

Neural networks unveiling the properties of gravitational wave background from supermassive black hole binaries

Matteo Bonetti^{1,2,3} , Alessia Franchini^{1,2,4} , Bruno Giovanni Galuzzi⁵ , and Alberto Sesana^{1,2} 

¹ Dipartimento di Fisica “G. Occhialini”, Università degli Studi di Milano-Bicocca, Piazza della Scienza 3, 20126 Milano, Italy
e-mail: matteo.bonetti@unimib.it

² INFN, Sezione di Milano-Bicocca, Piazza della Scienza 3, 20126 Milano, Italy

³ INAF – Osservatorio Astronomico di Brera, Via Brera 20, 20121 Milano, Italy

⁴ Institut für Astrophysik, Universität Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

⁵ IBFM – Institute of Bioimaging and Molecular Physiology, Via Fratelli Cervi, 93, 20054 Segrate, Italy
e-mail: brunogiovanni.galuzzi@ibfm.cnr.it

Received 30 October 2023 / Accepted 22 April 2024

ABSTRACT

Supermassive black hole binaries (SMBHBs) are binary systems formed by black holes with masses exceeding millions of solar masses, and are expected to form and evolve in the nuclei of galaxies. The extremely compact nature of these objects leads to the intense and efficient emission of gravitational waves (GWs), which can be detected by the Pulsar Timing Array (PTA) experiment in the form of a gravitational wave background (GWB); that is, a superposition of GW signals coming from different sources. The modelling of the GWB requires some assumptions as to the binary population, and exploration of the whole parameter space involved is hindered by the great computational cost involved. We trained two neural networks (NN) on a semi-analytical modelling of the GWB generated by an eccentric population of MBHBs that interact with the stellar environment. We then used the NN to predict the characteristics of the GW signal in regions of the parameter space that we did not sample analytically. The developed framework allows us to quickly predict the amplitude, shape, and variance of the GWB signals produced in different realisations of the universe.

Key words. black hole physics – gravitation – gravitational waves – methods: analytical – methods: numerical – galaxies: kinematics and dynamics

1. Introduction

Discovering supermassive black hole binaries (SMBHBs) is key for a complete understanding of the Universe, and in particular an understanding of how galaxies form and evolve through cosmic time. Their existence is strongly favoured and predicted by all the theoretical models of galaxy mergers (e.g., [Begelman et al. 1980](#); [Valtaoja et al. 1989](#); [Milosavljević & Merritt 2003](#); [Volonteri et al. 2003](#); [Merritt & Milosavljević 2005](#); [Dotti et al. 2012](#); [Colpi 2014](#); [Klein et al. 2016](#); [Gualandris et al. 2017](#); [Kelley et al. 2017b](#); [Tremmel et al. 2018](#); [Dayal et al. 2019](#); [Bonetti et al. 2019](#); [Chen et al. 2020](#); [Barausse et al. 2020](#); [Valiante et al. 2021](#), and references therein); however, SMBHBs still remain elusive, with only a number of candidates that are yet to be confirmed. A very promising and timely way to gain insight into this elusive SMBHB population is through the detection of a gravitational wave background (GWB) signal at low (i.e., 10^{-9} – 10^{-7} Hz) frequencies using millisecond pulsars as macroscopic clocks.

As pulsars are very precise clocks, tiny deviations from the expected time of arrival of radio pulses on Earth can reveal the passage of GWs ([Verbiest et al. 2016](#)). Pulsar Timing Array (PTA) experiments are sensitive to low-frequency GWs emitted by SMBHBs with masses of above $10^8 M_{\odot}$ and whose incoherent superposition is expected to form a stochastic GWB ([Lentati et al. 2015](#)). Very recently, all the PTA collaborations across the globe (i.e., European PTA, Indian PTA, Parkes PTA, North America Nanohertz Observatory for GWs (NANOGrav)

and Chinese PTA) found evidence of a stochastic process with common amplitude and spectral slope across many monitored pulsars with a statistical significance of between 2σ and 4σ depending on the number of pulsars and on the employed analysis technique ([Antoniadis et al. 2023a,b,c,d, 2024](#); [Smarra et al. 2023](#); [Agazie et al. 2023a,b,c](#); [Afzal et al. 2023](#); [Reardon et al. 2023](#); [Xu et al. 2023](#)). This discovery clearly opens a completely new window on the Universe, allowing us to deepen our knowledge of different phenomena and probe new astrophysical and cosmological sources. Several theoretical interpretations of this GW signal have been proposed since its discovery ([Antoniadis et al. 2024](#); [Afzal et al. 2023](#)), although possibly the most plausible is a SMBHB population origin ([Antoniadis et al. 2024](#); [Agazie et al. 2023d](#)). Ideally, the signal produced by a population of inspiralling SMBHBs would manifest as a stochastic Gaussian process characterised by a power-law Fourier spectrum of delays and advances of pulse arrival times, with the characteristic inter pulsar correlations identified by [Hellings & Downs \(1983\)](#). However, the overall signal might be dominated by a handful of massive, nearby sources that might result in a sufficiently strong signal to be individually resolved as continuous GWs (CGWs, [Sesana et al. 2009](#); [Babak & Sesana 2012](#); [Kelley et al. 2018](#)). Furthermore, the signal might deviate from the isotropy, Gaussianity, and stationarity that characterise signals of primordial origin in the Universe ([Ravi et al. 2012](#)).

The GWB encodes precious information about the elusive SMBHB population. In particular, the signal amplitude and spectral shape are deeply connected to the galaxy merger

rate and the dynamical properties of the emitting SMBHBs (Kocsis & Sesana 2011; Sesana 2013; Ravi et al. 2014). While current models try to describe the GWB in terms of normalisation and spectral shape, the actual SMBHB population, being formed by discrete objects, also imprints an intrinsic variance in the GWB. To model the GWB together with its variance, Monte Carlo realisations of the entire population (millions of SMBHBs) are needed (Sesana et al. 2008; Amaro-Seoane et al. 2010; Chen et al. 2017). Unfortunately, the computational cost required to explore any wide parameter space is prohibitive, meaning that only limited sampling of GW spectra at discrete points in the multi-dimensional binary population parameter space is feasible, which leaves a large fraction of the parameter space unexplored. It is therefore crucial to find a way to ‘interpolate’ the GWB spectra across the multi-dimensional space, ultimately extending the investigation over the whole parameter space. An early attempt in this direction was made by Taylor et al. (2017), who employed a Gaussian process (GP) emulator to interpolate over the parameter space represented by two main parameters: initial binary eccentricity e_0 and stellar density at influence radius ρ . Agazie et al. (2023d) extended the work of Taylor et al. (2017), introducing more parameters in the model, but always using a GP-based regression tool. However, these latter authors neglect the effect of SMBHB eccentricity and focus only on the effects of the environment.

Here, we propose a new framework, where we use two trained neural networks (NN) to evaluate the GWB in the whole parameter space, beyond the region we use for the training of the NN. The advantage of using NN resides in its deterministic nature and its lower computational cost. NNs are also highly flexible and can represent very complex functions with deep architectures. We model the SMBHB population, taking into account an agnostic parametrisation of the mass function shape and normalisation, the presence of non-negligible eccentricity of the binaries, and considering the influence of the stellar hardening on the SMBHB evolution. We trained our NN using numerous realisations of the GWB signals in different kinds of universes in order to efficiently assess their variance.

The paper is organised as follows. In Sect. 2 we summarise the theoretical framework that we use to model the SMBHB population, highlighting the novelties in our approach. In Sect. 3, we provide specific details of the NN algorithm used. In Sect. 4, we present the results of our approach, while in Sect. 5 we discuss the implications of our algorithm, envisage its forthcoming development and usage, and finally draw our conclusions.

2. Theoretical model

Assuming that the binaries in the Universe are circular and evolve only due to gravitational wave emission, we can write the GWB characteristic amplitude as a function of the number density of mergers (Phinney 2001):

$$h_c^2(f) = \frac{4G}{\pi f^2 c^2} \int_0^\infty dz \int_0^\infty d\mathcal{M} \frac{d^2n}{dzd\mathcal{M}} \frac{1}{(1+z)} \frac{dE_{\text{GW}}(\mathcal{M})}{d \ln f_r}, \quad (1)$$

where \mathcal{M} is the chirp mass, the $1/(1+z)$ term takes into account the redshift of gravitons, $d^2n/dz d\mathcal{M}$ is the comoving number density of GW events, usually inferred through numerical simulations or semi-analytical models, and E_{GW} is the energy generated by an event. The rest-frame frequency is $f_r = (1+z)f$, where f is the observed frequency, and $f_r = 2f_{\text{orb}}$, where f_{orb} is the rest-frame orbital frequency.

The comoving number density per unit of redshift and chirp mass can be written as

$$\frac{d^2n}{dzd\mathcal{M}} = \frac{d^3N}{dzd\mathcal{M}d \ln f_r} \frac{dz}{dV_c} \frac{d \ln f_r}{dt_r} \frac{dt_r}{dz}, \quad (2)$$

where the first term on the right hand side indicates the comoving number of binaries emitting in a given logarithmic frequency interval with chirp mass and redshift in the range $[\mathcal{M}, \mathcal{M} + d\mathcal{M}]$ and $[z, z + dz]$, respectively, dV_c is the comoving volume shell between z and $z + dz$, and t_r is the time measured in the source rest frame.

From cosmology (Hogg 1999), we can write (see also Chen et al. 2017)

$$\frac{dz}{dV_c} \frac{dt_r}{dz} = \frac{1}{(1+z)4\pi c d_M^2}, \quad (3)$$

where d_M is the proper-motion distance. For circular binaries, the gravitational radiation is emitted at twice the orbital frequency, and the sky- and polarisation-averaged strain amplitude is given by (Thorne 1987)

$$h = \frac{8\pi^{2/3}}{10^{1/2}} \frac{\mathcal{M}^{5/3}}{d_M(z)} f_r^{2/3}. \quad (4)$$

The temporal evolution of the emission frequency is expressed as

$$\frac{dt_r}{df_r} = \frac{5c^5 \pi^{-8/3}}{96G^{5/3}} \mathcal{M}^{-5/3} f_r^{-11/3}, \quad (5)$$

while the radiated energy per logarithmic frequency interval is given by (Thorne 1987)

$$\frac{dE_{\text{GW}}(\mathcal{M})}{d \ln f_r} = \frac{\pi^{2/3} G^{2/3}}{3} \mathcal{M}^{5/3} f_r^{2/3} = \frac{dt_r}{d \ln f_r} \pi^2 d_M^2(z) f_r^2 h^2. \quad (6)$$

Substituting Eqs. (2) and (6) in Eq. (1), we obtain

$$h_c^2(f) = \int_0^\infty dz \int_0^\infty d\mathcal{M} \frac{d^3N}{dzd\mathcal{M}d \ln f_r} h^2(f_r), \quad (7)$$

which states that the observed characteristic-squared amplitude of the GWB is given by the integral over all the sources emitting in the frequency bin $d \ln f_r$ multiplied by the squared strain of each source (Sesana et al. 2008). The above expression can be written with a normalization that depends on the details of the SMBHB population as (Jenet et al. 2006)

$$h_c(f) = h_{1\text{yr}} \left(\frac{f}{\text{yr}^{-1}} \right)^{-2/3}, \quad (8)$$

where $h_{1\text{yr}}$ depends on the assumed model.

2.1. Binary population

We chose a simple and agnostic model for the comoving number density per unit redshift z and binary total mass M (Middleton et al. 2016):

$$\frac{d^2n}{dzd \log_{10} M} = A \left(\frac{M}{10^7 M_\odot} \right)^{-\alpha} e^{-\left(\frac{M}{M_0}\right)^\beta} (1+z)^\gamma e^{-z/z_0}. \quad (9)$$

We assume the parameters $A, \alpha, M_0, \gamma, \beta$, and z_0 (i.e., the Universe parameters)¹, which characterise the number density of the

¹ In the following, we fix $\gamma = 1$ and $z_0 = 2$. This choice is motivated by the fact that those quantities weakly affect the mass function and fixing their value allows us to reduce the number of parameters that the NN model needs to account for.

Table 1. Ranges of the parameters that characterise a chosen universe.

Quantity	Range	Units
A	10^{-7} – 10^{-2}	$\text{Mpc}^{-3} \text{Gyr}^{-1}$
α	0–1.5	–
β	0.5–2	–
M_0	10^7 – 10^9	M_\odot
ρ	1 – 10^5	$M_\odot \text{pc}^{-3}$
e_0	0–0.99	–

GW sources that vary within a given range; see Table 1. From Eq. (2), we calculate the number of binaries in each frequency bin as

$$\frac{d^3 N}{dz d \log_{10} M_1 dq df_{\text{orb}}} = \frac{d^2 n}{dz d \log_{10} M} (1+z) 4\pi c d_M^2 \frac{dt_r}{df_{\text{orb}}}, \quad (10)$$

where

$$\frac{dt_r}{df_{\text{orb}}} = \left(\frac{df_{\text{orb}}}{dt_r} \Big|_{\text{GW}} + \frac{df_{\text{orb}}}{dt_r} \Big|_{\star} \right)^{-1} \quad (11)$$

takes into account the frequency evolution due to GW emission (effective at smaller scales) and stellar hardening (dominating the large-scale evolution).

In particular, the contribution due to GW emission for a population of eccentric binaries is given by

$$\frac{df_{\text{orb}}}{dt_r} \Big|_{\text{GW}} = \frac{96 G^{5/3} (2\pi)^{8/3}}{5c^5} \frac{q M_1^{5/3}}{(1+q)^{1/3}} f_{\text{orb}}^{11/3} F(e), \quad (12)$$

where

$$F(e) = \frac{1 + 73/24e^2 + 37/96e^4}{(1 - e^2)^{7/2}}. \quad (13)$$

The frequency evolution due to the presence of the stellar hardening is given by

$$\frac{df_{\text{orb}}}{dt_r} \Big|_{\star} = \frac{3G^{4/3} M^{1/3}}{(2\pi)^{2/3}} \frac{H\rho}{v_{\text{disp}}} f_{\text{orb}}^{1/3}, \quad (14)$$

where M is the binary mass, $H = 15$ is a dimensionless constant parametrising the efficiency of energy extraction by stars scattering on a SMBHB, ρ is the stellar density, and v_{disp} is the velocity dispersion, both quantities calculated at the influence radius of the binary (Sesana & Khan 2015).

Equating Eqs. (12) and (14), we define the decoupling frequency that marks the transition between the GW- and stellar-hardening-driven regimes:

$$f_{\text{dec}} = \left(\frac{64G^{1/3} (2\pi)^{10/3}}{5c^5} \frac{v_{\text{disp}}}{\rho H} \frac{M_1^2 q}{M^{2/3}} F(e_0) \right)^{-3/10}, \quad (15)$$

where M_1 is the mass of the primary, $q = M_2/M_1$ is the mass ratio, and e_0 is the eccentricity of the binary during the stellar hardening phase, which we assume to remain constant until the binary reaches the GW-driven regime. Above the decoupling frequency, the binary circularises because of GW emission. The orbital frequency and the eccentricity are bound by the following equation:

$$\frac{f_{\text{orb}}}{f_{\text{orb},0}} = \left(\frac{1 - e_0^2}{1 - e^2} \left(\frac{e}{e_0} \right)^{12/19} \left(\frac{1 + \frac{121}{304} e^2}{1 + \frac{121}{304} e_0^2} \right)^{870/2299} \right)^{-3/2}. \quad (16)$$

Table 2. Grid for binary population sampling.

Quantity	Range	Units	Number of bins
z	0–10	–	40
$\log_{10} m_1$	7–10.5	M_\odot	35
$\log_{10} q$	–2 to 0	–	25
f_{orb}	10^{-11} – 10^{-7}	Hz	360
e	0–0.999	–	100

Notes. Bins are uniformly spaced in their respective ranges, except for the redshift (log-uniform distributed from 0.005 to 2, while uniformly distributed from 2 to 10) and the orbital frequency (90 uniform bins covering each dex).

2.2. Discrete GWB signal construction

We have so far assumed that the binary population is described by a continuous differential distribution. However, in reality, the background is a superposition of discrete contributions from binaries drawn from that continuous distribution. This means that, in nature, the actual signal fluctuates depending on the specific draw realized in nature, and it is this intrinsic variance that we want to properly capture with our approach. To this end, for each value of the universe parameters ($A, \alpha, M, \beta, \rho, e_0$), we perform 100 realisations of the binary population, sampling the distribution in Eq. (10) in a Monte Carlo fashion.

We characterised the binary population as a function of redshift, primary mass, mass-ratio, orbital frequency and eccentricity. Specifically, we translated Eq. (10) into a numerical distribution function of $(z, M_1, q, f_{\text{orb}}, e)$ with finite size bins, as detailed in Table 2. We then sampled the discrete number of sources, drawing an integer number from a Poisson distribution with mean equal to the non-integer number of binaries in that bin predicted by Eq. (10).

In each multi-dimensional bin, we draw a random number between the lower and upper limit of $(M_1, q, z, f_{\text{orb}}, e)$ in order to assign the properties to a binary source. We repeat this procedure N times per bin, where N is the discrete number of binaries in that bin. In the bins that have $N > 50$, we only sample 50 binaries and we then multiply the resulting background by $N/50$.

For each population realisation, the GWB can be computed by summing the GW strain produced by each binary in our population. Circular binaries emit a GW signal at $2f_{\text{orb}}$, while eccentric ones also emit at multiple harmonics, $f_n = n f_{\text{orb}}$, where n is the harmonic number. The GW strain of each harmonic is given by (see e.g., Amaro-Seoane et al. 2010)

$$h_n(f_n) = 2 \sqrt{\frac{32}{5}} \frac{G^{5/3}}{c^4} \frac{M^{5/3}}{nd_M} \left(2\pi \frac{f_n}{n} \right)^{2/3} \sqrt{g(n, e)}, \quad (17)$$

where the dimensionless function $g(n, e)$ determines the fraction of the GW power that is emitted in each harmonic and reads

$$g(n, e) = \frac{n^4}{32} \left[\left(J_{n-2}(ne) - 2eJ_{n-1}(ne) + \frac{2}{n} J_n(ne) + 2eJ_{n+1}(ne) - J_{n+2}(ne) \right)^2 + (1 - e^2) \left(J_{n-2}(ne) - 2J_n(ne) + J_{n+2}(ne) \right)^2 + \frac{4}{3n^2} J_n^2(ne) \right], \quad (18)$$

with J_n representing the n th-order Bessel function of the first kind.

The relevant frequency band for pulsar timing observations is from $\Delta f = 1/T_{\text{obs}}$ – where $T_{\text{obs}} = 30$ yr is the assumed total observation time – to the Nyquist frequency $1/(2\Delta t)$, where Δt is the time between subsequent observations (around a couple of weeks). We assume the observed frequency to vary between $1/T_{\text{obs}}$ and 6×10^{-8} Hz, uniformly spaced by Δf , yielding $N_f = 56$ values. The GWB in the frequency bin $[f_j, f_{j+1}]$ is therefore given by the sum of the GW strain multiplied by the number of cycles that each binary makes in the observation time:

$$h_c^2(f_j) = \sum_{i=1}^N \sum_{n=1}^{\bar{n}} h_{n,i}^2(f_{n,i}) \frac{f_{n,i}}{\Delta f(1+z)} \Theta\left(\frac{f_{n,i}}{1+z}\right), \quad (19)$$

where the $1+z$ factor converts from rest frame to observing frequency, $f_{n,i} = n f_{\text{orb},i}$ is the orbital frequency of the i th binary, and $\Theta(f_{n,i}/(1+z)) = 1$ if $f_j \leq f_{n,i}/(1+z) \leq f_{j+1}$ and zero otherwise. We note that the i index runs over the sample of binaries, while the n index runs, for each binary, over the harmonics, with \bar{n} chosen such that a sufficient number of them is included in the computation of the signal. We take $\bar{n} = 8n_{\text{max}}$, where n_{max} is a numerical proxy of the harmonic number at which the maximum GW power is emitted and can be accurately approximated as (Hamers 2021)²

$$n_{\text{max}}(e) \approx 2 \left(1 + \sum_{k=1}^4 c_k e^k \right) (1 - e^2)^{-3/2}, \quad (20)$$

where $c_1 = -1.01678$, $c_2 = 5.57372$, $c_3 = -4.9271$, and $c_4 = 1.68506$.

Finally, we note that, as we generated 100 realisations of the binary population, we have 100 GWB signals for each choice of our universe parameters.

3. The neural network model

A large database of GWB realisations is required for each combination of the universe parameters ($A, \alpha, M, \beta, \rho, e_0$) in order to train a machine learning (ML) model for regression. We therefore generated approximately $N_{\text{tot}} \sim 5 \times 10^5$ GWB signals, corresponding to $N_{\text{model}} = 5120$ different universes and 100 realisations of the binary population within each universe. The parameter values were selected using Sobol sequences within the ranges specified in Table 1. The final dataset is therefore composed of $N_{\text{tot}} = N_{\text{model}} \times 100$ realisations of the GWB signal. For each combination of the universe parameters, we computed the mean μ and standard deviation σ for each of the N_{tot} distributions as a function of frequency. We then created an input–output dataset, where the input consists of the six universe parameters and the frequency, and the outputs are the μ and the σ values. To assess the performance and generalisation ability of our ML model, we randomly partitioned the dataset into a training set (80%) and a test set (20%). The training set was used to estimate the ML model parameters, while the test set was reserved for evaluating the final model performance. Additionally, because of variations in the scale of input parameters (also referred to as features), we standardised their values by centring each feature at a mean of 0 with a standard deviation of 1. This preprocessing step facilitates the learning process for ML models such as the NN we employ in the present work, as their parameters are estimated through a gradient-based optimisation algorithm. Ultimately, we applied the Box-Cox power transformation to the output values, aiming

² For $n > 1000$ we group the harmonics in batches of 20 to speed up the evaluation.

to obtain a normal distribution with a mean of 0 and a standard deviation of 1. This crucial preprocessing step mitigates the risk of training ML models with excessively low loss values, which could lead to significant approximation errors.

The two output parameters μ and σ exhibit significant differences, that is, of orders of magnitude. To address this variation and ensure effective training, we implemented two distinct ML models: one for predicting μ and another for predicting σ . Specifically, we employed two feed-forward NNs (Goodfellow et al. 2016). Each of these models defines a mapping $y = g(x, \theta)$, where x represents the input (i.e., the universe parameters and the frequency) and y represents the output (i.e., μ and σ , respectively). Here, θ denotes the set of parameters defining the ML model, which are estimated using the training set.

This model is referred to as feed-forward because information flows through the function from the input x to the output y without any feedback connections where the model outputs influence itself. The NN model is defined by simultaneously composing multiple functions g^i , with $i = 1, \dots, N$, arranged in a chain structure, resulting in $g(x, \theta) = g^N(g^{N-1}(\dots g^1(x, \theta_1) \dots), \theta_{N-1})$. The length of the chain, denoted N (where $N \geq 2$), is referred to as the network depth. The first function, g^1 , is known as the input layer, while the final function is the output layer, responsible for producing a value close to the target y . The intermediate layers are referred to as hidden layers, as their outputs are not explicitly observed in the training data. Each layer of the network typically represents a vector value. In our case, the input layer contains seven neurons, corresponding to the seven features. Each hidden layer i consists of n_{hidden} neurons and is a function of the preceding layer $i - 1$. Typically, this connection involves an affine transformation followed by a fixed non-linear transformation, known as the activation function. This process is expressed as follows:

$$\mathbf{h}^i = a^i(\mathbf{W}^i \mathbf{h}^{i-1} + \mathbf{b}^i). \quad (21)$$

Here, the matrix \mathbf{W}^i and \mathbf{b}^i are the parameters of the affine transformation, commonly referred to as weights and bias, respectively, and a^i is the activation function. It is worth noting that $\mathbf{h}^1(x) = x$, and $\mathbf{h}^N(x) = y$.

For the output layer of the two networks, we employed a linear activation function $a(x) = x$. Keeping the neural architecture fixed, which includes the depth N of the network, the width n_{hidden} , and the activation function a for various hidden layers, the network refines its parameters (i.e., the weights W^i and biases b^i for each layer) by minimising a loss function during the training phase. In our scenario, we chose a weighted version of the mean absolute error (MAE) between the true and predicted outputs as the loss function:

$$\sum_{k=1}^{N_{\text{tot}}} w_k(f) |y_{\text{test}}^k - y_{\text{pred}}^k|. \quad (22)$$

Here, y_{test}^k and y_{pred}^k are the true and predicted outputs, and $w_k(f)$ is a factor that weights the absolute error linearly as a function of the different frequencies f . This weighting leads the NN to focus more on better fitting the lower frequencies. The reason for this choice is twofold: firstly, the low-frequency spectrum is the part that is currently within the reach of PTA experiments, and secondly, the low-frequency end of the spectrum is less noisy and therefore easier to model, because it is less sensitive to fluctuations generated by ‘loud’ single sources.

We used the Adam optimiser (Kingma & Ba 2014) with a positive learning rate l_r to estimate the parameters of the two

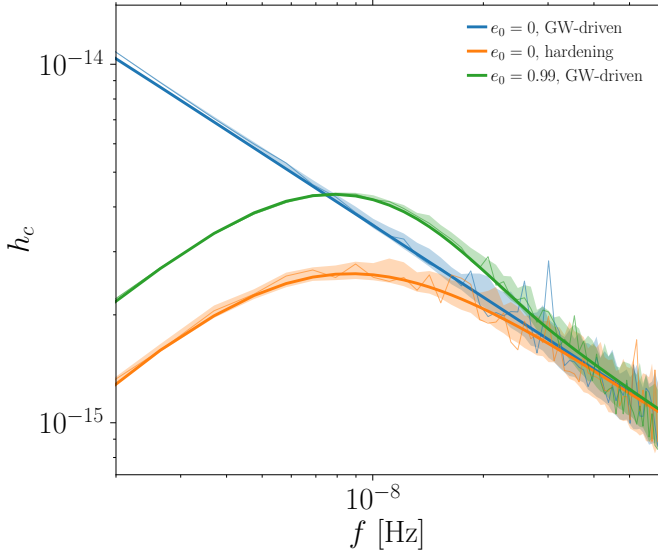


Fig. 1. Characteristic strain as a function of frequency for three different universe models ($\log_{10} A = -2$, $\alpha = 0.3$, $\beta = 1$, $\log_{10} M_0 = 8$) featuring: circular GW-driven binaries (blue), circular but stellar+GW-driven binaries ($\rho = 10^4 M_\odot \text{pc}^{-3}$, orange), and eccentric GW-driven binaries ($e_0 = 0.99$ at 5×10^{-12} Hz, green). Thick lines denote the theoretical GWB, while shaded areas denote the 10th and 90th percentiles of the characteristic strain when different discrete realisations of the population are considered. The thin line of the same colour denotes one such realisation for each model.

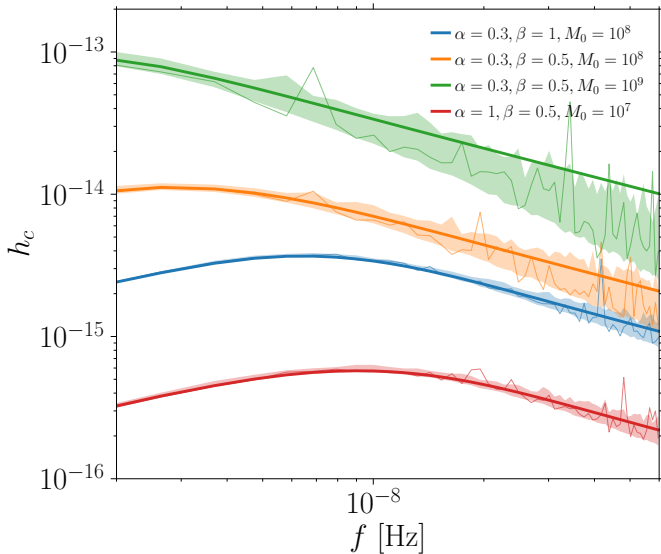


Fig. 2. Same as Fig. 1, but considering how the characteristic strain of the GWB varies by changing the parameters controlling the shape of the BH mass function (see labels). Higher cut-off masses (M_0), shallower power law ($\alpha < 1$), and a shallower exponential cutoff ($\beta < 1$) produce a more prominent GWB, which is also characterised by a larger variance.

networks. More specifically, we trained the NN for 100 epochs using a batch size of 128 and an early stopping criterion to prevent potential over-fitting, halting the training phase when the MAE on a validation set (20% of the training set) ceased to improve.

The hyper-parameters $a(x)$, n_{hidden} , and n_{neurons} of the neural architecture, and the learning rate l_r of the optimiser, were determined using a hyper-parameter strategy based on Bayesian

optimisation (BO). BO is a sequential model-based approach used to solve complex global optimisation problems based on expensive-to-evaluate black-box functions, but is also commonly used to tune the hyper-parameters of ML algorithms (Snoek et al. 2012; Galuzzi et al. 2020; Victoria & Maragatham 2021). At the end of the hyper-parametrisation optimisation, we obtained $a(x) = \tanh(x)$, $n_{\text{hidden}} = 2$, $n_{\text{neurons}} = 70$, and $l_r = 0.001$ for the NN associated with μ , and $a(x) = \text{Relu}(x) = \max(x, 0)$, $n_{\text{hidden}} = 2$, $n_{\text{neurons}} = 50$, and $l_r = 0.001$ for the NN associated with σ .

The training of the ML model and the hyper-parameter optimisation are implemented using the Keras library (Chollet 2015; O’Malley et al. 2019).

4. Results

Here, we first present the results we obtained from the MC generation of the GWB signals dataset, discussing its properties, and then we show the results of the training of the NN on our sample.

4.1. Dataset properties

As our dataset is relatively large, it is useful to summarise its main properties and comment on the main features that can be captured by varying different parameters. The parameter A acts as a normalisation, shifting the amplitude of the GWB up and down without affecting its shape. Nevertheless, by controlling the number of binaries in a universe, this parameter could in principle be linked to relevant astrophysical parameters. The effective modelling employed in our work can indeed be recast in terms of observable quantities, as done in Chen et al. (2017).

A more substantial role in shaping the GWB is instead played by the parameters e_0 and ρ . Both parameters affect the GWB spectral shape and are specifically responsible for low-frequency turnovers. However, disentangling the effects of stellar hardening and the presence of eccentric binaries is not straightforward, as the two cause a similar behaviour, that is, a turnover in the GWB amplitude at low frequencies. The attenuation of the GWB signal is due to the fact that both the presence of a stellar (or gaseous) component and a non-zero initial eccentricity cause the binary to spend less time emitting GWs in a given frequency range. In particular, the interaction of SMBHBs with their environment is important at frequencies $f \ll 1 \text{ yr}^{-1}$ where GW emission is still efficient (Kocsis & Sesana 2011). Figure 1 shows the signals generated from a population of circular GW-driven binaries (blue line), circular binaries that are affected by stellar hardening (orange line), and a population of extremely eccentric solely GW-driven binaries (green line). Although both stellar hardening and eccentricity produce a low-frequency turnover, there is a clear difference between the orange and green populations. While the signal affected by stellar hardening is simply lower than the circular GW-driven one, the GWB generated from eccentric binaries shows both a turnover at low frequencies and a bump, which for the specific model represented is above 10^{-8} Hz (Sesana 2013; Chen et al. 2017; Kelley et al. 2017a). We note here that the turnover at low frequency is significant for very high eccentricities only. Decreasing the eccentricity to $e_0 = 0.9$ leads to a much milder effect at low frequencies. The latter is due to the fact that eccentric binaries emit the GW power at multiple harmonics of the orbital frequency, effectively shifting the power from lower to higher frequencies instead of simply depleting the spectrum at lower frequencies by evolving faster, as is the case for stellar hardening. This behaviour of the eccentricity also implies that GWB generated by eccentric SMBHBs might

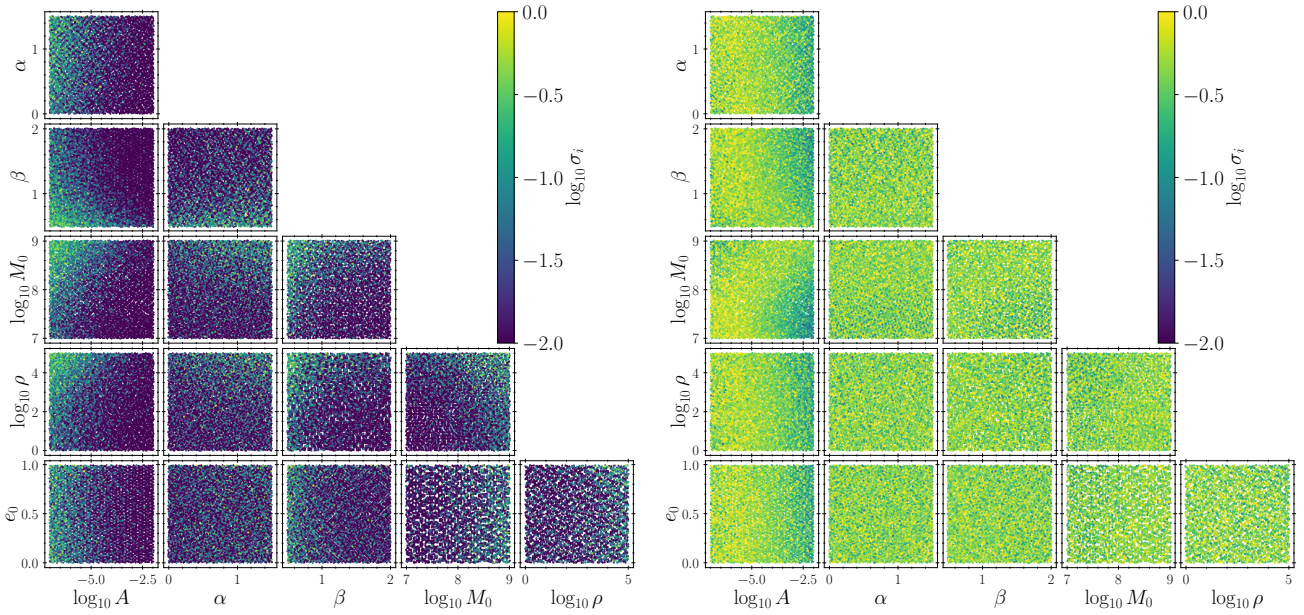


Fig. 3. Corner plot showing the dependence of GWB variance (colour scale) on the model parameters, considering the lowest (left) and highest (right) frequency bin. As can be inferred from the colour scale (more evident in the left panel), lower values of α and β combined with a high cutoff mass M_0 are linked to a larger scatter, which is caused by the increased probability of having more massive binaries that can produce spikes in the signal. Finally, the figure also shows that the sampling of the parameter space is sufficient to cover it uniformly.

show a certain degree of correlation among frequency bins, as the same binary can distribute the GW power in several of them.

The most important information for the GWB interpolation across the whole parameter space is how the variance changes as a function of the parameters that characterize the universe ($A, \alpha, \beta, M_0, \rho, e_0$). In Fig. 1, the shaded areas denote the intrinsic span of the signal generated by different realisations of the SMBHB population. The higher the frequency, the fewer the sources per bin and therefore the stronger the influence of the granularity of the population on the GWB spectrum, effectively increasing the variance of the signal, which possibly features strong spikes due to rare but loud binaries. The various effects of SMBHB population granularity are more evident from Fig. 2, which shows four different signals generated by changing the universe parameters that describe the SMBHB mass function, namely α, β , and M_0 , which are responsible for the GWB variance.

A SMBHB population characterised by a prominent high-mass tail (see green curve of Fig. 2) can drastically affect the GWB spectrum both in terms of normalisation and variance. We can see that if the power-law decay of the mass distribution is shallower (i.e., $\alpha < 1$) and correspondingly so is the exponential cutoff at high masses ($\beta < 1$ values), the variance between different realisations is larger, resulting in a very high scattering in the signal at different frequencies. The comparison between the green and orange curve shows the effect of increasing the mass cutoff value; an order-of-magnitude increase results in a significantly reduced variance and an order-of-magnitude-weaker signal. If the SMBHB mass function is characterised instead by a steeper exponential decay (blue line, $\beta = 1$) but at the same cutoff value (i.e., $10^8 M_\odot$), the variance is further reduced, the normalisation is lower and there is a much more prominent turnover at lower frequencies compared to the shallower exponential decay. Finally, a steeper power-law decay distribution coupled with a smaller mass cutoff value and a shallow exponential decay causes the variance in the signal to drop to a minimum, affecting only the GWB at high frequencies, as shown

by the red line. This is consistent with the fact that it is less likely to have very massive SMBHBs strongly influencing each GWB signal realisation. However, in this latter case, the amplitude of the signal might be too low to be detectable in the PTA band.

It is also interesting to observe how the GWB variance correlates among the model parameters. However, as all generated GWB have relatively different amplitudes, we cannot simply compare the different variances (at various frequencies) associated with a specific set of universe parameters ($A, \alpha, \beta, M_0, \rho, e_0$). However, as ρ and e_0 are the same for all simulations in Fig. 2, we can see from the comparison between the blue and red lines that the bending occurs at higher frequencies if the signal is strongly dominated by lighter systems (red lines). In order to get rid of the GWB normalisation and to focus on the spread of the signal, at each frequency, we divide the GWB signal of each realisation by the mean value (computed over the 100 realisations for a specific set of universe parameters). This allows us to compare the intrinsic spreads of GWBs with different strength levels. This information is shown in the corner plots in the left and right panel of Fig. 3, where the colour scale shows the ‘intrinsic’ variance of the GWB signal in the lowest and highest frequency bins, respectively. Darker (lighter) regions correspond to lower (higher) variance. In general, comparing the two panels of Fig. 3, it is clear that the variance is higher at higher frequencies (as expected). Moreover, we can also identify regions of the subpanels showing correlations. Specifically, as already noted above, universes characterised by shallower power laws (small α), shallower exponential cutoff (small β), and high-mass cutoff (large M_0) show a clearly larger intrinsic variance. Another interesting pattern is shown by the correlation of A and ρ , where universes with small A and large ρ seem to show a relatively large variance. This is because that combination of parameters quite drastically reduces the number of binaries per frequency bin, increasing therefore the granularity and making that frequency bin more susceptible to the influence of rarer but more massive systems, possibly producing spikes in the GWB.

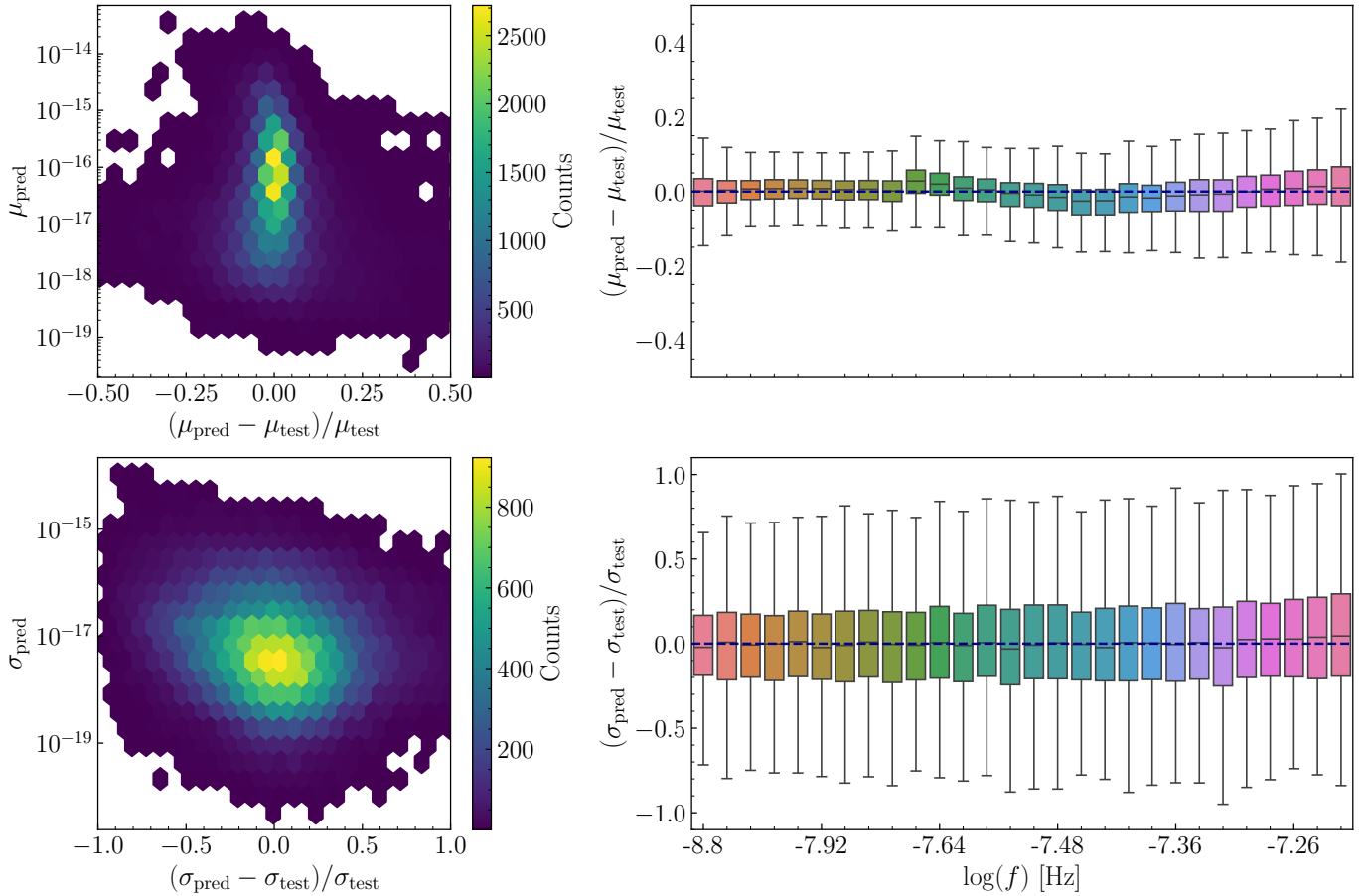


Fig. 4. Performance of our algorithm in the explored parameter space. Left panels: density contours of the relative error (i.e., the difference between the predicted value and the true value, which is the value of the test set) divided by the true value, as a function of the predicted quantity, either mean μ_{pred} (top) or standard deviation σ_{pred} (bottom). The colour scale denotes the number counts within each bin where the error was computed. Right panels: relative error on mean (top) and standard deviation (bottom) as a function of frequency. The box plot highlights the median error enclosed by the 25th and 75th percentiles of the error distribution at each frequency, with the error bars enclosing the full spread of values. The mean and standard deviation are computed in logscale. The blue dashed line marks the zero relative error.

4.2. Model prediction

We now focus on the performance of the NN model. The panels of Fig. 4 show the overall accuracy in reproducing the test set mean μ and standard deviation σ for each Universe. The left panels show the density contours of the relative error of each quantity as a function of the predicted quantity itself. The upper left panel represents the mean μ , while the lower left panel represents the standard deviation σ . We can clearly see that the distribution of the relative error is centred at zero for both the mean and standard deviation.

We achieve a remarkably low relative error for the μ value. Specifically, only 2.1% of our sample exhibits an error exceeding 25%. Conversely, the bottom left panel of Fig. 4 shows that there are data points with relatively large errors for the σ value, even at the lower end of the frequency range. However, only 12.6% of our sample has a relative error of larger than 50%.

The dependence of the model performance on frequency is better displayed in the right panels of Fig. 4, where the relative errors on the mean (upper right panel) and standard deviation (lower right panel) are plotted against frequency. At each frequency, the box plot shows the value of the median error limited by the 25th and 75th percentile (colour box), with error bars showing the full spread of these values. This provides important information about the distribution of the outliers, as the

Table 3. The MAE, mean absolute percentage error (MAPE), square correlation coefficient (R2), and Spearman correlation (SC) computed on the test set.

	MAE	MAPE	R2	SC
μ_{pred}	$9.86e-18$	0.013	0.9914	0.9992
σ_{pred}	$1.183e-17$	0.0373	0.7004	0.9772

box size is significantly larger at higher frequencies, where single loud sources dominate the GWB signal. We further note that the errors tend to be smaller at lower frequencies for the μ value, where we instructed the NN to achieve the best possible performance.

In Table 3, we report some standard accuracy indicators: the MAE, the mean absolute percentage error (MAPE), the square correlation coefficient (R2), and the Spearman correlation (SC), all of which are computed on the test set. All the indicators show the satisfactory performance of our NN model. Specifically, the small values of MAE and MAPE imply that the NN model predictions are in agreement with the true values of the test set, while the values of R2 and SC very close to unity for both the mean and standard deviation indicate that the performance of the model is overall very good.

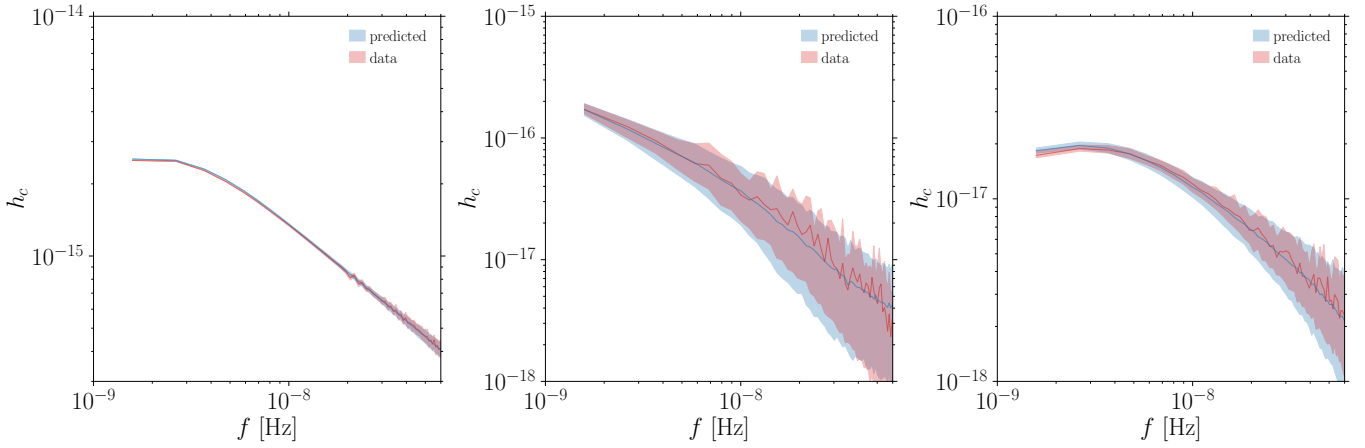


Fig. 5. GWB prediction of the NN model compared to three models of the test set of the database. Shaded areas denote the 10th and 90th percentiles of the characteristic strain, while the thin lines with the same colour represent one GWB realisation for each model. Left panel: universe parameters are $\log_{10} A = 2.359$, $\alpha = 0.343$, $\beta = 1.926$, $\log_{10} M_0 = 8.041$, $\log_{10} \rho = 2.267$, and $e_0 = 0.056$. Middle panel: universe parameters are $\log_{10} A = -6.599$, $\alpha = 0.197$, $\beta = 1.109$, $\log_{10} M_0 = 8.806$, $\log_{10} \rho = 0.562$, and $e_0 = 0.079$. Right panel: universe parameters are $\log_{10} A = -5.51$, $\alpha = 1.293$, $\beta = 0.578$, $\log_{10} M_0 = 7.533$, $\log_{10} \rho = 0.73$, and $e_0 = 0.957$.

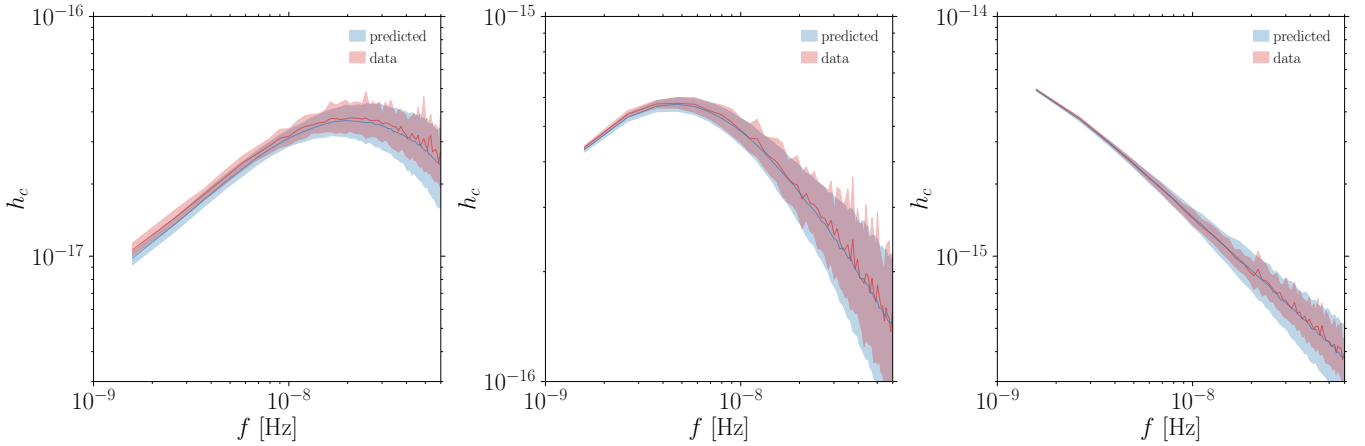


Fig. 6. Same as Fig. 5, but for different universe parameters. Left panel: universe parameters are $\log_{10} A = -4.674$, $\alpha = 0.655$, $\beta = 1.159$, $\log_{10} M_0 = 8.16$, $\log_{10} \rho = 4.312$, and $e_0 = 0.939$. Middle panel: universe parameters are $\log_{10} A = -3.743$, $\alpha = 0.549$, $\beta = 1.726$, $\log_{10} M_0 = 8.714$, $\log_{10} \rho = 4.389$, and $e_0 = 0.041$. Right panel: universe parameters are $\log_{10} A = -2.135$, $\alpha = 1.273$, $\beta = 0.842$, $\log_{10} M_0 = 8.905$, $\log_{10} \rho = 0.069$, and $e_0 = 0.918$.

Finally, we represent the predicted GWB for a sample of different Universes in Figs. 5 and 6. We show the comparison between the GWB constructed following the procedure outlined in Sect. 2.2 (red) and the signal predicted by our NN framework (blue) for six different choices of the Universe parameters. The shaded areas denote the 10th and 90th percentiles of the h_c distribution at each frequency, while the thin lines represent one realisation of the signal. The NN prediction is in very good agreement with the simulated data, that is, in terms of the mean, variance, and shape of the GWB for all the examples shown in Figs. 5 and 6.

Therefore, we conclude that, despite the possible dip in performance in a narrow region of the parameter space, the NN framework presented here accurately predicts the shape and amplitude of the diverse GWB signals.

5. Discussion and conclusions

In this work, we built a model composed of two NNs that efficiently interpolate the stochastic GWB emitted by SMBHBs across the wide parameter space that describes their popula-

tion. We concentrate in particular on the low-frequency part of the spectrum, as this is currently being surveyed by PTA experiments.

We generated a large dataset of GWB by considering an agnostic modelisation of the underlying SMBHB population. GWBs are generated from the discrete population of SMBHBs, significantly improving the simple power-law description of the GWB signal. We explored $N_{\text{model}} = 5120$ different universes (i.e., different parameter configurations) – computing 100 realisations of the GWB for each of them – in order to uniformly cover the possible deviations from the power-law prediction due to the eccentricity and hardening of the SMBHB population. This approach allowed us to trace not only the shape and amplitude but also the variance of the GWB signal, which is ultimately influenced by the discrete nature of the SMBHB population; the GWB signal itself can be mainly composed of strong signals coming from very massive and/or nearby binaries. We note here that we computed the sky-polarisation-averaged strain, essentially neglecting the effect of the inclination angle of binaries with respect to the line of sight. Taking into account the inclination could potentially introduce more variance, as face-on

binaries will produce a stronger signal compared to edge-on systems. However, we find this effect to be negligible for the background represented by both the green and blue line in Fig. 2.

We used the generated dataset to train the two NNs to efficiently explore the parameter space, therefore overcoming the bottleneck represented by the expensive MC sampling of realistic SMBHB populations for GWB generation. We find that our model performs very well, and is able to reproduce the shape, amplitude, and variance of the GWB signals of the test set. We show that our model performs better at lower frequencies, where by design it has been instructed to achieve a better accuracy, as current data are now available in that range and the signal is easier to model owing to the lack of a strong single-source contribution. Our trained NN model allows us to efficiently explore the parameter space of our agnostic modelisation. We plan to include the NN model in an end-to-end Bayesian inference pipeline that will be used to produce informed posterior distributions on our model parameters. This will allow us to make astrophysical inferences from the signal detected by PTA collaborations, possibly providing constraints on the elusive SMBHB populations. The *ab initio* inclusion of the GWB variance in our modelisation is crucial for the inference of the SMBHB parameters. As we have access to only one Universe (ours), a reliable assessment of the cosmic variance of our SMBHB population is of capital importance in order to correctly interpret the nature of the current and future PTA detections. We defer the application of our model to actual PTA data to a forthcoming publication.

Acknowledgements. We thank the anonymous referee for a constructive report that significantly improved the quality of the manuscript. AF and AS acknowledge financial support provided under the European Union’s H2020 ERC Consolidator Grant “Binary Massive Black Hole Astrophysics” (B Massive, Grant Agreement: 818691). MB acknowledges support provided by MUR under grant “PNRR – Missione 4 Istruzione e Ricerca – Componente 2 Dalla Ricerca all’Impresa – Investimento 1.2 Finanziamento di progetti presentati da giovani ricercatori ID:SOE_0163” and by University of Milano-Bicocca under grant “2022-NAZ-0482/B”.

References

- Afzal, A., Agazie, G., Anumarlapudi, A., et al. 2023, *ApJ*, 951, L11
- Agazie, G., Alam, M. F., Anumarlapudi, A., et al. 2023a, *ApJ*, 951, L9
- Agazie, G., Anumarlapudi, A., Archibald, A. M., et al. 2023b, *ApJ*, 951, L8
- Agazie, G., Anumarlapudi, A., Archibald, A. M., et al. 2023c, *ApJ*, 951, L10
- Agazie, G., Anumarlapudi, A., Archibald, A. M., et al. 2023d, *ApJ*, 952, L37
- Amaro-Seoane, P., Sesana, A., Hoffman, L., et al. 2010, *MNRAS*, 402, 2308
- Antoniadis, J., Arumugam, P., Arumugam, S., et al. 2023a, *A&A*, 678, A50
- Antoniadis, J., Arumugam, P., Arumugam, S., et al. 2023b, *A&A*, 678, A49
- Antoniadis, J., Arumugam, P., Arumugam, S., et al. 2023c, ArXiv e-prints [arXiv:2306.16226]
- Antoniadis, J., Babak, S., Bak Nielsen, A. S., et al. 2023d, *A&A*, 678, A48
- Antoniadis, J., Arumugam, P., Arumugam, S., et al. 2024, *A&A*, 585, A94
- Babak, S., & Sesana, A. 2012, *Phys. Rev D*, 85, 044034
- Barausse, E., Dvorkin, I., Tremmel, M., Volonteri, M., & Bonetti, M. 2020, *ApJ*, 904, 16
- Begelman, M. C., Blandford, R. D., & Rees, M. J. 1980, *Nature*, 287, 307
- Bonetti, M., Sesana, A., Haardt, F., Barausse, E., & Colpi, M. 2019, *MNRAS*, 486, 4044
- Chen, S., Sesana, A., & Del Pozzo, W. 2017, *MNRAS*, 470, 1738
- Chen, Y., Yu, Q., & Lu, Y. 2020, *ApJ*, 897, 86
- Chollet, F. 2015, *Keras*, <https://keras.io>
- Colpi, M. 2014, *Space Sci. Rev.*, 183, 189
- Dayal, P., Rossi, E. M., Shiralilou, B., et al. 2019, *MNRAS*, 486, 2336
- Dotti, M., Sesana, A., & Decarli, R. 2012, *Adv. Astron.*, 2012, 940568
- Galuzzi, B. G., Giordani, I., Candelieri, A., Perego, R., & Archetti, F. 2020, *Computat. Manage. Sci.*, 17, 495
- Goodfellow, I., Bengio, Y., & Courville, A. 2016, *Deep Learning* (MIT Press)
- Gualandris, A., Read, J. I., Dehnen, W., & Bortolas, E. 2017, *MNRAS*, 464, 2301
- Hamers, A. S. 2021, *Res. Notes. Am. Astron. Soc.*, 5, 275
- Hellings, R. W., & Downs, G. S. 1983, *ApJ*, 265, L39
- Hogg, D. W. 1999, ArXiv e-prints [arXiv:astro-ph/9905116]
- Jenet, F. A., Hobbs, G. B., van Straten, W., et al. 2006, *ApJ*, 653, 1571
- 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
- Kelley, L. Z., Blecha, L., Hernquist, L., Sesana, A., & Taylor, S. R. 2018, *MNRAS*, 477, 964
- Kingma, D. P., & Ba, J. 2014, ArXiv e-prints [arXiv:1412.6980]
- Klein, A., Barausse, E., Sesana, A., et al. 2016, *Phys. Rev. D*, 93, 024003
- Kocsis, B., & Sesana, A. 2011, *MNRAS*, 411, 1467
- Lentati, L., Taylor, S. R., Mingarelli, C. M. F., et al. 2015, *MNRAS*, 453, 2576
- Merritt, D., & Milosavljević, M. 2005, *Liv. Rev. Relat.*, 8, 8
- Middleton, H., Del Pozzo, W., Farr, W. M., Sesana, A., & Vecchio, A. 2016, *MNRAS*, 455, L72
- Milosavljević, M., & Merritt, D. 2003, *ApJ*, 596, 860
- O’Malley, T., Bursztein, E., Long, J., et al. 2019, *KerasTuner*, <https://github.com/keras-team/keras-tuner>
- Phinney, E. S. 2001, ArXiv e-prints [arXiv:astro-ph/0108028]
- Ravi, V., Wyithe, J. S. B., Hobbs, G., et al. 2012, *ApJ*, 761, 84
- Ravi, V., Wyithe, J. S. B., Shannon, R. M., Hobbs, G., & Manchester, R. N. 2014, *MNRAS*, 442, 56
- Reardon, D. J., Zic, A., Shannon, R. M., et al. 2023, *ApJ*, 951, L6
- Sesana, A. 2013, *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
- Smarra, C., Goncharov, B., Barausse, E., et al. 2023, *Phys. Rev. Lett.*, 131, 171001
- Snoek, J., Larochelle, H., & Adams, R. P. 2012, *Advances in Neural Information Processing Systems*, 25
- Taylor, S. R., Simon, J., & Sampson, L. 2017, *Phys. Rev. Lett.*, 118, 181102
- Thorne, K. S. 1987, *Gravitational Radiation* (Cambridge: Cambridge University Press), 330
- Tremmel, M., Governato, F., Volonteri, M., Quinn, T. R., & Pontzen, A. 2018, *MNRAS*, 475, 4967
- Valiante, R., Colpi, M., Schneider, R., et al. 2021, *MNRAS*, 500, 4095
- Valtaoja, L., Valtonen, M. J., & Byrd, G. G. 1989, *ApJ*, 343, 47
- Verbiest, J. P. W., Lentati, L., Hobbs, G., et al. 2016, *MNRAS*, 458, 1267
- Victoria, A. H., & Maragatham, G. 2021, *Evol. Syst.*, 12, 217
- Volonteri, M., Haardt, F., & Madau, P. 2003, *ApJ*, 582, 559
- Xu, H., Chen, S., Guo, Y., et al. 2023, *RAA*, 23, 075024