Learning How to Surf: Reconstructing the Propagation and Origin of Gravitational Waves with Gaussian Processes

Guadalupe Cañas-Herrera, Omar Contigiani, Valeri Vardanyan

Abstract

Soon, the combination of electromagnetic and gravitational signals will open the door to a new era of gravitational-wave (GW) cosmology. It will allow us to test the propagation of tensor perturbations across cosmic time and study the distribution of their sources over large scales. In this work, we show how machine-learning techniques can be used to reconstruct new physics by leveraging the spatial correlation between GW mergers and galaxies. We explore the possibility of jointly reconstructing the modified GW propagation law and the linear bias of GW sources, as well as breaking the slight degeneracy between them by combining multiple techniques. We show predictions roughly based on a network of Einstein Telescopes combined with a high-redshift galaxy survey ($z \lesssim 3$). Moreover, we investigate how these results can be rescaled to other instrumental configurations. In the long run, we find that obtaining accurate and precise luminosity distance measurements (extracted directly from the individual GW signals) will be the most important factor to consider when maximizing the constraining power.

Unified Astronomy Thesaurus concepts: Gravitational wave astronomy (675); Large-scale structure of the universe (902); Gaussian Processes regression (1930); Non-standard theories of gravity (1118)

1. Introduction

The first direct detection of gravitational waves (GWs) by Abbott et al. (2016) triggered a rapidly increasing interest in exploiting this new field for cosmological information. GWs alone are not particularly useful because the data provide only a sky position and a measure of the luminosity distance to the source. However, GW sources can serve as powerful cosmological probes when combined with electromagnetic (EM) data, from which redshifts can be extracted. This idea dates back to Schutz (1986) and has two main variations.

The first and simplest possibility is the observation of so-called standard sirens (Holz & Hughes 2005), which are GW sources with EM counterparts from which a redshift can be observed. As an example, a population of binary neutron stars, such as the already observed GW170817 (Abbott et al. 2017a), can be used to reconstruct the luminosity-redshift function and constrain cosmological observables. In this context, the aforementioned observation has already lead to promising results in constraining the Hubble constant $H_0 = 100 \, h \, \text{km} \, \text{s}^{-1} \, \text{Mpc}^{-1}$ (Abbott et al. 2017b; Chen et al. 2018).

The second possibility is to use dark sirens with a statistical counterpart. The likely counterpart is identified using the rough sky-localization offered by current GW detectors in combination with deep galaxy catalogs. This technique, sometimes combined with the first, has also been used to extract a measurement of $H_0$ (Soares-Santos et al. 2019; Abbott et al. 2021a), and a refined version of it is the focus of this work. Assuming that both galaxies and the hosts of GW mergers are biased tracers of the same cosmological structure, it is possible to measure a nonzero cross-correlation signal between the two (see Section 3). At the linear level, in particular, the description of this signal is especially simple. It should be noted that this idea has been explored by others before, sometimes using a different formalism (see, e.g., Namikawa et al. 2016; Mukherjee & Wandelt 2018; Scelfo et al. 2018), and that similar methods have already been employed to provide a measurement of $H_0$ (Finke et al. 2021).

In addition to these two methods, modeling of higher-order effects in the merger dynamics or knowledge of intrinsic source distribution can also be used to extract the luminosity distance-redshift relation from GW observations alone. Because GW signals provide a measurement of the redshifted chirp-mass of the system, features dependent only on the mass of the system (e.g., tidal effects) can be used to extract the cosmological redshift present in the signal (Messenger & Read 2012). Furthermore, expected features in the source distribution, such as the pair-instability mass gap (Farr et al. 2019) or the peak of the star formation rate (Ye & Fishbach 2021), can also be used to extract this relation from catalogs of GW events.

We also mention that, in principle, the same idea can be used in the absence of resolved events. Previous works, however, have shown that a detection of this effect from a stochastic background signal is unlikely to happen soon due to the low signal-to-noise ratio ($S/N$) attainable by current and proposed experiments covering the optimal wavelength range (Alonso et al. 2020; Cañas Herrera et al. 2020).

In this paper, we discuss the possibility of using the nonzero correlation between the distribution of EM galaxies and resolved GW mergers to jointly extract information about the two main quantities in the field of GW cosmology: the luminosity distance as a function of redshift, describing the propagation of GWs across cosmic time, and the linear bias of GW sources with respect to the underlying dark matter distribution of the Universe, describing their clustering properties. The linear bias in this context is a multiplicative factor relating the spatial correlation function of a source population to the correlation function of the matter distribution of the Universe.

It is already established that GWs carry the potential of constraining the fundamental laws of gravity. This is the case...
because propagation of GWs in modified gravity scenarios differs from predictions of the general theory of relativity (GR) in multiple ways (see, e.g., Deffayet & Menou 2007; Ezquiaga & Zumalacárregui 2020; Garoffolo et al. 2020). One of the clear signatures is the speed of tensor modes, which can be both sub- and superluminal as opposed to the GR case, where GWs propagate at the speed of light. The tight constraints on speed deviations imposed by the multimessenger observations of GW170817 (Abbott et al. 2017c), for example, have ruled out a wide parameter space of otherwise viable scalar-tensor theories of gravity (Lombriser & Taylor 2016; Baker et al. 2017; Creminelli & Vernizzi 2017; Ezquiaga & Zumalacárregui 2017; Sakstein & Jain 2017). Similarly, implications for bigravity models have also been demonstrated in the literature (see Max et al. 2017, 2018; Akrami et al. 2018; Belgacem et al. 2019).

Another striking difference between modified gravity and GR is the modified friction of GWs (Amendola et al. 2018; Belgacem et al. 2018). This feature arises in models with nonminimal coupling of a scalar field and gravity, which manifests itself as a redshift-dependent gravitational coupling. As a result, the inferred luminosity distance to GW sources differs from the corresponding EM luminosity distance. This interesting phenomenon has already been constrained using the aforementioned multimessenger detection of GW170817 (Arai & Nishizawa 2018; Lagos et al. 2019) and the mass distribution of existing GW catalogs (María Ezquiaga 2021). In this work, we investigate the possibility of testing this hypothesis using the spatial clustering of GW sources and galaxies.

The discovery of the first LIGO-Virgo binary black hole has been used to motivate alternative scenarios where the binary did not represent the endpoint of stellar evolution, but originated either as a pair of primordial black holes (PBH) or some exotic compact objects (see, e.g., Bird et al. 2016; Sasaki et al. 2016; Bustillo et al. 2021). In particular, the last half-decade has seen a resurgence in interest for PBHs (see, e.g., Clèses & García-Bellido 2017; Sasaki et al. 2018; Raccanelli et al. 2016, 2018). The main difference between the PBH and stellar evolution scenarios is the spatial distribution of GW mergers, measurable both in the redshift and sky distribution of the sources. In Section 3.1 we expand on how to model this difference through the linear bias and present our model.

In addition to presenting our formalism, we investigate a possible method to precisely reconstruct the redshift evolution of clustering and modified gravity effects in the upcoming decades. Our study is based on Gaussian processes (GPs), a well-known hyperparametric regression method (Rasmussen & Williams 2005). The structure and implementation of this pipeline are presented in Section 4.

GPs have been widely used in the literature to reconstruct the shapes of physical functions such as the dark energy equation of state $w(z)$ (Shafieloo et al. 2012; Gerardi et al. 2019), the primordial inflaton’s speed of sound (Cañas Herrera et al. 2021), or the mass function of the merging binary black hole systems (Li et al. 2021). GPs are very useful for such functional reconstructions due to their flexibility and simplicity, relying only on a handful of hyperparameters. In general, binned reconstructions (Crittenden et al. 2009, 2012; Zhao et al. 2012) as well as parametric reconstructions with high-degree polynomials are similar in spirit to GPs while also offering more flexibility in the extrapolation (Colgáin & Sheikh-Jabbari 2021). The paper is organized as follows. In Section 2 we summarize the essential concepts concerning the GW propagation in modified gravity models. In Section 3 we detail the modeling of the clustering correlations between the GW source population and galaxies. In Section 4 we explain our reconstruction pipeline for the GW luminosity distance and the bias. Our findings are presented in Section 5 and are further discussed in Section 6.

Unless stated otherwise, our fiducial cosmology is based on the best-fit results from Planck 2018 (Aghanim et al. 2020). In our analysis, we use COLOSSUS (Diemer 2018) and Astropy (Robitaille et al. 2013; Price-Whelan et al. 2018) for cosmological calculations, sklearn (Pedregosa et al. 2011) for the GP implementation, emcee (Foreman-Mackey et al. 2013) as our posterior sampler, and GetDist (Lewis 2019) to plot the final contours. Our analysis pipeline is made publicly available, see Vardanyan et al. (2021).

2. Gravitational Wave Propagation

In GR, and around a background Friedmann–Lemaître–Robertson–Walker (FLRW) metric, the amplitude of GWs evolves according to

$$h''_\alpha + 2\omega h'_\alpha - \frac{\omega^2}{c^2} h_\alpha = 0,$$

where $h_\alpha$ denotes the amplitude of either polarization ($\alpha \in \{\perp, \parallel\}$), primes denote derivatives with respect to the conformal time, and $\omega$ is the conformal Hubble function. In this equation, the prefactor of the Laplacian term controls the propagation speed, which we have set to coincide with the speed of light in $c = 1$ units.

The second term is the standard cosmic friction term that causes the strain amplitude to decay as $h_\alpha(z) \propto D_L^{-1}(z)$, with $D_L$ being the FLRW luminosity distance,

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

where the Hubble function $H(z)$ is given in terms of the Hubble constant $H_0$, present-day dark matter abundance $\Omega_m$, and dark energy abundance $\Omega_{DE}(z)$ as

$$H(z) = H_0[\Omega_m(1 + z)^3 + \Omega_{DE}(z)].$$

Throughout this paper, we assume a constant equation of state $w_0$ for dark energy, such that its energy density is given by

$$\Omega_{DE}(z) = (1 - \Omega_m)(1 + z)^{3(1+w_0)}.$$

The standard ΛCDM cosmology corresponds to $w_0 = -1$. It is now established that modifications of GR can affect the propagation of GWs. The important effect for us is the modified friction term with respect to the GR expectation in Equation (1),

$$h''_\alpha + [2 + \alpha_M(z)]\omega h'_\alpha - \frac{\omega^2}{c^2} h_\alpha = 0,$$

where we have again imposed the GW speed to be unity, as suggested by observations. The modified friction term introduces a new scaling $h_\alpha(z) \propto 1/D_{L,GW}(z)$, with $D_{L,GW}(z) \propto D_L(z)$ for nonzero $\alpha_M(z)$. The luminosity distance to GW events can be written as

$$\frac{D_{L,GW}}{D_{L,EM}}(z) = \exp\left\{-\frac{1}{2} \int_0^z \frac{dz'}{H(z')} \frac{\alpha_M(z')}{(1 + z')}\right\}.$$
In this work, we assume that the luminosity distance for EM sources $D_{L,\text{EM}}$ is unaffected and is equal to the expression in Equation (2). The function $\alpha_M$ corresponds to the running of the effective Planck mass, i.e.,

$$\alpha_M = \frac{d \log (M_{\text{eff}}/M_P)}{d \log a},$$

where $M_P$ is the Planck mass and $M_{\text{eff}}$ is its effective value at redshift $z = 1/a - 1$. This function encodes information about extensions of GR such as scalar-tensor theories (Horndeski 1974; Bellini & Sawicki 2014), or more broadly, quantum gravity (Calcagni et al. 2019). The modified friction term is also a natural prediction of nonlocal modifications of gravity (Dirian et al. 2016).

From an effective field theory point of view $\alpha_M(z)$ is a free function of order unity. In practical studies of modified gravity and dark energy, however, $\alpha_M$ is often assumed to take simple parametric forms. The main guiding principle is the assumption that its effects should be negligible in the early universe, which prompts us to choose $\alpha_M(z)$ to be proportional either to the dark energy abundance or simply to some power of the scale factor $a$.

These parameterizations make it possible to find a closed-form expression for the ratio in Equation (6) and have inspired a widely used parameterization of the ratio as a monotonically increasing function of order unity. In practical studies of modifications of gravity (Dirian et al. 2016),

$$\frac{D_{L,\text{GW}}(z)}{D_{L,\text{EM}}(z)} = \Xi_0 + \frac{1 - \Xi_0}{(1 + z)^n}.$$  

In this expression, $\Xi_0$ and $n$ are two constant parameters that are typically $\sim 1$.

### 3. Angular Power Spectra

#### 3.1. GW Sources

We consider GW mergers with a distribution in redshift written as

$$n_{\text{GW}}(z) = \frac{n_0}{1 + z},$$

where $n_0$ corresponds to the comoving number density of observed events as a function of redshift, and the term $(1 + z)$ takes into account the cosmological time dilation. In our analysis, we use a constant value of $n_0 \approx 3 \times 10^{-6} h^3 \text{Mpc}^{-3}$ (with $h$ denoting the usual normalized Hubble constant), motivated by current LIGO constraints (Abbott et al. 2021b).

For a given selection of sources along the line of sight, the average number of projected sources can be written using the comoving distance $\chi(z)$,

$$\bar{n}_{\text{GW}} = \int_0^\infty dz \frac{\chi^2(z)}{H(z)} S(z)n_{\text{GW}}(z).$$

The function $S$ encodes the selection and the scatter due to observational errors. In this paper, simple bins in a range $[D_{L,\text{min}}, D_{L,\text{max}}]$ are used, and we assume a lognormal distribution with fixed scatter $\sigma_{\ln D}$ for the individual sources (Oguri 2016). In this case, $S$ can be written as

$$S(z) = \frac{1}{2} \left[ x_{\text{min}}(z) - x_{\text{max}}(z) \right],$$

with

$$x_{\text{min}}(z) = \text{erfc} \left[ \frac{\ln D_{L,\text{min}} - \ln D_{L,\text{GW}}(z)}{\sqrt{2} \sigma_{\ln D}} \right].$$

and similarly for $x_{\text{max}}$. Including this effect makes $S$ resemble a top-hat function with damping tails dictated by $\sigma_{\ln D}$.

The angular power spectrum of these sources can be written using the Limber approximation,

$$C_{\text{GW}}(\ell) = \int_0^\infty dz \frac{H(z)}{\chi(z)} W_{\text{GW}}^2(z) \times b_{\text{GW}}^2(z) \frac{\ell + 1/2}{\chi(z)},$$

where $P(k, z)$ is the matter power spectrum at redshift $z$ and comoving scale $k$, $b_{\text{GW}}$ is the bias of the GW sources, and the window function can be written as

$$W_{\text{GW}}(z) = \frac{\chi^2(z)}{H(z)} \bar{n}_{\text{GW}} S(z).$$

For the purpose of illustration, we make use of a few simple parameterizations for the GW bias. We consider either a constant bias $b_{\text{GW}}$ with a value of order unity, or a more complex form,

$$b_{\text{GW}}(z) = b_0 \left( 1 + \frac{1}{D(z)} \right).$$

where $D(z)$ represents the growth factor. The first model, with its low constant value, mimics a PBH origin for the mergers (Bird et al. 2016; Raccanelli et al. 2016), while the second mimics the stellar evolution case by tracking the galaxy linear bias (Oguri 2016).

#### 3.2. Galaxies

Similarly to the GW population, we again assume a constant comoving number density of galaxies. Throughout our analysis, we fix

$$n_{\text{gal}}(z) = 10^{-3} h^3 \text{Mpc}^{-3},$$

and we write the autocorrelation signal of galaxies under the Limber approximation as

$$C_{\text{gal}}(\ell) = \int_0^\infty dz \frac{H(z)}{\chi(z)} W_{\text{gal}}^2(z) \times b_{\text{gal}}^2(z) \frac{\ell + 1/2}{\chi(z)},$$

In this expression, the definition of $W_{\text{gal}}$ is the same as $W_{\text{GW}}$ used in the previous section, except for using $n_{\text{gal}}(z)$, a different selection function, and $b_{\text{gal}}(z)$ is the linear galaxy bias. In our analyses, we assume a known galaxy bias in the form of

$$b_{\text{gal}}(z) = 1 + \frac{1}{D(z)}.$$  

In general, this function is expected to be accurately measured from the galaxy autocorrelation signal alone.

In this paper, we employ a top-hat selection function for $W_{\text{gal}}$, which assumes no uncertainty in galaxy redshift estimates. This choice mimics a spectroscopic galaxy survey.
or a general redshift survey with negligible uncertainties. As an example, another choice commonly found in the literature is a Gaussian distribution \(\mathcal{N}(z, \sigma_{\text{gal}})\), where \(\sigma_{\text{gal}}\) should be much larger than the expected redshift uncertainty for each individual galaxy.

By combining the distribution of GW sources and galaxies, one can construct a cross-correlation map. In our formalism, we write the cross-correlation between a GW bin \(i\) and a galaxy bin \(j\) (fully specified by their respective window functions) as

\[
C^i_j(\ell) = \int_0^\infty \frac{dz}{\chi^2(z)} W^i_{\text{GW}}(z) W^j_{\text{gal}}(z) \times b_{\text{GW}}(z)b_{\text{gal}}(z) \left( \frac{\ell + 1/2}{\chi(z)}, z \right).
\]

We conclude this section by pointing out that the power spectra in Equations (17), (13), and (19) do not include relativistic terms and do not capture the effects of evolution and lensing bias (see, e.g., Scelfo et al. 2018, 2020, for a detailed treatment). Specifically, while the effects of lensing are expected to be negligible at the redshifts considered here (\(z < 3\), see, e.g., Oguri 2016; Contigiani 2020; Mukherjee et al. 2020), the same is not true for relativistic effects. Therefore we choose not to consider low values of \(\ell\) in our analysis because the signal at these large angular scales is dominated by them.

4. Reconstructing GW Physics

The primary goal of the paper is to demonstrate the properties of GW propagation and source clustering can be reconstructed as a function of redshift. We do so by showing how an assumed fiducial model can be recovered by using mock angular power spectra with cosmic-variance or shot-noise limited uncertainties.

Our methodology hinges on the fact that by cross-correlating a GW luminosity distance bin with multiple galaxy redshift bins, we can determine the redshift of the GW sources by matching the clustering properties of the two at the true redshift (Oguri 2016; Bera et al. 2020).

We demonstrate this idea in Figure 1, where we have considered GW sources located at redshift \([0.9, 1.1]\) in a GR cosmology where \(D_{\text{L,GW}}(z) = D_{\text{L,EM}}(z)\). In this figure, we show the expected cross-correlation signal between the angular distribution of these sources and the angular distribution of galaxies located at various redshifts. As expected, in GR (\(\Xi_0 = 1\)) the signal peaks inside the correct redshift range (shaded area). However, as we depart from the GR luminosity distance relation, the location of this peak is affected.

4.1. Mock Data and Fiducial Model

In this section, we describe the recipe we used to generate the mock angular power spectra \((C_{\text{gal}}, C_{\text{GW}}\), and \(C_{\alpha}\)) that are fed into our reconstruction pipeline together with their error covariance matrix. When describing real data, these angular power spectra are extracted from the autocorrelation and cross-correlation maps representing the sky distribution of galaxies and GW sources. The recipe has three main ingredients: the details of the fiducial model, a description of the instrumental configuration, and a definition of the dominant source of error.

The first ingredient is the fiducial model. Our decision in this case is based on the results of Baker & Harrison (2021), where present-day constraints on the function \(\alpha_M\) appearing in Equation (6) are presented. As shown in Belgacem et al. (2019), the results of the \(\alpha_M \propto a\) parameterization found in that work can be mapped to the \(\Xi(z)\) function in Equation (8). Using this transformation, we find that the \(3\sigma\) upper limit roughly corresponds to

\[
\Xi_0 \lesssim 1.4,
\]
As for the last ingredient, we assume cosmic-variance or shot-noise limited uncertainties. In this case, we can write the covariance matrix of the autocorrelation and cross-correlation signals defined in Equations (13), (17), and (19) as

\[
\text{Cov} \{C^{ij}(\ell) C^{mn}(\ell')\} = \frac{\delta_{\ell\ell'}}{2(2\ell + 1)f_{\text{sky}}} \times \left( C^{im} C^{jn} + C^{in} C^{jm} \right),
\]

(21)

where the indices \(i, j, m, n\) can represent either galaxy or GW bins. The terms \(C^{im}\) contain the shot-noise contribution when they represent the autocorrelations in the same bin,

\[
C^{im}(\ell) = C^{mn}(\ell) + \frac{\delta_{im}}{\bar{n}},
\]

(22)

where \(\bar{n}\) is the average density of projected objects from Equation (10). In this work, we assume a survey covering a sky fraction equal to \(f_{\text{sky}} = 0.5\).

Let us point out that we do not use the cross-bin correlations for the bins of the same type as a signal. However, we properly take into account the \(C^{im}\) terms for overlaps between two different GW bins. Similar terms for galaxy bins are completely negligible as there are no overlaps between the spectroscopic redshift bins.

For the setup described in this section, we find a total S/N of the GW-gal and GW-GW angular power spectra of \(\sim 37\). This value is dominated by the GW-gal cross-correlations because the GW-GW autocorrelations are not well measured (S/N \(\lesssim 6\)).

To generalize our choices, in Section 5.2 we expand on the effect of different combinations of instrumental specifications on the precision of the reconstruction.

### 4.2. Reconstruction Pipeline

In this section, we describe how the mock data presented in the previous section can be used to reconstruct \(D_{L,GW}\) and \(b_{GW}\) as a function of \(\zeta\). For ease of interpretation and visualization, in our analysis we do not fit these functions directly, but instead focus on the ratios \(D_{L,GW}/D_{L,\text{EM}}(\zeta)\) and \(b_{GW}/b_{\text{gal}}(\zeta)\). We point out that we do not marginalize over different possibilities for \(b_{\text{gal}}(\zeta)\). This is because we are not interested in exploring the properties of the galaxy population, which are expected to be very well constrained by the galaxy-galaxy correlation signal alone.

As emphasized earlier, our approach makes use of a GP regression. This method is often used when the function of interest is directly measured at certain redshifts. These measurements are used as a training sample of the GP model, which can then predict the values of the reconstructed function at redshifts lacking any direct measurements (Belgacem et al. 2020; Levi Said et al. 2021; Mukherjee & Mukherjee 2021; Perenon et al. 2021; Renzi et al. 2021).

However, this is not directly applicable for our current problem as neither the \(D_{L,GW}/D_{L,\text{EM}}(\zeta)\) nor the \(b_{GW}/b_{\text{gal}}(\zeta)\) are directly observable. Instead, these functions determine the auto- and cross-angular power spectra, which constitute our direct observables. In order to use GPs for our problem, we consider a certain number of redshift nodes for the two functions, referred to as training nodes with a slight abuse of terminology. The amplitudes of the nodes are free, and given a node configuration, we consider GPs that pass through all of these nodes exactly. To render our scenario computationally feasible and not consider many functions for each node configuration, we use the GPs regressor of the python package sklearn to output the best fit and use this as our function.

Our use of GPs can be thought of as a binning of the functions of interest in redshift space and imposing a certain prior correlations between the bins. These correlations are specified by the GP kernel function, which in our case is chosen to be

\[
\kappa(z_i, z_j; L) \propto \exp \left\{ -\frac{1}{2} \left( \frac{|z_i - z_j|}{L} \right)^2 \right\},
\]

(23)

where \(L\) is the so-called correlation length. This kernel is flexible enough for our purposes, and we do not expect the detailed choice to have any significant impact on our results.

For computational purposes, we generate the GPs using a baseline around \(D_{L,GW}/D_{L,\text{EM}}(\zeta) = 1\). This baseline causes the GPs reconstruction to efficiently return to \(D_{L,GW}/D_{L,\text{EM}}(\zeta) = 1\) when not pushed toward other values by the training nodes.

This process is described in Figure 2 for two values of the correlation length \(L\). While pictured in this example, we exclude negative-valued functions for physical reasons in our analysis when we explore \(D_{L,GW}/D_{L,\text{EM}}(\zeta)\) and \(b_{GW}/b_{\text{gal}}(\zeta)\), and also nonmonotonic realizations of \(D_{L,GW}(\zeta)\).

The goal of our statistical analysis is to explore the possible constraints on the shape of both \(D_{L,GW}/D_{L,\text{EM}}(\zeta)\) and \(b_{GW}/b_{\text{gal}}(\zeta)\). To do this, we aim to sample the posterior distributions of the amplitudes of the nodes as well as the cosmological parameters. The correlation length \(L\) can in principle be fixed based on theoretical priors. Lacking such priors in our case, we only impose a wide uniform prior on \(L\) and consider it as a free parameter (see Table 1).

Each step of the sampling process consists of producing two GP curves—one for each \(D_{L,GW}/D_{L,\text{EM}}(\zeta)\) and \(b_{GW}/b_{\text{gal}}(\zeta)\)—given the current set of training node amplitudes. The curves are used in the calculation of theoretical auto- and cross-correlation power spectra described in Section 3. The theoretical power spectra enter the Gaussian likelihood together.
with the generated mock data. The theoretical predictions are computed using our python code, which is interfaced with the emcee sampler. The typical runs with varying cosmology take approximately 10 hr on a modern machine.

After obtaining the posterior distribution of the nodes, we reverse-engineer the problem to obtain confidence contours for each $D_{L, GW}/D_{L, EM}(z)$ and $b_{GW}/b_{gal}(z)$. For all the sampled node amplitudes we generate the corresponding GP profiles on finite but sufficiently many redshift points and calculate the 68% and 95% confidence intervals at each redshift using the statistical python package GetDist.

To assess the impact of different cosmological backgrounds and clustering properties, we include in our reconstruction three nuisance parameters: the dark matter abundance $\Omega_m$, the Hubble constant $H_0$, and the dark energy equation of state parameter $w_0$. A summary of our model parameters and the priors used in this work is presented in Table 1.

We would like to point out that upcoming galaxy surveys will measure these parameters with very high precision. While GWs alone might be able to provide competitive constraints, the focus of our analysis is not on constraining them. Rather, we would like to quantify how accurately the GW luminosity distance and source properties can be measured. These properties are not accessible to generic redshift surveys and can only be measured with the use of GW-specific observables. This reasoning justifies our tight Gaussian priors on the aforementioned cosmological parameters.

### 5. Results

#### 5.1. Reconstructions

In our reconstructions, we always impose the $D_{L, GW}/D_{L, EM}(z)$ to become unity at redshift zero by placing a fixed node at $z = 0$ with an amplitude of 1. In addition to this fixed node, we have four nodes for each of the reconstructed functions. We have arrived at this number by gradually increasing the number of nodes and monitoring the goodness of fit. In practice, we have monitored the Akaike information criterion (AIC) for one-, two-, and three-node setups for the GW luminosity distance. Our experiments suggest that as expected, the one-node configuration is significantly worse than the presented four-node setup. The two- and three-node setups have similar performance compared to the four-node case, but the latter still outperforms the former two. We then use the same number of nodes for the bias reconstruction.

In our main analysis, the redshift positions $z_i$ of the nodes are fixed. We have performed an experimental run, however, to assess the impact of letting them free in reduced uniform prior ranges. The result of this experimentation is that the node locations remain unconstrained, and the final posterior of, e.g., $D_{L, GW}/D_{L, EM}(z)$ does not change when the node redshift locations are being sampled as free parameters. This implies that the exact locations of the GP notes are unimportant because they are uniformly distributed in the redshift range of interest.

It is worth emphasizing that when our methodology is applied to real data, the number of nodes, as well as their exact redshift placements, should be constrained by performing similar, but more systematic experiments. Particularly, more accurate measures, such as Bayesian evidence ratios, should be employed. Moreover, if enough data are used, some possible constraints could be found by letting the training nodes be completely free in the entire redshift range of interest. As the resulting posteriors are expected to be multimodal, this should be investigated using nested sampling algorithms.

As mentioned earlier, we also explore the GP correlation lengths both for the bias and the luminosity distance. On physical grounds, we are interested in smooth GP functions and have imposed the minimum of the uniform prior range for $L$ to be on the order of the inter-node distance so that the smoothness is maintained. We find that as expected, both of the correlation lengths remain unconstrained within the imposed wide prior ranges.

The results of our joint reconstruction is presented in Figure 3. In the same figure, we also compare these constraints to different theoretical models. In the case of $D_{L, GW}/D_{L, EM}(z)$, we

---

Table 1

| Parameter   | Prior                        |
|-------------|------------------------------|
| Node amplitudes | [0, 11] (Uniform)     |
| Correlation length (L) | [1, 10] (Uniform)    |
| $\Omega_m$   | 1% (Gaussian)               |
| $h$          | 1% (Gaussian)               |
| $w_0$        | 5% (Gaussian)               |

Note. The GP hyperparameters (i.e., the two correlation lengths and the $4 \times 2$ amplitudes) are explored independently. The fiducial model is given by $\Xi_0 = 1.4$, $n = 1$, $\Omega_m = 0.31$, $h = 0.67$, and $w_0 = -1$.

---

Figure 3. Confidence intervals (68%, in black, and 95%, in lighter gray) of the jointly reconstructed functions $D_{L, GW}/D_{L, EM}(z)$ and $b_{GW}/b_{gal}(z)$. Together with the assumed fiducial model, we also plot the expectation for different models (see text for more details). The vertical lines mark the fixed location of the nodes used in the GP reconstruction.
we use the parameterization

$$\alpha_M(z) = \alpha_0 \left[ \frac{H_0}{H(z)} \right]^2, \quad (24)$$

where we use Equation (3) with $a_0 = -1$ to obtain the plotted lines (Belgacem et al. 2019). On the other hand, for $b_{GW}/b_{gal}(z)$, we plot the lines corresponding to constant values of $b_{GW}(z)$ while keeping the galaxy bias fixed to the expression in Equation (18).

In principle, the output of our sampling can also be used to reconstruct the function $\alpha_M(z)$ by calculating the numerical derivative of $D_{L,GW}/D_{L,EM}(z)$. For the purposes of this paper, however, we chose not to do this. The kernel in Equation (23) can be interpreted as a smoothness prior, and the value of $\alpha_M(z)$ is directly affected by it. Because of this, if one is interested in inferring $\alpha_M(z)$, GPs should be used to sample this function directly.

As expected, we observe that the fiducial models for both $D_{L,GW}/D_{L,EM}(z)$ and $b_{GW}/b_{gal}(z)$ are well encoded within the reconstructed confidence contours in both panels of Figure 3. The constraints at higher redshift ($z \approx 3$) for both reconstructions are broader. This is an effect that could not be seen if a parametric function were used for $D_{L,GW}/D_{L,EM}(z)$, for instance, as the parameterization would have fixed the behavior similarly at low and higher redshifts.

Finally, in Figure 4 we show the correlation between these functions. We observe weak but nonzero correlations between the GW bias and the luminosity distance. In general, parametric models might induce nonphysical correlations. GPs are expected to behave better in this regard, but they can still induce spurious correlations due to finite correlation lengths. For consistency, we have also performed a reconstruction using a redshift binning approach and concluded that we can recover a very similar correlation structure with either method.

5.2. Signal-to-noise Scaling

The constraining power of our method crucially depends on a number of observational specifications. The most relevant parameters are 1) the angular sensitivity, specified by the maximum multipole $\ell_{\text{max}}$ of the angular power spectra; 2) the number of GW sources, which is specified by the comoving number density $n_{GW}$; and 3) the precision of the GW luminosity distance measurements $\sigma_{\nu_D}$. In the case of $n_{GW}$, we adjust the value of $n_0$ in Equation (9) as a way to explore different values of the total number of observed GW events, $N = 4\pi f_{\nu_{GW}} b_{GW}$. In principle, this should include selection effects that are not captured by our formalism. Obviously, for a given experimental configuration, the mentioned three variables are not independent, but it is still interesting to find the dependence of our results on each one of them separately. This allows us to reach conclusions without relying on specific experiments, and to suggest potential design guidelines for future GW detectors.

To attain such insights, in this subsection we consider constraints on the parametric expression in Equation (8), as well as on the parametric GW bias given by Equation (15). For simplicity, we fix $n = 1$ and only constrain the parameter $\Xi_0$.

When we varied $\ell_{\text{max}}$ and $N$, we kept the remaining configuration fixed (including the luminosity distance binning). Each case of $\sigma_{\nu_D}$, on the other hand, is accompanied by an adjustment in the number of luminosity distance bins. This is done to be consistent with our binning strategy, namely that the luminosity distance width of each bin is at least $\mathcal{O}(3)$ times wider than $\sigma_{\nu_D}$.

Our results are summarized in Figure 5, where we plot the anticipated uncertainties in $\Xi_0$ (upper panel) and $b_0$ (lower panel) as a function of the $S/N$ of the cross-correlation in Equation (19).

For a fixed $\sigma_{\nu_D}$, the constraining power on $\Xi_0$ and $b_0$ is almost completely determined by the cross-correlation $S/N$. This fact suggests that regardless of how the given S/N is realized (either by increasing the number of sources or by improving the angular sensitivity), the expected constraints are the same. This implies that the results presented in this paper can be easily scaled to different configurations. As expected, we find that the constraints scale as $1/S/N$.

The situation is somewhat different for the case of varying $\sigma_{\nu_D}$ and the number of luminosity distance bins. The constraints on the bias still follow the same form (see the lower panel), but the scaling of the $\Xi_0$ constraints, on the other hand, is much steeper than in the cases of varying $\ell_{\text{max}}$ and $n_{GW}$, roughly $1/S/N^3$. This fact can be qualitatively understood by recalling the importance of the relative positions of GW and galaxy window functions demonstrated in Figure 1. Sampling this relation with a larger number of window functions increases the precision of our reconstruction.

The results presented in this section quantify the importance of accurate luminosity distance measurements and demonstrate the benefit that lower values of $\sigma_{\nu_D}$ can bring to a binned approach.
6. Discussion and Conclusions

In this paper, we have shown how the combination of GW observations and redshift surveys can be exploited in the era of GW cosmology. We have identified two essential quantities that characterize this new research field: 1) the luminosity distance relation, a clear imprint of modifications to the propagation of tensor modes (see Section 2), and 2) the bias of the sources, regulating their spatial distribution and betraying their origin.

Proposed GW detectors such as the Einstein Telescope (Maggiore et al. 2020) or Cosmic Explorer (Reitze et al. 2019) are expected to probe a sizeable fraction of the visible Universe and produce large statistical samples. In this context, we point out that the number of sources assumed for our main analysis, $10^5$, is particularly conservative and differs from expectations by at least an order of magnitude (Maggiore et al. 2020). This difference is primarily due to our assumption of a constant comoving density of events. While we do not explore other assumptions for the distribution of GW sources over cosmic time, we have investigated similar effects in Section 5.2, where we have shown how our results can be rescaled to other instrument configurations or to a different number of observed sources.

Our formalism, based on binned angular power spectra and sky maps, is optimal for a large number of sources with no known counterpart. Its main advantages are related to the simple modeling of the theoretical signals and their data covariance matrix. Because no reconstruction of the underlying density field is necessary, the predictions display a clear separation of scales. For example, the angular scales that we have considered are all within the linear regime ($k \lesssim 0.1 \text{ Mpc}^{-1}$). Furthermore, because this formalism is well established, our shot-noise-limited covariance matrix can be easily generalized to include additional sources of (co-) variance.

Although a comprehensive comparison between multiple approaches is beyond the scope of this work, it is worth discussing how our results compare to others found in the literature. We preface this by saying that one-to-one comparisons, however, are often complicated either by significantly different assumptions or the impossibility of directly translating these assumptions from one prescription to another. Despite this, here we draw a parallel between our method and two other methods.

The first method is the one used in Mukherjee et al. (2021), which has also been shown to be extremely successful in measuring both $p_{GW}(z)$ and $D_{L, GW}(z)$ using parametric models. Similarly to this work, the information is also extracted from the cross-correlation with redshift sources, but no binning of the GW data is performed. In this case, we have verified that such methods perform significantly better than our map-based approach in the case of a low number density of GW sources and large uncertainty in the measured $D_{L, EM}(z)$. These features, in particular, make it especially useful for near-future samples of a few dozen objects.

The second promising method for measuring $D_{L, GW}(z)$ that has been proposed in the literature is offered by GW sources with known counterparts. These observations give direct access to $D_{L, GW}$ as a function of redshift and can be combined with similar measurements in the EM spectrum to obtain $D_{L, GW}/D_{L, EM}(z)$. The analysis of Belgacem et al. (2020) is based on this methodology, and similarly to ours, also employs GPs to reconstruct this ratio from an Einstein Telescope sample with $\sim 10^2$ sources.

Ultimately, we expect this counterpart-based formalism and the one described in this work to be complementary: a direct measurement of $D_{L, GW}(z)$ can be used to break the degeneracy between bias and luminosity distance shown in Figure 4. However, because the fraction of events with known counterparts that will be observed is heavily dependent on both the GW source distribution and multiple instrumental setups, we do not attempt to combine the two methods here.

In conclusion, the combination of GW resolved events and the clustering of galaxies is expected to improve our current knowledge of the physical properties of the Universe. Our work shows how to reconstruct these properties as a function of redshift in a generic way, and it highlights the need for accurate and precise measurements of $D_{L, GW}$. This will require control over the instrument calibration uncertainties (Cahillane et al. 2017), but also the degeneracy between the inclination of the source and its luminosity distance (Ghosh et al. 2016). In the future, we aim to apply our current analysis pipeline to the next generation of large-scale structure surveys and incoming GW observations.

We thank Eoin Ó Colgáin, Masamune Oguri, and Masahiro Takada for useful conversations. G.C.H. and O.C. are supported by a de Sitter fellowship from the Dutch Research Council (NWO). G.C.H. acknowledges support from the Delta Institute for Theoretical Physics (D-ITP consortium), a program...
by the NWO. The work of V.V. is financed by the WPI Research Center Initiative, MEXT, Japan, and by JSPS KAKENHI grant number 20K22348.

Software: COLOSSUS (Diemer 2018), Astropy. (Robitaille et al. 2013; Price-Whelan et al. 2018), sklearn (Pedregosa et al. 2011), emcee (Foreman-Mackey et al. 2013), GetDist (Lewis 2019), GW-Cosmo (Vardanyan et al. 2021).

ORCID iDs
Guadalupe Cañas-Herrera @ https://orcid.org/0000-0003-2796-2149
Omar Contigiani @ https://orcid.org/0000-0001-7699-6834
Valeri Vardanyan @ https://orcid.org/0000-0002-8496-5859

References
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016, PhRvL, 116, 061102
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017a, PhRvL, 119, 161101
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017b, Natur, 551, 85
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017c, ApJL, 848, L13
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2021a, ApJ, 909, 218
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2021b, ApJL, 913, L7
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2021c, ApJL, 883, L42
Finke, A., Foffa, S., Iacovelli, F., Maggiore, M., & Mancarella, M. 2021, JCAP, 2021, 026
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Garoffolo, A., Tasinato, G., Carbone, C., Bertacca, D., & Matarrese, S. 2020, JCAP, 2020, 040
Gerardi, F., Martinelli, M., & Silvestri, A. 2019, JCAP, 2019, 042
Ghosh, A., Del Pozzo, W., & Ajith, P. 2016, PhRvD, 94, 104070
Hall, E. D., & Evans, M. 2019, CQGra, 36, 225002
Holz, D. E., & Hughes, S. A. 2005,ApJ, 629, 15
Horndeski, G. W. 1974, JETP, 10, 363
Lagos, M., Fishbach, M., Landry, P., & Holz, D. E. 2019, PhRvD, 99, 083504
Levi Said, J., Misfsud, J., Sultana, J., & Adami, K. Z. 2021, JCAP, 2021, 015
Lewis, A. 2019, arXiv:1910.13970
Li, Y.-J., Wang, Y.-Z., Han, M.-Z., et al. 2021, ApJ, 917, 33
Lombriser, L., & Taylor, A. 2016, JCAP, 2016, 031
Maggiore, M., Van Den Broeck, C., Bartolo, N., et al. 2020, JCAP, 2020, 050
Maria Ezquiaga, J. M. 2021, arXiv:2104.05139
Max, K., Platscher, M., & Smirnov, M. 2017, PhRvL, 119, 111101
Max, K., Platscher, M., & Smirnov, M. 2018, PhRvD, 97, 064009
Messinger, C., & Read, J. 2012, PhRvL, 108, 091101
Mukherjee, P., & Mukherjee, A. 2021, MNRAS, 504, 3938
Mukherjee, S., & Wandelt, B. D. 2018, arXiv:1808.06615
Mukherjee, S., Wandelt, B. D., & Silk, J. 2020, MNRAS, 494, 1956
Mukherjee, S., Wandelt, B. D., & Silk, J. 2021, MNRAS, 502, 1136
Namikawa, T., Nishizawa, A., & Taruya, A. 2016, PhRvD, 94, 024013
Oguri, M. 2016, PhRvD, 93, 083511
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, J. Mach. Learn. Res., 12, 2852, http://jmlr.org/papers/v12/pedregosa11a.html
Perenon, L., Martinelli, M., Ilic, S., et al. 2021, arXiv:2105.06163 Price-Whelan, A. M., Sipőcz, B. M., G"unther, H. M., et al. 2018, AJ, 156, 123
Raccanelli, A., Kovetz, E. D., Bird, S., Cholis, I., & Mounoz, J. B. 2016, PhRvD, 94, 023516
Raccanelli, A., Vidotto, F., & Verde, L. 2018, JCAP, 2018, 003
Rasmussen, C. E., & Williams, C. K. I. 2005, Gaussian Processes for Machine Learning. Adaptive Computation and Machine Learning (Cambridge, MA: The MIT Press)
Reitze, D., Adhikari, R. X., Ballmer, S., et al. 2019, BAAS, 51, 35
Renzi, F., Hogg, N. B., Martinelli, M., & Nesseris, S. 2021, PDU, 32, 100824
Robitaille, T. P., Tolliver, E. J., Greenfield, P., et al. 2013, A&A, 558, A33
Sadeghian, A., Jain, B. 2017, PhRvL, 119, 251303
Sasaki, M., Suyama, T., Tanaka, T., & Yokoyama, S. 2016, PhRvL, 117, 061101
Sasaki, M., Suyama, T., Tanaka, T., & Yokoyama, S. 2018, CQGra, 35, 063001
Sclafino, G., Bellomo, N., Raccanelli, A., Matarrese, S., & Verde, L. 2018, JCAP, 2018, 039
Sclafino, G., Boco, L., Lapi, A., & Viel, M. 2020, JCAP, 2020, 045
Schutz, B. F. 1986, Natur, 323, 310
Shafeefoo, A., Kim, A. G., & Linder, E. V. 2012, PhRvD, 85, 23530
Soares-Santos, M., Palmese, A., Hartley, W., et al. 2019, ApJL, 876, L7
Vardanyan, V., Cañas Herrera, G., & Contigiani, O. 2021, GW-cosmo, v1.0.0, Zenodo, doi:10.5281/zenodo.4905445
Weltman, A., Bull, P., Camera, S., et al. 2020, PASA, 37, e002
Ye, C., & Fishbach, M. 2021, PhRvD, 104, 043507
Zhao, G.-B., Crittenden, R. G., Pogosian, L., & Zhang, X. 2012, PhRvL, 109, 171301

Ezquiaga, J. M., & Zumalacárregui, M. 2020, PhRvD, 102, 124048
Farr, W. M., Fishbach, M., Ye, J., & Holz, D. E. 2019, ApJL, 883, L42
Finke, A., Foffa, S., Iacovelli, F., Maggiore, M., & Mancarella, M. 2021, JCAP, 2021, 026