Constraining astrophysical observables of galaxy and supermassive black hole binary mergers using pulsar timing arrays

Siyuan Chen,[1,2,3,4]★ Alberto Sesana[1,5] and Christopher J. Conselice[6]

[1]*Institute of Gravitational Wave Astronomy and School of Physics and Astronomy, University of Birmingham, Birmingham B15 2TT, UK*
[2]*Station de Radioastronomie de Nançay, Observatoire de Paris, PSL University, CNRS, Université d'Orléans, F-18330 Nançay, France*
[3]*FEMTO-ST Institut de recherche, Department of Time and Frequency, UBFC and CNRS, ENSMM, F-25030 Besançon, France*
[4]*Laboratoire de Physique et Chimie de l'Environnement et de l'Espace, LPC2E UMR7328, Université d'Orléans, CNRS, F-45071 Orléans, France*
[5]*Dipartimento di Fisica 'G. Occhialini', Universita' degli Studi di Milano-Bicocca, Piazza della Scienza 3, I-20126 Milano, Italy*
[6]*Centre for Astronomy and Particle Theory, University of Nottingham, Nottingham NG7 2RD, UK*

ABSTRACT

We present an analytic model to describe the supermassive black hole binary (SMBHB) merger rate in the Universe with astrophysical observables: galaxy stellar mass function, pair fraction, merger time-scale, and black hole–host galaxy relations. We construct observational priors and compute the allowed range of the characteristic spectrum h_c of the gravitational wave background (GWB) to be $10^{-16} < h_c < 10^{-15}$ at a frequency of $f = 1 \, \mathrm{yr}^{-1}$. We exploit our parametrization to tackle the problem of astrophysical inference from pulsar timing array (PTA) observations. We simulate a series of upper limits and detections and use a nested sampling algorithm to explore the parameter space. Corroborating previous results, we find that the current PTA non-detection does not place significant constraints on any observables; however, either future upper limits or detections will significantly enhance our knowledge of the SMBHB population. If a GWB is not detected at a level of $h_c(f = 1 \, \mathrm{yr}^{-1}) = 10^{-17}$, our current understanding of galaxy and SMBHB mergers is disfavoured at a 5σ level, indicating a combination of severe binary stalling, overestimating of the SMBH–host galaxy relations, and extreme dynamical properties of merging SMBHBs. Conversely, future detections of a Square Kilometre Array (SKA)-type instrument will allow to constrain the normalization of the SMBHB merger rate in the Universe, the time between galaxy pairing and SMBHB merging, the normalization of the SMBH–host galaxy relations and the dynamical binary properties, including their eccentricity and density of stellar environment.

Key words: black hole physics – gravitational waves – methods: data analysis – pulsars: general – galaxies: evolution – galaxies: formation.

1 INTRODUCTION

It is well established that supermassive black holes (SMBHs) reside at the centre of massive galaxies (see e.g. Event Horizon Telescope Collaboration et al. 2019), and that their masses correlate with several properties of the hosts (see Kormendy & Ho 2013, and references therein). In the hierarchical clustering scenario of structure formation (White & Rees 1978), the SMBHs hosted in merging galaxies sink to the centre of the merger remnant because of dynamical friction, eventually forming a bound SMBH binary (SMBHB) at parsec scales (Begelman, Blandford & Rees 1980). The binary subsequently hardens because of (hydro)dynamical

interaction with the dense background of stars and gas (see Dotti, Sesana & Decarli 2012, for a review), until gravitational wave (GW) emission takes over at sub-parsec separations, leading to the final coalescence of the system (Peters & Mathews 1963). Upon coalescence, the frequency emitted by an SMBHB of mass M at the last stable orbit is $f_{\mathrm{LSO}} \approx 4/M_9 \mu\mathrm{Hz}$, where $M_9 = M/10^9 \, \mathrm{M}_\odot$, making inspiralling SMBHBs the loudest sources in the Universe of sub-μHz GWs. This frequency regime is accessible via precise timing of millisecond pulsars (MSPs). The most massive systems closest to Earth might be powerful enough to be detected as individual deterministic sources at nano-Hz frequencies (Sesana, Vecchio & Volonteri 2009; Mingarelli et al. 2017). There are, however, $\approx 10^{10}$ massive galaxies in the Universe. If each of them experienced one or more major merger in its lifetime and if the resulting SMBHB emits GWs at nano-Hz frequencies for ≈ 1 Myr (Mingarelli et al. 2017),

★ E-mail: siyuan.chen@cnrs-orleans.fr

then there are, at any time ≈ 1 million SMBHBs emitting GWs in the frequency band probed by pulsar observations, resulting in an unresolved stochastic GW background (GWB, see e.g. Rajagopal & Romani 1995; Jaffe & Backer 2003; Sesana, Vecchio & Colacino 2008; Ravi et al. 2012).

A GWB affects the time of arrivals (TOAs) of radio pulses emitted by an ensemble of MSPs in a characteristic and correlated fashion (Hellings & Downs 1983). Pulsar timing arrays (PTA, Foster & Backer 1990) search for GWs using this Hellings & Downs correlation. Although a GWB has not been detected yet, the three currently leading PTAs – the European Pulsar Timing Array (EPTA, Desvignes et al. 2016), the North American Nanohertz Observatory for Gravitational Waves (NANOGrav, Arzoumanian et al. 2018), and the Parkes Pulsar Timing Array (PPTA, Reardon et al. 2016) – already produced stringent upper limits (Shannon et al. 2015; Lentati et al. 2015; Arzoumanian et al. 2018). The three PTAs work together under the aegis of the International Pulsar Timing Array (IPTA, Verbiest et al. 2016), with the goal of building a larger TOA data set to improve sensitivity. With the contribution of emerging PTAs in India, China, and South Africa, a detection is expected within the next decade (Rosado, Sesana & Gair 2015; Taylor et al. 2016b; Kelley et al. 2017b).

Besides detecting the low-frequency GWB, the final goal of PTAs is to extract useful astrophysical information from their data to address the 'inverse problem'. Since the GWB shape and normalization depends on the statistical properties of the SMBHB population and on the dynamics of individual binaries in their late inspiral (see Sesana 2013a, for a general discussion of the relevant processes), stringent upper limits, and eventually a detection, will allow to gain invaluable insights in the underlying relevant physical processes. In fact, the most stringent upper limits to date have been already used to place tentative constrains on the population of SMBHBs (Simon & Burke-Spolaor 2016; Taylor, Simon & Sampson 2017; Middleton et al. 2018). Generally speaking, the normalization of the GWB depends on the cosmic SMBHB merger rate, and its shape on the typical SMBHB eccentricity and on the effectiveness of energy and angular momentum loss to the dense environment of gas and stars surrounding the binary. Arzoumanian et al. (2016) investigated the implications of the NANOGrav nine-year upper limits on several astrophysical ingredients defining the underlying SMBHB population model. They found that the data prefer low SMBHB merger rate normalization, light SMBHs for a given galaxy mass (i.e. a low normalization of the SMBH–host galaxy scaling relation), eccentric binaries, and dense stellar environment. Their analysis should be taken as a proof of concept, since each parameter was investigated separately, keeping all the other fixed. In a subsequent extension of the work, Simon & Burke-Spolaor (2016) have shown that it is possible to use that limit to constrain simultaneously the parameters describing the $M_{BH}-M_{bulge}$ and the typical SMBHB merger time-scale, but still keeping other relevant parameters within a narrow prior range and assuming a GWB characteristic strain, h_c, described by a $f^{-2/3}$ power law, appropriate for circular, GW-driven binaries (thus not considering the detailed SMBHB dynamics). Taylor et al. (2017) focused on the determination of the parameters driving the dynamical evolution of individual binaries, showing with detailed GWB simulations interpolated by means of Gaussian processes, that eccentricity and density of the stellar environment can be constrained for a specific choice of the SMBHB merger rate. Finally, a more sophisticated astrophysical inference investigation was conducted in Arzoumanian et al. (2018), including model selection between different

SMBHB population models from the literature, and constrains on the SMBHB eccentricity and environment density for different $M_{BH}-M_{bulge}$ scaling relations.

In Middleton et al. (2016), we started a long-term project of creating a general framework for astrophysical inference from PTA data. In Chen, Sesana & Del Pozzo (2017b), we presented a fast and flexible way to compute the stochastic GWB shape for a general parametrization of the SMBHB merger rate and the relevant properties defining the SMBHB dynamics. We demonstrated the versatility of our model on synthetic simulations in Chen et al. (2017a) and eventually applied it to the most stringent PTA limit to date in Middleton et al. (2018). This latter study, in particular, was instrumental in demonstrating that PTA upper limits are not in tension with our current understanding of the cosmic galaxy and SMBH build-up.

Although previous work has focused on particular bits of physics contributing to the amplitude and shape of the GWB spectrum, we combine here all ingredients of the SMBHB merger rate into one overall model to simultaneously contrain the entire parameter space without keeping certain aspects fixed. We make in this paper an important step towards this goal by rewriting our model of the SMBHB merger rate as a parametric function of astrophysical observables rather then considering a purely phenomenological form. In fact, as shown in Sesana (2013b), the SMBHB merger rate can be derived from the galaxy stellar mass function (GSMF), the galaxy pair fraction, the SMBHB merger time-scale, and the scaling relation connecting SMBHs and their hosts. By expressing the SMBHB merger rate as a function of simple analytical parametrizations of these ingredients – constrained by independent observations –, we build a GWB model that allows to use PTA observations to constrain a number of extremely relevant astrophysical observables.

The paper is organized as follows. Section 2 summarizes the model to compute the characteristic strain of the GWB and highlights the changes introduced in this paper. Section 3 derives the parametric formulation of the SMBHB merger rate as a function of all the relevant observational parameters describing the properties of merging galaxies and their SMBHs. In Section 4, we briefly described how the PTA signal is constructed, the simulation set-up of the different investigated PTAs, and the Bayesian method used in the analysis. Observationally motivated prior distributions for all model parameters are given in Section 5. Detailed results are presented and discussed in Section 6, and in Section 7, we summarize our main findings and outline future research directions.

Unless stated otherwise, we use the standard Λ cold dark matter (ΛCDM) as our cosmology with the Hubble parameter $h_0 = 0.7$ and constant $H_0 = 70$ km Mpc^{-1}s^{-1} and energy density ratios $\Omega_M = 0.3$, $\Omega_k = 0$, and $\Omega_\Lambda = 0.7$.

2 GWB STRAIN MODEL

Deviations from an unperturbed spacetime arising from an incoherent superposition of GW sources (i.e. a stochastic GWB) are costumarily described in terms of characteristic strain h_c, which represents the amplitude of the perturbation per unit logarithmic frequency interval. We compute h_c following Chen et al. (2017b, Paper I hereafter). The model allows for the quick calculation of h_c given the chirp mass $\mathcal{M}$, redshift z, and eccentricity e at decoupling of any individual binary. The total strain of the GWB can then be computed by integrating over the population $\frac{d^2n}{dzd\mathcal{M}}$, giving the main

equation of Paper I:

$$h_c^2(f) = \int \mathrm{d}z \int \mathrm{d}\mathcal{M} \frac{\mathrm{d}^2 n}{\mathrm{d}z\mathrm{d}\mathcal{M}} h_{c,\,\mathrm{fit}}^2 \left(f\frac{f_{\mathrm{p},0}}{f_{\mathrm{p,t}}} \right)$$

$$\times \left(\frac{f_{\mathrm{p,t}}}{f_{\mathrm{p},0}} \right)^{-4/3} \left(\frac{\mathcal{M}}{\mathcal{M}_0} \right)^{5/3} \left(\frac{1+z}{1+z_0} \right)^{-1/3} \tag{1}$$

where $h_{c,\,\mathrm{fit}}$ is the strain of a reference binary with chirp mass $\mathcal{M}_0$, redshift z_0, eccentricity e_0, and $f_{\mathrm{p},0}$ is the peak frequency of the spectrum (see equation 13 in Paper I and relative discussion therein). The main concept of equation (1) is to use the self-similarity of the characteristic strain of a purely GW emission driven binary to go from the reference spectrum $h_{c,\,\mathrm{fit}}$ with fixed parameters to the emitted spectrum of a binary h_c with arbitrary parameters via shifts in frequency, chirp mass, and redshift.

As in Paper I, we assume that the evolution of the binary is driven by hardening in a stellar environment before GW emission takes over at a transition frequency given by (equation 21 in Paper I):

$$f_{\mathrm{t}} = 0.356\,\mathrm{nHz} \left(\frac{1}{F(e)} \frac{\rho_{i,100}}{\sigma_{200}} \zeta_0 \right)^{3/10} \mathcal{M}_9^{-2/5}, \tag{2}$$

where

$$F(e) = \frac{1 + (73/24)e^2 + (37/96)e^4}{(1 - e^2)^{7/2}} \tag{3}$$

(Peters & Mathews 1963), $\mathcal{M}_9 = \mathcal{M}/(10^9\,\mathrm{M}_\odot)$ is the rescaled chirp mass, $\rho_{i,100} = \rho_i/(100\,\mathrm{M}_\odot\mathrm{pc}^{-3})$ is the density of the stellar environment at the influence radius of the SMBHB, $\sigma_{200} = \sigma/(200\,\mathrm{km\,s}^{-1})$ is the velocity dispersion of stars in the galaxy, and ζ_0 is an additional multiplicative factor absorbing all systematic uncertainties in the estimate of $\rho_{i,\,100}$. In fact, as extensively described in Paper I, the stellar density of the host galaxy bulge follows a Dehnen profile (Dehnen 1993) with a fiducial inner density slope $\gamma = 1$. This specific profile choice, together with an empirical estimate of scale radius, fixes ρ_i for a given stellar bulge mass. Galaxies can, however, be more/less compact and have steeper/shallower density profiles, thus resulting in ρ_i that can be different by orders of magnitude from this value. We thus capture this possibility by introducing the multiplicative factor ζ_0. If, for example, ζ_0 is measured to be ≈ 10, this means that massive galaxies have on average higher central densities than implied by a standard Dehnen profile. Note that both ρ_i and σ enter in the calculation to the 3/10 power. Although difference in ρ_i can be significant, massive galaxies have generally 250 km s^{-1} < σ < 350 km s^{-1}. We thus keep σ constant in our calculation, since we found that such small range does have a negligible impact on the shape of the spectrum. Note however that ζ_0 can be considered as a multiplicative factor absorbing systematics in the determination of ρ_i and σ.

Finally, the spectrum described by equation (1) is corrected by including an a high-frequency drop related to an upper mass limit calculated, at each frequency, via (equation 39, Paper I)

$$N_{\Delta f} = \int_{f-\Delta f/2}^{f+\Delta f/2} \mathrm{d}f \int_{\bar{\mathcal{M}}}^{\infty} \mathrm{d}\mathcal{M} \int_0^\infty \mathrm{d}z \frac{\mathrm{d}^3 N}{\mathrm{d}f\mathrm{d}z\mathrm{d}\mathcal{M}} = 1, \tag{4}$$

This upper mass limit $\bar{\mathcal{M}}$ takes into account that, particularly at high frequencies, there is less than 1 binary above $\bar{\mathcal{M}}$ contributing to the signal within a frequency bin $\Delta f = 1/T$. Statistically, this means that in a given realization of the Universe, there will be either one or zero loud sources contributing to the signal. In the case the source is present, it can be removed from the GWB computation since it

will be likely resolvable as an individual deterministic GW source (see discussion in Sesana et al. 2008).

In Paper I, we used a phenomenological parametric function to describe the SMBHB merger rate $\mathrm{d}^2 n/(\mathrm{d}z\mathrm{d}\mathcal{M})$, and introduced an extra parameter e_0 to allow for eccentric binaries at f_{t}. The quantity $\mathrm{d}^2 n/(\mathrm{d}z\mathrm{d}\mathcal{M})$, however, cannot be directly measured from observations. It can be either computed theoretically from galaxy and SMBH formation and evolution models (e.g. Sesana et al. 2008; Ravi et al. 2012; Kelley et al. 2017b) or it can be indirectly inferred from observations of other astrophysical quantities, such as the galaxy mass function, pair fraction, typical merger time-scales, and the SMBH–host galaxy relation. Parametrizing the SMBHB merger rate as a function of astrophysical observables would therefore allow to reverse engineer the outcome of current and future PTA observations to obtain useful constrains on those observables. With this goal in mind, in this paper we expand the model from Paper I in two ways:

(i) we introduce an extra parameter ζ_0, see equation (2), to allow for variations from the fiducial values of the density of the stellar environment;

(ii) we cast the phenomenological SMBHB merger rate $\mathrm{d}^2 n/(\mathrm{d}z\mathrm{d}\mathcal{M})$ in terms of astrophysical observables, such as galaxy mass function and pair fraction, galaxy−black hole relations, etc., as we detail next in Section 3.

3 PARAMETRIC MODEL OF THE SMBHB MERGER RATE

As detailed in Sesana (2013a) and Sesana et al. (2016), the differential galaxy merger rate per unit redshift, mass, and mass ratio, can be written as

$$\frac{\mathrm{d}^3 n_{\mathrm{G}}}{\mathrm{d}z'\mathrm{d}M\mathrm{d}q} = \frac{\Phi(M,z)}{M\ln 10} \frac{\mathcal{F}(M,z,q)}{\tau(M,z,q)} \frac{\mathrm{d}t}{\mathrm{d}z} \tag{5}$$

where $\Phi(M, z)$ is the redshift-dependent GSMF, $\mathcal{F}(M, z, q)$ is the differential pair fraction with respect to the mass ratio q (see equation 12 below), and $\tau(M, z, q)$ is the merger time-scale. M is the mass of the primary galaxy, z is the redshift of the *galaxy pair*, and q is the mass ratio between the two galaxies. It is important to note that a pair of galaxies at redshift z will merge at redshift $z' < z$. The time-scale $\tau(M, z, q)$ is used to convert the pair fraction of galaxies at z into the galaxy merger rate at $z' < z$ (Mundy et al. 2017). The merger redshift is obtained by solving for z' the implicit equation

$$\int_{z'}^{z} \frac{\mathrm{d}t}{\mathrm{d}\bar{z}}\mathrm{d}\bar{z} = \tau(M,z,q), \tag{6}$$

where, assuming a flat ΛCDM model,

$$\frac{\mathrm{d}t}{\mathrm{d}\bar{z}} = \frac{1}{H_0(1+\bar{z})\sqrt{\Omega_{\mathrm{M}}(1+\bar{z})^3 + \Omega_{\mathrm{k}}(1+\bar{z})^2 + \Omega_\Lambda}}. \tag{7}$$

The GSMF $\Phi(M, z)$ can be written as a single Schechter function (Conselice et al. 2016)

$$\Phi(M, z) = \frac{\mathrm{d}n_{\mathrm{G}}}{\mathrm{d}\log_{10} M} = \ln 10\,\Phi_0(z)$$

$$\times \left(\frac{M}{M_0(z)} \right)^{1+\alpha_0(z)} \exp\left(-\frac{M}{M_0(z)} \right), \tag{8}$$

where $\Phi_0(z), M_0(z)$, and $\alpha_0(z)$ are phenomenological functions of redshift of the form (Mortlock et al. 2015):

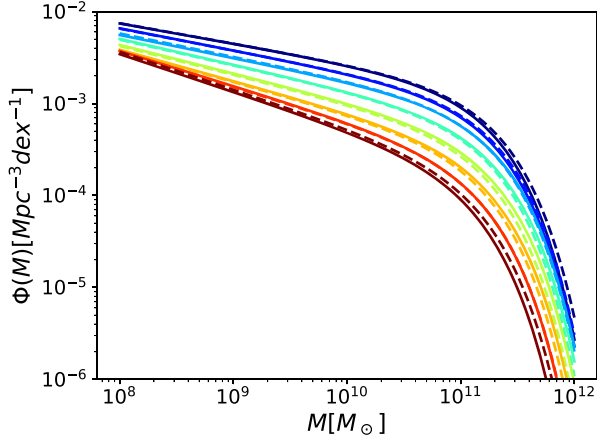$$\log_{10}\Phi_0(z) = \Phi_0 + \Phi_{\mathrm{I}}z \tag{9}$$

Figure 1. Comparison between measured and computed GSMFs for eight redshift bins in the range $0.4 < z < 3$, where blue represents lower and red higher redshift values. The solid lines represent the original mass functions reported by Conselice et al. (2016) and Mortlock et al. (2015), the dashed lines are best fits obtained using equation (8) with the parametric functions $\Phi_0(z)$, $M_0(z)$, and $\alpha(z)$ for the central values of the eight corresponding redshift bins.

$$M_0(z) = M_0 \tag{10}$$

$$\alpha_0(z) = \alpha_0 + \alpha_I z \tag{11}$$

The five parameters Φ_0, Φ_I, M_0, α_0, and α_I are sufficient to fit the original Schechter functions at any redshift; an example is shown in Fig. 1. To simplify the notation, in the following Φ_0, M_0, and α_0 will implicitly denote their corresponding redshift-dependent functions $\Phi_0(z)$, $M_0(z)$, and $\alpha_0(z)$.

The differential pair fraction as a function of q is given by

$$\mathcal{F}(M, z, q) = \frac{df_{pair}}{dq} = f_0' \left(\frac{M}{aM_0} \right)^{\alpha_f} (1+z)^{\beta_f} q^{\gamma_f} \tag{12}$$

where $aM_0 = 10^{11} \, \mathrm{M_\odot}$ is an arbitrary reference mass. Note that, in the literature, pair fractions are usually given as a function of primary galaxy mass and redshift only (e.g. Mundy et al. 2017), such that

$$f_{pair} = f_0 \left(\frac{M}{aM_0} \right)^{\alpha_f} (1+z)^{\beta_f}, \tag{13}$$

i.e. integrated over the mass ratio of the pairs. The integral of equation (12) over q gives

$$f_{pair} = f_0' \left(\frac{M}{aM_0} \right)^{\alpha_f} (1+z)^{\beta_f} \int q^{\gamma_f} dq, \tag{14}$$

which becomes equivalent to equation (13) by setting

$$f_0 = f_0' \int q^{\gamma_f} dq. \tag{15}$$

Equation (15) allows to map an observational prior of the form of equation (13) into the four parameters of our model f_0', α_f, β_f, and γ_f.

We use an analogue parametrization for the merger time-scale:

$$\tau(M, z, q) = \tau_0 \left(\frac{M}{bM_0} \right)^{\alpha_\tau} (1+z)^{\beta_\tau} q^{\gamma_\tau} \tag{16}$$

where $bM_0 = 0.4/h_0 \times 10^{11} \, \mathrm{M_\odot}$, and τ_0, α_τ, β_τ, and γ_τ are four further model parameters. Equation (16) has originally been derived to describe the *galaxy* merger time-scale (Snyder et al. 2017). A further delay is, however, expected between the galaxy merger and the SMBHB final coalescence. In fact, after dynamical friction has merged the two galaxies and has brought the two SMBHs in the nuclear region, the newly formed SMBHB has to harden via energy and angular momentum losses mediated by either stars or gas, before GW emission eventually takes over (see Dotti et al. 2012, for a review). Depending on the details of the environment, this process can take up to several Gyrs, and even cause the binary to stall (Sesana & Khan 2015; Vasiliev, Antonini & Merritt 2015; Kelley, Blecha & Hernquist 2017a). For simplicity, we assume here that this further delay can be re-absorbed in equation (16), which we then use to describe the time elapsed between the observed galaxy pair and the final *SMBHB* coalescence.

Substituting equations (8), (12), and (16) into equation (5) gives

$$\frac{d^3 n_G}{dz' dM dq} = n_{eff} \left(\frac{M}{M_0} \right)^{\alpha_{eff}} e^{-M/M_0} (1+z)^{\beta_{eff}} q^{\gamma_{eff}} \frac{dt}{dz}, \tag{17}$$

where the effective parameters are

$$n_{eff} = \frac{\Phi_0 f_0'}{M_0 \tau_0} \frac{b^{\alpha_\tau}}{a^{\alpha_f}} = \frac{\Phi_0 f_0'}{M_0 \tau_0} \left(\frac{0.4}{h_0} \right)^{\alpha_\tau} \left(\frac{10^{11}}{M_0} \right)^{\alpha_\tau - \alpha_f}$$

$$\alpha_{eff} = \alpha_0 + \alpha_f - \alpha_\tau$$

$$\beta_{eff} = \beta_f - \beta_\tau$$

$$\gamma_{eff} = \gamma_f - \gamma_\tau \tag{18}$$

Equation (17) is still a function of the merging galaxy stellar masses, which needs to be translated into SMBH masses. The total mass of a galaxy M can be converted into its bulge mass M_{bulge}, using assumptions on the ellipticity of the galaxy: more massive galaxies are typically elliptical and have higher bulge to total stellar mass ratio. We use a phenomenological fitting function (Bernardi et al. 2014; Sesana et al. 2016) to link the bulge mass to the total stellar mass of a galaxy:

$$\frac{M_{bulge}}{M} = \begin{cases} \frac{\sqrt{6.9}}{(\log M - 10)^{1.5}} \exp\left(\frac{-3.45}{\log M - 10} \right) + 0.615 & \text{if } \log M > 10 \\ 0.615 & \text{if } \log M < 10. \end{cases} \tag{19}$$

Note that this fit is appropriate for ellipticals and spheroidals, whereas spiral galaxies usually have smaller bulge-to-total mass ratio. In Sesana (2013a) different scaling relations were used for blue and red galaxy pairs (under the assumption that blue pairs are predominantly spirals and red pairs predominantly elliptical). The result was that the GW signal is completely dominated by red pairs. We have checked on Sesana (2013a) data that approximating all galaxies as spheroidals affects the overall signal by less than 0.05 dex. We therefore apply equation (30) to all galaxies, independent on their colour or morphology.

We can then apply a scaling relation between the galaxy bulge mass M_{bulge} and black hole mass M_{BH} of the form (see e.g. Kormendy & Ho 2013)

$$M_{BH} = \mathcal{N} \left\{ M_* \left(\frac{M_{bulge}}{10^{11} \, \mathrm{M_\odot}} \right)^{\alpha_*}, \epsilon \right\}, \tag{20}$$

where $\mathcal{N}\{x, y\}$ is a lognormal distribution with mean value x and standard deviation y, to translate galaxy mass M into black hole mass M_{BH}. Note that the galaxy mass ratio q is in general different from the black hole mass ratio $q_{BH} = q^{\alpha_*}$. Finally, the galaxy merger rate

Table 1. List of the 18 parameters in the model, including their description, standard, and extended prior distribution ranges. Squared brackets indicate flat uniform distributions, while $\pm$ signs indicate the median and 90 per cent credible intervals for the distributions, as shown in Figs 4 and 5 of Section 5.

Parameter	Description	Standard	Extended
Φ_0	GSMF norm	$-2.77^{+0.27}_{-0.29}$	$-2.77^{+0.27}_{-0.29}$
Φ_I	GSMF norm redshift evolution	$-0.27^{+0.23}_{-0.21}$	$-0.27^{+0.23}_{-0.21}$
$\log_{10}M_0$	GSMF scaling mass	$11.24^{+0.20}_{-0.17}$	$11.24^{+0.20}_{-0.17}$
α_0	GSMF mass slope	$-1.24^{+0.16}_{-0.16}$	$-1.24^{+0.16}_{-0.16}$
α_I	GSMF mass slope redshift evolution	$-0.03^{+0.16}_{-0.14}$	$-0.03^{+0.16}_{-0.14}$
f_0	Pair fraction norm	$[0.02,0.03]$	$[0.01,0.05]$
α_f	Pair fraction mass slope	$[-0.2,0.2]$	$[-0.5,0.5]$
β_f	Pair fraction redshift slope	$[0.6,1]$	$[0,2]$
γ_f	Pair fraction mass ratio slope	$[-0.2,0.2]$	$[-0.2,0.2]$
τ_0	Merger time norm	$[0.1,2]$	$[0.1,10]$
α_τ	Merger time mass slope	$[-0.2,0.2]$	$[-0.5,0.5]$
β_τ	Merger time redshift slope	$[-2,1]$	$[-3,1]$
γ_τ	Merger time mass ratio slope	$[-0.2,0.2]$	$[-0.2,0.2]$
$\log_{10}M_*$	$M_{\rm bulge}-M_{\rm BH}$ relation norm	$8.17^{+0.35}_{-0.32}$	$8.17^{+0.35}_{-0.32}$
α_*	$M_{\rm bulge}-M_{\rm BH}$ relation slope	$1.01^{+0.08}_{-0.10}$	$1.01^{+0.08}_{-0.10}$
ϵ	$M_{\rm bulge}-M_{\rm BH}$ relation scatter	$[0.3,0.5]$	$[0.2,0.5]$
e_0	Binary eccentricity	$[0.01,0.99]$	$[0.01,0.99]$
$\log_{10}\zeta_0$	Stellar density factor	$[-2,2]$	$[-2,2]$

n_G (17) can be converted into the SMBHB merger rate n:

$$\frac{{\rm d}^3n}{{\rm d}z'{\rm d}M_{\rm BH}{\rm d}q_{\rm BH}} = \frac{{\rm d}^3n_G}{{\rm d}z'{\rm d}M{\rm d}q}\frac{{\rm d}M}{{\rm d}M_{\rm BH}}\frac{{\rm d}q}{{\rm d}q_{\rm BH}}. \tag{21}$$

Equation (20) adds three further parameters to the model: M_*, α_*, and ϵ. Lastly, it is convenient to map $M_{\rm BH}$, $q_{\rm BH}$ into the SMBHB chirp mass $\mathcal{M} = M_{\rm BH}q_{\rm BH}^{3/5}/(1 + q_{\rm BH})^{1/5}$, by performing a variable change and integrate over the black hole mass ratio to produce an SMBHB merger rate as a function of chirp mass and redshift only:

$$\frac{{\rm d}^2n}{{\rm d}z'{\rm d}\mathcal{M}} = \int \frac{{\rm d}^3n_G}{{\rm d}z'{\rm d}M_{\rm BH}{\rm d}q_{\rm BH}}\frac{{\rm d}M_{\rm BH}}{{\rm d}\mathcal{M}}{\rm d}q_{\rm BH}. \tag{22}$$

Summarizing, the SMBHB merger rate ${\rm d}^2n/({\rm d}z'{\rm d}\mathcal{M})$ is described as a function of 16 empirical parameters that are related to astrophysical observables: (Φ_0, Φ_I, M_0, α_0, α_I) for the GSMF, (f'_0, α_f, β_f, γ_f) for the pair fraction, (τ_0, α_τ, β_τ, γ_τ) for the merger time-scale, and (M_*, α_*, ϵ) for the galaxy–SMBH scaling relation. Further, the first three sets of parameters can be grouped into the four effective parameters given by equation (18). The two extra parameters (e_0, ζ_0) enter the computation of the shape of the GW spectrum via the transition frequency f_t given in equation (2). We can therefore express the stochastic GWB in equation (1) as a function of 18 phenomenological parameters, listed in Table 1.

4 GWB SIMULATIONS AND ANALYSIS

As in Chen et al. (2017a, Paper II hereafter), we compute the signal-to-noise ratio (S/N) $\mathcal{S}$ of a detection of a GWB in the frequency domain as (Moore, Taylor & Gair 2015; Rosado et al. 2015):

$$\mathcal{S}^2 = 2\sum_{i=1,N}\sum_{j>i}T_{\rm ij}\int \frac{\Gamma_{\rm ij}^2 S_{\rm h}^2}{(S_{\rm n}^2)_{\rm ij}}{\rm d}f, \tag{23}$$

where $\Gamma_{\rm ij}$ are the Hellings–Downs coefficients (Hellings & Downs 1983):

$$\Gamma_{\rm ij} = \frac{3}{2}\gamma_{\rm ij}\ln\left(\gamma_{\rm ij}\right) - \frac{1}{4}\gamma_{\rm ij} + \frac{1}{2} + \frac{1}{2}\delta_{\rm ij}, \tag{24}$$

where $\gamma_{\rm ij} = [1 - \cos(\theta_{\rm ij})]/2$ and $\theta_{\rm ij}$ is the relative angle between pulsars i and j. $S_{\rm h}$ and $S_{\rm n}$ in equation (23) are spectral densities of the signal and noise respectively, and $S_{\rm n}$ includes the 'self-noise' contribution of the pulsar term (see equation 11 in Paper II for details).

We can simplify equation (23) by assuming that all pulsars are identical (except for their position in the sky), i.e. all pulsars have the same properties: rms $\delta_{\rm i} = \delta$, observation time $T_{\rm ij} = T$ and observation cadence Δt. Furthermore, we also assume that there is a sufficient number of pulsars N, uniformly distributed in the sky, so that each individual coefficient $\Gamma_{\rm ij}$ can be replaced by the rms computed across the sky $\Gamma = \sqrt{\langle\Gamma_{\rm ij}^2\rangle} = 1/(4\sqrt{3})$, and the double sum over all pairs of pulsars $\sum_{i=1,N}\sum_{j>i}$ becomes $N(N-1)$. For an observation time T, the spectrum of the GWB is resolved into a Fourier series of frequencies $1/T, 2/T, \ldots, (k+1)/T$ with an equal bin width $\Delta f = 1/T$ and central frequencies $f_k = (2k+1)/(2T)$. The total S/N in equation (23) can thus be split into frequency bin components $\mathcal{S}_k$:

$$\mathcal{S}_k^2 \approx T\Gamma^2 N(N-1)\frac{S_{\rm h}^2}{S_{\rm n}^2}\Delta f. \tag{25}$$

In the strong signal regime ($S_{\rm h} \gg S_{\rm n}$), equation (25) can further be reduced to the approximate total S/N of a strong detection in M frequency bins

$$\mathcal{S} = \left(\sum_k \mathcal{S}_k^2\right)^{1/2} = \left(\frac{\Gamma^2}{1+\Gamma^2}MN(N-1)\right)^{1/2} \approx \Gamma N M^{1/2}, \tag{26}$$

where we used the fact that $\Gamma \ll 1$ and $N \gg 1$. Equation (26) is a drastic simplification, still it provides the relevant scaling between $\mathcal{S}$, number of pulsars in the array, and frequency range in which the signal is resolved. For $\Gamma \approx 0.14$, to achieve an S/N ≈ 5 in the lowest few frequency bins, an array of about 20 equally good pulsars is needed (see also Jenet et al. 2006).

PTA data are simulated as in Paper II. For a signal h_c with amplitude $A_{{\rm det},k}$ in the kth frequency bin, the detection S/N $\mathcal{S}_k$
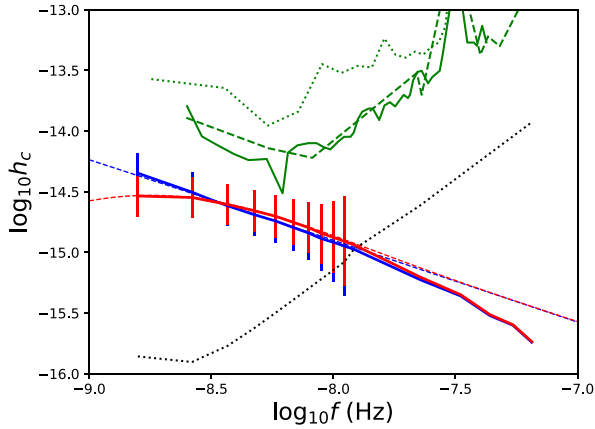
Figure 2. Examples of simulated detections for two different spectral shapes. Signal models correspond to the default SMBHB population with parameters defined in Section 5 and high eccentricity ($e_t = 0.9$, red) and almost circular ($e_t = 0.01$, blue). For each model, solid lines are the theoretical spectra including the high-frequency steepening due to the mass upper limit defined by equation (4), and dashed lines depict spectra excluding this feature (therefore with $h_c \propto f^{-2/3}$ at high frequency) for comparison. Error bars centred around the model value are the observed amplitudes with associated uncertainties when $\mathcal{S}_k > 1$. The black dotted line represents the nominal 1σ model sensitivity curve of the PTA for the *IPTA30* case (h_n as computed from equation 20 in Paper II). Green lines in the upper part of the figure are current EPTA (dotted), NANOGrav (dashed), and PPTA (solid) upper limits.

is related to the detection uncertainty $\sigma_{det, k}$ via (see equation 18 in Paper II)

$$\sigma_{det,k} = \frac{1}{\mathcal{S}_k}. \tag{27}$$

4.1 Simulated data sets

Besides adding a future and an ideal upper limit, we use the same simulation setup as in Paper II, with the simplifying assumptions that all pulsars are observed with the same cadence Δt for the same duration of T and have the same rms of δ. These assumptions only affect the S/N of the detection, thus setting the error bars $\sigma_{det, k}$. This is purely a choice of convenience that does not affect the general validity of our results. We expand upon the four cases from Paper II by adding two more upper limit cases to get a total of six fiducial cases (three upper limits and three detections):

(i) case *PPTA15*: we use the upper limit curve of the most recent PPTA analysis, as given by Shannon et al. (2015), which is representative of current PTA capabilities and results in a GWB upper limit of $A = 10^{-15}$;[1]

(ii) case *PPTA16*: we shift the *PPTA15* curve down by one order of magnitude, which is representative of an upper limit of $A = 10^{-16}$, reachable in the Square Kilometre Array (SKA) era;

(iii) case *PPTA17*: we shift the *PPTA15* curve down by two orders of magnitude, which is representative of an upper limit of $A = 10^{-17}$. Although a two orders of magnitude leap in sensitivity might require decades of timing with the full SKA, we use this scenario to infer

what conclusions can be drawn by a non-detection at a level well below currently predicted GWB values;

(iv) case *IPTA30*: $N = 20$, $\delta = 100$ ns, $T = 30$ yr, and $\Delta t = 1$ week. This PTA results in a detection S/N ≈ 5–10 and is based on a future extrapolation of the current IPTA, without the addition of new telescopes;

(v) case *SKA20*: $N = 100$, $\delta = 50$ ns, $T = 20$ yr, and $\Delta t = 1$ week. This PTA results in a high significance detection with S/N ≈ 30–40, which will be technically possible in the SKA era;

(vi) case *ideal*: $N = 500$, $\delta = 1$ ns, $T = 30$ yr, and $\Delta t = 1$ week. This theoretically possible ideal PTA provides useful insights of what might be achievable in principle.

4.2 Data analysis method

We apply Bayes' theorem to perform inference on our model M, given some data d and a set of parameters θ:

$$p(\theta|d, M) = \frac{p(\theta|M)p(d|\theta, M)}{p(d|M)}, \tag{28}$$

where $p(\theta|d, M)$ is the posterior distribution coming from the analysis of the PTA measurement, $p(\theta|M)$ is the prior distribution and accounts for any beliefs on the constraints of the model parameters (prior to the PTA measurement), $p(d|\theta, M)$ is the likelihood of producing the data for a given model and parameter set, and $p(d|M)$ is the evidence, which is a measure of how likely the model is to produce the data.

To simulate detections, we apply the likelihood from Paper II

$$p_{det}(d|A_{trial}(f)) \propto \exp\left\{ -\frac{\left[\log_{10} A_{trial}(f) - \log_{10} A_{det}(f)\right]^2}{2\sigma_{det}(f)^2} \right\}, \tag{29}$$

to each frequency bin for which $\mathcal{S}_k > 1$, and then sum over the frequency bins to obtain the total likelihood. For the upper limit analyses, we use the directly derived likelihood from the PPTA upper limit, as described in appendix A.3 of Middleton et al. (2018).

Prior distributions are taken from independent theoretical and observational constrains, as described in Section 5. The parameter space is sampled using cpnest (Del Pozzo & Veitch 2015), which is a parallel implementation of the nested sampling algorithm in the spirit of Veitch et al. (2015) and Skilling (2004). Nested sampling algorithms do not only provide posterior distributions, but also the total evidence. This allows us to compute Bayes factors for model comparisons. Each simulation has been run with 1000 livepoints, producing ~ 2500 independent posterior samples.

5 DEFINING THE PRIOR RANGES OF THE MODEL PARAMETERS

There is a vast literature dedicated to the measurement of the GSMF, galaxy pair fraction, merger time-scale, and SMBH–host galaxy scaling relations. We now described how independent observational and theoretical work translates into constrained prior distributions of the 18 parameters of our model. A summary of all the prior ranges is given in Table 1.

5.1 Galaxy stellar mass function

At any given redshift, the GSMF is usually described as a Schechter function with three parameters (Φ_0, M_0, α_0). The parameters, however, are independently determined at any redshift. Depending

---

[1]A represents the amplitude of the GWB at a reference frequency of $f = 1\,\mathrm{yr}^{-1}$ under the assumption that its spectrum is described by a single power law with $h_c \propto f^{-2/3}$, appropriate for circular, GW-driven binaries.
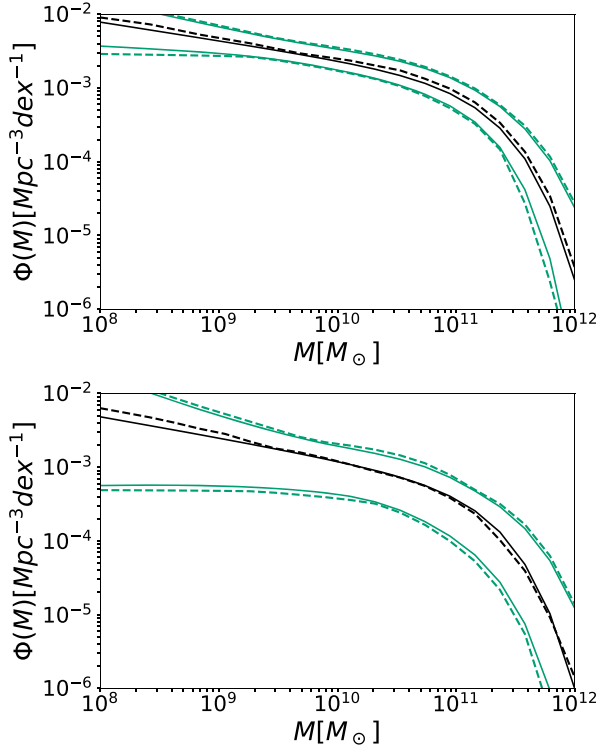
Figure 3. Comparison between the allowed region of GSMFs for different redshift bins computed from Conselice et al. (2016, three dashed lines) and the region of GSMFs recovered by using the five GSMF parameters (Φ_0, Φ_I, M_0, α_0, α_I) (three solid lines). Black lines represent the median, whilst green lines represent the upper and lower bounds of the 99 per cent confidence region. The top and bottom panels show the GSMF in the $0 < z < 1$ and $1 < z < 3$ redshift bins, respectively.

Figure 4. Corner plot showing the prior distributions of the five GSMF parameters (Φ_0, Φ_I, M_0, α_0, α_I) used in this work.

on the number of redshift bins n to be considered in the computation, this can easily lead to a very large number of parameters $3n$. To reduce the dimensionality of the problem from $3n$ to five, we note that the parameters (Φ_0, α_0) show clear linear trends with redshift, whilst M_0 is fairly constant (see Mortlock et al. 2015). This allows for a reparametrization as a function of the five parameters (Φ_0, Φ_I, M_0, α_0, α_I) performed in Section 3.

A comprehensive list of published values for the parameters (Φ_0, M_0, α_0) for various redshift bins can be found in Conselice et al. (2016), which forms the basis of our prior distribution. We compute $\Phi(M, z)$ between 10^9 and $10^{12}\,M_\odot$ for all sets of (Φ_0, M_0, α_0), dividing the sample in two redshift bins: $0 < z < 1$ and $1 < z < 3$. This gives a range of values for $\Phi(M, z)$, shown in Fig. 3. We then take uniform distributions of ($\Phi_0 \in [-3.4, -2.4]$, $\Phi_I \in [-0.6, 0.2]$, $M_0 \in [11, 11.5]$, $\alpha_0 \in [-1.5, -1.]$, and $\alpha_I \in [-0.2, 0.2]$), compute the $\Phi(M, z)$ for each sample and redshift bins and compare them against the allowed range. If the value is within the range, the sample is accepted, otherwise it is rejected. The resulting prior distributions are shown in Fig. 4.

5.2 Pair fraction

Constraints on the pair fraction have been derived by counting the numbers of paired and merged galaxies in various surveys with a number of different photometric and spectroscopic techniques (see e.g. Conselice et al. 2003; Xu et al. 2012; Robotham et al. 2014;
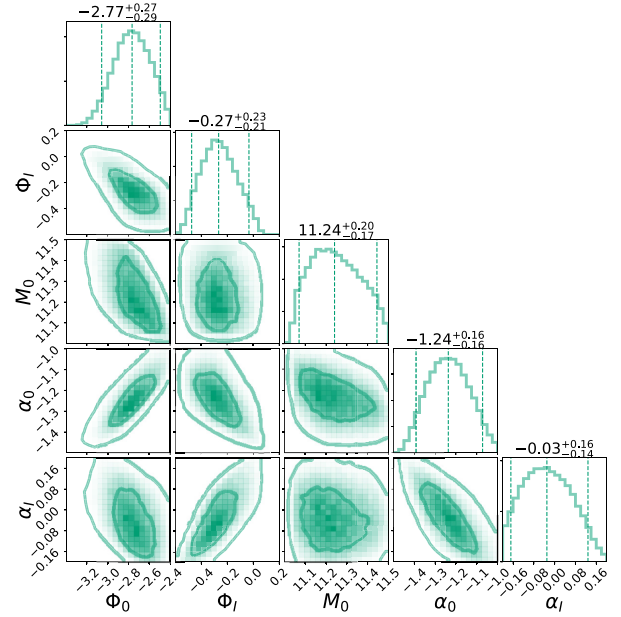
Keenan et al. 2014; Duncan et al. 2019). Recently, Mundy et al. (2017) have combined data from several surveys to produce an overall up-to-date constrain. We base our prior range on the results reported in table 3 of their paper, for the All+GAMA + D17 survey combination and galaxy separation of 5–30 kpc:

$$f_{\mathrm{pair}} = \begin{cases} 0.028 \pm 0.002 \times (1+z)^{0.80\pm0.09} & \text{for } \log M > 10 \\ 0.024 \pm 0.004 \times (1+z)^{0.78\pm0.20} & \text{for } \log M > 11. \end{cases} \quad (30)$$

This is one of the flatter redshift dependences within the Mundy et al. (2017) compilation. It is, however, likely the more accurate measurement, coming from a combination of deep surveys. Moreover, while stronger redshift dependences are common for Milky Way-size galaxies, most f_{pair} measurements of galaxies with $M > 10^{11}\,M_\odot$ have a relatively flat redshift dependence. Most of the GWB will come from SMBHBs hosted in those massive galaxies, this justifies our choice. Noting that both sets of parameters and uncertainties in equation (30) are similar, we use flat priors for $f_0 \in [0.02, 0.03]$ and $\beta_{\mathrm{f}} \in [0.6, 1]$ for all galaxy masses. Steeper redshift dependences are allowed in our set of 'extended priors', introduced in Section 5.6 below. Mundy et al. (2017) also find no significant dependency on galaxy mass, thus we pick α_{f}, $\gamma_{\mathrm{f}} \approx 0$, adding the possibility of a mild deviation by imposing a flat prior α_{f}, $\gamma_{\mathrm{f}} \in [-0.2, 0.2]$.

5.3 Merger time-scale

We define the merger time-scale, as the time elapsed between the observation of a galaxy pair at a given projected separation (usually 20 or 30 kpc) and the final coalescence of the SMBHB, thus including the time that it takes for the two galaxies to effectively merge, plus the time required for the SMBHs to sink to the centre, form a binary and harden via stellar scattering. Galaxy merger time-scales have been computed both for simulations of isolated galaxy mergers (Lotz et al. 2011) and from ensemble of haloes and galaxies extracted from large cosmological simulations (Kitzbichler & White

2008; Snyder et al. 2017), resulting in a large dynamical range, typically between 0.1 and 1 Gyr. On the other hand, the SMBHB merger time-scale has been estimated by means of *N*-body and special purpose Monte Carlo codes (e.g. Khan et al. 2012; Vasiliev et al. 2015; Sesana & Khan 2015). All studies show that three-body scattering is efficient in driving the binary to final coalescence withing a Gyr.

We therefore choose the parametrization given by equation (16) with wide uniform prior ranges $\tau_0 \in [0.1, 2]$ Gyr and $\beta_\tau \in [-2, 1]$, which is sufficiently generic to cover the observation based range of possible effects influencing the total merger time. The mass dependencies are generally found to be milder, playing a minor role. We therefore choose flat prior ranges $\alpha_\tau, \gamma_\tau \in [-0.2, 0.2]$. These conservative prior ranges are extended in Section 5.6 to include possibe physical effect further delaying the merger time-scale, such as inefficient replenishment of the loss cone slowing down the binary hardening. Note that in this latter case all SMBHBs forming in galaxy pairs observed at $z < \sim 2$ would not merge by $z = 0$, thus being effectively 'stalled' for the sake of our analysis.

5.4 $M_{\rm bulge}-M_{\rm BH}$ relation

Since SMBHs are thought to have an important impact on the formation and evolution of their host galaxy and vice versa, the relation between their mass and several properties of the host galaxy has been studied and constrained by a number of authors (see Kormendy & Ho 2013, for a comprehensive review). Here, we use the tight relation between the SMBH mass and the stellar mass of the spheroidal component (i.e. the bulge) of the host galaxy, which has been described as a power law of the form of equation (20) with some intrinsic scattering. Although non-linear functions have been proposed in the literature (see e.g. Graham & Scott 2012; Shankar et al. 2016), the non-linearity is mostly introduced to describe the (observationally very uncertain) low-mass end of the relation. Since the vast majority of the GWB is produced by SMBH with masses above 10^8 M$_\odot$ (Sesana et al. 2008), we do not consider here those alternative parametrizations.

Similarly, we do also not consider the possibility of a redshift dependent $M_{\rm bulge}-M_{\rm BH}$ relation (see Li, Ho & Wang 2011, and references therein). Recent findings strenghten the view that there is no evidence for a cosmic evolution (Schulze & Wisotzki 2014) or only a very weak one (Salviander, Shields & Bonning 2015). This additional weak redshift dependence would likely not have a significant impact on our results and would be in any case covariant with other redshift dependences, and thus unlikely to be constrained by our analysis.

To construct the prior distributions, we apply the same method as in Section 5.1. We define the allow region of the $M_{\rm BH}-M_{\rm bulge}$ relation as the one enclosed within a compilation of relations collected from the literature in Middleton et al. (2018). We then draw relations from a uniform distribution of $\log_{10} M_* \in [7.75, 8.75]$ and $\alpha_* \in [0.9, 1.1]$ and accept them if they fall within the region allowed by observations. Additionally, we assume a flat distribution for the scattering $\epsilon \in [0.3, 0.5]$. Fig. 5 shows the obtained prior distributions for $(M_*, \alpha_*, \epsilon)$.

5.5 Eccentricity and stellar density

The last two parameters deal with the properties of the individual binary. As the eccentricity at decoupling is not well constrained (see e.g. Sesana & Khan 2015; Mirza et al. 2017), we choose an uninformative flat prior $e_0 \in [0.01, 0.99]$. The other additional
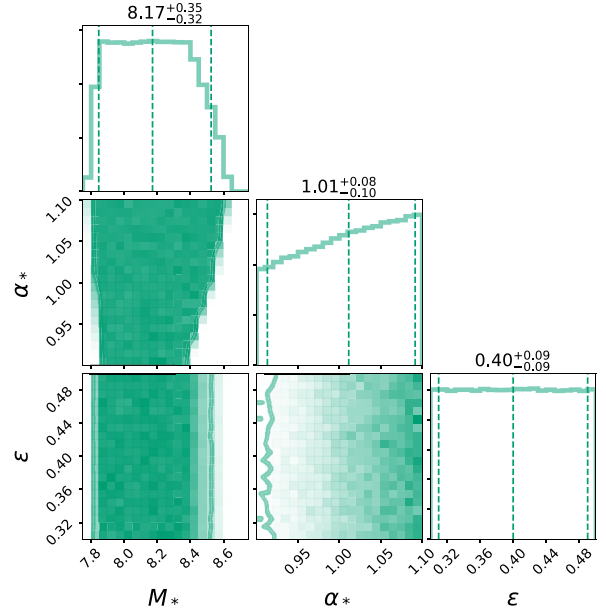

Figure 5. Corner plot showing the prior distributions of the three $M_{\rm bulge}-M_{\rm BH}$ relation parameters $(M_*, \alpha_*, \epsilon)$ used in this work.

parameter describes the stellar density around the SMBHB (see Section 2). ζ_0 is a multiplicative factor added to the density at the SMBHB influence radius, $\rho_{\rm i,100}$, calculated by using the fiducial Dehnen profile defined in Paper I. This has an impact on the frequency of decoupling, as a higher density of stars in the galactic centre means more efficient scattering. The SMBHB thus experiences a faster evolution, reaching a higher $f_{\rm t}$ before transitioning to the efficient GW emission stage. We choose to include densities that are between 0.01 and 100 times the fiducial value, aiming at covering the large variation of stellar densities observed in cusped versus cored galaxies (Kormendy et al. 2009). This translates into a flat prior $\log_{10} \zeta_0 \in [-2, 2]$.

5.6 Extended prior ranges

Unless otherwise stated, the prior ranges just described are used in our analysis. However, we also consider 'extended' prior ranges for some of the parameters. Although observational determination of the galaxy mass function is fairly solid, identifying and counting galaxy pairs in large galaxy surveys is a delicate endeavour, especially beyond the local Universe. We therefore also consider extended prior ranges $f_0 \in [0.01, 0.05]$, $\alpha_{\rm f} \in [-0.5, 0.5]$, and $\beta_{\rm f} \in [0, 2]$, allowing for more flexibility in the overall normalization, redshift, and mass evolution of the galaxy pair fraction. Likewise, SMBHB merger time-scales are poorly constrained. The prior range adopted in Section 5.3 is rather wide, but notably does not allow for stalling of low-redshift binaries (the maximum allowed merger time-scale being 2 Gyr). Also in this case we consider extended prior ranges $\tau_0 \in [0.1, 10]$ Gyr, $\alpha_\tau \in [-0.5, 0.5]$ and $\beta_\tau \in [-3, 1]$, allowing the possibility of SMBHB stalling at any redshift. Finally, we also consider a wider prior on the scatter of the $M_{\rm BH}-M_{\rm bulge}$ relation $\epsilon \in [0.2, 0.5]$, mostly because several authors find $\epsilon \approx 0.3$, which is at the edge of our standard prior. All standard and extended priors are listed in Table 1.

6 RESULTS AND DISCUSSION

Having defined the mathematical form of the signal, the prior ranges of all the model parameters, the simulated data and the form of the likelihood function, we performed our analysis on the six limits and detections described in Section 4.1. In this section, we present the results of our simulations and discuss their astrophysical consequences in detail. We first present the implications of current and future upper limits and then move onto discussing the different cases of detection. Note that, although all 18 parameters are left free to vary within their respective priors, we will present posteriors only for the subset of parameters that can be significantly constrained via PTA observations. Those are the overall normalization of the merger rate n_{eff}, the parameters defining the merger time-scale τ_0, α_τ, β_τ, the parameters defining the $M_{\text{BH}} - M_{\text{bulge}}$ relation M_*, ϵ, the eccentricity at the transition frequency e_0, and the normalization of the stellar density ζ_0. Because the large number of parameters and the limited information enclosed in the GWB shape and normalization, other parameters are generally unconstrained. Corner plots including all 18 parameters for all the simulated upper limits and detections are presented in Appendix A, available in electronic form. All runs are performed using the standard prior distributions derived in Section 5, unless stated otherwise.

6.1 Predicted GWB strain

A direct product of combining the GWB model described in Section 2 to the astrophysical priors presented in Section 5 is a robust update to the expected shape and normalization of the signal. Thus, before proceeding with the analysis of our PTA simulations, we present this result. In Fig. 6 the predicted strain of the GWB using our standard prior is compared to the *ALL* model from Middleton et al. (2018). The shapes and normalization of the two predictions, shown in the top panel, are fairly consistent. At $f = 1 \, \text{yr}^{-1}$, our model predicts $10^{-16} < h_c < 10^{-15}$ at 90 per cent confidence, which is slightly more restrictive than the *ALL* model. This has to be expected since model *ALL* from Middleton et al. (2018) is constructed following the method of Sesana (2013a). The latter, in fact, gave equal credit to all measurements of the galaxy mass function, pair fractions and SMBH–galaxy scaling relations, without considering any possible correlation between their underlying parameters. Our detailed selection of the prior range takes correlations between different parameters into account (see e.g. Fig. 4) and is likely more restrictive in terms of galaxy pair fraction.

The bottom panel of Fig. 6 shows the predicted h_c range assuming circular, GW driven binaries and no high-frequency drop, hence producing the standard $f^{-2/3}$ spectral shape. In this simplified case $h_c(f = 1 \, \text{yr}^{-1})$ is a factor of ≈ 2 higher, spanning from 2×10^{-16} to 2×10^{-15}. Still, most of the predicted range lies below current PTA upper limits, as well as being consistent with other recent theoretical calculations (Dvorkin & Barausse 2017; Kelley et al. 2017b; Bonetti et al. 2018).

6.2 Upper limits

6.2.1 *Current upper limit at* A(f = f = 1 yr⁻¹) = 1 × 10⁻¹⁵

First, we discuss the implication of current PTA upper limits. Here, we use the PPTA upper limit, nominally quoted as $A(f = 1 \, \text{yr}^{-1}) = 1.0 \times 10^{-15}$, which represents the integrated constraining power over the entire frequency range assuming a $f^{-2/3}$ power law. As it has been recently pointed out by Arzoumanian et al. (2018),
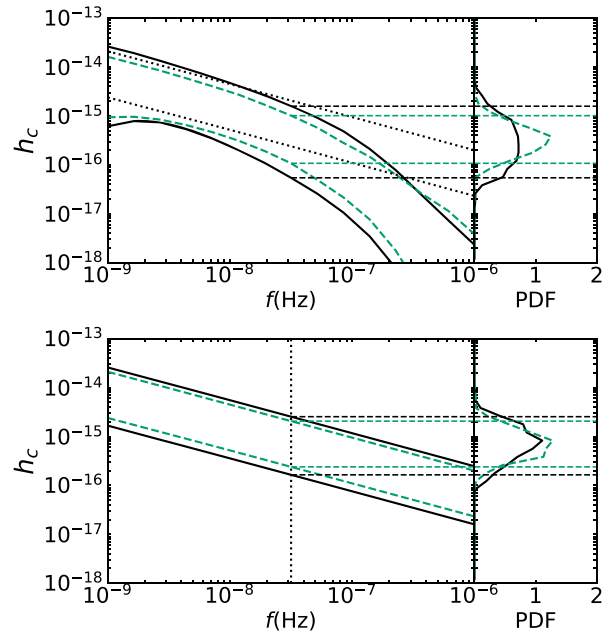

Figure 6. Comparison of the predicted characteristic strain, h_c, predicted by our new model (green dashed lines) compared to the *ALL* model from Middleton et al. (2018, solid black lines). The top panel shows the predicted strain from the full model, while the bottom panel restricts the model to circular SMBHBs without the drop at high frequencies. For comparison between the top and bottom panels, the equivalent of the bottom panel black solid lines are plotted in the top panel as black dotted lines. The left-hand panels show the frequency–strain plot, while the right-hand panels show the posterior density function (PDF) at $f = 1 \, \text{yr}^{-1}$.

the sensitivity of PTAs has become comparable to the uncertainty in the determination of the solar system ephemeris (SSE) – the knowledge of which is required to refer pulse TOAs collected at the telescopes to the solar system baricentre. Thus, it has become necessary to include an extra parametrized model of the SSE into the GWB search analysis pipelines. This leads to a more robust albeit higher upper limit, as part of the constraining power is absorbed into the uncertainty of the SSE. A robust upper limit including this effect has recently been placed by the NANOGrav Collaboration at $A(f = 1 \, \text{yr}^{-1}) = 1.45 \times 10^{-15}$, which is higher but of the same order as the PPTA upper limit. We therefore consider the PPTA upper limit in this analysis, with the understanding that the recent NANOGrav upper limit would lead to very similar implications. Since the NANOGrav and PPTA upper limits are in fact obtained at each frequency independently, our analysis takes advantage of this by using the constraining power for the GWB spectrum at each frequency separately.

Fig. 7 shows that upper limits add very little knowledge to our understanding of the SMBHB population as constrained by the priors on our model parameters. This is in agreement with there being no tension between the current PTA non-detection of the GWB and other astrophysical observations, as extensively discussed in Middleton et al. (2018). The range of characteristic strain of the GWB predicted by the prior ranges of our model $10^{-16} < h_c < 10^{-15}$, shown in the upper left plot of Fig. 7, is only mildly reduced by current PTA observations. Therefore, PTAs are starting to probe the interesting, astrophysical region of the parameter space, without yet being able to rule out significant areas, as can be seen in the posterior
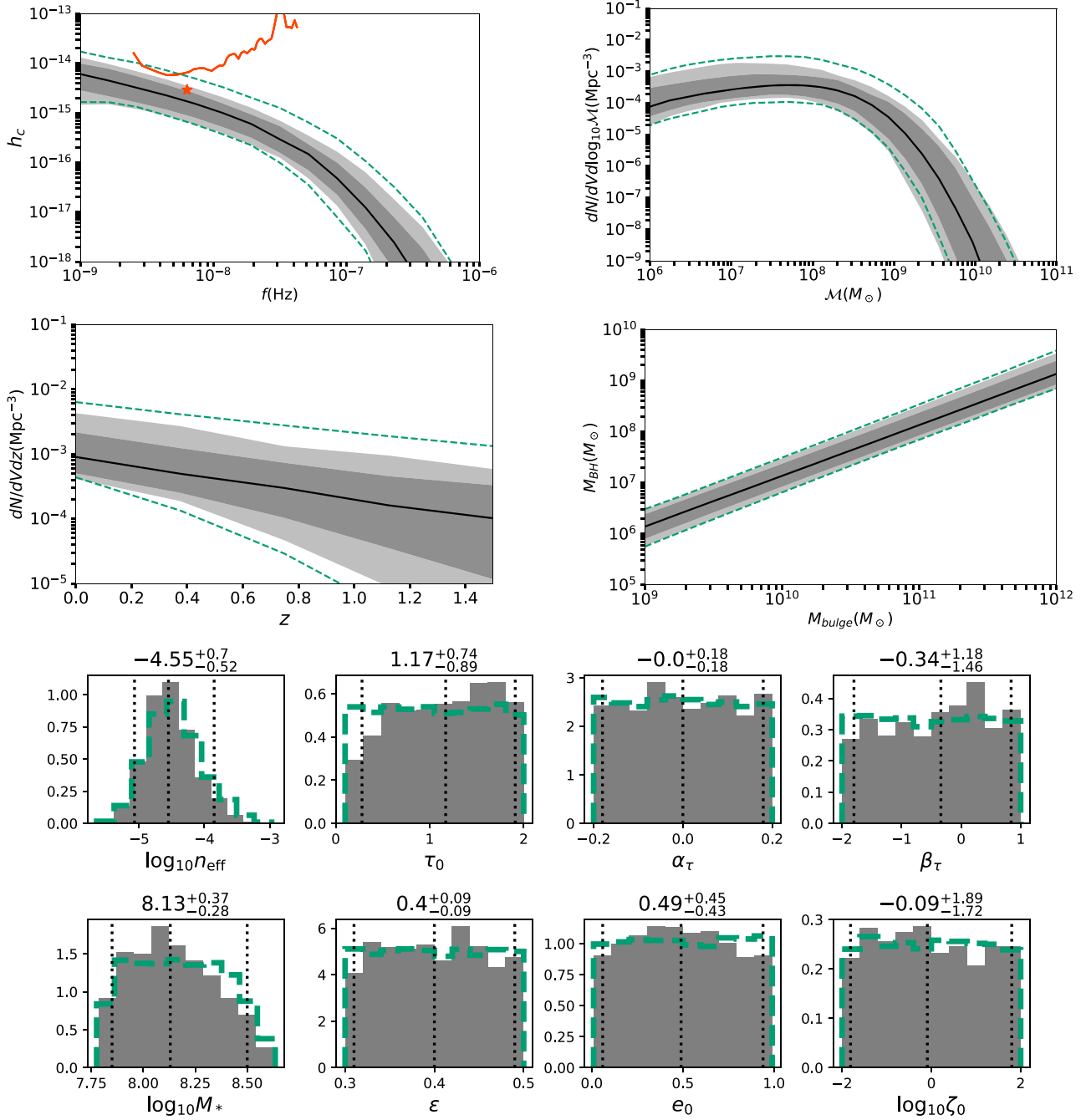
Figure 7. Implication of a 95 per cent upper-limit of $A(f = 1\,\mathrm{yr}^{-1}) = 1 \times 10^{-15}$, corresponding to the most stringent PTA upper limit to date. The posterior for the spectrum (top left), mass (top right), redshift functions (centre left), and $M_{\mathrm{bulge}}-M_{\mathrm{BH}}$ relations (centre right) are shown as shaded areas, with the central 68 per cent and 90 per cent confidence regions indicated by progressively lighter shades of grey, and the solid black line marking the median of the posterior. While the prior is indicated by green dashed lines. The solid orange line and star in the top left panel indicate the frequency dependent and nominal frequency integrated 95 per cent upper limit from Shannon et al. (2015) respectively. The bottom row shaded histograms show the marginalized posteriors for selected model parameters with the prior distributions indicated with green dashed lines, see Section 5 and Table 1.

distribution of the model parameters shown at the bottom of Fig. 7. This results into a logarithmic Bayesian evidence $\log_e \mathcal{Z}(10^{-15}) = -0.55$. The evidence is normalized so that an ideal reference model that is unaffected by the measurement has $\log_e = 0$. The log evidence can therefore be directly interpreted as the Bayes factor against such a model. In this specific case, we find $e^{-0.55} = 0.58$, indicating that

current upper limits do not significantly disfavour the prior range of our astrophysical model. This can also be seen in the bottom row posteriors of Fig. 7 where the posterior and prior distributions are almost identical, e.g. the effective merger rate (top left histogram) has an upper limit of $n_{\mathrm{eff}} \sim 1.4(2.1) \times 10^{-4}\,\mathrm{Mpc}^{-3}\,\mathrm{Gyr}^{-1}$ for the posterior(prior) respectively.

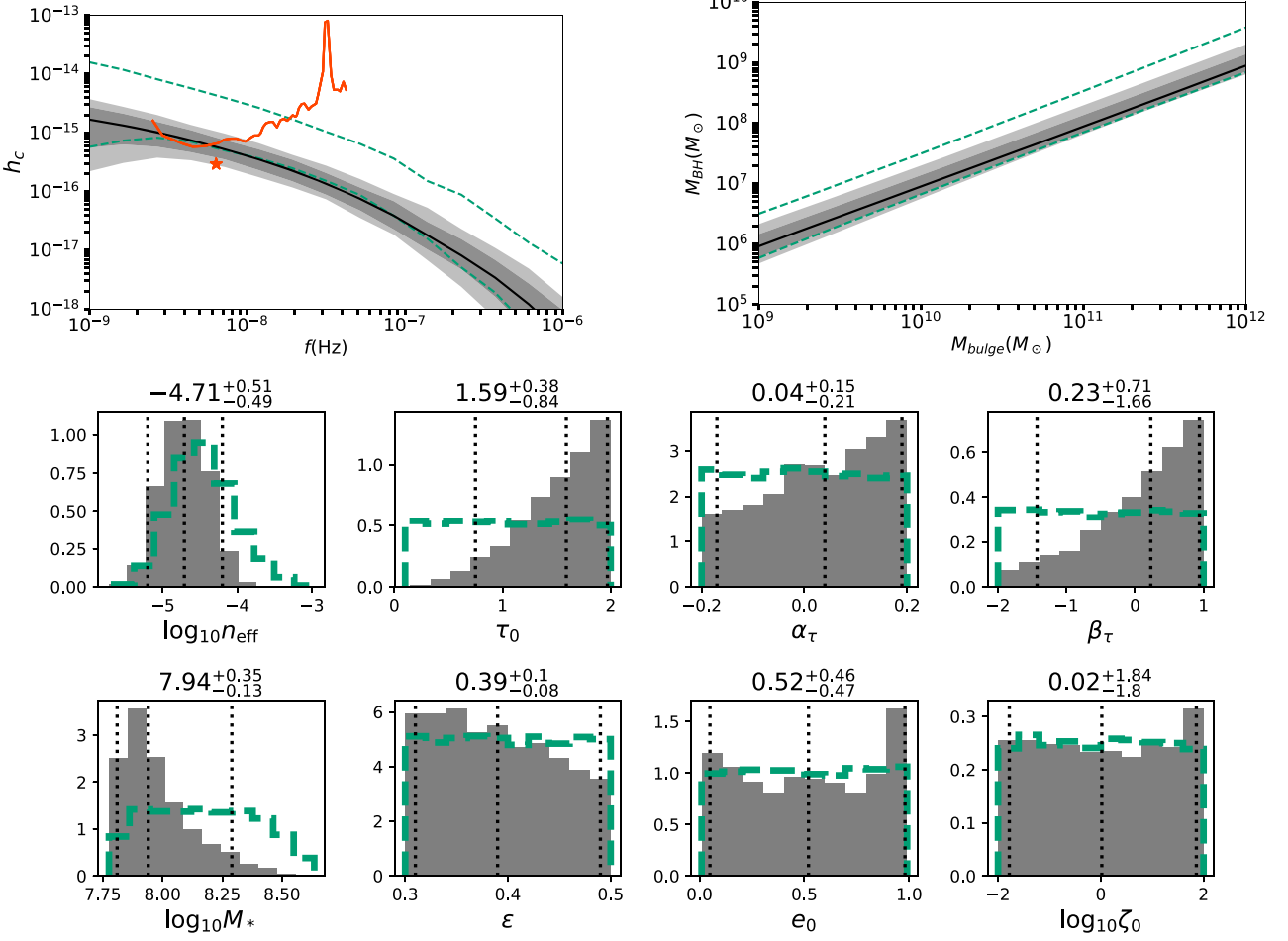**Figure 8.** Same as Fig. 7 without the posterior distributions for the mass and redshift functions, but for an upper limit of $A(f = 1 \, \mathrm{yr}^{-1}) = 1 \times 10^{-16}$.

6.2.2 Future upper limit at $A(f = 1 \, yr^{-1}) = 1 \times 10^{-16}$

To investigate what useful information on astrophysical observables can be extracted by future improvements of the PTA sensitivity, we have shifted the upper limit down by an order of magnitude to $A(f = 1 \, \mathrm{yr}^{-1}) = 1.0 \times 10^{-16}$, indicative of the possible capabilities in the SKA era (Janssen et al. 2015).

Results are shown in Fig. 8. Unlike the current situation, a future upper limit can put significant constraints on the allowed parameter space, also reflected in value of the Bayesian evidence $\log_e \mathcal{Z}(10^{-16}) = -4.32$. The odds ratio compared to a reference model untouched by the limit is now $e^{-4.32} = 0.013$, indicating that our astrophysical prior would be disfavoured a 2.5σ level. The Bayes factor $\mathcal{B} = \exp \mathcal{Z}(10^{-15}) / \exp \mathcal{Z}(10^{-16}) \approx 43$ provides evidence that there is tension between current constraints on astrophysical observables (defining our prior) and a PTA upper limit of 10^{-16} on the GWB level. The top left panel of Fig. 8 shows that h_c is relegated at the bottom of the allowed prior range, and the top right panel indicates that a low normalization to the $M_{\mathrm{BH}}-M_{\mathrm{bulge}}$ relation is preferred. The bottom row posteriors in Fig. 8 show significant updates with respect to their prior distributions. A more restrictive upper limit on the effective merger rate (top left histogram) at $n_{\mathrm{eff}} \sim 6.3 \times 10^{-5} \, \mathrm{Mpc}^{-3} \, \mathrm{Gyr}^{-1}$ can be placed and the distribution of all parameters defining the merger time-scale are skewed towards high values, meaning that longer merger time-scales, i.e. fewer

mergers within the Hubble time, are preferred. Besides favouring lower merger rates, light SMBH are also required, as shown by the posterior of the M_* parameter. Lastly, there is a slight preference for SMBHBs to be very eccentric and in dense stellar environments, although the whole prior range of these parameters is still possible.

6.2.3 Ideal Upper limit at $A(f = 1 \, yr^{-1}) = 1 \times 10^{-17}$

Pushing the exercise to the extreme, we shift the future upper limit down by another order of magnitude to $A(f = 1 \, \mathrm{yr}^{-1}) = 1.0 \times 10^{-17}$, which might be reached in the far future by a post-SKA facility (Janssen et al. 2015). None the less, this unveils what would be the consequences of a severe non-detection, well below the level predicted by current SMBHB population models. Fig. 9 compares inference on model parameters for the *P*PTA17 run, assuming either standard or extended prior distributions.

If we assume standard priors, constraints are pushed to the extreme compared to those derived in the *P*PTA16 case. The Bayesian evidence is now $\log_e \mathcal{Z}(10^{-17}) = -13.69$. The odds ratio compared to a reference model untouched by the limit becomes $e^{-13.69} \approx 10^{-6}$, indicating that our astrophysical prior would be severely disfavoured at a 5σ level. This would rule out the vast majority of our current constraints on the GSMF, pair fraction, merger time-scale, and $M_{\mathrm{bulge}}-M_{\mathrm{BH}}$ relation. Although the effective merger rate is only limited to be smaller than $n_{\mathrm{eff}} \sim 5.6 \times 10^{-5} \, \mathrm{Mpc}^{-3} \, \mathrm{Gyr}^{-1}$, all
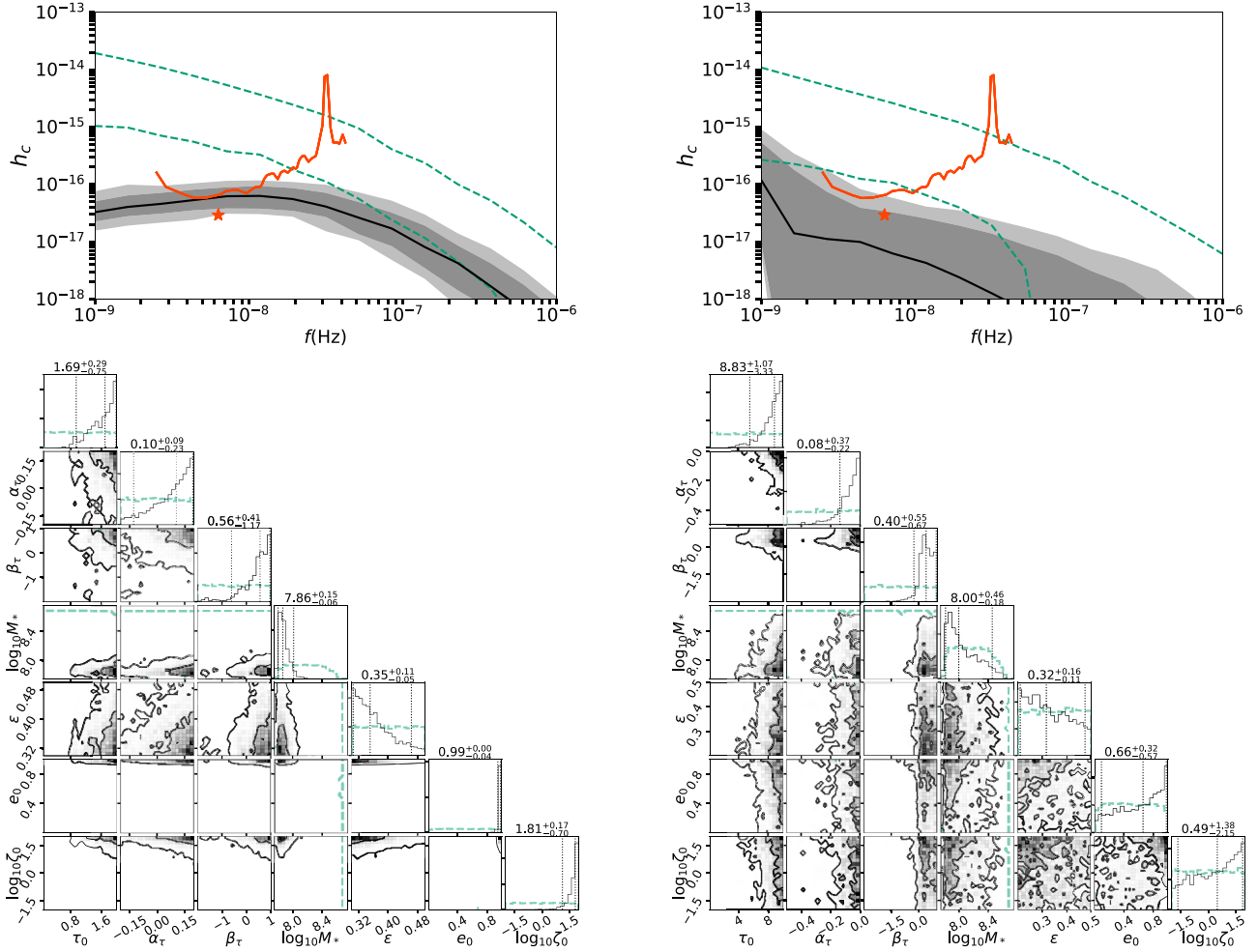
Figure 9. Implications of a 95 per cent upper-limit of $A(f = f = 1\,\mathrm{yr}^{-1}) = 1 \times 10^{-17}$ with standard (left-hand column) and extended (right-hand column) prior distributions. In each column, the top panel shows the posterior of the recovered GWB spectrum with the central 68 per cent and 90 per cent confidence regions indicated by progressively lighter shades of grey, and the solid black line marking the median of the posterior. While the prior is indicated by green dashed lines. The bottom corner plots show the (1)2D posteriors for each model parameter pairs as shaded area, the different levels of shading indicate the 5 per cent, 50 per cent, and 95 per cent confidence regions. In each sub-panel, the green dashed lines show the 100 per cent confidence levels for the prior.

other parameters in the bottom row corner plots in Fig. 9 show rather extreme posterior distributions. Since our standard prior does not allow stalling of low-redshift SMBHBs (the maximum normalization of the local merger time-scale being 2 Gyr), skewing the merger time-scale to extreme values is not sufficient to explain the non-detection. Further, the normalization to the $M_{\mathrm{BH}} - M_{\mathrm{bulge}}$ is severely pushed to the low end, at $M_* < 10^8\,\mathrm{M}_\odot$, thus completely ruling out several currently popular relations (e.g. Kormendy & Ho 2013; McConnell & Ma 2013). Even with the smallest possible $M_{\mathrm{BH}} - M_{\mathrm{bulge}}$, a non-detection at $A(f = 1\,\mathrm{yr}^{-1}) = 1.0 \times 10^{-17}$ requires a very high-frequency turnover of the GWB (see upper left panel of Fig. 9), which can be realized only if all binaries have eccentricity $e_0 > 0.95$ and reside in extremely dense environments (at least a factor of 10 larger than our fiducial Dehnen profile).

As mentioned above, our standard prior on the total merger time-scale (see Section 5.3), implies that stalling hardly occurs in nature. Although this is backed up by recent progresses in *N*-body simulations and the theory of SMBHB hardening in stellar environments (see e.g. Sesana & Khan 2015; Vasiliev et al. 2015), we want to keep all possibilities open and check what happens when arbitrary long merger time-scales, and thus stalling, are allowed.

We note, however, that such a model is intrinsically inconsistent, because when very long merger time-scales are allowed, one should also consider the probable formation of SMBH triplets, due to subsequent galaxy mergers. Triple interactions are not included in our models but they have been shown (Bonetti et al. 2018; Ryu et al. 2018) to drive about one-third to the stalled SMBHBs to coalescence in less than 1 Gyr. Therefore, we caution that actual constrains on model parameters would likely be more stringent than what described in the following. The extended prior distributions relaxes the strong evidence of -13.69 to $\log_e \mathcal{Z}(10^{-17}) = -4.56$ and the Bayes factor becomes comparable to the *P*PTA16, this is mainly due to allowing binaries to stall as the merger time-scale increases to $\tau_0 > 5.5$ Gyr. The extreme constraints on the other parameters are consequently loosened, although posterior distributions of M_*, e_0, and ζ_0 indicate that light SMBHBs are favoured, along with large eccentricities and dense environments. The stalling of a substantial fraction of SMBHB pushes the effective merger rate to drop below $n_{\mathrm{eff}} \sim 1.1 \times 10^{-5}\,\mathrm{Mpc}^{-3}\,\mathrm{Gyr}^{-1}$.

Table 2 summarizes the increasing constraining power as the upper limits are lowered. As they become more restrictive, fewer mergers are allowed. The effective merger rate is therefore pushed

Table 2. List of bounds for selected parameters and evidences for the upper limit cases. The 95 per cent upper bounds for the effective merger rate, the 95 per cent lower bounds for the merger time norm, and the evidences are reported in the columns. The five rows list the values for the standard prior, current, future, and ideal upper limit posteriors from top to bottom.

Parameter	$\log_{10} n_{eff}$	τ_0	$\log_e \mathcal{Z}$
Standard prior: no upper limit	< -3.68	>0.2	0
Standard prior: $A = 1 \times 10^{-15}$	< -3.85	>0.28	-0.55
Standard prior: $A = 1 \times 10^{-16}$	< -4.2	>0.75	-4.32
Standard prior: $A = 1 \times 10^{-17}$	< -4.25	>0.94	-13.69
Extended prior: $A = 1 \times 10^{-17}$	< -4.96	>5.5	-4.56

to be as low as possible with long merger time-scales, low SMBHB masses, large eccentricities, and dense environments. Bayes factors comparing the current observational constraints, i.e. the prior ranges, with posterior constraints can be calculated from the evidences. These, however, show that the tension increases from 0.6σ with the current upper limit of $A(f = 1 \text{ yr}^{-1}) = 1 \times 10^{-15}$ to 5σ with an ideal upper limit at $A(f = 1 \text{ yr}^{-1}) = 1 \times 10^{-17}$. Relaxing the upper bound on the merger time norm and other constraints (see Section 5.6) can alleviate the tension between current observations and such a upper limit to 2.6σ (although this does not take into account for triple-induced mergers, as mentioned above).

6.3 Simulated detections

Although it is useful to explore the implication of PTA upper limits, it is more interesting to consider the case of a future detection, which is expected within the next decade (Rosado et al. 2015; Taylor et al. 2016b; Kelley et al. 2017b). We therefore turn our attention at simulated detections and their potential to put further constraints on the astrophysics of galaxy evolution and SMBHB mergers. To simulate a detection, the GWB strain is computed for a specific set of parameters, i.e. the injected signal, which is detected at the computed values $A_k = h_c(f_k)$ with an uncertainty of $\sigma_{det, k}$ given by equation (27). As these simulated detections are very ideal, effects that could pollute the strength the GWB detection are mostly neglected. However, we include an empirical term in the computation of S_n (see equations 9 and 10 in Paper II) to account for the flattening of the sensitivity at low frequencies.

The amplitude of the simulated GWB is defined by the 16 parameters describing the SMBHB merger rate. We fix those as follows: $(\Phi_0, \Phi_I, M_0, \alpha_0, \alpha_I, f_0, \alpha_f, \beta_f, \gamma_f, \tau_0, \alpha_\tau, \beta_\tau, \gamma_\tau, M_*, \alpha_*, \epsilon) = (-2.6, -0.45, 10^{11.25}, -1.15, -0.1, 0.025, 0.1, 0.8, 0.1, 0.8, -0.1, -2, -0.1, 10^8, 1, 0.3)$. The low-frequency turnover is defined by the two extra parameters (e_0, ζ_0). We fix $\zeta_0 = 1$ and we produce two GWB spectra distinguished solely by the assumed value of the eccentricity: $e_i = 0.01$ (circular case) and $e_i = 0.9$ (eccentric case). This set of parameters is chosen such that it results in a GWB strain of $h_c = 5.0 \times 10^{-16}$ at $f = 1 \text{ yr}^{-1}$ (i.e. well within current upper limits), whilst being consistent with the current constraints of all the relevant astrophysical observables:

(i) GSMF: the values for $(\Phi_0, \Phi_I, M_0, \alpha_0, \alpha_I)$ are chosen, such that they accurately reproduce the currently best measured GSMF, i.e. they are close to the best-fitting values of the reparametrization described in Section 5.1;

(ii) merger time-scale: $\tau_0 = 0.8$ Gyr is chosen to match the predicted merger time-scales found in simulations by Lotz et al. (2010), while $\beta_\tau = -2$ is chosen to match the expected redshift evolution of the merger time-scale from Snyder et al. (2017);

(iii) $M_{bulge} - M_{BH}$ relation: $(M_*, \alpha_*, \epsilon)$ have been chosen to produce the injected characteristic strain amplitude, consistent with the allowed prior shown in Fig. 5.

The other parameters are chosen to be close to the centre of their prior ranges, except for the eccentricity, as mentioned above.

6.3.1 Circular case

Fig. 10 shows a comparison of the results of the *IPTA30* (left-hand column) and *SKA20* (right-hand column) set-ups for the circular case ($e_0 = 0.01$). In the *IPTA30*(*SKA20*) case the GWB has been detected in 10(14) frequency bins up to frequencies of $\sim 1(2) \times 10^{-8}$ Hz, for a total detection S/N $\mathcal{S} \approx 20(100)$. Qualitatively, both detections provide some extra constraints on selected prior parameters. The injected spectrum, mass and redshift function are recovered increasingly better as the S/N increases. Still, a broad portion of the initial parameter space is allowed, especially for the redshift evolution of the SMBHB merger rate. It should be noted that PTAs have the most constraining power around the bend of the mass function, at the SMBHB chirp mass $\mathcal{M} \approx 3 \times 10^8 \, M_\odot$. The posterior panels at the bottom of Fig. 10 show that there is not much additional information gained compared to the prior knowledge for most of the parameters (full corner plots shown in Appendix A, available in electronic form), with three notable exceptions:

(i) merger time-scale: τ_0 is marginally constrained around the injected value (0.8 Gyr) in the *IPTA30* case, the constraint becomes better in the *SKA20* case. β_τ is also skewed towards low values (consistent with the $\beta_\tau = -2$ injection). A clean PTA detection thus potentially allow to constrain the time-scale of SMBHB coalescence, which can help in understanding the processes driving the merger;

(ii) $M_{bulge} - M_{BH}$ relation: the M_* panels show a tightening of the M_* distribution with increasing S/N. A detection would thus also allow to constrain the $M_{bulge} - M_{BH}$ relation;

(iii) eccentricity and stellar density: the posterior distributions for e_0 and $\log_{10} \zeta_0$ show some marginal update. In particular in the *SKA20* case, extreme eccentricities, above $e_0 > 0.9$ can be safely ruled out. Note that the absence of a low-frequency turnover also favours small value of ζ_0, fully consistent with the injected value $\zeta_0 = 1$.

6.3.2 Eccentric case

The results for the *IPTA30* and *SKA20* eccentric cases are shown in Fig. 11, with full corner plots reported in Appendix A, available in electronic form. In general results are comparable to the circular case shown above, as the only difference is in the injected eccentricity parameter. The left-hand column (*IPTA30* case) of Fig. 11 shows nearly identical posterior distributions to its circular counterpart reported in Fig. 10, this also translates into similar recovered spectrum, mass, and redshift functions.

However, in the *SKA20* case, the detection S/N is high enough to allow a clear detection of the spectrum turnover in the lowest frequency bins. Which is not the case for *IPTA30*, as can be seen in the top row spectra plots of Fig. 11. This has important consequences for astrophysical inference since an observable turnover is only possible if binaries are significantly eccentric and/or evolve in very dense environments. This is shown in the e_0 and ζ_0 posterior distributions at the bottom right of Fig. 11: eccentricities $e_0 < 0.4$ are excluded and densities higher than what predicted by the
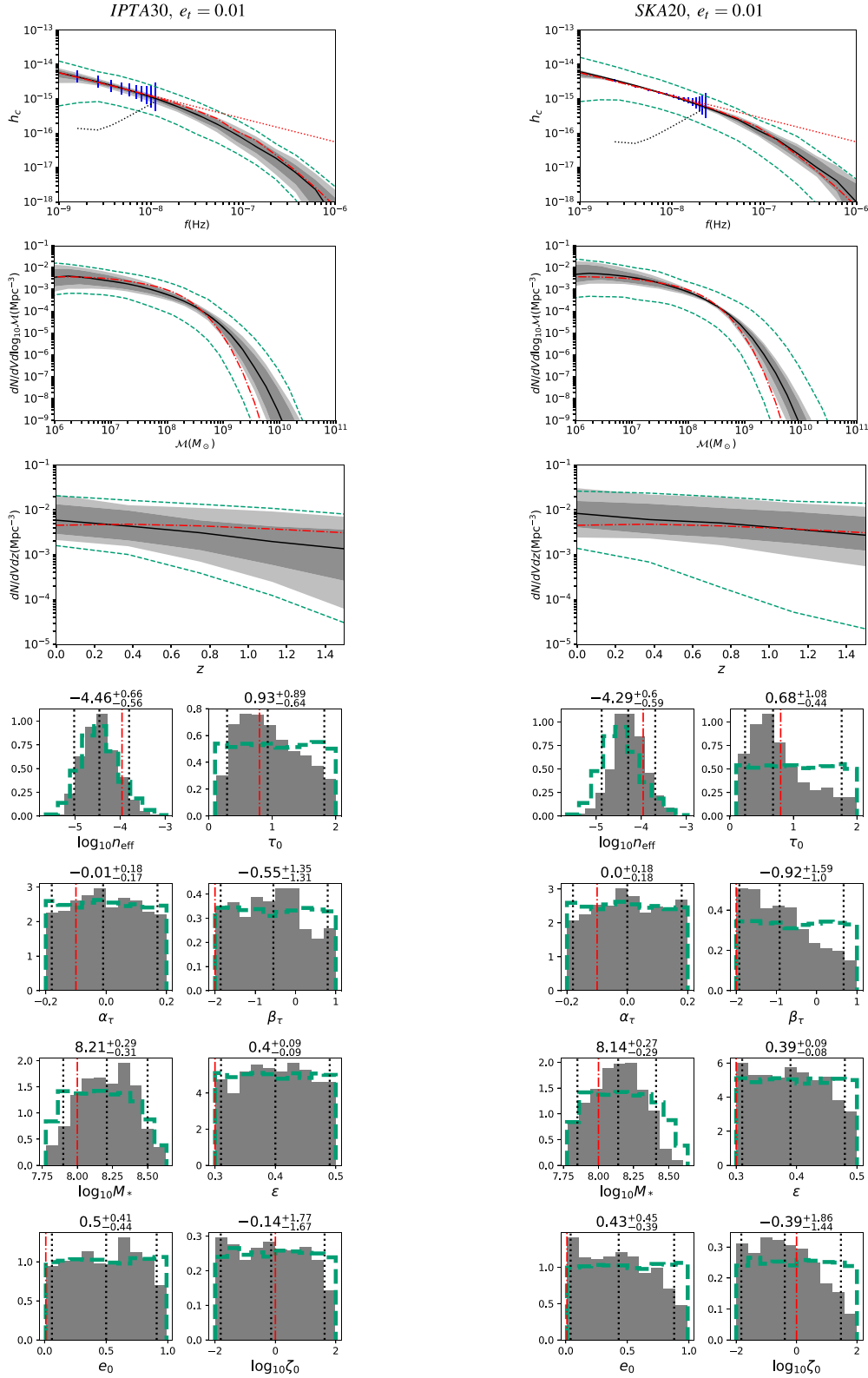
Figure 10. Implications of a PTA detection at a low (S/N ≈ 20, left-hand column) and moderate (S/N ≈ 100, right-hand column) significance, assuming an SMBHB population with default parameters and almost circular ($e_t = 0.01$) at decoupling. As in Fig. 7, the posteriors for the spectrum, mass, and redshift functions (in descending order from the top) are shown as shaded areas. In each of those panels, the dashed green lines indicate the prior and the dash–dotted red line indicates the injected model. In the top panels, the vertical blue bands indicate the 68 per cent confidence interval of the observed signal amplitude at each frequency bin. The dotted line is the nominal 1σ sensitivity of the considered PTA. The dotted red line shows the simulated spectrum assuming no drop due to missing sources at high frequencies. The bottom row histograms in shades of grey show the marginalized posteriors for selected model parameters with the prior distributions indicated by dashed green lines. The injected parameter values are marked by red dashed–dotted lines.
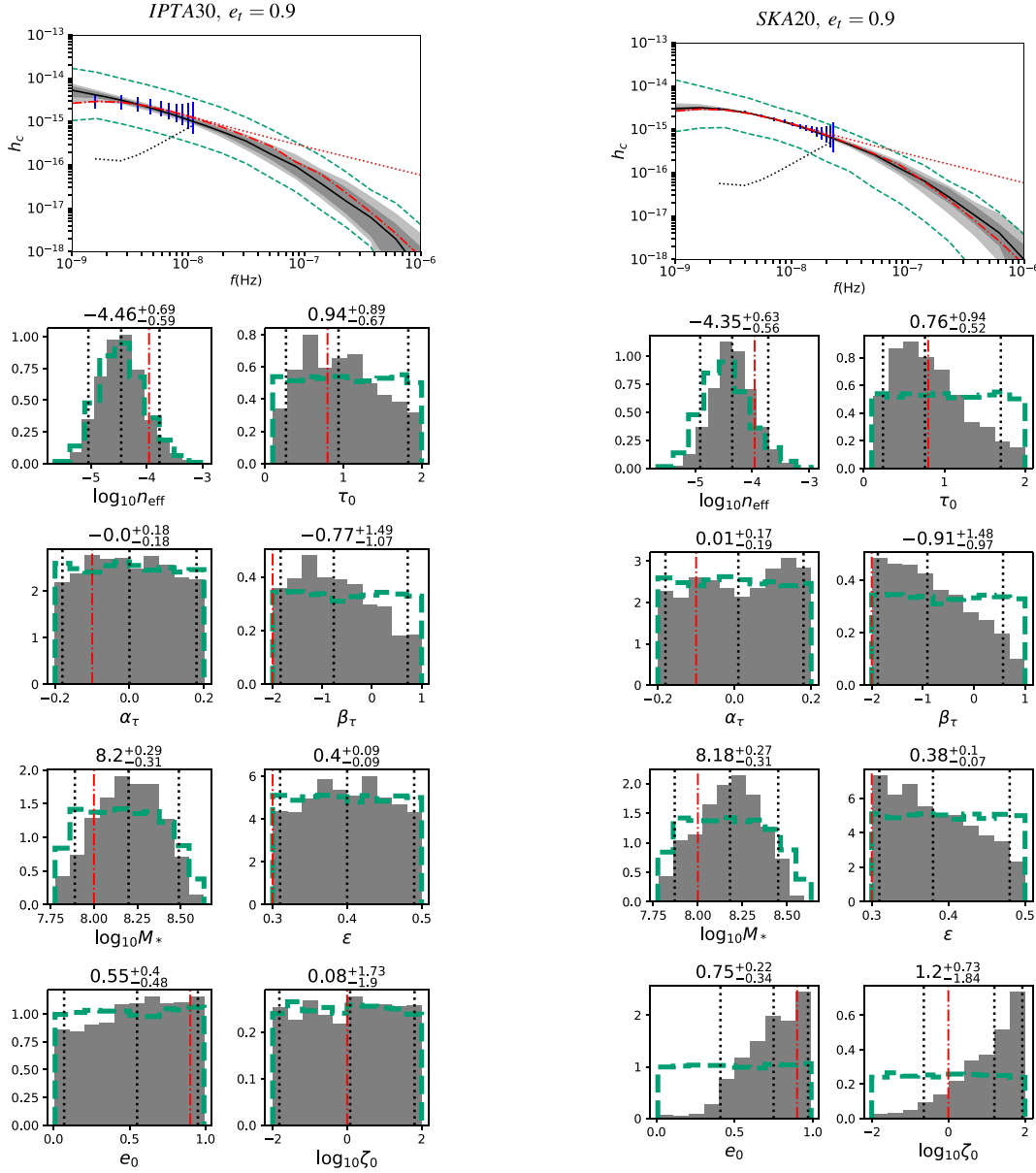
Figure 11. Same as Fig. 10, but assuming decoupling eccentricity of $e_t = 0.9$. The posterior panels of the mass and redshift functions are analogue to those in the circular case and thus not shown here.

fiducial Dehnen model are strongly favoured. The full corner plot A1 reported in the Supporting Information A, available in electronic form, also highlights the $e_0 - \zeta_0$ degeneracy, as a low-frequency turnover can be caused by either parameters; very eccentric binaries in low-density stellar environments pruduce a turnover at the same frequency as more circular binaries in denser stellar environments. Additionally, a large region in the $e_0 - \zeta_0$ plane has been ruled out($e_0 > 0.41$ and $\log_{10}\zeta_0 > -0.63$). This also prompts some extra constrain in the $M_{BH}-M_{bulge}$ relation, as can be seen in the trends in the α_* and ϵ distributions.

Summarizing, little extra astrophysical information (besides the non-trivial confirmation that SMBHBs actually do merge) can be extracted in the *IPTA30*, whereas many more interesting constrains emerge as more details of the GWB spectrum are unveiled in the *SKA20* case. Although posteriors on most of the parameters remain broad, the typical SMBHB coalescence time-scale can be

constrained around the injected value; the posterior distributions of n_{eff} and M_* are tightened, providing some extra information on the SMBHB merger rate and on the $M_{BH}-M_{bulge}$ scaling relation; significant constrains onto the SMBHB eccentricity and immediate environment can be placed if a low-frequency turnover is detected.

6.3.3 Ideal case

We show ideal detections for both the circular and eccentric cases in Fig. 12. Although, such detection may not be achievable by PTAs in the foreseeable future, these results show what might be constrained in principle by combining astrophysical prior knowledge to precise measurements of the amplitude and shape of the nano-Hz stochastic GWB.
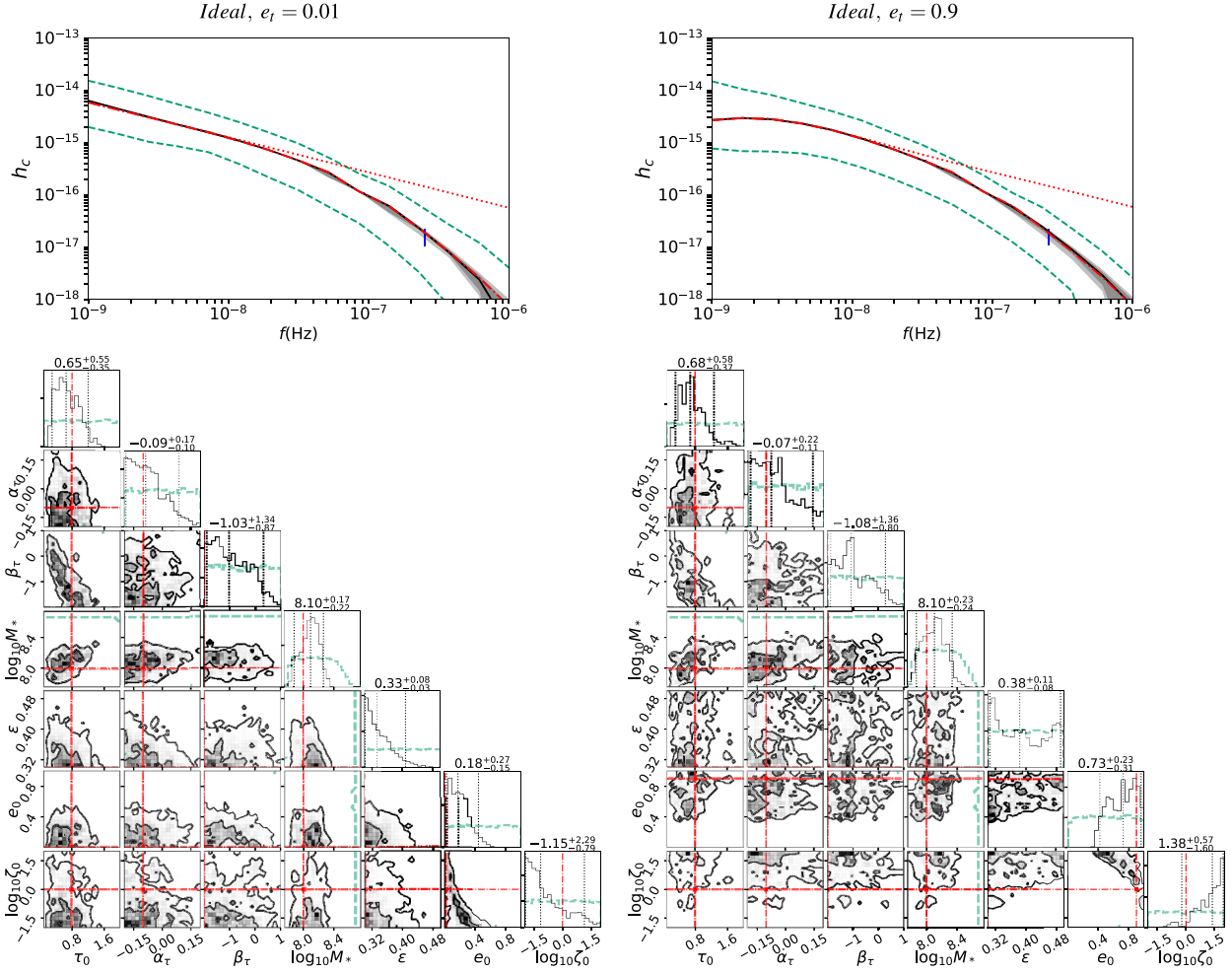
Figure 12. Implications of an ideal detection with 500 MSPs timed at sub-ns precision for 30 yr. The injected model has default parameters with $e_t = 0.01$ (left-hand column) and $e_t = 0.9$ (right-hand column). Panel sequence and style as in Fig. 9.

The spectra, mass, and redshift functions (not shown in the figure) are recovered extremely well in both cases. Both corner plots also show interesting constrains on some key parameters. The typical merger time-scale τ_0 is correctly measured and constrained within less than 1 Gyr uncertainty, and clear trends in α_τ and β_τ provide some extra information on the merger time-scale evolution with galaxy mass and redshift. Note that those are parameters defining the *SMBHB* coalescence time which are unlikely to be measured by any other means. The normalization of the $M_{\rm bulge} - M_{\rm BH}$ relation is also significantly constrained, as shown by the tight M_* posterior distributions. Again we see both in the circular and eccentric cases the degeneracy between eccentricity e_0 and stellar density ζ_0, as in the *SKA20* eccentric case above. The posterior regions contain the injected values and exclude a large area from the prior: $e_0 > 0.42$, $\log_{10}\zeta_0 > -0.22$ for the circular and $e_0 < 0.45$, $\log_{10}\zeta_0 < 1.14$ for the eccentric case (95th percentile). Although, the *ideal* eccentric detection has a vastly larger S/N than its *SKA20* analogue, the constraints on e_0 and $\log_{10}\zeta_0$ are comparable due to the degeneracy between the two parameters. Table 3 shows the increasing constraining power on selected key parameters as the detection S/N improves for the eccentric case.

7 CONCLUSIONS AND OUTLOOK

We have presented an analytic parametrized model for the SMBHB merger rate in terms of astrophysical observables, including: GSMF, pair fraction, merger time-scale, and black hole–host galaxy relations. We described each individual ingredient with a simple analytic function and exploited our state of the art knowledge from observations, theory, and simulations to define the prior range of each free parameter in the model. We then sampled the allowed parameter space (18 parameters in total) to produce an updated measure of the expected amplitude of the stochastic GWB across the frequency range. At $f = 1\,{\rm yr}^{-1}$ our model with the prior selection from Section 5 results in a characteristic strain $10^{-16} < h_c < 10^{-15}$, confirming recent findings (e.g. Middleton et al. 2018). We used our model to interpret current and future PTA upper limits and detections, linking the outcome of PTA observations to constraints on interesting observed quantities describing the cosmic population of merging galaxies and SMBHs.

Consistent with our previous results (Middleton et al. 2018), we find that current PTA upper limits can only add very little to the prior knowledge of the physical parameters as determined by current observations and simulations. However, as the sensitivity of PTA improves over time, upper limits can become stringent

Table 3. List of credible intervals for selected parameters of the model. Each column reports the median value together with the errors bracketing the 90 per cent confidence regions for selected parameters. The five rows list the boundaries defined by the prior distributions, the posterior distributions as measured in the *IPTA30*, *SKA20*, and *ideal* cases, and the injected values from top to bottom.

Parameter	$\log_{10}n_{\mathrm{eff}}$	τ_0	$\log_{10}M_*$	e_0	$\log_{10}\zeta_0$
Prior	$-4.47^{+0.79}_{-0.61}$	$1.04^{+0.86}_{-0.85}$	$8.17^{+0.36}_{-0.32}$	$0.50^{+0.44}_{-0.44}$	$0.01^{+1.79}_{-1.80}$
IPTA30	$-4.46^{+0.70}_{-0.59}$	$0.94^{+0.89}_{-0.67}$	$8.20^{+0.29}_{-0.31}$	$0.55^{+0.40}_{-0.48}$	$0.08^{+1.73}_{-1.90}$
SKA20	$-4.35^{+0.63}_{-0.56}$	$0.76^{+0.94}_{-0.52}$	$8.18^{+0.26}_{-0.31}$	$0.75^{+0.22}_{-0.34}$	$1.20^{+0.74}_{-1.83}$
Ideal	$-4.24^{+0.38}_{-0.38}$	$0.68^{+0.58}_{-0.37}$	$8.10^{+0.23}_{-0.24}$	$0.73^{+0.23}_{-0.31}$	$1.38^{+0.57}_{-1.60}$
Injection	-4.0	0.8	8.0	0.9	0.0

enough to probe interesting regions of the prior parameter space. The more stringent the upper limit becomes, the more extreme the conditions for the SMBHB population must be. Longer merger time (maybe even stalling) of binaries, less massive black holes and a spectral turnover at $f > 10$ nano-Hz, all contribute to reduce the characteristic strain of the GWB in the PTA observable band. A upper limit at $A(f = 1\,\mathrm{yr}^{-1}) = 1.0 \times 10^{-16}$ indicates moderate tension (at a nominal 2.5σ level) between PTA observations and current astrophysical constraints. Pushing it down to $A(f = 1\,\mathrm{yr}^{-1}) = 1.0 \times 10^{-17}$ would imply a strong 5σ tension with our current knowledge of the process of SMBHB formation and dynamics. Explaining a GWB below this level requires invoking a combination of SMBHB stalling, overestimate of the SMBH–host galaxy scaling relations, extreme eccentricities, and dense environments.

Although exploring progressively stringent upper limits is a useful exercise, we are particularly interested in addressing the astrophysical significance of a future PTA detection. A weak initial detection at S/N $\mathcal{S} \approx 5\text{–}10$ will only put marginally better constraints on the underlying astrophysics of galaxies and SMBHs. As the detection significance increases, so do the constraints, as shown in Table 3. A full SKA-type array, detecting the GWB at $\mathcal{S} \approx 30 - 40$, will enable us to place important constrains on the normalization of the cosmic SMBHB merger rate, the time elapsed between galaxy pairing and SMBHB mergers, the normalization of the SMBH–host galaxy relations and the dynamical properties of the merging SMBHBs. Since there is limited information in the GWB amplitude and spectral shape, even an ideal detection, reconstructing the GWB almost perfectly, will allow to place constrains only on a subset of the 18 parameters of the model. In particular, we have identified four quantities that can be well constrained with PTA observations on the GWB: the merger time-scale of the SMBHB, the $M_{\mathrm{bulge}}-M_{\mathrm{BH}}$ relation, the eccentricity – density of the stellar environment and the overall effective merger rate of galaxies. This can be understood in terms of the distinctive features of the GWB spectrum. The observation of a low-frequency turnover constrains the dynamics of individual SMBHBs, providing information about their eccentricity and the effectiveness of the hardening mechanism driving the merger process (i.e. the density of the stellar environment). The high -frequency drop is determined by the high mass tail of the SMBH mass function, which is directly connected to the $M_{\mathrm{BH}}-M_{\mathrm{bulge}}$ relation. Whether a GWB is detected or not, immediately put a (loose) constrain on the SMBHB merger time-scale. And the general amplitude of the strain allows to refine the measurement of the SMBHB merger time-scale as well as determining the overall cosmic merger rate.

We stress that our model is still idealized in many ways. In particular, we employ a deterministic relation between model parameters and GWB spectrum. In reality, the GWB has some intrinsic variance due to the specific statistical realization of the SMBHB population occurring in nature. This is particularly important because the GWB strain is dominated by the most massive SMBHBs in the Universe, which are intrinsically rare. Including a self-consistent computation of the variance in the model requires extensive Monte Carlo simulations, making the computation of the likelihood function prohibitively expensive for a direct nested sampling exploration of the parameter space. This difficulty can be overcome in the future by combining targeted simulations, sparsely sampling the parameter space with dedicated interpolation processes, which was demonstrated by Taylor et al. (2017) on a parameter space of reduced complexity.

Although the introduction of intrinsic variance will likely degrade the inference on astrophysical observables, we also stress that we are still not using all the information encoded in the GW signal. In particular, information extracted from the shape and normalization of the GWB should be complemented with the statistics and properties of individually resolvable sources, which will provide precious extra information about the most massive SMBHBs and their physical properties (e.g. their eccentricity, Taylor et al. 2016a). Likewise, non-stationarity of the GWB will be indicative of highly eccentric binaries, allowing to disentangle eccentricity from extreme environments as the cause of a putative low-frequency turnover. A comprehensive inference model from PTA observations will have to simultaneously combine all this information. Although there is still a lot of work to do, this study constitutes an important step forward in this endeavour.

REFERENCES

Arzoumanian Z. et al., 2016, ApJ, 821, 13
Arzoumanian Z. et al., 2018, ApJS, 235, 37
Begelman M. C., Blandford R. D., Rees M. J., 1980, Nature, 287, 307
Bernardi M., Meert A., Vikram V., Huertas-Company M., Mei S., Shankar F., Sheth R. K., 2014, MNRAS, 443, 874
Bonetti M., Sesana A., Barausse E., Haardt F., 2018, MNRAS, 477, 2599
Chen S., Middleton H., Sesana A., Del Pozzo W., Vecchio A., 2017a, MNRAS, 468, 404 (Paper II)
Chen S., Sesana A., Del Pozzo W., 2017b, MNRAS, 470, 1738 (Paper I)

Conselice C. J., Bershady M. A., Dickinson M., Papovich C., 2003, AJ, 126, 1183
Conselice C. J., Wilkinson A., Duncan K., Mortlock A., 2016, ApJ, 830, 83
Dehnen W., 1993, MNRAS, 265, 250
Del Pozzo W., Veitch J., 2015, CPNest: Parallel Nested Sampling in Python. Available at: https://github.com/johnveitch/cpnest
Desvignes G. et al., 2016, MNRAS, 458, 3341
Dotti M., Sesana A., Decarli R., 2012, Adv. Astron., 2012, 3
Duncan K. et al., 2019, ApJ, 876, 110
Dvorkin I., Barausse E., 2017, MNRAS, 470, 4547
Event Horizon Telescope Collaboration et al., 2019, ApJ, 875, L1
Foster R. S., Backer D. C., 1990, ApJ, 361, 300
Graham A. W., Scott N., 2012, preprint (arXiv:1211.3199)
Hellings R. W., Downs G. S., 1983, ApJ, 265, L39
Jaffe A. H., Backer D. C., 2003, ApJ, 583, 616
Janssen G. et al., 2015, Proc. Sci., Gravitational Wave Astronomy with the SKA. SISSA, Trieste, PoS(AASKA14)037
Jenet F. A. et al., 2006, ApJ, 653, 1571
Keenan R. C. et al., 2014, ApJ, 795, 157
Kelley L. Z., Blecha L., Hernquist L., 2017a, MNRAS, 464, 3131
Kelley L. Z., Blecha L., Hernquist L., Sesana A., Taylor S. R., 2017b, MNRAS, 471, 4508
Khan F. M., Berentzen I., Berczik P., Just A., Mayer L., Nitadori K., Callegari S., 2012, ApJ, 756, 30
Kitzbichler M. G., White S. D. M., 2008, MNRAS, 391, 1489
Kormendy J., Ho L. C., 2013, ARA&A, 51, 511
Kormendy J., Fisher D. B., Cornell M. E., Bender R., 2009, ApJS, 182, 216
Lentati L. et al., 2015, MNRAS, 453, 2576
Li Y.-R., Ho L. C., Wang J.-M., 2011, ApJ, 742, 33
Lotz J. M., Jonsson P., Cox T. J., Primack J. R., 2010, MNRAS, 404, 575
Lotz J. M., Jonsson P., Cox T. J., Croton D., Primack J. R., Somerville R. S., Stewart K., 2011, ApJ, 742, 103
McConnell N. J., Ma C.-P., 2013, ApJ, 764, 184
Middleton H., Del Pozzo W., Farr W. M., Sesana A., Vecchio A., 2016, MNRAS, 455, L72
Middleton H., Chen S., Del Pozzo W., Sesana A., Vecchio A., 2018, Nat. Commun., 9, 573
Mingarelli C. M. F. et al., 2017, Nat. Astron., 1, 886
Mirza M. A., Tahir A., Khan F. M., Holley-Bockelmann H., Baig A. M., Berczik P., Chishtie F., 2017, MNRAS, 470, 940
Moore C. J., Taylor S. R., Gair J. R., 2015, Class. Quantum Gravity, 32, 055004
Mortlock A. et al., 2015, MNRAS, 447, 2
Mundy C. J., Conselice C. J., Duncan K. J., Almaini O., Häußler B., Hartley W. G., 2017, MNRAS, 470, 3507
Peters P. C., Mathews J., 1963, Phys. Rev., 131, 435
Rajagopal M., Romani R. W., 1995, ApJ, 446, 543
Ravi V., Wyithe J. S. B., Hobbs G., Shannon R. M., Manchester R. N., Yardley D. R. B., Keith M. J., 2012, ApJ, 761, 84
Reardon D. J. et al., 2016, MNRAS, 455, 1751
Robotham A. S. G. et al., 2014, MNRAS, 444, 3986
Rosado P. A., Sesana A., Gair J., 2015, MNRAS, 451, 2417
Ryu T., Perna R., Haiman Z., Ostriker J. P., Stone N. C., 2018, MNRAS, 473, 3410
Salviander S., Shields G. A., Bonning E. W., 2015, ApJ, 799, 173
Schulze A., Wisotzki L., 2014, MNRAS, 438, 3422
Sesana A., 2013a, Class. Quantum Gravity, 30, 224014
Sesana A., 2013b, MNRAS, 433, L1
Sesana A., Khan F. M., 2015, MNRAS, 454, L66
Sesana A., Vecchio A., Colacino C. N., 2008, MNRAS, 390, 192
Sesana A., Vecchio A., Volonteri M., 2009, MNRAS, 394, 2255
Sesana A., Shankar F., Bernardi M., Sheth R. K., 2016, MNRAS, 463, L6
Shankar F. et al., 2016, MNRAS, 460, 3119
Shannon R. M. et al., 2015, Science, 349, 1522
Simon J., Burke-Spolaor S., 2016, ApJ, 826, 11
Skilling J., 2004, in Fischer R., Preuss R., von Toussaint U., eds, AIP Conf. Proc., Vol. 735, Bayesian Inference and Maximum Entropy Methods in Science and Engineering. Am. Inst. Phys., New York, p. 395
Snyder G. F., Lotz J. M., Rodriguez-Gomez V., Guimarães R. d. S., Torrey P., Hernquist L., 2017, MNRAS, 468, 207
Taylor S. R., Huerta E. A., Gair J. R., McWilliams S. T., 2016a, ApJ, 817, 70
Taylor S. R., Vallisneri M., Ellis J. A., Mingarelli C. M. F., Lazio T. J. W., van Haasteren R., 2016b, ApJ, 819, L6
Taylor S. R., Simon J., Sampson L., 2017, Phys. Rev. Lett., 118, 181102
Vasiliev E., Antonini F., Merritt D., 2015, ApJ, 810, 49
Veitch J. et al., 2015, Phys. Rev. D, 91, 042003
Verbiest J. P. W. et al., 2016, MNRAS, 458, 1267
White S. D. M., Rees M. J., 1978, MNRAS, 183, 341
Xu C. K., Zhao Y., Scoville N., Capak P., Drory N., Gao Y., 2012, ApJ, 747, 85

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

Figure A1. Full corner plot of the 18 parameters for the PPTA15 case.
Figure A2. Full corner plot of the 18 parameters for the PPTA16 case.
Figure A3. Full corner plot of the 18 parameters for the PPTA17 case.
Figure A4. Full corner plot of the 18 parameters for the PPTA17 extended prior case.
Figure A5. Full corner plot of the 18 parameters for the IPTA30 circular case.
Figure A6. Full corner plot of the 18 parameters for the SKA20 circular case.
Figure A7. Full corner plot of the 18 parameters for the IPTA30 eccentric case.
Figure A8. Full corner plot of the 18 parameters for the SKA20 eccentric case.
Figure A9. Full corner plot of the 18 parameters for the ideal circular case.
Figure A10. Full corner plot of the 18 parameters for the ideal eccentric case.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

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