) on the Earth, or *TianQin* (Mei et al. 2020), *Taiji* (Luo et al. 2020), and the *Laser Interferometer Space Antenna* (LISA) (Amaro-Seoane et al. 2017) in space. The latter instrument is the focus of the present investigation. In what follows, we will briefly introduce LISA and review previous studies of LISA’s capability to do cosmological analyses using standard sirens. More details on how to extract cosmology from GWs by statistically matching with galaxy catalogs will be given in Sec. 2.

1.2 Cosmology with LISA

LISA is a space mission designed to detect GWs. It has been selected by the European Space Agency (ESA) for the L3 slot of its *Cosmic Vision* program, with launch expected in the early 2030s (Amaro-Seoane et al. 2017). By using inter-

ferometric technology already tested in space (Armano et al. 2016, 2018), with a much larger arm-length baseline (2.5×10^6 km) compared to present ground-based detectors (\sim few km), LISA aims at measuring GWs in the mHz frequency band, which is expected to be populated by GWs emitted by many different sources.

Expected GW sources include: massive black hole binary (MBHB) mergers (Klein et al. 2016), with masses ranging from $10^4 M_\odot$ to $10^7 M_\odot$ and detectable up to redshift ~ 20 ; the inspiral of stellar-origin black hole binaries (SOBHs) (Sesana 2016), with masses ranging from a few tens up to $\sim 100 M_\odot$, the merger of which will be detectable by ground-based interferometers; extreme mass-ratio inspirals (EMRIs) Babak et al. (2017), which are BBH systems formed by a massive black hole (MBH) and a stellar-origin black hole (SOBH); Galactic and Local-Group binaries (Breivik et al. 2018; Korol et al. 2017, 2018; Lau et al. 2019), i.e., compact stellar binaries in the Milky Way and nearby galaxies; and stochastic backgrounds of GWs (Caprini et al. 2016; Bartolo et al. 2016; Caprini & Figueroa 2018), of both cosmological and astrophysical origin. The measurement and analysis of all these sources will yield unprecedented astrophysical and fundamental physics information, and will allow us to test General Relativity in as yet unprobed regimes.

The GW sources that LISA will observe at cosmological distances can be used as standard sirens. These include MBHBs, EMRIs, and SOBHs. Unfortunately, only for the first of these types of sources are EM counterparts plausibly expected to be produced and observed by future EM facilities (Tamanini et al. 2016). MBHBs are in fact expected to emit a large amount of EM radiation in different bands at merger or during long-lasting (\sim weeks/months) afterglows (see, e.g., Palenzuela et al. 2010; Dotti et al. 2012; Giacomazzo et al. 2012; Moesta et al. 2012), and possibly even through pre-merger signals (Kocsis et al. 2008; O’Shaughnessy et al. 2011; Kaplan et al. 2011; Haiman 2017; Dal Canton et al. 2019). If sufficiently accurate sky localisation can be attained from the GW parameter estimation analysis and if the EM counterpart is sufficiently powerful to be spotted by EM telescopes, then we expect to identify the host galaxy of up to a few LISA MBHB mergers per year (Tamanini et al. 2016; Tamanini 2017). These golden sources can then be used as high-redshift standard sirens to map the expansion of the Universe up to $z \sim 10$. Although the low number of expected EM counterparts and the high redshift of MBHB mergers are not ideal to test standard cosmological models such as Λ CDM or to place constraints on late-time dark energy (DE) (Tamanini et al. 2016; Tamanini 2017; Belgacem et al. 2019b), they can efficiently be used to probe deviations from Λ CDM at earlier cosmological epochs, specifically in the interval $3 \lesssim z \lesssim 10$ (Caprini & Tamanini 2016; Cai et al. 2017; Belgacem et al. 2019b; Speri et al. 2020).

At lower redshift, LISA will provide other GW sources that can be used as standard sirens. SOBHs will be mainly detected at redshifts $z \lesssim 0.1$, while EMRIs might be observed up to $z \sim 4$, with a broad peak around $z \approx 1$. Unfortunately, the most widely accepted formation channels for these types of sources do not predict associated EM counterparts (see, e.g., Belczynski et al. 2002; Amaro-Seoane et al. 2007; Rodriguez et al. 2016), implying that statistical matching with galaxy catalogs will be necessary in order to extract cosmological information from them. A few investigations have al-

readily assessed the potential of LISA to test the Hubble law by statistically matching the sky localisation region of SOBHBs with galaxy catalogs (Kyutoku & Seto 2017; Del Pozzo et al. 2018). It was found that constraints on the Hubble constant can reach at best a few %, though uncertainties on the expected sensitivity of LISA at high-frequency could well undermine this result (Moore et al. 2019). No thorough investigation has so far been performed considering EMRIs as possible standard siren sources. The only analysis that can be found in the literature (MacLeod & Hogan 2008) suggests that ~ 20 EMRIs detected at $z \sim 0.5$ could be used to constrain the Hubble constant at the 1% level. However, beside considering an old configuration of LISA, this study was highly idealized: the authors assumed only linear cosmic expansion neglecting the acceleration of the Universe, they employed a simplified statistical framework and did not perform parameter estimation over the GW signals, using approximate relations only. A complete cosmological investigation with LISA EMRIs is currently missing in the literature.

1.3 Outline

The scope of the present investigation is to provide an in-depth and up-to-date analysis of the prospects of using EMRIs detected by LISA as GW standard sirens. As already stressed and cited (above), the standard capture scenario to generate EMRIs does not predict observable EM counterparts. We thus employ a statistical method, assigning to each galaxy within the LISA 3D error volume a probability of being the host of the EMRI within a Bayesian framework. We will start by presenting the statistical methodology that we use to infer cosmological parameters by cross-matching GW sky localisation regions with galaxy catalogs (Sec. 2). We will then describe the catalogs of EMRIs observed by LISA that we will use in our analysis (Sec. 3), and outline the procedure for matching sky localisations with galaxy catalogs (Sec. 4). We will subsequently present the results of our investigation (Sec. 5), discuss them (Sec. 6), and finally we will draw our conclusions (Sec. 7).

2 GRAVITATIONAL-WAVE COSMOLOGY WITHOUT EM COUNTERPARTS

To infer the value of the cosmological parameters, we operate within the framework of Bayesian inference (Jaynes 2003). The starting point of our analysis is: i) a list of GW EMRI observations D , combined with ii) a catalog of galaxies that are potential hosts of each individual EMRI. We will describe the EMRI and galaxy catalogs in Sec. 3 and 4, respectively. Here we provide the mathematical details of the analysis, given those populations.

Our inference model is constructed starting from Del Pozzo (2012), with updates from Del Pozzo et al. (2017); Del Pozzo et al. (2018) to account for the uncertainty on the redshift of potential counterparts to GW events. We will, however, use a different notation from both references. In the Bayesian framework, all knowledge about the cosmological parameters we are interested in, $\Omega \equiv \{H_0, \Omega_m, w_0, w_a\}$, is summarised by the *posterior* probability distribution,

$$p(\Omega | D \mathcal{H} I) = p(\Omega | \mathcal{H} I) \frac{p(D | \Omega \mathcal{H} I)}{p(D | \mathcal{H} I)}, \quad (2.1)$$

where \mathcal{H} is the cosmological model, that defines the relation between redshift, distance, and cosmological parameters, I represents all background information relevant for the inference of Ω , and $D \equiv \{D_1, \dots, D_N\}$ is the set of GW observations, with D_i the data from the i 'th EMRI event. In a realistic setting, this data would be the strain time series corresponding to the event observed by the LISA detector, so different EMRI events can overlap in time and therefore share this time series. For the current analysis, we will however represent the observed data instead by $D_i = \{\hat{d}_L, \hat{\theta}_{gw}, \hat{\phi}_{gw}\}_i$, that is, point estimates of the event luminosity distance and sky position in ecliptic coordinates, with associated uncertainties, and we assume these estimates for each event are independent.

The terms on the right hand side of Eq. (2.1) are the *prior* probability distribution $p(\Omega | \mathcal{H} I)$, the *likelihood* function $p(D | \Omega \mathcal{H} I)$, and the *evidence* $p(D | \mathcal{H} I)$. Since we are not interested in doing model selection, at this stage the evidence is considered as a normalization constant for the posterior, so we will only define the prior and the likelihood, eventually renormalizing Eq. (2.1) at the end of the computation.

To start with, we formalize our prior information and cosmological model in the following propositions:

I: “A GW event can be hosted by only one galaxy. Not all the galaxies within the comoving volume are visible. Observed and non-observed galaxies obey the same cosmology. Galaxies are highly clustered with each other. The host galaxy catalog is considered to be complete”.

H: “The cosmological model obeys a flat ($k = 0$) Friedmann-Lemaître-Robertson-Walker (FLRW) metric (Weinberg 1972) with negligible energy contribution from radiation ($\Omega_r = 0$), predicting that the luminosity distance d_L is a function of the redshift z_{gw} and of the cosmological parameters Ω : $d_L = d(\Omega, z_{gw})$ (Peebles 1993; Hogg 1999; Weinberg 2008)”.

With assumption *I*, we are stating that the catalogue of galaxies is representative of the whole galaxy distribution, and it is equivalent to assuming that the EMRI lies in one of the galaxies in the catalogue. EMRIs hosted in galaxies not in the catalogue are assumed to be sufficiently close to galaxies in the catalogue that this assumption holds. The likelihood for every event is determined by the same catalogue of galaxies, but for ease of computation we only use a subset of possible galaxies to analyse each event. Possible hosts are any galaxies that lie within the 2σ credible region for the GW direction and distance, for at least one value of the cosmological parameters within the prior range. This selects a 3D co-moving volume in direction and redshift (referred to as an “error box”) containing $N_{g,i}$ possible hosts, each with sky position (θ_j, ϕ_j) and redshift z_j .

We will further specify our model \mathcal{H} investigating two different *cosmological scenarios*, that will define the subset of Ω which we will be interested in. In this work we consider two models:

(i) A Λ CDM scenario characterized by a parameter space (h, Ω_m) to be explored ($h \equiv H_0/100 \text{ km}^{-1} \text{ s Mpc}$, while Ω_Λ is determined through the boundary condition $\Omega_m + \Omega_\Lambda = 1$, which holds assuming spatial flatness). We consider a prior range $h \in [0.6, 86]$ and $\Omega_m \in [0.04, 0.5]$, with fiducial values

dictated by the Millennium run (Springel et al. 2005) (i.e. $h = 0.73$, $\Omega_m = 0.25$, $\Omega_\Lambda = 0.75$ ¹).

(ii) A dark energy (DE) scenario in which we assume $(h, \Omega_m, \Omega_\Lambda)$ to be pre-determined by other probes at the values of the Millennium run (Springel et al. 2005) and we search on the parameters (w_0, w_a) defining the DE equation of state (Linder 2003) $w(z) = w_0 + z/(1+z)w_a$. We draw from the prior range $w_0 \in [-3, -0.3]$, $w_a \in [-1, 1]$, with fiducial values $w_0 = -1$ and $w_a = 0$, corresponding to the cosmological constant Λ .

We assume each GW event is statistically independent from any other event. Astrophysically, every EMRI event occurs independently of the others since we will most likely never observe multiple EMRIs occurring in the same galaxy. There is, however, coupling between parameter estimation for the events because they will be overlapping in the LISA data set and hence subject to the same noise fluctuations. We expect the observed EMRIs to be essentially orthogonal to each other and therefore they will be susceptible to different components of the instrumental noise. So, mutual independence should be a good approximation. Thus, the likelihood function simplifies to the product of the likelihoods for each individual observation,

$$p(D | \Omega \mathcal{H} I) = \prod_{i=1}^N p(D_i | \Omega \mathcal{H} I). \quad (2.2)$$

Therefore, we only need to determine how to construct the likelihood for a single GW event, whose main ingredients we introduce in what follows. The relevant quantities for the inference of the cosmological parameters are the GW luminosity distance d_L (which is directly measured) and redshift z_{gw} (which is inferred through the cosmology priors)². Assuming that the correlation between sky localisation and other parameters is negligible, a multiple application of marginalisation and product rule lead to a single-event likelihood written as:

$$p(D_i | \Omega \mathcal{H} I) = \int dd_L dz_{gw} p(d_L | z_{gw} \Omega \mathcal{H} I) \cdot p(z_{gw} | \Omega \mathcal{H} I) p(D_i | d_L z_{gw} \Omega \mathcal{H} I), \quad (2.3)$$

where the integrals on d_L and z_{gw} go, in principle, from 0 to ∞ .

Once we know the redshift z_{gw} along with the values of the cosmological parameters (assuming their priors) and the cosmological model \mathcal{H} , we have (Del Pozzo 2012)

$$p(d_L | z_{gw} \Omega \mathcal{H} I) = \delta(d_L - d(\Omega, z_{gw})), \quad (2.4)$$

where under our background information I the function $d(\Omega, z_{gw})$ is given by (Hogg 1999):

$$d(\Omega, z_{gw}) = \frac{c(1+z_{gw})}{H_0} \int_0^{z_{gw}} \frac{dz'}{E(z')}, \quad (2.5)$$

with the function $E(z')$ given by:

$$E(z') = \sqrt{\Omega_m(1+z')^3 + \Omega_\Lambda g(z', w_0, w_a)}, \quad (2.6)$$

¹ Although these values of cosmological parameters are outdated, this is irrelevant for the purpose of testing our statistical methodology for constraining cosmology with standard sirens.

² To ease the notation, when there is no possibility of confusion we will be dropping the EMRI index i , as in the quantities d_L and z_{gw} .

where

$$g(z', w_0, w_a) = (1+z')^{3(1+w_0+w_a)} e^{-3\frac{w_0 z'}{1+z'}}, \quad (2.7)$$

and as already noted in the definition of \mathcal{H} we neglect the radiation density Ω_r , since $\Omega_r \ll \Omega_m, \Omega_\Lambda$.

Our prior information I prescribes that each GW is hosted by a galaxy; we are assuming that the distribution of z_{gw} is the same as that of the host galaxies. Thus we have

$$p(z_{gw} | \Omega \mathcal{H} I) \propto \sum_{j=1}^{N_{g,i}} w_j \exp \left\{ -\frac{1}{2} \left(\frac{z_j - z_{gw}}{\sigma_{z_j}} \right)^2 \right\}, \quad (2.8)$$

where we have introduced relative weights w_j for each galaxy host (this is different from MacLeod & Hogan (2008), where all galaxies within the error-box were assumed to be equally likely hosts). The uncertainty σ_{z_j} includes the contribution due to the peculiar velocity of the galaxy j , and the sum goes over all the possible galaxy hosts $N_{g,i}$ associated to the i -th event. As in Del Pozzo et al. (2018), we assume a typical redshift error of $\Delta z = 0.0015$ corresponding to a rms peculiar velocity of 500 km/s. We assign the relative weights w_j to each galaxy from the marginal distribution over the sky position angles computed in each galaxy, as defined by Eq. (4.4) and described in detail in Sec. 4. We note that the product $w_j p(D_i | d_L z_{gw} \Omega \mathcal{H} I)$ should equal the full likelihood evaluated at the sky location of the galaxy, which is assumed to be known perfectly. Therefore w_j depends on both d_L and on the observed data, which is not explicit in the preceding equation. Here we approximate w_j as a constant, given by marginalising the posterior over luminosity distance at the sky location of the galaxy, and similarly assume that the projected distance posterior does not depend on sky location. This is equivalent to assuming that the joint likelihood on sky location and distance factorises. Assigning weights in this way we are treating all galaxies as *a priori* equally likely hosts of observed EMRIs, but weightings can also be used to reflect the relative probability that a galaxy is the GW host, for example based on the galaxy luminosity. The expressions look the same, but the weights are then proportional to the product of the sky-location contribution and the pre-assigned weight.

The remaining term in Eq. (2.3) is:

$$p(D_i | d_L z_{gw} \Omega \mathcal{H} I) = p(D_i | d_L z_{gw} \mathcal{H} I), \quad (2.9)$$

which is the likelihood for the GW data. As described above, we approximate the GW data as point estimates of the relevant parameters, with Gaussian uncertainties, and so replace this term by a quasi-likelihood specified as a multivariate Gaussian, with covariances estimated from the Fisher Matrix (see Sec. 3), and also accounting for the weak lensing uncertainty $\sigma_{WL,i}$ as modelled in Tamanini et al. (2016), see Eq. (7.3) therein (see also Hirata et al. (2010); Cusin & Tamanini (2020)).

The final expression for the single-event likelihood is:

$$p(D_i | \Omega \mathcal{H} I) = \frac{1}{2\pi} \int dz_{gw,i} \sum_{j=1}^{N_{g,i}} \frac{w_j}{\sigma_{z_j} \sqrt{\sigma_{d_L,i}^2 + \sigma_{WL,i}^2}} \cdot \exp \left\{ -\frac{1}{2} \left[\frac{(z_j - z_{gw,i})^2}{\sigma_{z_j}^2} + \frac{(\hat{d}_{L,i} - d(\Omega, z_{gw,i}))^2}{\sigma_{d_L,i}^2 + \sigma_{WL,i}^2} \right] \right\}. \quad (2.10)$$

As mentioned above, $(\hat{d}_L, \cos \hat{\theta}_{gw}, \hat{\phi}_{gw})_i$ are the point estimates of the parameters, and the uncertainties, $\sigma_{d_L,i}$ etc.,

are determined from the Fisher matrix evaluated at the true parameter values. We have also assumed that the joint likelihood on sky location and distance factorises as mentioned above.

Until now, we have assumed that all EMRIs that occur during the observation period are included in the analysis. However, in practice our detectors have limited sensitivity and we will only include in the analysis the events that are successfully “detected” by our analysis pipelines, that is, those for which some ranking statistic computed from the data is above some predetermined threshold. Whether or not an event is found is a property of the data only, and assuming that the total number of events does not convey any information the selection effects can be accounted for (Mandel et al. 2019) by replacing $p(D_i | \Omega \mathcal{H} I)$ with

$$\hat{p}(D_i | \Omega \mathcal{H} I) = \frac{p(D_i | \Omega \mathcal{H} I)}{\int_{D > \text{threshold}} p(D_i | \Omega \mathcal{H} I) dD}, \quad (2.11)$$

where the integral is over all data sets that would give detection statistics above the threshold and hence be included in a cosmological analysis. In practice, the detectability of a GW event is primarily determined by its signal-to-noise ratio (SNR), which in turn is primarily determined by luminosity distance. In the quasi-likelihood model used here we could thus use a cut on the observed luminosity distance, $\hat{d}_L > d_{\text{crit}}$, as a proxy to represent selection effects. As the choice of cosmological parameters is varied, the centres of the Gaussian distributions in the likelihood will move inside and outside the selection cut. Although a small number of the Gaussian distributions span the boundary, the selection function normalisation can be well approximated by the number of host galaxies that remain consistent with being inside the luminosity distance horizon. In our analysis we found that this number changed by very little over the range of cosmological parameters consistent with our priors, and so we approximated this normalisation as a constant. However, we also verified that our results were insensitive to that approximation.

This completes the definition of our inference problem. We will further discuss the limitations and approximations of our framework in Sec. 3, 7, and Appendix A.

We explore the posterior distribution in Eq. (2.1) using `cpnest` (Veitch et al. 2020), a parallel nested sampling algorithm implemented in `Python`. We make use of the `LALCosmologyCalculator` library in `LAL` (LIGO Scientific Collaboration 2018) and of `numpy` (van der Walt, Stéfan et al. 2011), `scipy` (Virtanen et al. 2020), `matplotlib` (Hunter 2007), `cython` (Behnel et al. 2011), and the plotting utilities in Foreman-Mackey (2016). The inference code utilised in this paper is publicly available in Del Pozzo & Laghi (2020).

3 DETECTING EMRIS WITH LISA

3.1 Expected properties of the EMRI population

The catalogs of EMRIs detected by LISA which we use in our work are based on the analysis of Babak et al. (2017). Here we review how these catalogs have been constructed and outline their main properties. For more information the reader is referred to Babak et al. (2017) (see also Babak et al. 2015; Gair et al. 2017).

Because of their small mass-ratio, EMRIs present a very

slow inspiral, producing many cycles (10^4 - 10^5) within LISA’s sensitivity band. For the same reason, the detailed dynamics of these systems strongly differ from equal mass compact binaries. The motion of the stellar BH can in fact be approximated by geodesics of the massive BH, with small but relevant corrections due to its own *self-force* (for a review, see, e.g., Poisson et al. 2011). Unfortunately ongoing perturbative calculations, which exploit the extreme difference in masses of these systems, are not yet at the level needed to calculate the emitted GW waveform with the accuracy required for LISA observations (i.e., by keeping track of the phase over the entire inspiral). The analysis of Babak et al. (2017) considered thus an approximate analytical model to estimate the waveform generated by EMRIs, the so-called *analytical kludge* model (Barack & Cutler 2004). This approximation was used to produce waveforms under two different endpoints for the dynamical evolution of the system: until the Schwarzschild last stable orbit (more pessimistic) and until the Kerr last stable orbit (more optimistic). In our study we will only consider results obtained with the latter of these assumptions (denoted “AKK” in Babak et al. (2017)).

In order to simulate the response of LISA and estimate the uncertainty on the parameters of the GW waveform generated by an EMRI, the investigation made by Babak et al. (2017) employed a *Fisher matrix* approach, useful to quickly analyse many different events. Their results were obtained setting a threshold of SNR=20 for LISA detection and considering a LISA sensitivity curve as specified by the LISA White Paper written in response to ESA’s call for the L3 mission slot (Amaro-Seoane et al. 2017). The results of this paper will thus be based on that sensitivity curve as well. We note here that although the laser stability requirement has been relaxed since, this has an impact to the detector performance at $f > 0.01$ Hz, which is unlikely to significantly affect EMRIs.

As pointed out by Babak et al. (2017), the main uncertainties in forecasting how many EMRIs will be detected by LISA are of astrophysical origin. The expected rate of EMRIs depends in fact on several astrophysical assumptions, including:

- the MBH population in the accessible LISA mass range (from $10^4 M_\odot$ to $10^7 M_\odot$), the redshift evolution of their mass function, and their spin distribution;
- the fraction of MBHs hosted in dense stellar cusps, which constitute the nurseries for the formation of EMRIs;
- the EMRI rate per individual MBH, and the mass and eccentricity distribution of the inspiralling compact object.

In the analysis of Babak et al. (2017), by considering 12 different combinations of prescriptions for the assumptions listed above, the authors produced forecasts for 12 different scenarios. Among all 12 scenarios, the rates for LISA detections using the AKK model always fall in between 10% to 20% of the total EMRI population, and in absolute numbers they span a range from one to a few thousands of detections per year, with similar properties between the 12 EMRI populations corresponding to each scenario. In particular, irrespective of the astrophysical scenario, EMRIs detected by LISA will come from MBHs with masses from $3 \times 10^4 M_\odot$ to $3 \times 10^6 M_\odot$, over a redshift range that is broadly peaked between $0.5 < z < 2$, with tails usually reaching $z \sim 5$.

In this paper we concentrate on three EMRI models, span-

ning the bulk of the uncertainty range in the number of EMRI detections, corresponding to models M1, M5, and M6 of Babak et al. (2017) (cf. their Table I):

- our *fiducial* model is based on M1, which depends on rather standard assumptions for the MBH mass function and EMRI properties;
- our *pessimistic* model is based on M5, featuring a downturn of the MBH mass function at $M < 10^7 M_\odot$, which strongly suppresses the occurrence of EMRIs;
- our *optimistic* model is based on M6. This is similar to M1, but ignores the effect of cusp erosion following galaxy mergers, which boosts the EMRI rate by a factor of ~ 2 .

Although there are more optimistic and pessimistic EMRI models in Babak et al. (2017), those are based on rather ad-hoc assumptions, in particular about the relative occurrence of plunges and EMRIs. In fact, when a BH is scattered on an almost radial orbit, chances are that it plunges directly onto the MBH, rather than forming an EMRI. N-body simulations of realistic stellar cusps suggest that there are few such direct plunges for each EMRI (Merritt 2015). The models listed above assume a rather conservative ratio of 10 plunges per EMRI. We note that, as the number of useful events N increases further, one can reasonably expect that our final results, i.e., the constraints that we obtain on the cosmological parameters, scale approximately with \sqrt{N} (cf. Sec. 6).

Each EMRI model predicts different events, thus it can be thought of as a different data set D . For each data set, we will adopt the cosmological *inference* models \mathcal{H} detailed in Sec. 2.

3.2 EMRI selection

Because of the complexity of their waveform, a relatively large SNR is required for EMRI detection. As already noted, the threshold is customarily set to SNR= 20, which results in several hundreds-to-thousands of EMRIs detected over 10 years for the three models considered here, as reported in the second column of Table 1.

Although such an abundance of sources is in principle an asset for our analysis, a number of considerations have to be made when designing a practical implementation of the algorithm. In fact, the speed of computation of Eq. (2.10) strongly depends on the dimension of the EMRI catalog. This imposes a limitation on the number of events that we can analyse in a reasonable time. At the current state, catalogs of hundreds of EMRIs are prohibitive, as they require $\mathcal{O}(\text{months})$ to be analysed (especially with the limited computational resources available to this project). This is due to two main reasons: i) the increase in the number of single-event likelihoods that have to be computed and ii) the large number of hosts $N_{g,i}$ that some events will have, resulting in a significant slowdown of each likelihood evaluation.

Point ii) suggests some useful guidelines for devising an EMRI selection strategy. As shown for our fiducial model M1 by the green histograms in Fig. 1, the majority of EMRIs will be observed at $z > 1$. As z increases, so do the errors in the estimate of their parameters and the associated LISA 3D localisation volume ΔV . When ΔV gets too large, the number of candidate galaxy hosts within it, $N_{g,i}$, increases dramatically, washing out the information enclosed in their

clustering properties, which is of paramount importance for the success of our technique.

This suggests two approaches to the selection of the events. One possible strategy, that we indicate as our first selection procedure, is to pre-select EMRIs with a good LISA 3D sky location estimate up to some maximum redshift. Following this route, we select only EMRIs with sky localisation error $\Delta\Omega = 2\pi \sin\theta_{gw} \sqrt{\Delta\theta_{gw} \Delta\phi_{gw} - (\Sigma^{\theta_{gw}\phi_{gw}})^2} < 2 \text{ deg}^2$, being θ_{gw} and ϕ_{gw} the latitude and longitude of the source in ecliptic coordinates and $\Sigma^{\theta_{gw}\phi_{gw}}$ their correlation extracted by the 3D marginalised correlation matrix of the Fisher analysis. We further require a relative error in the luminosity distance determination: $\Delta d_L/d_L < 0.1$. Moreover, an additional constraint stems from the fact that, in order to keep the number of candidate hosts within a maximum of a few thousands, we use the mock Millennium sky to select galaxy hosts up to $z = 1$. This implies that we keep only those events for which the furthest candidate host compatible with the adopted priors in the cosmological parameters is at $z < 1$. In fact, the cut in our galaxy catalogue at $z = 1$ implies that we discard some of the events at $z \gtrsim 0.7$ for which the furthest candidate compatible with the range of cosmological priors lies at $z > 1$. We note that this cut depends on the adopted cosmological model, and yields a slightly different selection for the Λ CDM and the DE cases. The number of events selected according to this procedure for the Λ CDM and DE models are given in the third and fourth columns of Tab. 1.

Although legitimate, this selection procedure is rather contrived, as it combines a number of ad hoc choices in the selection process. A more straightforward alternative is to observationally select events by simply imposing an SNR threshold. This is illustrated for our fiducial model M1 by the orange histograms in Fig. 1. In fact, high-SNR events are expected to be those yielding a better determination of the EMRI parameters (distance and sky location) and are also preferentially at relatively low redshifts: this is evident looking at panels a), b), and c). This combination of properties automatically limits the number of candidate hosts – see panel d) – making them the most informative sources for our analysis. We found empirically that imposing a LISA SNR threshold SNR>100 returns only events with $N_g \lesssim 1000$ out to $z \approx 0.75$, as shown in panel a). The imposition of an SNR threshold, together with the cut at $z = 1$ dictated by the galaxy catalog that is subsequently used (which is sensitive to the assumed cosmological model), yields the number of events reported in the fifth and sixth columns of Table 1. The first selection procedure produces a larger number of EMRIs (see the third and fourth columns of Tab. 1), among which we found several events with relatively low SNR and $N_g > 1000$. Tests performed on these samples produced reasonable estimates of the cosmological parameters, but also displayed some undesired behaviour related to the limitations of our likelihood implementation, which became particularly evident when the majority of the selected events had low SNR. These aspects will be further discussed later on and in particular in Appendix A2. Therefore, unless otherwise stated, results reported in this work follow the SNR>100 selection criterion.

MODEL	detected (SNR>20)	$\Delta d_L/d_L < 0.1$ & $\Delta\Omega < 2 \text{ deg}^2$		SNR>100	
		ΛCDM	DE	ΛCDM	DE
M1 (<i>fid</i>)	2941	180	202	32	33
M5 (<i>peas</i>)	472	20	21	6	6
M6 (<i>opt</i>)	4788	260	281	72	73

Table 1. Number N of EMRIs observed by LISA in 10 years of operation. For each model (first column) we report the total number of detections (second column), those additionally localized within $\Delta d_L/d_L < 0.1$ & $\Delta\Omega < 2 \text{ deg}^2$, referred in the text as first selection procedure, condition that depends on the chosen cosmological model (third and fourth columns), and those with SNR> 100 (fifth and sixth columns). In both selection procedures we require no potential host galaxy above $z = 1$.

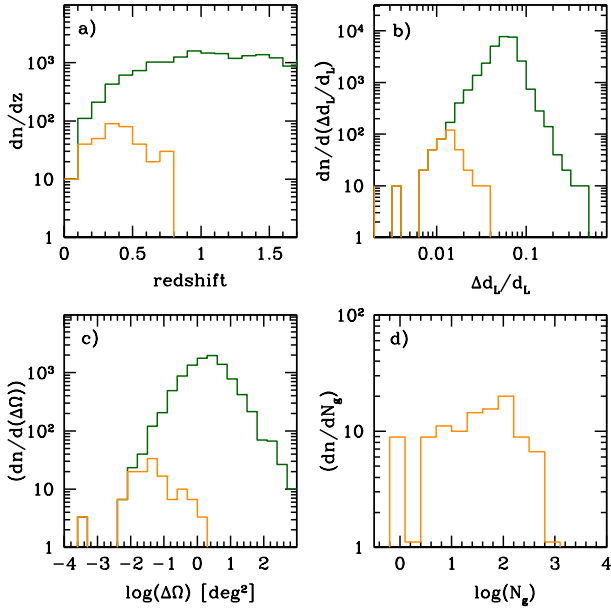


Figure 1. Relevant properties of the events used in the analysis, taken from model M1. Green histograms in panels a), b), and c) show the redshift, distance error and sky location error distributions for the entire population of EMRIs detected by LISA in 10 years of operations. For the purpose of our analysis we select only systems with SNR>100. Those are shown in panels a), b) and c) as orange histograms. For these latter systems, panel d) shows the distribution of the number of candidate hosts within the 3D 2σ error volume, averaged over the three realisations of M1 (see section 4 for details).

3.3 Caveats and limitations

Selecting only high-SNR events allows us to avoid some limitations and complications related to a number of features that are not (as of yet) modeled in our pipeline.

First, we do not fold into our analysis the possible incompleteness of the galaxy catalogs. As we will see in the next section, galaxy catalogs are constructed starting from the flux-limited mock sky realizations of Henriques et al. (2012) based on the Millennium simulation (Springel et al. 2005). As reported in our prior information I , we are going to assume that each GW event is hosted by one of the galaxies within the reconstructed co-moving volume, regardless of whether the galaxy is luminous enough to be listed in the mock catalog or not. We assume that EMRIs are hosted by galaxies

with stellar mass $M_* > 10^{10} M_\odot$, for which we found that the Millennium sky maps are complete out to $z \approx 0.5$. The large majority of events we select are at $z < 0.5$ (see Fig. 1). Thus, the mass threshold imposed in constructing our catalogs should not yield a relevant Malmquist bias.

Another possible source of bias may be given by evolution effects, i.e., the fact that the host galaxy population evolves within the allowed redshift range for the host. Also in this case, the high-SNR cut preferentially selects low- z events with precise distance measurement, thus limiting the redshift of candidate hosts within a window that is narrow enough that evolution effects can be neglected.

Last, this selection leaves out events with a large number of candidate hosts, that are not expected to make significant contributions to the inference. When testing our pipeline, we found that less informative, low SNR events not only do not add much information to the inference, but tend to produce a bias in the estimate of cosmological parameters, especially those that are weakly constrained. More details about this bias are given in App. A2. Its origin is still under investigation and will be a subject of future work. Although some level of bias can also be seen in our main results (see the posterior for w_a in Fig. 5), it does not affect the estimate of the constrained parameters.

4 STATISTICAL MATCHING WITH GALAXY CATALOGUES

Having selected a sample of “useful” EMRIs, the next step is to simulate realistic distributions for candidate host galaxies. To this end, we closely follow the procedure outlined in Del Pozzo et al. (2018), with some improvements, as we now summarize.

We use the simulated observed sky of Henriques et al. (2012) constructed from the Millennium run (Springel et al. 2005) which, critically for this work, reproduces the observed clustering properties of galaxies. The catalog, that we use up to $z = 1$, lists a number of galaxy properties, including mass, observed redshift z_{obs} and cosmological redshift z_{cos} , which differ due to the extra redshift contribution imprinted on the former by galaxy peculiar velocities, $\Delta z_{v_p} \equiv z_{\text{obs}} - z_{\text{cos}} = (1 + z_{\text{cos}})v_p/c$ ($v_p \ll c$), v_p being the peculiar velocity.

The simulation assumes a ΛCDM cosmology with $h = 0.73$, $\Omega_m = 0.25$, $\Omega_\Lambda = 0.75$, which will thus correspond to the fiducial values for our analysis as well. Although those parameters are outdated (Ade et al. 2016), this is irrelevant for the illustrative purpose of our analysis.

The LISA 3D error volume is approximated by a multivariate Gaussian distribution:

$$p(D_i|\Theta) = \frac{1}{\sqrt{(2\pi)^3|\Sigma|}} \exp\left\{-\frac{1}{2}\tilde{\Theta}_i^T \Sigma^{-1}\tilde{\Theta}_i\right\}, \quad (4.1)$$

where $\tilde{\Theta}_i \equiv \Theta - \hat{\Theta}_i(D_i)$, and the 3D parameter vector includes the source luminosity distance, the cosine of its declination and its right ascension, i.e. $\Theta = (d_L, \cos\theta_{gw}, \phi_{gw})$. The vector $\hat{\Theta}_i(D_i) = (\hat{d}_L, \cos\hat{\theta}_{gw}, \hat{\phi}_{gw})_i$ defines the best measured values of the parameters (i.e. the center of the Gaussian), which depend on the observed data. We assume that these are the only quantities measured from the data

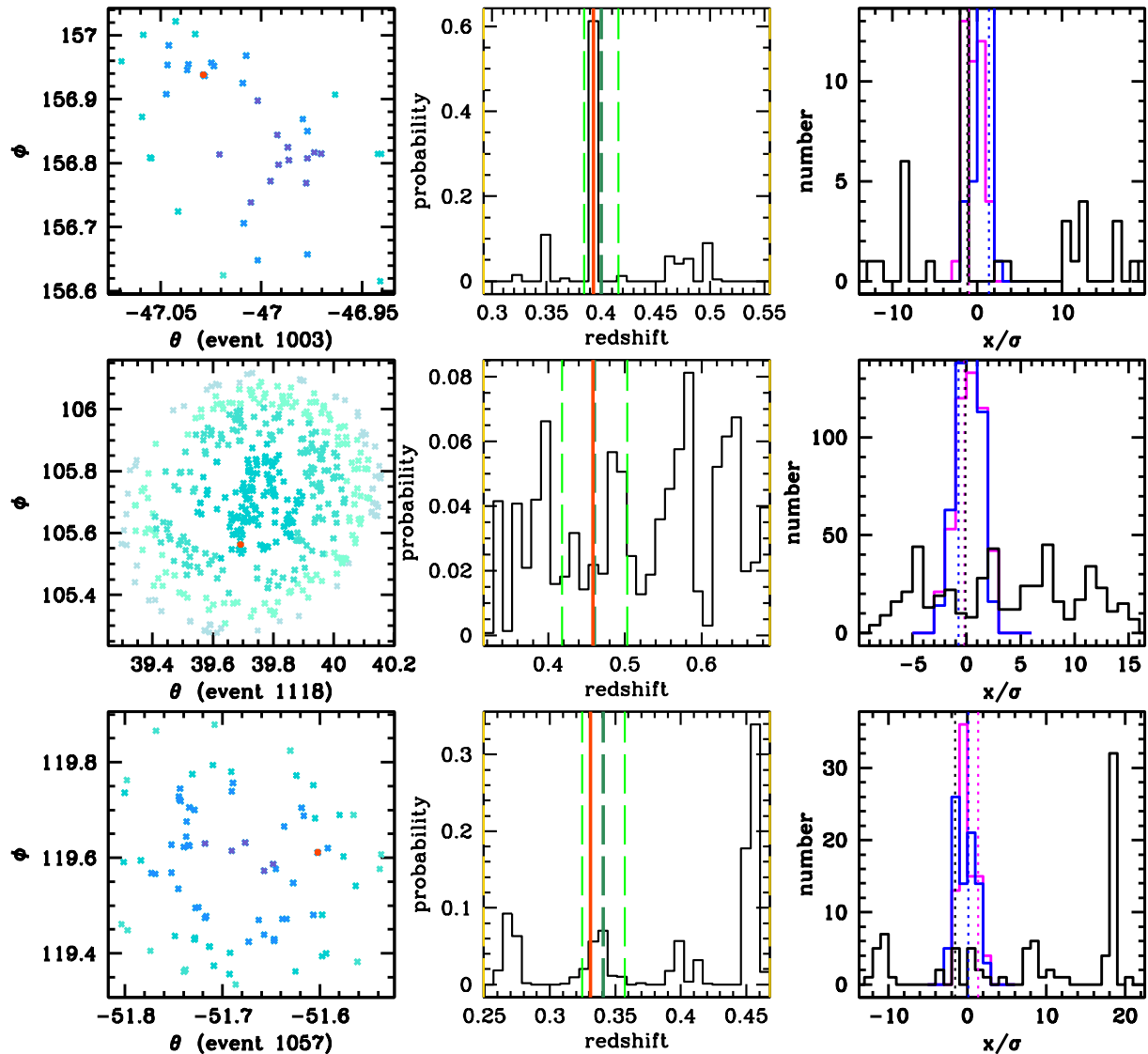


Figure 2. Each row of three panels represents the properties of the galaxy host candidates for a selected event extracted from model M1_2. The left panels show a scatter plot of candidate hosts in the $\theta - \phi$ plane; the shades of blue, from lighter to darker, are in order of increasing probability of hosting the EMRI, according to the LISA sky localisation in the celestial sphere. The red dot is the true host, which is generally not in the center of the sky localisation area. The middle panels show the redshift probability distributions of the candidates hosts (black histograms), i.e. the distribution of galaxies, each weighted by its own probability of being the host according to its sky location. The dark-green vertical line is the fiducial redshift measured by LISA and the two light-green lines bracket the 2σ redshift interval implied by the LISA measurement error in the luminosity distance alone, assuming the true cosmology. The vertical yellow lines at the edges bracket the redshift interval uncertainty when the cosmological parameters are allowed to vary within the assumed priors (here Λ CDM is assumed). The redshift of the true host is marked by the vertical red line. Finally, the right panels represent the deviation of the x quantities, θ (magenta), ϕ (blue), and d_L (black), from the best measured value x_t normalized to the respective LISA measurement errors. The vertical dotted lines are the values of the true hosts. Note that the d_L of candidate galaxies can deviate from the best measured value by many σ , as measured by LISA, due to the prior uncertainties on the cosmological parameters.

and drop the explicit dependence on D_i in subsequent equations. The measurement uncertainties and correlations are described by the 3D correlation matrix Σ , which we extract from the full fifteen-dimensional correlation matrix of the EMRI parameter estimation carried out in Babak et al. (2017), by marginalising over all other parameters. In practice, these posterior widths could depend on both the true parameters of the EMRI and on the particular realisation of the noise in the LISA data. For the well localised EMRIs we select in our analysis, these dependencies are likely

to be weak. Therefore, we assume here that the measurement uncertainties are fixed and known for each event. The first task is to associate a true host to each EMRI, consistent with these errors. To pick the true host, we first leave $(\cos\theta_{gw}, \phi_{gw})$ free, we compute d_L for all galaxies in the Millennium sky and then evaluate the marginalised luminosity

distance likelihood given by:

$$p(d_L|D_i) = \frac{\exp\left\{-\frac{1}{2} \frac{(d_L - \hat{d}_L(D_i))^2}{\sigma_{d_L}^2 + \sigma_{WL}^2}\right\}}{\sqrt{2\pi(\sigma_{d_L}^2 + \sigma_{WL}^2)}}, \quad (4.2)$$

at the luminosity distance of each galaxy. A host galaxy is then randomly selected from the catalogue with probability proportional to this marginalised likelihood. Leaving $\cos\theta_{gw}$ and ϕ_{gw} free ensures that the pool of galaxies from which we select the host is large enough that the selection process reflects the clustering properties of galaxies at the measured distance of the event, while the luminosity distance of the selected host, $d_{L,H}$, is fully consistent with the LISA measurement error distribution by construction³. The host is described by the parameter vector $\Theta_H = (d_{L,H}, \cos\theta_{gw,H}, \phi_{gw,H})$. We then fix $d_L = d_{L,H}$ in Eq. (4.1) and draw a pair $(\cos\theta'_{gw}, \phi'_{gw})$ from the 2D slice of the $p(\Theta)$ distribution. The LISA error volume is then re-positioned so that $\cos\hat{\theta}_{gw} = \cos\theta_{gw,H} - \cos\theta'_{gw}$ and $\hat{\phi}_{gw} = \phi_{gw,H} - \phi'_{gw}$, which, together with \hat{d}_L , redefine the vector $\hat{\Theta}$. This procedure ensures that the true host is drawn consistently with the clustering properties of galaxies, and that its position relative to the best LISA measured values $\hat{\Theta}$ is drawn according to the 3D volume error described by Eq. (4.1). Now that we have identified a true host and have placed the LISA error volume appropriately with respect to it, we can truncate the galaxy catalogue by sub-selecting all galaxies g_j so that $z_{\text{obs},j}$ is consistent with the LISA measurement, including the uncertainties in the cosmological parameters. In principle, the sum over galaxies in Eq. (2.10) includes all galaxies in the catalogue up to $z = 1$. However, galaxies lying in the tail of the Gaussian distribution for all choices of cosmological parameters in the prior are exponentially suppressed in the likelihood. Therefore, it is more efficient to remove these from the catalogue. We retain those galaxies for which $z^- < z_{\text{obs},j} < z^+$, where z^- and z^+ are implicitly given by

$$d_L \pm 2\sigma_{d_L} = (1 + z^\pm)d_0 \int_0^{z^\pm} \frac{dz'}{F^\pm(z')}. \quad (4.3)$$

Here $d_0 = c/H_0$ is the Hubble distance, and $F(z) = hE(z)$, with $E(z)$ given in Eq. (2.6). $F^\pm(z)$ are the realizations of $F(z)$ that minimize and maximize the $d_L - z$ conversion within the assumed prior on the cosmological parameters. In practice Eq. (4.3) extends the Δz due to the 2σ error in the LISA measurement of d_L as much as allowed by the prior range on the cosmology. Finally, we need to keep in mind that due to peculiar velocities, the true cosmological redshift of each galaxy $z_{\text{cos}} \neq z_{\text{obs}}$, the difference between the two being Δz_{v_p} . We thus consider all galaxies with $z_{\text{obs}} \in [z^- - \Delta z_{v_p}^-, z^+ + \Delta z_{v_p}^+] \equiv \Delta z_{\text{tot}}$, where $\Delta z_{v_p}^-, \Delta z_{v_p}^+$ are simply Δz_{v_p} calculated at z^+, z^- , taking a characteristic

value of $v_p = 500 \text{ km s}^{-1}$, consistent with the standard deviation of the galaxy radial peculiar velocity distribution in the Millennium run (Springel et al. 2005).

We compute the weights w_j , appearing in Eq. (2.10), for each galaxy g_j within Δz_{tot} by marginalization of Eq. (4.1) over d_L (assuming uniform priors):

$$p(D_i|\cos\theta_j, \phi_j) = \int dd_L p(D_i|\Theta) \equiv w_j, \quad (4.4)$$

evaluated at $(\cos\theta_j, \phi_j)$ (where w_j must not be confused with the DE equation of state parameters w_0 and w_a). We do so for all galaxies falling into a sky region $\Delta\Omega_{\text{tot}}$ covering 2σ in the determination of $\cos\theta_{gw}$ and ϕ_{gw} . In practice we assign probabilities only based on the marginalized 2D sky location error, discarding further information included in the correlation of $\cos\theta_j$ and ϕ_j with d_L , which is a conservative assumption.

In summary, to assess the power of EMRI cosmology with LISA, we take three EMRI population models (M1, M5, and M6) from Babak et al. (2017), and we consider the two cosmology scenarios presented in Sec. 2: Λ CDM and DE. We select only EMRIs with $\text{SNR} > 100$ and associate to each event candidate hosts according to the procedure outlined above. For each model, the procedure of drawing the true host is repeated three times for each EMRI, so that we have three independent realizations of the galaxy host candidates, making our analysis robust with respect to statistical fluctuations in the galaxy distribution. We identify the independent realizations with a number following the name of the model, so that for model M1 we refer to M1_1, M1_2, and M1_3, the same being for M2 and M3. Finally, for each model, we investigate the impact of the mission lifetime by considering observation of EMRIs carried out in 4 years and 10 years of LISA operations.

An example of the outcome of the procedure outlined above is shown in Fig. 2, where we depict the properties of the selected galaxies within the error box $\Delta z_{\text{tot}} \times \Delta\Omega_{\text{tot}}$ for three EMRI events selected from model M1_2 under the assumption of a Λ CDM cosmology. The $\cos\theta - \phi$ correlation is visible in the left column panels, together with the clustering pattern of the galaxies in the sky. Note that the true host, which differently from MacLeod & Hogan (2008) we do not remove from the group of possible hosts and is marked by the red dot in the left panels and the red vertical line in the middle panels, is generally offset from the center of the error box by construction, which mimics a realistic situation. The various elements of the Δz_{tot} construction procedure can be appreciated in the central panels. The dark-green line marks the redshift corresponding to the best LISA distance measurement \hat{d}_L . The light-green lines bracket the Δz interval associated to the $2\sigma_{d_L}$ LISA error assuming the true (i.e., the Millennium run) cosmology. Finally, the yellow vertical lines bracket the interval Δz_{tot} , allowed by the prior range on the cosmological parameters and accounting for the errors due to peculiar velocities (which are generally subdominant, unless $z \ll 0.1$). The black histograms show the probability distributions of observed redshifts z_{obs} of the hosts in the error volume, where the probability of each individual host, w_j , is given by Eq. (4.4). As it should be, z_{obs} of the true host generally falls within the light-green lines, although this is not always strictly the case due to peculiar velocities. Finally, the right panels show sanity checks of the offset distri-

³ This is necessary because the EMRI catalogues used in this work were generated independently from the Millennium sky and its clustering properties. By allowing hosts to be selected within the whole $4/3\pi[(d_L + 2\sigma_{d_L})^3 - (d_L - 2\sigma_{d_L})^3]$, we ensure that at least the hosts probability reflects the clustering of the Millennium sky within that luminosity distance range.

butions of d_L , $\cos\theta$, and ϕ of the candidate hosts from the best-measured LISA values normalized to the respective measurement errors. The distributions for d_L (black histograms) extend to several σ due to the extra contribution allowed by varying the cosmological parameters within the prior range.

The top panels in Fig. 2 show a “very good EMRI”, for which the putative hosts cluster in the Δz range allowed by the true cosmology. The middle panels show a “non informative EMRI”, displaying an essentially flat (although very noisy) probability distribution of the hosts across the whole redshift range allowed by the cosmological prior. Finally, the bottom panels show a “bad EMRI”. In this latter case, the true host was picked in a small group of galaxies at $z \approx 0.35$, and by chance, in the same sky region, there is a much larger group of galaxies at $z \approx 0.45$, which skews the probability distribution of the hosts towards a cosmology different from the true, although still within the allowed prior. Events of this type are expected to be subdominant, and by stacking the posteriors of several events, the true cosmological parameters naturally emerge from the analysis, as we now show.

5 RESULTS

In this section we report the results of our investigation following the procedure described in the previous sections. We will first show the results for the Λ CDM scenario (Sec. 5.1) and then for the DE scenario (Sec. 5.2). For both scenarios we will consider 4 and 10 years of LISA observational time.

Now we describe how we obtain the posteriors reported in this section. Concerning the 4-year results, in order to accumulate enough statistical evidence to produce reliable results for a 4-year LISA mission, for each EMRI model we generated three 4-year catalogs from each of the three realisations of the 10-year catalog described in Sec. 3.2 and 4. The number of events in each 4-year catalog was obtained as follows: i) we randomly picked an integer $n_{4\text{yr}}$ by drawing from a Poisson distribution with mean equal to the 4/10 of the total number of events (see Table 1); ii) we then selected randomly $n_{4\text{yr}}$ events from the original catalog; iii) finally we applied the $\text{SNR} > 100$ cut to obtain the final set of EMRIs for the analysis. In this way we have a total of nine 4-year realisations per EMRI scenario (for example, in case of M1 we produced three realisations for M1_1, three for M1_2, and three for M1_3), each of which has been analysed separately. The posteriors for the cosmological parameters are then averaged (Del Pozzo et al. 2018) over the different realisations. This procedure is applied for both the Λ CDM and the DE cosmological inference models. Our results with 10 years of LISA mission lifetime for each EMRI scenario are simply obtained averaging the analysis of the three different catalog realisations (again, for M1 it will be the average of M1_1, M1_2, and M1_3). Such catalogs contain the exact number of EMRI events as given in the fifth and sixth columns of Table 1, depending on the assumed cosmological model.

For illustrative purposes in Fig. 3 we show a distance-redshift diagram as obtained from our analysis in the Λ CDM scenario for the M1_2 realisation only of our fiducial EMRI model, assuming 10 years of LISA mission lifetime. The credible regions pictured in Fig. 3 for each EMRI event are the posterior distributions resulting from combining the LISA distance measurement with the combined information col-

lected from all galaxies within the sky localisation region associated to that particular event. For each EMRI, to compare the prior range for z_{gw} with the corresponding range of galaxy host redshifts, in Fig. 3 we show the redshift z_j of each galaxy j within the sky localisation region of each EMRI event, displayed as w_j -weighted dots at a fixed luminosity distance set, for convenience, equal to \hat{d}_L . Note also how measurements at low redshift are more accurate due to the lower number of galaxies contained within the reconstructed volume, while events at higher redshift possess on average less constraining power due to the large number of galaxies within their sky error boxes.

5.1 Λ CDM

Results for the Λ CDM scenario are reported in Fig. 4 for our three selected EMRI models: M5 (pessimistic), M1 (fiducial), and M6 (optimistic). We show corner-plots with 2D posteriors in the h - Ω_m parameter space, together with marginalized 1D posteriors on both h and Ω_m . Results with both 4 years (upper-row) and 10 years (lower-row) of LISA mission operation are shown. The constraints reported in the figure indicate 90% confidence levels around the measured median value. In what follows we will comment in detail on these two scenarios.

5.1.1 4 years of LISA observations

As described above, the joint posteriors reported in Fig. 4 are averaged over the nine independent realizations of each EMRI scenario.

As reported in Fig. 4, the constraints we obtain on h with 4 years of LISA observations range from a maximum of 0.026 (3.6%), within the M5 scenario, to a minimum of 0.012 (1.6%), within the M6 scenario. The constraints on Ω_m go instead from 0.150 (60.0%) in the M5 scenario, to 0.081 (32.0%) in the M6 scenario. In general, the predominance of low-redshift events in the distance-redshift diagram is such that they are more efficient at constraining H_0 rather than Ω_m . Note that the posterior on Ω_m in the M5 (pessimistic) case is actually dominated by our prior, implying that in this scenario the LISA data do not yield statistically relevant information on Ω_m . Also note that, as expected, our results obtained with M1 (our fiducial scenario) are superseded by those obtained with M6 (optimistic), as in this case the total number of useful EMRIs is more than doubled (cf. Table 1).

5.1.2 10 years of LISA observations

As shown by the lower-row plots in Fig. 4, the constraints on h now range from a worst-case result of 0.019 (2.6%) to a best-case result of 0.008 (1.1%), while the constraints on Ω_m range from 0.107 (42.8%) to 0.058 (23.2%). Differently from the 4-year case, the posterior on Ω_m in the M5 scenario (pessimistic) is not dominated by the prior, meaning that in this scenario 10 years of mission duration could start yielding meaningful information on Ω_m .

The results obtained with M5, our pessimistic scenario, might seem too optimistic if one considers the fact that only 6 EMRI events are counted in the 10-year catalog that we used (see Table 1). However, we have selected the loudest

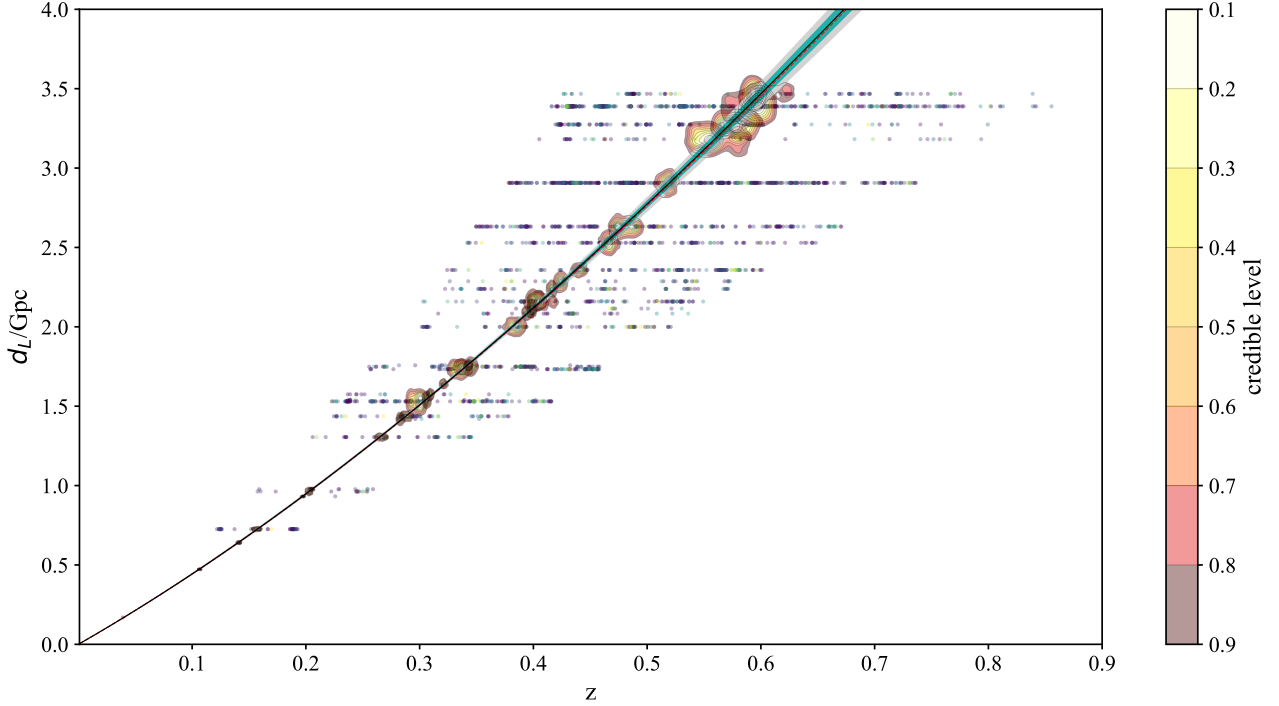


Figure 3. $d_L - z$ regression line for M1_2, one of the three realisations of our fiducial scenario M1 in the 10-year mission case: median (solid black) and 68% and 90% credible regions in light seagreen and light gray, respectively. The dashed line corresponds to the Millennium Simulation fiducial values of h and Ω_m . The coloured regions show the posterior distribution for z_{gw} and d_L for each EMRI event. The horizontal dots show the redshift of each candidate galaxy host for that particular EMRI. For illustrative purpose, we assigned to each galaxy a luminosity distance equal to \hat{d}_L . The dots are also colour-coded from light to dark for increasing values of the weights w_j .

events, which should be the most informative from a parameter estimation point of view. An analogous consideration could be made for the 4-year results, where the number of EMRI events in the employed M5 catalogs are on average only $n_{4\text{yr}} = 2.4$.

We have moreover checked that the 10-year M5 catalog contains a low-redshift event which in the three versions of the 10-year catalog happens to have only 1, 3, and 8 galaxies within its 3D localisation error box. In the catalog where this EMRI has a single host, the inference of its redshift is evidently optimal. In the catalog where it has three galaxies, two of these galaxies are spot-on the true redshift value, while the third one is less relevant, because it lies at the margins of the sky localisation region and thus, according to how our analysis is implemented (cf. Sec. 2), it is weighted with a lower probability with respect to the others. This event alone therefore provides a highly accurate measurement of the Hubble constant at low redshift, which dominates the results reported in the first column of Fig. 4. We note that although the detection of such an advantageous event is a question of luck, the actual probability of finding a similar event in the few EMRI events included in our catalogs is actually non-negligible, in that our analysis automatically selects for well-localized events at high SNR. In general, any EMRI event of this kind will substantially drive our results, while less well-localized, low-SNR events are not expected to contribute (but see App. A2 for a test-case in which we *only* use the faintest events of our fiducial model).

5.2 Dark Energy

Results for the DE cosmological scenario are reported in Fig. 5, again for all three EMRI models. We show corner-plots with 2D posteriors in the w_0 - w_a parameter space, together with marginalized 1D posteriors on both w_0 and w_a , for both 4 years (upper-row) and 10 years (lower-row) of LISA mission operations. Again constraints shown in Fig. 5 define 90% confidence levels around median values. In what follows we will address in more details the two observational scenarios.

As already noted in Sec. 4, the total number of useful events for each EMRI model slightly varies according to the cosmological model under consideration (cf. Table 1); thus the DE catalogs have a slightly different total number of useful EMRIs with respect to the ones used for Λ CDM.

5.2.1 4 years of LISA observations

From the top-row panels of Fig. 5 we can notice that constraints on w_0 range from a maximum of 0.123 (12.3%) to a minimum of 0.068 (6.8%) for the M5 (pessimistic) and M6 (optimistic) models, respectively. On the other hand, the measurements on w_a are largely uninformative as they yield 90% confidence intervals of ~ 0.7 (corresponding to $\sim 70\%$ of the prior) irrespective of the EMRI cosmological scenario. This implies that although LISA EMRIs might be able to tell us something interesting on the current value of the DE equation of state (w_0), they are likely to not have the statistical power to tell us anything about its current time evolution (w_a).

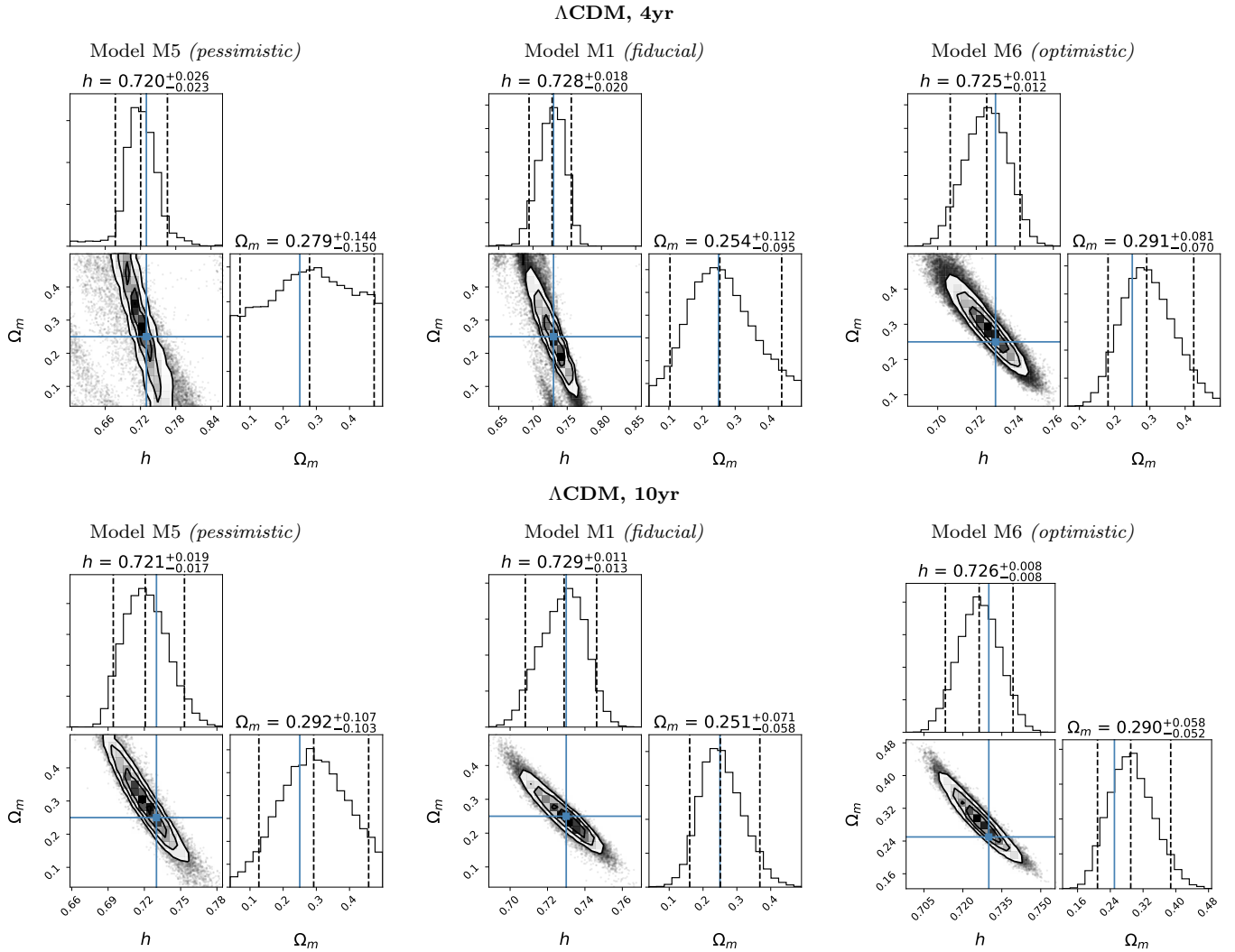


Figure 4. Corner plots of the posteriors for the parameters h and Ω_m in the Λ CDM scenario. In each plot the lower-left panel reports the contours of the joint posterior, while the upper and right panels show the same posterior after marginalization over one parameter. In each panel, the cyan lines mark the fiducial values and the dashed lines indicate the median and 90% credible interval extracted from the marginalized posterior.

5.2.2 10 years of LISA observations

Results are reported in the bottom row of Fig. 5. We can see that with 10 years of observations LISA EMRIs will yield constraints on w_0 ranging from 0.083 (8.3%) to 0.059 (5.9%), slightly improving over the 4-year results. Results for w_a are again largely uninformative, as they still produce 90% confidence constraints around 0.6-0.7, which still constitute a non-negligible fraction of the prior. However, we notice the apparent railing of w_a in case of M6, which becomes prominent in the 10-year case. We will comment on this in App. A1.

6 DISCUSSION

Overall the results we obtained from our investigation show that EMRIs detected by LISA will have an interesting cosmological potential. Assuming the Λ CDM model, the Hubble constant will be probed at the percent level, while in the evolving DE scenario the equation of state of DE will be

tested with an accuracy better than 10%. Some constraint, even though not particularly strong ($\sim 20\%$), can also be put on Ω_m . The constraints on these three parameters, for all three EMRI models and both 4 and 10 years of LISA observations, have been summarised in Fig. 6. On the other hand, only a marginal gain will be achieved on w_a (for DE), which usually is better constrained by measurements at high redshift, which are not so numerous in our catalogs.

6.1 Comparison with EM observations

How do our results compare with current and future EM observations? Current EM measurements of the Hubble constant have reached percent levels of accuracy with observations collected from the CMB (Ade et al. 2016; Aghanim et al. 2018) and from local-Universe distance indicators (Riess et al. 2016, 2019), in particular type-Ia supernovae (SNIa). However, as mentioned in the Introduction, these two techniques yield different values of the Hubble constant which are now

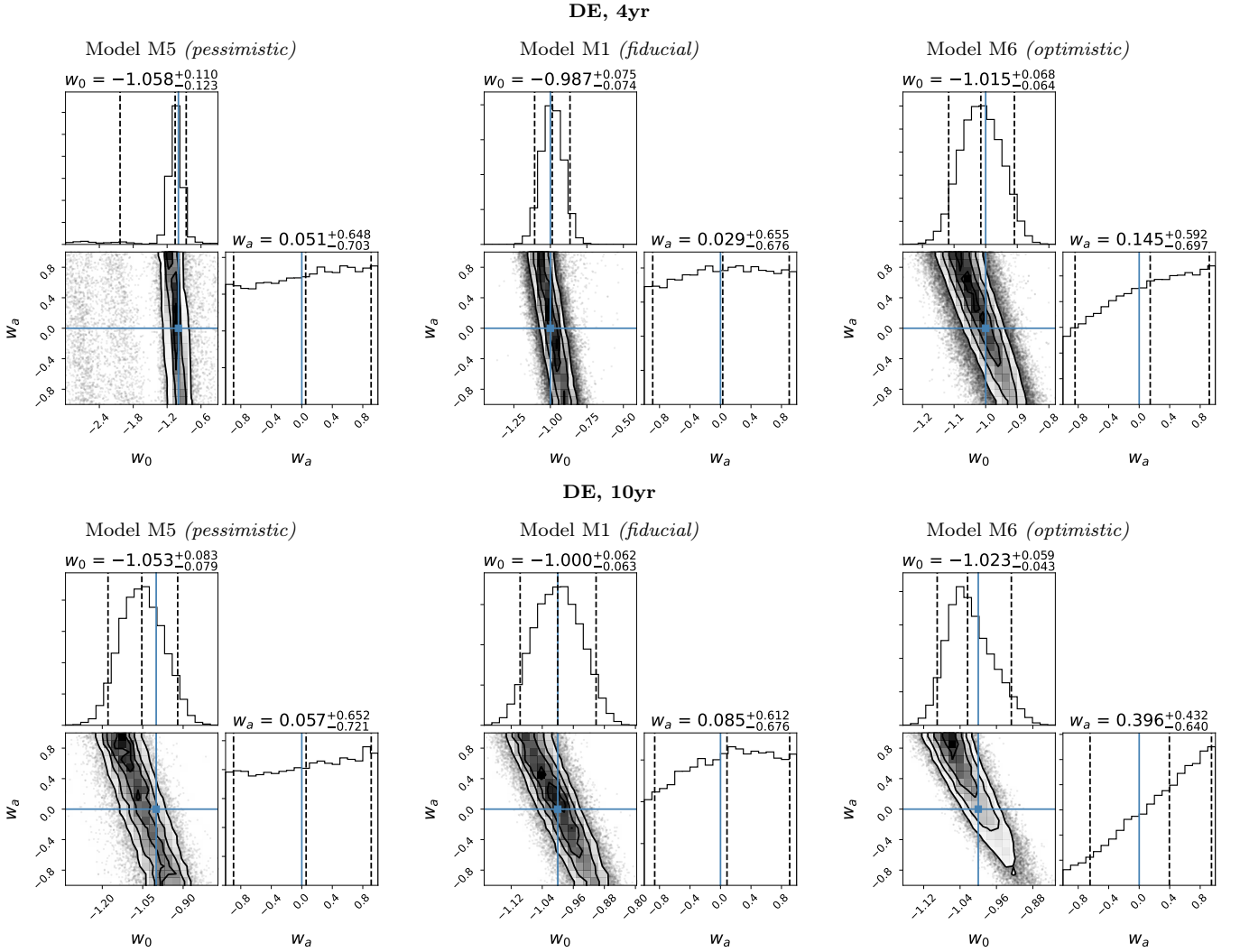


Figure 5. Corner plots of the posteriors for the parameters w_0 and w_a in the DE scenario. Same as Fig. 4. See the Appendix for a discussion on the M6 model.

in tension with each other at more than the 4σ statistical level (Riess et al. 2019). Whether this tension is due to systematics in one of the two measurements or to new physics beyond Λ CDM, it will be only decided by future observations, including the ones performed with GW standard sirens (Chen et al. 2018). The advantage of standard sirens in this respect is that their measurements present completely different systematics with respect to both CMB and local EM measurements, and will thus deliver completely independent information on the Hubble constant, which will hopefully point towards its correct value. Similarly, the interesting aspect of the estimates that we obtained on H_0 from our analysis is not that we will reach a level of accuracy comparable with current and future EM probes, but that EMRIs detected with LISA will provide yet another complementary measurement of H_0 , useful to corroborate results from both EM observations and ground-based GW cosmological measurements. In the likely scenario in which by the time LISA flies we will already have solved the Hubble tension, the measurements of H_0 with LISA EMRIs will help consolidate our control over all systematics, including the ones affecting GWs mea-

surements with ground-based interferometers, and our confidence over the results obtained, especially in the case in which physics beyond Λ CDM is discovered.

We can also compare our results for the dynamical DE cosmological scenario with current measurements and forecasts for the future. At present, distance measurements taken with SN Ia alone, i.e., not combined with CMB observations, constrain w_0 at the ~ 0.2 ($\sim 20\%$) level (Scolnic et al. 2018) (this number is obtained by fitting simultaneously for Ω_m , but not for w_a). Future EM observations collected with the Euclid mission are expected to improve the constraints up to $\Delta w_0 \simeq 0.06$ and $\Delta w_a \simeq 0.26$ (Amendola et al. 2018). The results we obtained for w_0 in Sec. 5 reach the same level of accuracy of present EM observations in the worst case scenario (M5), while they match the expected results from the Euclid mission in the best case scenario (M6). This implies that cosmological measurements with LISA EMRIs will be of great interest to probe the DE equation of state, not only because they will reach constraints comparable with EM probes, but most importantly because, as stressed above, they will provide this level of accuracy with completely different system-

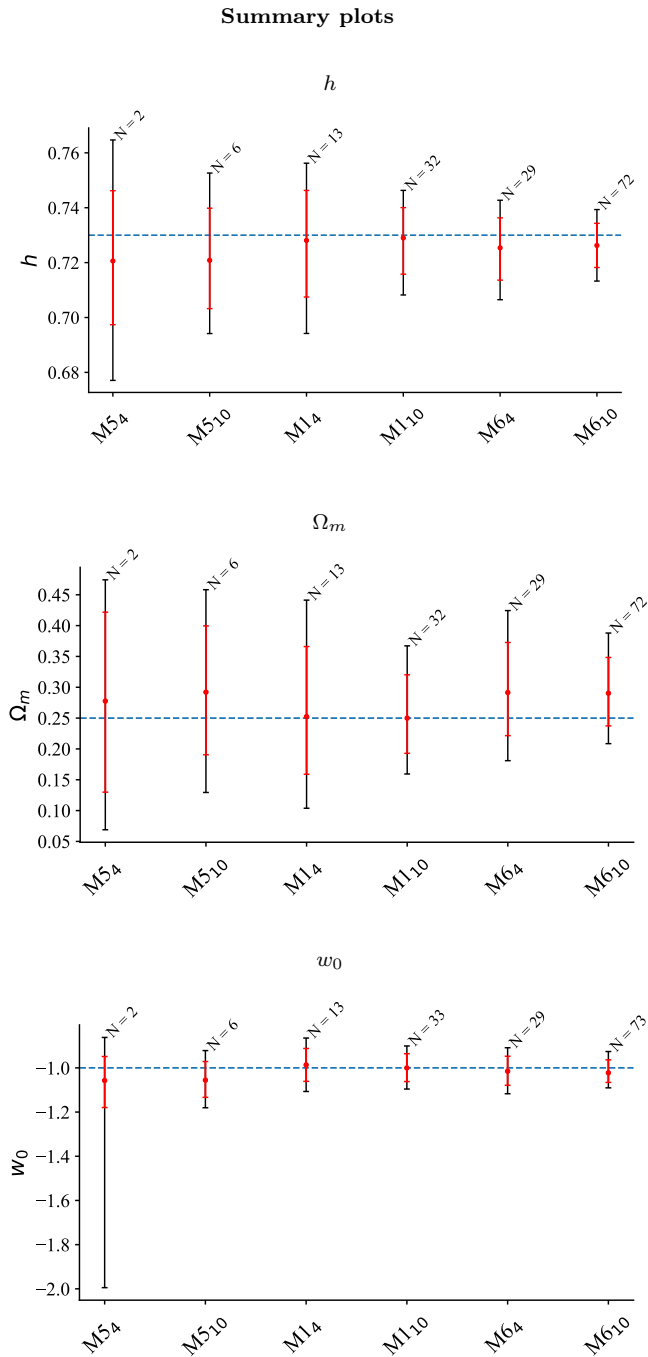


Figure 6. Summary of results. We show 90% (black) and 68% (red) percentiles, together with the median (red dot) for both h , Ω_m , and w_0 for each EMRI model (M5, M1, M6) and LISA observational scenario (4 and 10 years) considered. The blue dashed horizontal line denotes the true cosmology. For each data point, we also report the average number N of EMRIs considered in the analysis, see Table. 1.

atics, increasing our confidence on such measurements (especially if the cosmological constant value $w_0 = -1$ appears to be ruled out).

The constraints obtained on the other two parameters considered in our analysis, namely Ω_m for Λ CDM and w_a for DE, seem not to be comparable with the level of precision

of current and future EM observations (Scolnic et al. 2018; Amendola et al. 2018). This means that although they will be affected by completely different systematics, most likely they will not be able to provide useful additional information with respect to EM observations. In other words, the situation will be similar to the H_0 constraints that we have today with GWs (Abbott et al. 2017c): even if they constitute a complementary measurement of H_0 , affected by completely different systematics, the level of precision of current GW observations does not offer a result comparable to EM measurements, implying that interesting cosmological insight beyond broad consistency between different measurements cannot be obtained.

Better constraints on cosmological parameters beyond H_0 and w_0 might be obtained by extending our analysis to EM-RIs with lower SNR, since ideally we would have additional points to place on our $d_L - z$ plot. However, as noticed in Sec. 3 and as discussed in App. A2, the statistical identification method presents some difficulties when applied to high-redshift EMRI events having a large number of potential hosts. In our present framework the modelling and inference of galaxy evolution features as well as the inclusion of low-SNR EMRIs, which could allow to better probe further cosmological parameters, are not taken into account and will be investigated in future studies. A more promising strategy nevertheless may be to combine measurements from different GW sources detected by LISA, as discussed hereafter.

6.2 Comparison with other GW cosmological measurements

As we mentioned in Sec. 1, EMRIs are not the only LISA GW sources that can be used as standard sirens. Other investigations considering SOBHBs and MBHBs detected by LISA have already produced cosmological estimates. How do our results compare with those?

LISA SOBHBs will be mainly detected at low redshift ($z \simeq 0.1$) and thus will be only useful to constrain H_0 . Recent analyses (Kyutoku & Seto 2017; Del Pozzo et al. 2018) suggest that SOBHBs could be used to probe H_0 at the few % level, similarly to the results we obtained with EMRIs. SOBHBs can however reach this level of accuracy only if the detection rate will fall in the optimistic range, which will crucially depend on the yet uncertain sensitivity that LISA will be able to achieve at high frequencies (Moore et al. 2019). Note also that an “archival” search for SOBHBs will be possible: detecting the black hole merger with the third generation of ground-based GW detectors and then search those sources in the (archival) LISA data with the reduced prior (Ewing et al. 2021). On the other hand, EMRIs rely on the LISA sensitivity around the mHz (middle-band) which is guaranteed to be at the level considered in our analysis (Amaro-Seoane et al. 2017), if not better. Nevertheless as we have seen, EMRI expected rates are affected by other uncertainties of astrophysical and theoretical nature. It is important to notice that LISA will fail in deliver compelling low-redshift cosmological results on H_0 only if both the LISA high-frequency sensitivity does not perform as expected and the true astrophysical population of EMRIs falls on the very pessimistic side, thus not providing enough detections. In all other scenarios we might expect LISA to deliver useful low-redshift constraints on H_0 , from either SOBHBs, EMRIs, or both.

LISA MBHB mergers, which are expected to produce observable EM counterparts, will instead probe the expansion of the Universe at much higher redshift, providing in this way complementary constraints to the ones obtained at low redshift with both SOBHBs and EMRIs. LISA MBHBs might yield constraints on H_0 at the few % level (Tamanini et al. 2016; Tamanini 2017; Belgacem et al. 2019b), comparable to SOBHB and EMRI results. This implies that LISA will deliver accurate measurements of H_0 from data sets at different redshift ranges, possibly providing further insights into the current Hubble tension between local measurements (low- z) and CMB observations (high- z). LISA MBHBs will moreover be useful to probe other cosmological parameters which cannot be constrained at low redshift, for example Ω_m for Λ CDM. The integration of low- and high-redshift standard sirens will thus allow LISA to test the expansion of the Universe with GWs from early-to-late cosmological times, without the need to combine it with other probes. This makes LISA a unique cosmological observatory (Tamanini 2017), whose thorough implications are currently being investigated and will be presented in a future study (Tamanini et al. 2021).

Finally, we can compare our cosmological results with the ones forecast for future Earth-based GW interferometers. The current LIGO-Virgo network of detectors, with further improvements and the addition of more interferometers, is expected to produce constraints on H_0 at the few % level only if the rate of multi-messenger detections will lie on the optimistic side Chen et al. (2018). Future third-generation detectors, such as ET (Punturo et al. 2010; Maggiore et al. 2020), are instead expected to precisely probe the expansion of the Universe up to $z \sim 2$ (Sathyaprakash et al. 2010; Taylor & Gair 2012; Del Pozzo et al. 2017; Belgacem et al. 2019a). They will provide complementary results with respect to LISA in approximately the same time period (2030s), implying that cross-checking GW cosmological measurements from space and from the ground will increase our confidence in the obtained results and will help us to get a handle on the expected systematics.

6.3 More optimistic EMRI scenarios

Because of the reasons mentioned in Sec. 3, in our cosmological investigation we selected and analysed only three out of the twelve EMRI models considered in Babak et al. (2017). In first approximation, we can expect that our results can be extrapolated to the other EMRI models according to the square-root of the number of detected events. In other words, the constraints we found on h and w_0 should improve by a factor inversely proportional to \sqrt{N} (Schutz 1986), where N is the total number of EMRIs useful for cosmology, reported in Sec. 3 for each model.

Unfortunately, it is difficult, if not impossible, to estimate this number for the other EMRI models of Babak et al. (2017) without performing the detailed analysis we carried out for the models M1, M5, and M6. For this reason, in what follows we use the total number of detected EMRI events, as given by Table I in Babak et al. (2017) (AKK column), as a rough proxy for this number. We must keep in mind however that different EMRI scenarios will of course yield a population of EMRIs with different properties, which will certainly translate in a different number of cosmologically useful events after one applies the threshold we defined in Sec. 3. The estimates

that follow should hence only be taken as simple indicative numbers useful to quickly explore other possible EMRI scenarios, rather than statistically robust results.

The majority of the EMRI scenarios considered in Babak et al. (2017) present a number of detections comparable to M1, our fiducial model. We can thus expect that for all these scenarios we would achieve cosmological constraints comparable to the ones we obtained with M1. Model M8 yields a number of detections similar to M5, and thus for M8 we expect roughly the same cosmological results as the more pessimistic ones we found in our analysis. The most pessimistic model of Babak et al. (2017), namely model M11, provides only one detected event per year, way below the detections achieved with M5, our pessimistic scenario. Needless to say, we would not expect any useful cosmological results with M11.

On the other hand, there are three models which predict more detections than those obtained in our optimistic scenario M6. We can thus extrapolate results starting from the detections of M6 by using our simple \sqrt{N} estimate. Model M3 presents roughly twice the detections obtained with M6, which translates into improved 4-year (10-year) constraints (90% C.I.) as $\Delta h \simeq 0.0092$ (0.0062) and $\Delta w_0 \simeq 0.052$ (0.045), corresponding to relative errors of 1.3% (0.84%) and 5.2% (4.5%), respectively. The two most optimistic scenarios are however models M7 and M12, which deliver an order of magnitude more EMRI detections with respect to M6. Within M7 we estimate improved 4-year (10-year) constraints as $\Delta h \simeq 0.005$ (0.034) and $\Delta w_0 \simeq 0.029$ (0.025), corresponding to 0.69% (0.46%) and 2.9% (2.5%) relative uncertainties, respectively. For M12 instead we find $\Delta h \simeq 0.0040$ (0.0027) and $\Delta w_0 \simeq 0.023$ (0.020), corresponding to 0.55% (0.37%) and 2.3% (2.0%).

According to these rough estimates, we could reach sub-percent constraints on H_0 and two-percent constraints on DE (w_0) if the rate of EMRIs detected by LISA is at the most optimistic end of expectations. From a cosmological perspective these would constitute outstanding results, which even EM probes might not be able to achieve in the near future. Having said that, we must stress again that these are only simple estimates which rely on extremely optimistic EMRI scenarios and do not take into account all the complexities of the true detected population of EMRIs.

6.4 Future prospects

The analysis performed in the present study can be improved in several ways. First of all, the numbers and properties of the EMRI population detected by LISA could be better characterized by considering enhanced astrophysical modelling and by refining the LISA response. In the error estimation we have used the long-wavelength approximation to the response which takes into account the amplitude and the phase (Doppler) modulation of the signal due to LISA's motion around the Sun. This modulation is what gives us the sky position used in this paper. The full response also includes the sky-dependent time delays due to GW propagation across satellites in the constellation. Taking the full response, therefore, could improve the sky localisation for EMRIs with $M < \text{few} \times 10^5 M_\odot$. In addition, one could decisively improve the data analysis treatment by starting from better EMRI waveforms, possibly including inputs from self-force calculations, and by performing a full Bayesian parameter estimation.

tion with a better model for the instrumental noise. Although such improvements would require further theoretical developments and costly numerical computations, they would permit to consider full non-Gaussian posteriors in both distance and sky localisation, which will be important in making our analysis more realistic, especially for low-SNR events.

Another important addition that might be taken into consideration is modelling galaxy time evolution in order to characterize departures from homogeneity and uniformity in the employed galaxy catalog (see Sec. 2). This would be a more realistic representation of the Universe, but the price to pay is the need to perform simultaneous inference of both the cosmological parameters and the properties of the galaxy evolution, which on the one hand is computationally costly and might degrade cosmological results, but on the other hand will provide complementary astrophysical insight.

Another improvement would be to account for incompleteness of the galaxy catalogue. In the current analysis we have ignored this and assumed that out-of-catalogue hosts will be close to hosts that are included in the catalogue. This becomes increasingly problematic for more distant events, where a greater fraction of potential hosts will be missing. Incompleteness can be accounted for in the analysis, for example by weighting galaxies in the catalogue by the number of nearby galaxies that are missing, or by adding an appropriate number of missing galaxies into the assumed redshift distribution (Abbott et al. 2020d). The impact of such corrections should be explored in the future.

One can also think of using galaxy observational features, such as luminosity, mass, metallicity, and others, to weight galaxies differently within each GW sky localisation region. Moreover, one could use empirical relations, for example between the mass of the MBH located at the center of the galaxy, which is inferred from EMRI parameter estimation, and the mass or luminosity of galaxies in order to identify more likely host galaxy candidates. Although these methods would reduce the number of possible host galaxies for each EMRI event, they would also introduce a dependency upon astrophysical modelling into the analysis, possibly introducing new systematics, or the need to marginalise over additional astrophysical parameters.

We conclude this section by noting a few aspects of our inference model that could be improved. First of all, we decided to neglect the intrinsic EMRI population evolution in the prior assignment over z_{gw} . The possibility of some bias in the estimation of the cosmological parameters due to a rapid evolution of the rate of EMRIs with redshift, which means that the weighting within each box should not be uniform in z , has been already pointed out in MacLeod & Hogan (2008), although they did not account for this effect in their analysis. Just like for the cosmological evolution of the galaxy population, inclusion of this additional feature might degrade the overall inference over Ω . However, in scenarios where the number of detected EMRIs is large, this term would impose additional constraints on the whole EMRI population through the time dependence of the merger rate, thus potentially increasing the amount of information contained in each posterior. Moreover, we would be able to relax the arbitrary SNR cutoff of 100, using more faintest and possibly farther events, thus exploring a larger co-moving volume of the Universe.

We stress that the present investigation constitutes a first

simple attempt at assessing the cosmological potential of LISA EMRIs. Further studies, improving the analysis along the lines outlined above, will be needed in order to provide more reliable forecasts and to prepare for the developing of the pipelines needed to analyse the expected data from LISA.

7 CONCLUSION

In this article we investigated the cosmological potential of EMRIs detected by LISA. By statistically matching the sky localisation region of the loudest EMRI events (as given by the analysis of Babak et al. (2017)) with the position and redshift of galaxies within a given catalog (in our case based on the Millennium run), we extracted constraints on the parameters characterizing the background evolution of two cosmological models: Λ CDM and a dynamical DE scenario. Our results show that interesting cosmological insight can be gained from EMRIs as standard sirens. Over three different EMRI models and two LISA mission durations, constraints on H_0 and w_0 (the equation of state of DE) can respectively reach the $\sim 1.1\%$ and $\sim 4.3\%$ levels in the best case scenario, and be degraded to maximum uncertainties of $\sim 3.6\%$ and $\sim 12.3\%$ in the worst case scenario (cf. Fig. 6). In particular, for our fiducial model M1, EMRI observations are expected to constrain H_0 and w_0 to $\sim 2.5\%$ (1.5%) and $\sim 7.4\%$ (6.2%) respectively, assuming a four (ten) year LISA mission. As we discussed at length in Sec. 6, these results will be of great value for cosmology, and will increase our confidence on other EM and GW measurements.

Our results are largely compatible with those reported in MacLeod & Hogan (2008), the only EMRI cosmological analysis available in the literature to date. In that analysis it was claimed that 1% accuracy on H_0 can be reached with as few as 20 EMRIs at redshift $z < 0.5$. However, that analysis assumed parameter estimation accuracies appropriate for a more optimistic LISA design. It was shown in Gair et al. (2017) that 7/1/8 EMRI events satisfying these optimistic parameter constraints were found in two years of observation for models M1/M5/M6, respectively. Therefore, we would expect to achieve comparable precision with ten years of observation of models M1 and M6, as we find here. The fact that we do not do better, even with the slightly larger number of well-localized events and the inclusion of additional less well-localized events in the analysis, is most likely due to the simplifications employed in the analysis in MacLeod & Hogan (2008), such as the use of a linear cosmic expansion model with H_0 as the parameter.

Finally, as in MacLeod & Hogan (2008), we let each EMRI source contribute on an equal footing to the likelihood, while, differently from MacLeod & Hogan (2008), we do not assume equal probability over the whole sky localisation region, but weight galaxies by the marginalised likelihood, see Eqs. (2.8) and (4.4).

As a general final remark, we stress that LISA, a space mission dedicated to GW science, will reveal itself as a unique cosmological probe, through which we will be able to map the expansion of the Universe using different GW sources as standard sirens at different redshifts, including, as thoroughly demonstrated by our study, EMRIs. Future more-in-depth investigations may deliver cosmological forecast analyses with

LISA EMRIs extended at higher redshift, thus allowing us to explore the high-redshift Universe with dark sirens.

ACKNOWLEDGEMENTS

D.L. and W.D.P. thank Stefano Rinaldi for discussions. N.T. acknowledges partial support from the COST Action CA16104 “Gravitational waves, black holes and fundamental physics” (GWverse), supported by COST (European Cooperation in Science and Technology). A.S. is supported by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program ERC-2018-COG under grant agreement No 818691 (B Massive).

APPENDIX A: DISCUSSION OF CAVEATS AND SYSTEMATICS

Here we discuss two aspects of the results presented in the paper: (i) the railing of the parameter w_a observed in the 10-year M6 scenario; (ii) the consequences of imposing a cutoff in SNR on the Λ CDM and DE analyses.

A1 Railing w_a

The right panels of Fig. 5 show the results for w_0 and w_a in the 4 and 10-year optimistic scenarios (M6). It can be noted that while w_0 is correctly measured (being the true value always well within the 90% credible intervals), w_a , although uninformative, shows a railing against the upper prior bound. This is particularly evident in the 10-year analysis (bottom right panel of Fig. 5, where the median value is 40% off the true value and the correlation between w_0 and w_a tends to push for $w_0 < -1$). It is worth to focus on this specific scenario and see what is the possible cause of this behaviour.

We use the same set of 73 loudest EMRI events used in that scenario, this time associating to each event only one galaxy host, chosen to be the nearest in redshift to the EMRI. This analysis may be viewed as representative of the physical scenario in which all our events have an EM counterpart, so it can be used also to investigate what the DE paradigm would predict in that case. In our likelihood, Eq. (2.10), this choice corresponds to assigning $w_j = 1$ to the nearest host and $w_j = 0$ to the others. The scope of this test is to see whether the railing seen for w_a is due to the particular nature of the EMRI catalog or to our formulation of the problem. The results, shown in Fig. A1, seem to indicate that indeed this is imputable to the latter. The railing is now absent, showing posteriors for w_0 and w_a that are fully consistent with the fiducial values. Thus, cross-matching the EMRIs with their nearest-in- z hosts gives substantially unbiased posteriors. This seems to point to some limitations on our formulation concerning the way in which the galaxy hosts are treated.

Looking at Fig. A1, it is also interesting to note that the CLs on the measure of w_0 are comparable to those of the general case – where we include all galaxy hosts – reported in Fig. 5: this shows that even if one were to observe EM counterparts to these EMRIs, the inference would not substantially improve.

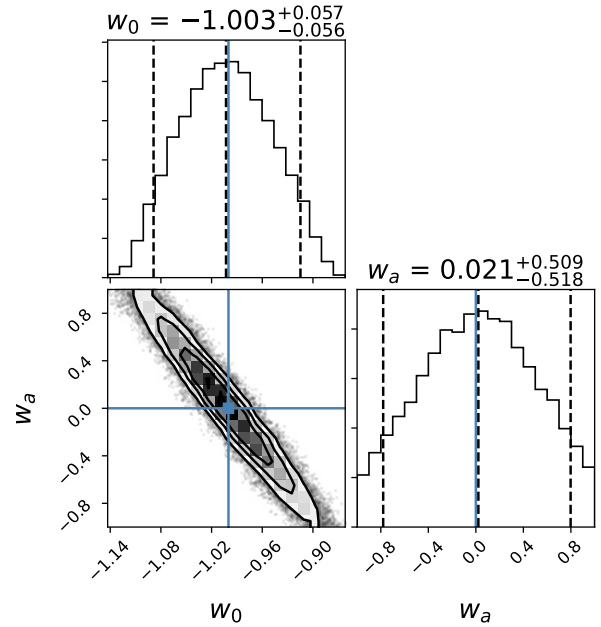


Figure A1. M6 DE 10 years, obtained analysing the same set of events used to produce the analogous corner plot in Fig. 5, but now keeping only the nearest-in- z host for each EMRI. As can be seen, the railing behaviour of w_a has now disappeared.

A2 Low-SNR events

Even though we limited our analysis to high-SNR events, it is interesting to estimate the contribution of low-SNR events to the inference of the cosmological parameters under the assumption of likelihood (2.10). As an illustrative example, let us consider the EMRI catalogs for our fiducial model M1 in the 10-year mission case. We will first adopt our first selection procedure detailed in Sec. 3, which impose constraints on the measurement precision of the sky position and the luminosity distance (still associating to each EMRI hosts drawn from our galaxy catalog up to $z = 1$). Then, to limit ourselves to low-SNR events, we impose an upper cutoff $\text{SNR} < 40$ and analyse the events assuming our two cosmological models. In case of Λ CDM, the number of events is 30, while for DE there are 39 events, see Table A1.

The posteriors for the cosmological parameters are shown in the upper-row panels of Fig. A2: using only the quietest EMRIs does not yield uninformative posteriors. In particular, the correlation between the two parameters is mainly lost, as expected for low-SNR events, while we observe a preference for high values of the parameters.

To exclude the possibility of sampling issues or that the low-SNR events show some unaccounted-for systematic difference from the rest of the EMRI population, we analysed the same set of low-SNR events, but keeping only the nearest host in redshift to the EMRI, in a manner analogous to what we did in Sec. A1 for the loudest events of M6 in the DE scenario. As already noted in that section, such an analysis may be interpreted as the scenario in which we observe an EM counterpart to all our low-SNR EMRIs. The resulting posteriors are shown in the bottom-row panels of Fig. A2. The correlations between the parameters is restored and the railing has now disappeared, resulting in fairly informative

MODEL	detected (SNR>20)	$\Delta d_L/d_L < 0.1$ & $\Delta\Omega < 2 \text{ deg}^2$		SNR<40	
		Λ CDM	DE	Λ CDM	DE
M1 (<i>fid</i>)	2941	180	202	30	39

Table A1. Number N of EMRIs observed by LISA in 10 years of operation imposing an upper cut in SNR, as done for the results presented in Fig. A2. The upper cutoff SNR < 40 is imposed in order to study the effect of the faintest events on the inference problem.

posteriors for h and w_0 , a mildly informative posterior for Ω_m , and a substantially uniform posterior for w_a . The posteriors indicate that the low-SNR population of EMRIs is consistent with the general one and that the inclusion of multiple potential hosts in the analysis is likely the culprit. We have thus another indication that our treatment of multiple galaxy hosts is conditioned on our main assumptions, which thus deserve further investigation. Hence, one should be aware of the caveats to this likelihood in a general analysis where also low-SNR events are included. It is also true that the faintest events are expected to be less relevant to the inference problem, so they should not contribute in a relevant way. This is due to the large number of hosts that are typically associated to each EMRI. The results presented in this work are based on a different selection criterion (see Sec. 5) which do not suffer from any relevant bias in the estimate of the main parameters we are interested in, that is, h , Ω_m , and w_0 . We plan to further investigate the inclusion of low-SNR events in the cosmological inference elsewhere.

REFERENCES

Abbott B. P., et al., 2016a, *Phys. Rev. Lett.*, 116, 061102
 Abbott B. P., et al., 2016b, *Phys. Rev. Lett.*, 116, 221101
 Abbott B. P., et al., 2017a, *Classical and Quantum Gravity*, 34, 044001
 Abbott B. P., et al., 2017b, *Phys. Rev. Lett.*, 119, 161101
 Abbott B. P., et al., 2017c, *Nature*, 551, 85
 Abbott B. P., et al., 2017d, *Astrophys. J.*, 848, L13
 Abbott B. P., et al., 2019a, *Phys. Rev.*, X9, 031040
 Abbott B. P., et al., 2019b, *Phys. Rev. Lett.*, 123, 011102
 Abbott B. P., et al., 2019c, *Astrophys. J.*, 882, L24
 Abbott B. P., et al., 2019d, *Phys. Rev.*, D100, 104036
 Abbott B. P., et al., 2020a, preprint ([arXiv:2001.01761](https://arxiv.org/abs/2001.01761))
 Abbott R., et al., 2020b, preprint ([arXiv:2010.14527](https://arxiv.org/abs/2010.14527))
 Abbott R., et al., 2020c, preprint ([arXiv:2010.14529](https://arxiv.org/abs/2010.14529))
 Abbott B. P., et al., 2020d, preprint ([arXiv:1908.06060](https://arxiv.org/abs/1908.06060))
 Ade P. A. R., et al., 2016, *Astron. Astrophys.*, 594, A13
 Aghanim N., et al., 2018, preprint ([arXiv:1807.06209](https://arxiv.org/abs/1807.06209))
 Amaro-Seoane P., Gair J. R., Freitag M., Miller M. C., Mandel I., Cutler C. J., Babak S., 2007, *Classical and Quantum Gravity*, 24, R113
 Amaro-Seoane P., et al., 2017, preprint ([arXiv:1702.00786](https://arxiv.org/abs/1702.00786))
 Amendola L., et al., 2018, *Living Rev. Rel.*, 21, 2
 Armano M., et al., 2016, *Phys. Rev. Lett.*, 116, 231101
 Armano M., et al., 2018, *Phys. Rev. Lett.*, 120, 061101
 Babak S., Gair J. R., Cole R. H., 2015, *Fund. Theor. Phys.*, 179, 783
 Babak S., et al., 2017, *Phys. Rev.*, D95, 103012
 Barack L., Cutler C., 2004, *Phys. Rev.*, D69, 082005
 Barack L., et al., 2019, *Class. Quant. Grav.*, 36, 143001
 Bartolo N., et al., 2016, *JCAP*, 1612, 026

Behnel S., Bradshaw R., Citro C., Dalcin L., Seljebotn D. S., Smith K., 2011, *Computing in Science & Engineering*, 13, 31
 Belczynski K., Kalogera V., Bulik T., 2002, *ApJ*, 572, 407
 Belgacem E., Dirian Y., Foffa S., Howell E. J., Maggiore M., Regimbau T., 2019a, *JCAP*, 08, 015
 Belgacem E., et al., 2019b, *JCAP*, 1907, 024
 Breivik K., Kremer K., Bueno M., Larson S. L., Coughlin S., Kalogera V., 2018, *Astrophys. J.*, 854, L1
 Cai R.-G., Tamanini N., Yang T., 2017, *JCAP*, 1705, 031
 Caprini C., Figueroa D. G., 2018, *Class. Quant. Grav.*, 35, 163001
 Caprini C., Tamanini N., 2016, *JCAP*, 1610, 006
 Caprini C., et al., 2016, *JCAP*, 1604, 001
 Chen H.-Y., Fishbach M., Holz D. E., 2018, *Nature*, 562, 545
 Cusin G., Tamanini N., 2020, preprint ([arXiv:2011.15109](https://arxiv.org/abs/2011.15109))
 Dal Canton T., Mangiagli A., Noble S. C., Schnittman J., Ptak A., Klein A., Sesana A., Camp J., 2019, *Astrophys. J.*, 886, 146
 Dalal N., Holz D. E., Hughes S. A., Jain B., 2006, *Phys. Rev. D*, 74, 063006
 Del Pozzo W., 2012, *Phys. Rev.*, D86, 043011
 Del Pozzo W., Laghi D., 2020, *cosmolisa*, <https://github.com/wdpozso/cosmolisa>
 Del Pozzo W., Li T. G. F., Messenger C., 2017, *Phys. Rev. D*, 95, 043502
 Del Pozzo W., Sesana A., Klein A., 2018, *Mon. Not. Roy. Astron. Soc.*, 475, 3485
 Dotti M., Sesana A., Decarli R., 2012, *Adv. Astron.*, 2012, 940568
 Ewing B., Sachdev S., Borhanian S., Sathyaprakash B. S., 2021, *Phys. Rev. D*, 103, 023025
 Ezquiaga J. M., Holz D. E., 2020, preprint ([arXiv:2006.02211](https://arxiv.org/abs/2006.02211))
 Farr W. M., Fishbach M., Ye J., Holz D., 2019, *Astrophys. J. Lett.*, 883, L42
 Feeney S. M., Peiris H. V., Williamson A. R., Nissanke S. M., Mortlock D. J., Alsing J., Scolnic D., 2019, *Phys. Rev. Lett.*, 122, 061105
 Fishbach M., et al., 2019, *Astrophys. J.*, 871, L13
 Foreman-Mackey D., 2016, *The Journal of Open Source Software*, 24
 Gair J. R., Babak S., Sesana A., Amaro-Seoane P., Barausse E., Berry C. P. L., Berti E., Sopuerta C., 2017, *J. Phys. Conf. Ser.*, 840, 012021
 Giacomazzo B., Baker J. G., Miller M. C., Reynolds C. S., van Meter J. R., 2012, *Astrophys. J.*, 752, L15
 Gray R., et al., 2020, *Phys. Rev. D*, 101, 122001
 Haiman Z., 2017, *Phys. Rev.*, D96, 023004
 Henriques B. M. B., White S. D. M., Lemson G., Thomas P. A., Guo Q., Marleau G.-D., Overzier R. A., 2012, *MNRAS*, 421, 2904
 Hirata C. M., Holz D. E., Cutler C., 2010, *Phys. Rev. D*, 81, 124046
 Hogg D. W., 1999, arXiv e-prints, [pp-astro-ph/9905116](https://arxiv.org/abs/pp-astro-ph/9905116)
 Holz D. E., Hughes S. A., 2005, *Astrophys. J.*, 629, 15
 Hunter J. D., 2007, *Computing in Science & Engineering*, 9, 90
 Jaynes E. T., 2003, *Probability Theory: The Logic of Science*. Cambridge University Press, [doi:10.1017/CBO9780511790423](https://doi.org/10.1017/CBO9780511790423)
 Kaplan D. L., O’Shaughnessy R., Sesana A., Volonteri M., 2011, *Astrophys. J.*, 734, L37
 Klein A., et al., 2016, *Phys. Rev.*, D93, 024003
 Kocsis B., Haiman Z., Menou K., 2008, *Astrophys. J.*, 684, 870
 Korol V., Rossi E. M., Groot P. J., Nelemans G., Toonen S., Brown A. G. A., 2017, *Mon. Not. Roy. Astron. Soc.*, 470, 1894
 Korol V., Koop O., Rossi E. M., 2018, *Astrophys. J.*, 866, L20
 Kyutoku K., Seto N., 2017, *Phys. Rev.*, D95, 083525
 LIGO Scientific Collaboration 2018, LIGO Algorithm Library - LALSuite, free software (GPL), [doi:10.7935/GT1W-FZ16](https://doi.org/10.7935/GT1W-FZ16)
 Lau M. Y. M., Mandel I., Vigna-Gómez A., Neijssel C. J., Stevenson S., Sesana A., 2019, preprint ([arXiv:1910.12422](https://arxiv.org/abs/1910.12422))
 Linder E. V., 2003, *Phys. Rev. Lett.*, 90, 091301
 Luo Z., Guo Z., Jin G., Wu Y., Hu W., 2020, *Results in Physics*, 16, 102918

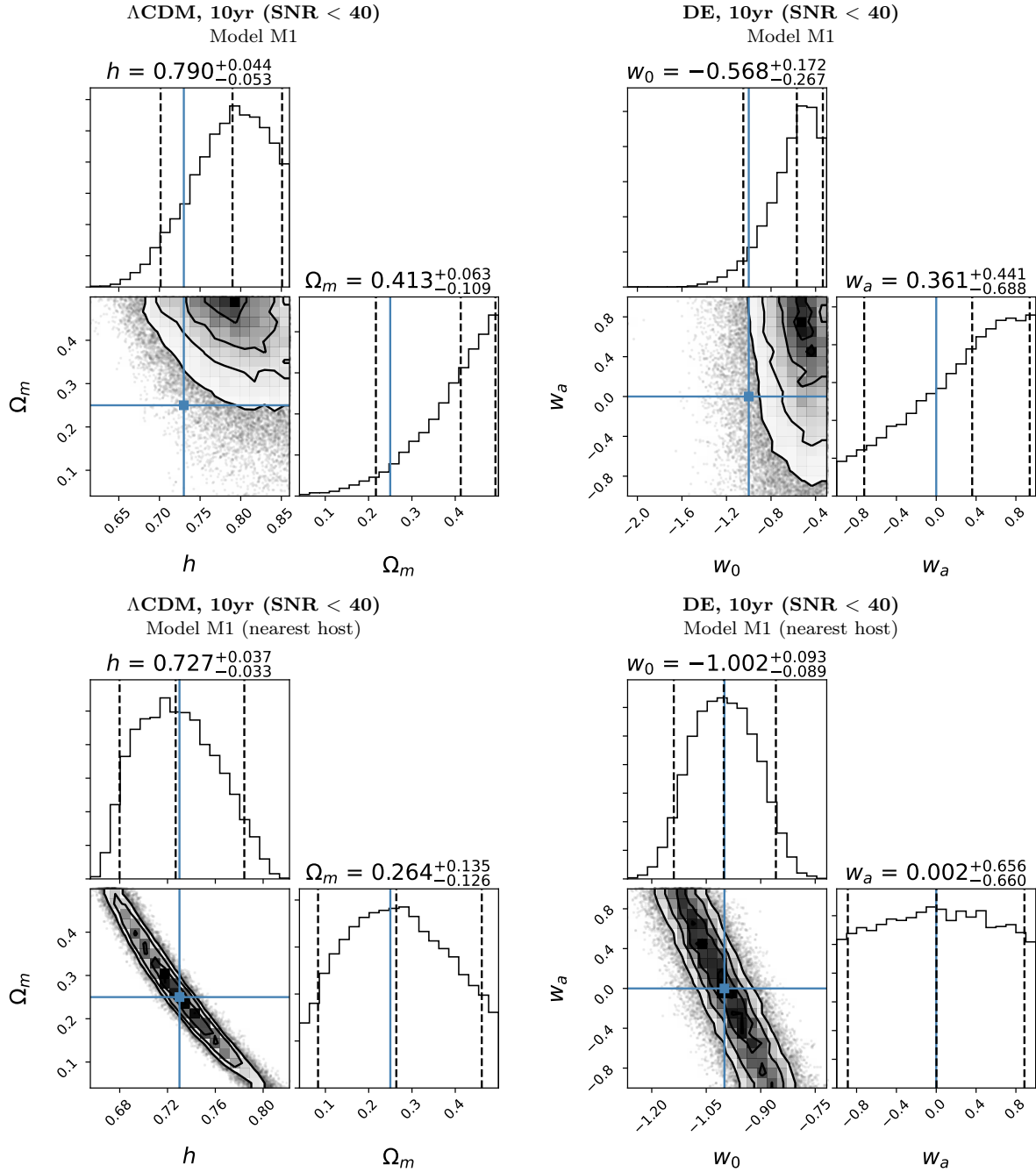


Figure A2. Corner plots of the posteriors on the parameters h , Ω_m , w_0 , and w_a in the Λ CDM and DE scenario, respectively, for our fiducial model M1 using the faintest events, selected imposing an upper threshold SNR < 40.

MacLeod C. L., Hogan C. J., 2008, *Phys. Rev.*, D77, 043512
 Maggiore M., et al., 2020, *JCAP*, 03, 050
 Mandel I., Farr W. M., Gair J. R., 2019, *Monthly Notices of the Royal Astronomical Society*, 486, 1086–1093
 Mei J., et al., 2020, *Progress of Theoretical and Experimental Physics*
 Merritt D., 2015, *ApJ*, 814, 57
 Moesta P., Alic D., Rezzolla L., Zanotti O., Palenzuela C., 2012, *Astrophys. J.*, 749, L32
 Moore C. J., Gerosa D., Klein A., 2019, *Mon. Not. Roy. Astron. Soc.*, 488, L94
 Mukherjee S., Wandelt B. D., 2018, preprint ([arXiv:1808.06615](https://arxiv.org/abs/1808.06615))
 Mukherjee S., Wandelt B. D., Nisanke S. M., Silvestri A., 2020a,

preprint ([arXiv:2007.02943](https://arxiv.org/abs/2007.02943))
 Mukherjee S., Wandelt B. D., Silk J., 2020b, preprint ([arXiv:2012.15316](https://arxiv.org/abs/2012.15316)), doi:10.1093/mnras/stab001
 Mukherjee S., Wandelt B. D., Silk J., 2020c, *Phys. Rev. D*, 101, 103509
 Mukherjee S., Wandelt B. D., Silk J., 2020d, *Mon. Not. Roy. Astron. Soc.*, 494, 1956
 Mörtzell E., Dhawan S., 2018, *JCAP*, 1809, 025
 Nisanke S., Holz D. E., Hughes S. A., Dalal N., Sievers J. L., 2010, *Astrophys. J.*, 725, 496
 Nisanke S., Holz D. E., Dalal N., Hughes S. A., Sievers J. L., Hirata C. M., 2013, preprint ([arXiv:1307.2638](https://arxiv.org/abs/1307.2638))
 O’Shaughnessy R., Kaplan D. L., Sesana A., Kamble A., 2011,

- [Astrophys. J.](#), 743, 136
- Oguri M., 2016, [Phys. Rev. D](#), 93, 083511
- Palenzuela C., Lehner L., Liebling S. L., 2010, [Science](#), 329, 927
- Peebles P. J. E., 1993, Principles of physical cosmology. Princeton University Press
- Petiteau A., Babak S., Sesana A., 2011, [Astrophys. J.](#), 732, 82
- Poisson E., Pound A., Vega I., 2011, [Living Rev. Rel.](#), 14, 7
- Punturo M., et al., 2010, [Class. Quant. Grav.](#), 27, 194002
- Reitze D., et al., 2019a, preprint ([arXiv:1903.04615](#))
- Reitze D., et al., 2019b, preprint ([arXiv:1907.04833](#))
- Riess A. G., et al., 2016, [Astrophys. J.](#), 826, 56
- Riess A. G., Casertano S., Yuan W., Macri L. M., Scolnic D., 2019, [Astrophys. J.](#), 876, 85
- Rodriguez C. L., Chatterjee S., Rasio F. A., 2016, [Phys. Rev. D](#), 93, 084029
- Sathyaprakash B. S., Schutz B. F., Van Den Broeck C., 2010, [Class. Quant. Grav.](#), 27, 215006
- Schutz B. F., 1986, [Nature](#), 323, 310
- Scolnic D. M., et al., 2018, [Astrophys. J.](#), 859, 101
- Sesana A., 2016, [Phys. Rev. Lett.](#), 116, 231102
- Soares-Santos M., et al., 2019, [Astrophys. J.](#), 876, L7
- Speri L., Tamanini N., Caldwell R. R., Gair J. R., Wang B., 2020, preprint ([arXiv:2010.09049](#))
- Springel V., et al., 2005, [Nature](#), 435, 629
- Tamanini N., 2017, [J. Phys. Conf. Ser.](#), 840, 012029
- Tamanini N., Caprini C., Barausse E., Sesana A., Klein A., Petiteau A., 2016, [JCAP](#), 1604, 002
- Tamanini N., et al., 2021, *in preparation*.
- Taylor S. R., Gair J. R., 2012, [Phys. Rev. D](#), 86, 023502
- Veitch J., Del Pozzo W., Cody Pitkin M., 2020, cplest, doi:10.5281/zenodo.835874, <https://doi.org/10.5281/zenodo.835874>
- Virtanen P., et al., 2020, [Nature Methods](#), 17, 261
- Weinberg S., 1972, Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity. John Wiley and Sons, New York
- Weinberg S., 2008, Cosmology. Cosmology, OUP Oxford
- van der Walt, Stéfan Colbert S. C., Varoquaux G., 2011, [Computing in Science & Engineering](#), 13, 22

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.