

Modeling non-Gaussianities in pulsar timing array data analysis using Gaussian mixture models

Mikel Falxa^{*} and Alberto Sesana

*Dipartimento di Fisica “G. Occhialini,” Università degli Studi di Milano-Bicocca,
Piazza della Scienza 3, I-20126 Milano, Italy and INFN,
Sezione di Milano-Bicocca, Piazza della Scienza 3, I-20126 Milano, Italy*



(Received 13 August 2025; accepted 28 January 2026; published 23 February 2026)

In pulsar timing array (PTA) data analysis, noise is typically assumed to be Gaussian, and the marginalized likelihood has a well-established analytical form derived within the framework of Gaussian processes. However, this Gaussianity assumption may break down for certain classes of astrophysical and cosmological signals, particularly for a gravitational wave background (GWB) generated by a population of supermassive black hole binaries (SMBHBs). In this work, we present a new method for testing the presence of non-Gaussian features in PTA data. We go beyond the Gaussian assumption by modeling the noise or signal statistics using a Gaussian mixture model (GMM). An advantage of this approach is that the marginalization of the likelihood remains fully analytical, expressed as a linear combination of Gaussian PTA likelihoods. This makes the method straightforward to implement within existing data analysis tools. Moreover, this method extends beyond the free spectrum analysis by producing posterior probability distributions of higher-order moments inferred from the data, which can be incorporated into spectral refitting techniques. We validate the model using simulations and demonstrate the sensitivity of PTAs to non-Gaussianity by computing the Bayes factor in favor of the GMM as a function of the injected excess moments. We apply the method to a more astrophysically motivated scenario where a single SMBHB is resolved on top of a Gaussian GWB and show that significant non-Gaussianities are introduced by the individual source. Finally, we test our model on a realistic GWB generated from a simulated population of SMBHBs.

DOI: [10.1103/PhysRevD.113.043047](https://doi.org/10.1103/PhysRevD.113.043047)

I. INTRODUCTION

Pulsar timing arrays (PTAs) provide a unique observational tool to explore gravitational waves (GWs) at nanohertz frequencies, far below those accessible to ground or space-based interferometers [1–3]. In this low-frequency band, the most promising sources of GWs are supermassive black hole binaries (SMBHBs), which are expected to form naturally as a consequence of galaxy mergers over cosmic time [4,5]. When two massive galaxies collide, their central black holes are brought together, eventually forming bound binary systems that lose energy through GW emission and spiral inward until merger. The combined GW emissions of an astrophysical population of such binaries result in a stochastic gravitational wave background (GWB). The spectral shape and amplitude of this background are

directly connected to the merger history of massive galaxies and the dynamical environment of their central black holes [5]. Additional sources of nanohertz frequency GWs may include cosmological phenomena such as first-order phase transitions, cosmic strings, or other relics from the early Universe [6].

The GWB produced by SMBHBs is expected to manifest as a low-frequency red noise process in pulsar timing data, spatially correlated between millisecond pulsars according to the Hellings-Downs (HD) curve when sources are isotropically distributed in the sky [7,8]. Several PTA collaborations have recently reported strong evidence for a common-spectrum process consistent with the properties of a GWB with the expected HD spatial correlations [9–13]. However, standard PTA analysis techniques assume that the signal and noise are both Gaussian stochastic processes [14]. This assumption is well motivated when the background is generated by a large number of unresolved sources, but may break down when only a few bright binaries contribute significantly, or when deterministic sources such as continuous gravitational waves (CGWs) are superimposed on the background [15]. In these cases, the statistical distribution of the signal can deviate

^{*}Contact author: mikel.falxa@unimib.it

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

from Gaussianity, introducing higher-order structure such as skewness or excess kurtosis in the data.

In this work, we develop a method to incorporate non-Gaussian statistics into PTA data analysis by modeling the Fourier coefficients of the signal as drawn from a Gaussian mixture model (GMM). This approach allows for analytical marginalization of the likelihood while introducing flexibility to capture heavy tails or excess higher-order moments in the signal distribution, complementing previous studies [16]. We explore the method's performance on simulated datasets including individual pulsar noise, common red noise with excess kurtosis, and composite signals formed by the superposition of a Gaussian background and a small number of CGWs. Additionally, we apply our model to search for non-Gaussianities in a realistic signal that is generated from the incoherent sum of CGWs produced by a simulated population of SMBHBs. The method enables the inference of posterior distributions for higher-order moments (e.g., kurtosis), which can be used in spectral refitting techniques to gain additional insight into the statistical structure and astrophysical origin of the observed signals. By providing a principled extension to the Gaussian framework, this work offers a practical pathway to account for and exploit non-Gaussian features in PTA data analysis.

The paper is organized as follows. First, we present the standard Gaussian likelihood that is traditionally used in PTA data analysis. Then, we show how GMMs can model non-Gaussian behaviors and be readily incorporated in current data analysis techniques to produce posterior distributions of higher-order moments. Finally, we test the model on simulated data to compare a Gaussian model with the non-Gaussian model. The latter is tested in an astrophysically motivated scenario where a CGW is resolved on top of either a purely Gaussian GWB or a realistic GWB.

II. METHOD

In Bayesian analysis, our knowledge of the parameters in a given model is updated by inferring their posterior probability distribution given some prior knowledge and the outcome of the experiment (cf. Bayes theorem [17]). This inference requires specifying a prior probability distribution for each parameter (i.e., a prior assumption about their behavior): for example, assuming Gaussian noise corresponds to assigning normally distributed priors to the noise parameters.

A. Gaussian noise

Consider a dataset δt with uncorrelated diagonal Gaussian noise matrix N . The data contain time-correlated noise that is modeled using its spectral decomposition on a discrete Fourier basis F with a vector of Gaussian distributed coefficients $\vec{a} \sim \mathcal{N}(0, \Phi)$ where the covariance matrix

Φ is a function of hyperparameters $\vec{\lambda}$. The posterior distribution $p(\vec{a}, \vec{\lambda} | \delta t)$ can be written as the product of, respectively, the Gaussian likelihood $\mathcal{L}(\delta t | \vec{a})$, the prior probability $p(\vec{a} | \vec{\lambda})$ on coefficients $\vec{a}$ and the prior probability $\Pi(\vec{\lambda})$ on hyperparameters $\vec{\lambda}$:

$$\ln p(\vec{a}, \vec{\lambda} | \delta t) = \ln \mathcal{L}(\delta t | \vec{a}) + \ln p(\vec{a} | \vec{\lambda}) + \ln \Pi(\vec{\lambda}). \quad (1)$$

When the assumption of Gaussian noise is enforced, we set a Gaussian prior probability distribution denoted $\Pi(\vec{a} | \vec{\lambda})$. In this case, this expression becomes

$$\begin{aligned} \ln p(\vec{a}, \vec{\lambda} | \delta t) = & -\frac{1}{2} [\delta t - F\vec{a}]^\top N^{-1} [\delta t - F\vec{a}] - \frac{1}{2} \ln \det 2\pi N \\ & - \frac{1}{2} \vec{a}^\top \Phi^{-1} \vec{a} - \frac{1}{2} \ln \det 2\pi \Phi \\ & + \ln \Pi(\vec{\lambda}), \end{aligned} \quad (2)$$

with $\Phi = \text{diag}\{S(f_n, \vec{\lambda}) \Delta f_n\}$ and $S(f_n, \vec{\lambda})$ the one-sided power spectral density (PSD), $f_n = n/T$ for $n = [1, 2, \dots, N_f]$ and $\Delta f_n = f_{n+1} - f_n$. Setting Φ as diagonal means assuming stationary noise (even though spectral leakage due to the finite observation time window can introduce spurious off-diagonal terms [18]).

This expression can be marginalized with respect to the Fourier coefficients $\vec{a}$ (see Appendix B) to give the marginalized likelihood expression commonly used in PTA data analysis [14]:

$$\begin{aligned} \ln \int p(\vec{a}, \vec{\lambda} | \delta t) d\vec{a} = & -\frac{1}{2} \delta t^\top C^{-1} \delta t - \frac{1}{2} \ln \det \{2\pi C\} + \ln \Pi(\vec{\lambda}) \\ = & \ln \mathcal{L}(\delta t | \vec{\lambda}) + \ln \Pi(\vec{\lambda}), \end{aligned} \quad (3)$$

where $C = N + F^\top \Phi F$ and $\mathcal{L}(\delta t | \vec{\lambda})$ is the marginalized Gaussian likelihood.

In PTA specifically, the F matrix is the concatenation of the Fourier basis of all the noise sources present in the data with the timing model design matrix. The latter accounts for first-order errors in the pulsar timing fit that is performed to obtain the timing residuals δt [19].

B. Non-Gaussian noise

In the case of non-Gaussian noise, we can write the same posterior as in Eq. (1), with the difference that $\vec{a} \sim p_a(\vec{\lambda}_{NG})$, meaning that the Fourier coefficients are distributed according to $p_a(\vec{\lambda}_{NG})$, a non-Gaussian probability distribution that is a function of hyperparameters $\vec{\lambda}_{NG}$. The problem is that the marginalization cannot be performed analytically [16]. Still, we can model any probability distribution using a GMM as

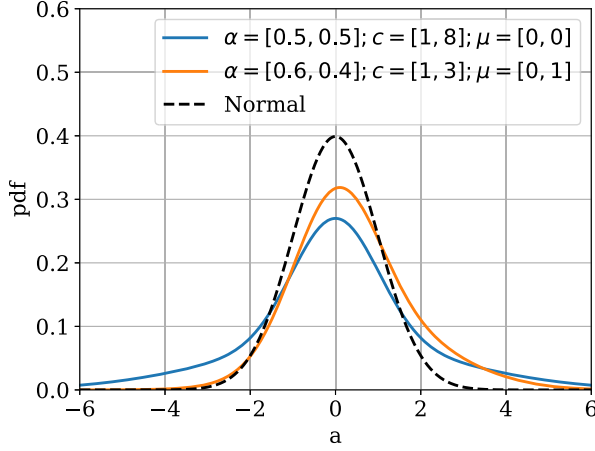


FIG. 1. Example of distributions obtained with a GMM of two Gaussians with $\Phi = 1$ [see Eq. (4)] and different sets of parameters α , μ and c introducing heavier tails or asymmetries (i.e., kurtosis and skewness).

$$p_a(a|\Phi, \vec{\alpha}, \vec{\mu}, \vec{c}) = \sum_{i=1}^N \alpha_i \frac{\exp\{-\frac{1}{2}[a - \mu_i]^2(c_i\Phi)^{-1}\}}{\sqrt{2\pi(c_i\Phi)}} \\ = \sum_{i=1}^N \alpha_i \varphi(a, \mu_i, \sqrt{c_i\Phi}), \quad (4)$$

with $\sum_i \alpha_i = 1$ to ensure the normalization of the distribution. We denote as $\varphi(a, \mu_i, \sqrt{c_i\Phi})$ the probability density function of a Gaussian distribution with mean μ_i and variance $c_i\Phi$ (see Fig. 1).

The advantage of writing the prior in this form is that the marginalization becomes fully analytical, as in Eq. (3). We consider the class of symmetric distributions centered on zero by setting $\vec{\mu} = 0$. For a single Fourier coefficient a , following the probability distribution in Eq. (4), the marginalized likelihood is

$$\ln \int p(a, \Phi, \vec{\alpha}, \vec{c} | \delta t) da = \ln \sum_{i=1}^N \alpha_i \mathcal{L}(\delta t | c_i\Phi) \\ + \ln \Pi(\Phi) + \ln \Pi(\vec{c}) + \ln \Pi(\vec{\alpha}), \quad (5)$$

where deviations from Gaussianity are controlled by the parameters α_i and c_i . In Table I we give the first central moments for a mixture of two Gaussian distributions.

One obvious caveat is that the more Gaussian distributions we use in the mixture, the more likelihood evaluations are required, making the overall computation slower (see Sec. IID). In particular, if the hypothesis of uncorrelated frequencies is required across the considered spectrum, the number of likelihood evaluations grows exponentially as $\sim N^{2N_f}$. We can however check Gaussianity frequency bin per frequency bin.

TABLE I. First four central moments for a mixture of two Gaussian distributions with $\vec{\alpha} = [1 - \alpha, \alpha]$, $\vec{c} = [1, c]$ and $\vec{\mu} = [0, \mu_0]$ [see Eq. (4)]. We have defined $E\{a\} = \bar{a}$. Note that for $\alpha = 0$, the moments are identical to those of the standard Gaussian distribution. For more details, see Appendix A.

q	$M_q = E\{(a - \bar{a})^q\}$
1	0
2	$\mu_0^2 \alpha [1 - \alpha] + \Phi [1 + \alpha(c - 1)]$
3	$\mu_0^3 [\alpha(1 - \alpha)^3 - \alpha^3(1 - \alpha)] + \Phi 3(1 - \alpha)\alpha\mu_0(c - 1)$
4	$\mu_0^4 [\alpha^4(1 - \alpha) + \alpha(1 - \alpha)^4] + 6\mu_0^2 \Phi [(1 - \alpha)\alpha^2 + \alpha(1 - \alpha)^2 c] + 3\Phi^2 [1 + \alpha(c^2 - 1)]$

For $\vec{\mu} = [0, 0]$ the $2q$ th central moment has a very simple expression:

$$M_{2q} = (2q - 1)!! \Phi^q [1 + \alpha(c^q - 1)]. \quad (6)$$

In this case, only the even moments M_{2q} are nonzero because the distribution is symmetric; thus, $M_{2q+1} = 0$.

If we denote $\mathcal{M}_q(\sigma)$ the q th moment of a centered Gaussian distribution with variance σ^2 , we can quantify deviations from Gaussianity by calculating the relative excess moments:

$$\Delta \bar{M}_q = \frac{M_q - \mathcal{M}_q(\sqrt{M_2})}{\mathcal{M}_q(\sqrt{M_2})}. \quad (7)$$

$\Delta \bar{M}_q$ tells us how much the probability distribution $p(a)$ is different from a Gaussian that has the same M_2 , i.e., the same variance. If $\Delta \bar{M}_q \neq 0$ for $q > 2$, $p(a)$ is not shaped like a Gaussian.

C. Common correlated noise: Higher-order statistics

If the observed signal is non-Gaussian and of gravitational origin, it is possible to calculate its higher-order correlators. In PTA, a Gaussian GWB is entirely characterized by its second moment and the associated 2-point correlation function $\chi_{IJ}^{(2)}$ between 2 pulsars is the well-known HD curve [7], describing the overlap between their response to an isotropic GW signal. Considering N_p pulsars and the n th Fourier frequency, the prior probability distribution for the vector of random Fourier coefficients $\vec{a}_n$ of length N_p is

$$\Pi(\vec{a}_n) = \frac{\exp\{-\frac{1}{2} \sum_{IJ} a_{I,n} [\chi_{IJ}^{(2)} \Phi_n]^{-1} a_{J,n}\}}{\sqrt{\det 2\pi \chi^{(2)} \Phi_n}}, \quad (8)$$

where $\chi^{(2)}$ is the matrix of the correlation coefficients between $\vec{a}_n$.

This expression yields correlated noise between pulsars that is responsible for off-diagonal terms in the marginalized likelihood expression (see Appendix B). In [20,21], we see that higher-order statistics necessitate the calculation of at least the 4-point correlation functions $\chi_{IJKL}^{(4)}$ characterizing the excess kurtosis (4th cumulant) of the noise statistics, that is

$$\langle a_{I,n} a_{J,n} a_{K,n} a_{L,n} \rangle \propto \chi_{IJKL}^{(4)} h_n^{(4)}, \quad (9)$$

where $h_n^{(4)}$ is the level of excess kurtosis and the brackets denote the ensemble average.

However, incorporating higher-order correlators can become expensive in PTA analysis, because the number of pulsars that make up the data is usually large. Consequently, the number of possible combinations and the associated numerical cost grow quickly. As a first approximation, we could ignore those interpulsar higher-order correlations $\chi_{IJKL}^{(4)}$ for $I \neq J \neq K \neq L$ and still model the non-Gaussianities using lower-order statistics, because the presence of excess kurtosis still introduces effects for the $I = J = K = L$ terms. Conceptually, this is the analog of modeling a common uncorrelated noise instead of an HD correlated one. We retain the statistics of the Fourier spectrum while neglecting the higher-order correlation signatures, hence losing some of the information. For example, if the signal is generated by an astrophysical population of SMBHBs, the underlying Poissonian statistics should still be detectable in the spectrum without accounting for $\chi_{IJKL}^{(4)}$ [22,23].

We show in Appendix C how we can add the contributions of higher-order statistics by expanding the prior probability distribution $\Pi(\vec{a}_n)$ in terms of the Gram-Charlier A series (or similarly using the Edgeworth series). This approach being computationally less efficient, it is not further explored in this work.

D. Non-Gaussian likelihood

When noise is assumed to be Gaussian and stationary, the likelihood takes a well-established form that is widely used in PTA data analysis [14]. The noise is described by its Fourier coefficients $\vec{a}$ and the assumption of stationarity requires that different frequencies are uncorrelated. In terms of prior probability, this translates into

$$p(\vec{a}) = \prod_{n=1}^{N_f} p_n(a_n^c, \Phi_n^c, \vec{a}, \vec{\mu}, \vec{c}) p_n(a_n^s, \Phi_n^s, \vec{a}, \vec{\mu}, \vec{c}), \quad (10)$$

where $p_n(a_n)$ is the prior probability for the n th coefficient and N_f is the number of frequency bins. There are $2N_f$ terms in total to account for a_n^c and a_n^s , respectively the real

and imaginary components of a_n . We reiterate that assuming each frequency to be uncorrelated is only an approximation when we have a finite observation time that might bias the inference [18].

If the $p_n(a_n)$ are modeled as a GMM of N Gaussian distributions, the total probability $p(\vec{a})$ becomes a sum of N^{2N_f} Gaussian distributions. Then, the evaluation of the marginalized non-Gaussian likelihood would require the evaluation of N^{2N_f} Gaussian likelihoods, making the process very computationally inefficient, although fully analytical. We propose two work-arounds for two limiting cases: (i) using the unmarginalized likelihood form and sampling the Fourier coefficients and (ii) using the marginalized likelihood by setting non-Gaussian behavior in one single frequency bin. The first approach is practical for single pulsar noise analysis, assuming that the non-Gaussian behavior of the noise is consistent across frequencies and that enough frequencies are probed to identify deviations from Gaussianity. The second approach can only be used for common noise between pulsars as it allows one frequency bin to deviate from Gaussianity. It does however require a larger number of pulsars to pick up the non-Gaussian behavior.

1. Method 1. Sampling the Fourier coefficients

In the case of single pulsar noise analysis, we directly sample the coefficient $\vec{a}$ using the likelihood

$$\mathcal{L}(\delta t | \vec{a}, \vec{\Phi}, \vec{\mu}, \vec{\alpha}, \vec{c}) = \mathcal{L}(\delta t | \vec{a}) p(\vec{a} | \vec{\Phi}, \vec{\mu}, \vec{\alpha}, \vec{c}). \quad (11)$$

This method is the most straightforward as it requires no analytical marginalization. In fact, it does not even require a GMM for the prior, as long as the distribution is known [16]. The main caveat is that the dimensionality of the parameter space may be very high. Previous work showed how this type of problem can be optimized using Gibbs sampling [24].

2. Method 2. Per-frequency marginalized non-Gaussian likelihood

For the per-frequency marginalized non-Gaussian likelihood, we only model non-Gaussian behavior at one frequency. Then, the prior probability distribution for the Fourier coefficients is given by Eq. (10) for $N_f = 1$, where the $p_n(a_n^c, \Phi_n^c, \vec{a}, \vec{\mu}, \vec{c})$ and $p_n(a_n^s, \Phi_n^s, \vec{a}, \vec{\mu}, \vec{c})$, modeled as GMMs, are given by Eq. (4). The marginalization of Eq. (11) with respect to $\vec{a}$ yields

$$\mathcal{L}(\delta t | \vec{\Phi}, \vec{\mu}, \vec{\alpha}, \vec{c}, \vec{\lambda}) = \sum_{i,j=1}^N \alpha_i \alpha_j \mathcal{L}(\delta t | c_i \Phi_n^c, c_j \Phi_n^s, \mu_i, \mu_j, \vec{\lambda}), \quad (12)$$

where $\mathcal{L}(\delta t | c_i \Phi_n^c, c_j \Phi_n^s, \mu_i, \mu_j, \vec{\lambda})$ is the marginalized likelihood derived in Sec. II B and $\vec{\lambda}$ is the vector of all other

parameters that are not related to the non-Gaussian signal (i.e., Gaussian prior widths, timing model parameters, individual pulsar noise). For $\mu_i = \mu_j = 0$, the likelihoods in the sum are identical to the commonly used PTA likelihood [14]. Then, the computation of the non-Gaussian likelihood can be parallelized by computing each term in the sum in parallel. Moreover, this method is very well suited for reversible jump Markov Chain Monte Carlo (MCMC) techniques where one could dynamically add or remove Gaussians to the mixture to find their optimal number and properties. This will be left for future studies.

In this work, we will only consider the case with $\Phi_n^c = \Phi_n^s$, $\vec{\mu} = 0$ and a mixture of two Gaussians. Using Eq. (4), we construct the probability distribution for the non-Gaussian noise,

$$p_n(a|\Phi_n, \alpha, c) = (1 - \alpha) \times \varphi(a, 0, \sqrt{\Phi_n}) + \alpha \times \varphi(a, 0, \sqrt{c\Phi_n}), \quad (13)$$

that is fully characterized by three parameters Φ_n, α and c . This yields a likelihood of the form

$$\begin{aligned} \mathcal{L}(\delta t|\Phi_n^c, \Phi_n^s, \alpha, c, \vec{\lambda}) &= (1 - \alpha)^2 \mathcal{L}(\delta t|\Phi_n, \Phi_n, \vec{\lambda}) \\ &+ \alpha^2 \mathcal{L}(\delta t|c\Phi_n, c\Phi_n, \vec{\lambda}) \\ &+ (1 - \alpha)\alpha \mathcal{L}(\delta t|\Phi_n, c\Phi_n, \vec{\lambda}) \\ &+ \alpha(1 - \alpha) \mathcal{L}(\delta t|c\Phi_n, \Phi_n, \vec{\lambda}). \end{aligned} \quad (14)$$

This method could be generalized to any known probability distribution $p_n(a, \vec{\lambda}_{NG})$ by first fitting a GMM to $p_n(a, \vec{\lambda}_{NG})$ and fixing the recovered parameters $\vec{a}, \vec{\mu}$ to their best fit values to get an analytical approximation of the marginalized non-Gaussian likelihood. If a mapping function $\vec{\lambda}_{NG} \rightarrow \vec{a}, \vec{\mu}$ is defined, one could infer the parameters $\vec{\lambda}_{NG}$ from the data through direct sampling or spectral refitting [25–27].

E. Free spectrum

In the PTA literature, a widely used data analysis technique is the free spectrum analysis [9–12, 19], that consists of estimating the rms value ρ of the noise (or signal) in the data at a given frequency bin. This estimate is directly proportional to the PSD of the stochastic processes and gives crucial information about its physics. However, this estimate always assumes that the noise in that frequency bin is Gaussian, which is not necessarily true. To produce an unbiased estimator of the PSD, we need to account for the real statistic of the Fourier bins $\vec{a}$ of that process. Given that we know the distribution of a_n at frequency n for both its real a_n^c and imaginary parts a_n^s (or equivalently cosine and sine components), the free spectrum is given by

$$\rho_n^2 = S(f_n)\Delta f = \frac{1}{2} \langle |a_n^c|^2 + |a_n^s|^2 \rangle, \quad (15)$$

where the brackets denote the ensemble average, following the definition of the PSD for stochastic processes that is related to the variance of a_n .

The free spectrum is obtained from the posterior samples of parameters characterizing the probability distribution of a_n . For Gaussian noise, it is directly given by the posterior distribution of its standard deviation $\sqrt{\Phi_n}$, since the term in the brackets behaves as a chi-square distribution with 2 degrees of freedom, as $\rho_n^2 = \Phi_n \langle \chi^2 \rangle / 2 = \Phi_n$.

When a_n is described as a 2 Gaussian GMM, the result is given by Table I. For more Gaussian in the GMM, we found no closed form expression of the ensemble average for the sum of the squares. However, the result can be easily obtained numerically by sampling the GMM. For a set of non-Gaussianity parameters Φ_n, α and c , given that we have inferred the posterior distribution $p(\Phi_n, \alpha, c|\delta t)$ from data δt , the probability distribution $p(a_n, \Phi_n, \vec{a}, \vec{c}|\delta t)$ is given by

$$p(a_n, \Phi_n, \alpha, c|\delta t) = p(a_n|\Phi_n, \alpha, c)p(\Phi_n, \alpha, c|\delta t), \quad (16)$$

that is the prior probability defined in Eq. (4) weighted by the posterior probability of parameters Φ_n, α, c . The probability distribution of a_n marginalized over the non-Gaussianity parameters can be obtained from the posterior samples of Φ_n, α, c . For N_D samples drawn from $p(\Phi_n, \alpha, c|\delta t)$, we have

$$\begin{aligned} p(a_n|\delta t) &= \int p(a_n, \Phi_n, \alpha, c|\delta t) d\Phi_n d\alpha dc \\ &\approx \frac{1}{N_D} \sum_{j=1}^{N_D} p(a_n|\Phi_j, \alpha_j, c_j). \end{aligned} \quad (17)$$

We can also construct posterior samples of the $2q$ th moment estimators ρ_n^{2q} of $p(a_n|\delta t)$ as

$$\rho_{n,j}^{2q} \approx \frac{1}{2N_s} \sum_{i=1}^{N_s} [|a_{n,j,i}^c|^{2q} + |a_{n,j,i}^s|^{2q}], \quad (18)$$

where the $a_{n,j,i}^c, a_{n,j,i}^s$ are sampled from $p(a_n, \Phi_{n,j}, \alpha_j, c_j|\delta t)$ and N_s is the total number of samples used to numerically estimate the average for the j th posterior sample of $p(\Phi_n, \alpha, c|\delta t)$. Here, we discard uneven moments because we only consider symmetric distributions. In this work, since we consider uncorrelated a_n^c and a_n^s modeled as a 2 Gaussian GMM, $\rho_{n,j}^{2q}$ and $\rho_{n,j}^{4q}$ can be directly obtained from the posterior samples of $\Phi_{n,j}, \alpha_j, c_j$ using the expressions in Table I.

The histogram of the recovered $\rho_{n,j}^{2q}$ corresponds to the posterior distribution of the inferred $2q$ th moments. This

additional information could be used in spectral refitting techniques to constrain the statistical properties of astrophysical signals like the GWB produced by a population of SMBHBs [21–23]. For $q = 1$, the distribution gives the already known Gaussian free spectrum that is an estimator of the PSD.

III. RESULTS

To test the performance of the model we will study two cases: the single pulsar noise analysis and the search of non-Gaussianities at a given frequency for a common uncorrelated pulsar noise. We consider a PTA of $N_p = 100$ pulsars with $T_{\text{obs}} = 10$ years of observation, $N_t = 500$ data points (≈ 1 week cadence) and $\sigma = 10^{-7}$ s timing uncertainty level. The data are simulated using `fakepta`¹ and the Gaussian likelihoods are computed using `ENTERPRISE`² [28]. The timing residuals δt_a of pulsar a can be written as

$$\delta t_I = w_I + \sum_n X_{I,n} \sin(2\pi f_n t) + Y_{I,n} \cos(2\pi f_n t), \quad (19)$$

with $w_I \sim \mathcal{N}(0, \sigma^2 \mathcal{I}_{N_t})$ where $\mathcal{I}_{N_t}$ is the identity matrix of rank N_t and $\vec{X}_n, \vec{Y}_n \sim p(\alpha, c, S(f_n) \Delta f)$ where $S(f_n)$ is the one-sided PSD of the non-Gaussian time-correlated process and α, c the parameters characterizing the non-Gaussianities of the probability distribution p . The PSD of the process $S(f)$ is chosen to be a power law of the form

$$S(A, \gamma, f) = \frac{A^2}{12\pi^2} \left(\frac{f}{f_{\text{yr}}} \right)^{-\gamma}, \quad (20)$$

with A the amplitude, γ the spectral index and $f_{\text{yr}} = 1/\text{yr}$ the frequency corresponding to a one-year period.

For each pulsar we consider a timing model accounting for the spin-down rate of that pulsar described by a quadratic polynomial in time and we marginalize over the first-order errors in the timing model parameters [14,19]. The full likelihood for common uncorrelated noise is written as a product of N_p individual pulsar likelihoods given by Eq. (14)

$$\mathcal{L}(\delta t | \Phi_n^c, \Phi_n^s, \alpha, c, \vec{\lambda}) = \prod_{k=1}^{N_p} \mathcal{L}_k(\delta t_k | \Phi_n^c, \Phi_n^s, \alpha, c, \vec{\lambda}_k). \quad (21)$$

A. Single pulsar noise

We test the model by injecting a time-correlated noise that has a consistent statistic across all frequency bins. In

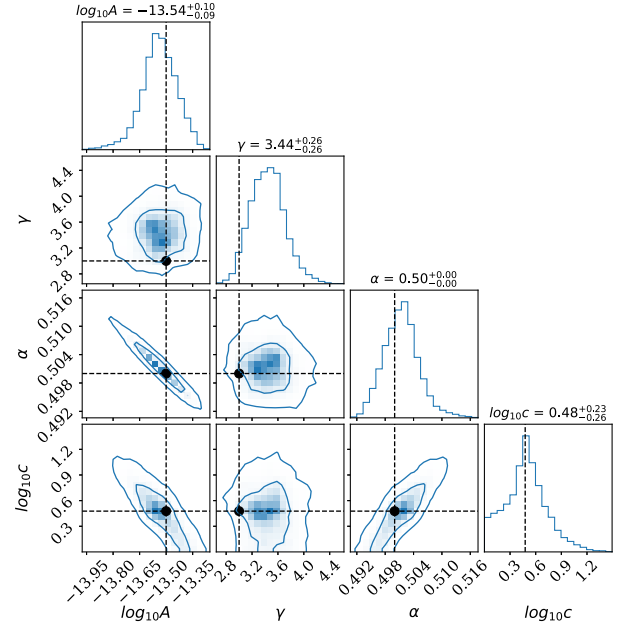


FIG. 2. Corner plot of the recovered parameters for the non-Gaussian model in single pulsar noise analysis. The Fourier coefficients are also sampled but not shown here for simplicity. The dashed line show the injected parameter values. The shaded areas show the 1 and 2 σ credible regions.

this example, we consider a noise that has Fourier coefficients a distributed according to GMM of 2 Gaussians with $\alpha = 0.5$ and $c = 3$, following a power law as in Eq. (20) with $\log_{10} A = -13.5$ and $\gamma = 3$ across 50 frequency bins that are the 100 first harmonics of $1/T_{\text{obs}}$. Recent works have shown the advantages of directly sampling the Fourier coefficients [24,29]. The only disadvantage comes from the high number of coefficients there are to sample, but the recovered posteriors for the unmarginalized parameters are identical to the traditionally marginalized likelihood.

The result of one simulation is shown in Fig. 2 to verify that we indeed recover the injected values of the parameters. We only show the hyperparameters characterizing the noise, while numerically marginalizing over all Fourier coefficients by sampling them. We have used 50 frequencies, for a total of 100 Fourier coefficients that we fit for. This approach is straightforward as it requires no analytical marginalization. However, the sampling is affected by the high dimensionality of the parameter space and many samples are required to converge to the proper distribution. Still, in the case of GMMs, the nonmarginalized (method 1) and marginalized (method 2) likelihoods presented in Sec. IID should respond equally to non-Gaussianities present in the data. This can be seen in Fig. 3 where we plot the Bayes factor $\mathcal{B}_G^{NG}$ comparing the non-Gaussian and Gaussian models as a function of the excess kurtosis $\Delta \bar{M}_4$. As $\Delta \bar{M}_4$ increases, the Bayes factor increases accordingly. Note that for $\Delta \bar{M}_4 = 0$ (i.e., Gaussian noise is injected), we

¹<https://github.com/mfalxa/fakepta>.

²<https://github.com/nanograv/enterprise>.

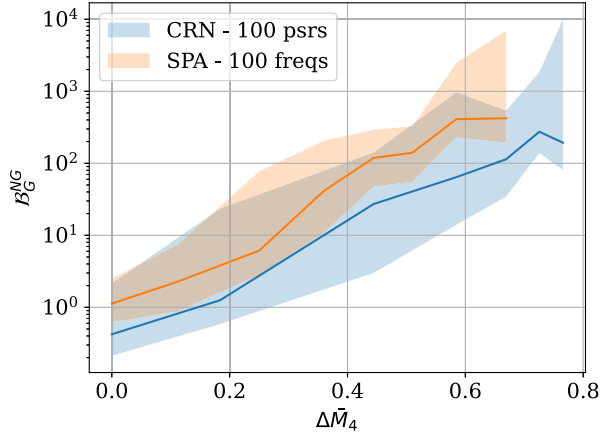


FIG. 3. Bayes factor $\mathcal{B}_G^{NG}$ between the non-Gaussian and Gaussian models as a function of the injected relative excess kurtosis $\Delta\bar{M}_4$ calculated using Eq. (7) for many simulations. The solid line is the median and the shaded region is bounded by the 16th and 84th percentiles. We show the two cases: (i) SPA (Single Pulsar Analysis) with 100 frequencies powerlaw red noise, (ii) CRN (Common red noise) with 100 pulsars.

have $\mathcal{B}_G^{NG} \approx 1$ meaning that the non-Gaussian model is not preferred over the Gaussian model.

B. Common red noise

We apply the per-frequency marginalized non-Gaussian likelihood method to infer the statistics of the injected common noise in the 100-pulsar simulated dataset. Our goal is to test that the Bayes factor correctly favors the injected non-Gaussianities over a Gaussian model while recovering the injected parameters. We simulate an uncorrelated common red noise where each frequency bin n follows a 2 Gaussian GMM $\vec{X}_n, \vec{Y}_n \sim p(\alpha, c, S(A, \gamma, f_n) \Delta f)$. The noise spectrum has characteristics similar to what is expected for a GWB produced by a population of SMBHBs in circular orbit with characteristic strain $h_c = A(f/f_{yr})^\beta$ with $\beta = -2/3$ and $A = 10^{-15}$. The conversion from h_c to the PSD of the induced timing residuals is given by $S(f) = h_c^2 / (12\pi^2 f^3)$ [30]; hence, $\gamma = 3 - 2\beta = 13/3$.

We fix $\alpha = 0.5$, $\gamma = 13/3$ and vary c between 1 and 15 to control the excess moments of the distribution. We sample the posterior distribution of Eq. (1) constructed from the likelihood in Eq. (14) using MCMC techniques. We assess the significance of non-Gaussianities by calculating the Bayes factor for the non-Gaussian model versus the Gaussian model $\mathcal{B}_G^{NG}$. The latter is estimated using hypermodeling [31,32] (product-space sampling) or alternatively, with the Savage-Dickey density since the Gaussian model is obtained by fixing $\alpha = 0$ [33,34]. We perform per-frequency free spectrum fits, modeling non-Gaussian features characterized by parameters Φ , α and c . The prior ranges of the parameters are listed in Table II. We

TABLE II. Uniform prior ranges for the parameters that are used in this work. The parameters $\log_{10} A$ and γ are used when we assume a power law spectrum. The free spectrum fits use the parameter $\log_{10} \sqrt{\Phi}$.

$\log_{10} A$	$[-18, -11]$
γ	$[1, 7]$
α	$[0, 1]$
$\log_{10} \sqrt{c}$	$[0, 2]$
$\log_{10} \sqrt{\Phi}$	$[-10, -4]$

produced a probability-probability plot to assess the good calibration of the simulations (see Fig. 11 in Appendix D).

In Fig. 3 we show $\mathcal{B}_G^{NG}$ as a function of the injected excess kurtosis $\Delta\bar{M}_4$ estimated using Eq. (7). We see a clear increase of the Bayes factor for growing $\Delta\bar{M}_4$, indicating that non-Gaussianities can be detected in Bayesian model selection.

The kurtosis increase can be seen in Fig. 4 where the 2nd and 4th moment posterior distributions given by Eq. (18) are shown for the Gaussian and non-Gaussian models. We see that the second moment is the same for both, meaning that the data inform the correct PSD under the hypothesis of Gaussian noise. However, the latter is unable to detect higher-order statistics that could contain interesting information about the signal. The relative excess kurtosis in Fig. 4 is calculated using Eqs. (7) and (18) as $\Delta\bar{M}_4 = [\rho^4 - 3(\rho^2)^2] / 3(\rho^2)^2$. In Fig. 5, we show the reconstruction of the Fourier coefficient statistic using

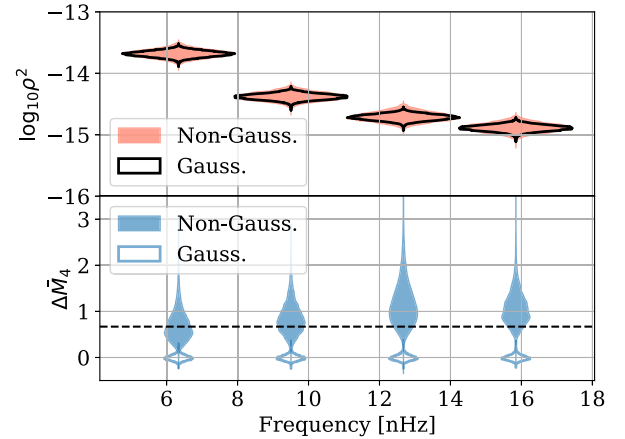


FIG. 4. Recovered free spectrum and excess kurtosis for the Gaussian model and non-Gaussian model at the 2, 3, 4, 5 harmonics of 10 yr for the simulated data containing a non-Gaussian CRN following a power law spectrum with $\log_{10} A = -15$, $\gamma = 13/3$, $\alpha = 0.5$ and $c = 10$. Top panel: violin plot of the estimated free spectrum; the two models are very consistent. Bottom panel: violin plot of the excess kurtosis posterior distribution giving zero for the Gaussian model; the dashed line shows the injected excess kurtosis obtained with Eq. (7), that is the same for all frequencies.

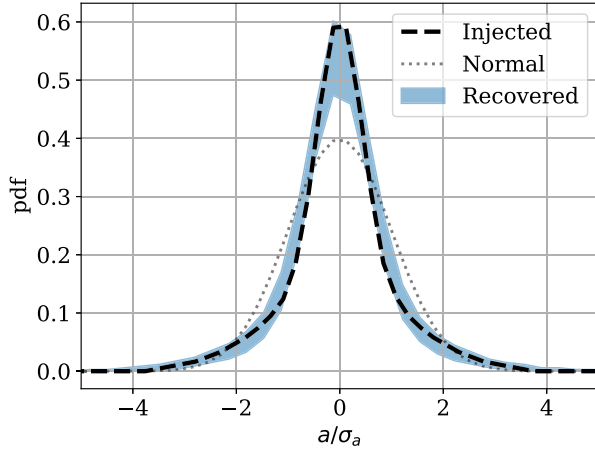


FIG. 5. Comparison between the injected and recovered probability distribution of the Fourier coefficients a , divided by the standard deviation σ_a , showing the deviation from Gaussianity. The blue shaded region corresponding to the binned $1\text{-}\sigma$ error obtained with 1000 reconstructions of $p(a_n, \Phi_n, \alpha, c|\delta t)$ using samples from $p(\Phi_n, \alpha, c|\delta t)$ according to Eq. (16). The dotted line shows a standard normal distribution for comparison.

Eq. (16) by drawing samples from the posterior distribution of parameters Φ_n , α and c . We correctly recover the injected probability distribution within $1\text{-}\sigma$ errors. We see that it has thinner body and heavier tails than a Gaussian, matching the expected shape for excess kurtosis.

C. Background plus few sources

An astrophysically interesting scenario where non-Gaussianities can be observed is when the signal in the data consists of a Gaussian background noise plus a few loud individual sources at some specific frequency [35]. This is, in fact, a reasonable proxy for what is expected from a realistic population of SMBHBs, since some binaries will be closer to us or more massive, appearing in the data as relatively loud CGWs [36–40]. If these binaries are identified and subtracted, the background population should be isotropic enough to produce a Gaussian noise, as expected by the law of large numbers [35]. In the following, we investigate the non-Gaussian statistics emerging from the superposition of a Gaussian GWB with a few CGWs. In order to apply method 2 of Sec. II D, we study the case where the CGWs fall within the same frequency bin. In PTA, an individual, circular SMBHB produces timing residuals given by the difference between the signal seen on Earth and at the pulsar,

$$s_a(t) = h \sum_{A=+, \times} F_A(\hat{p}_a, \hat{k}) [s_A(t) - s_A(t - \tau_a)], \quad (22)$$

where h is the GW amplitude, F_A the antenna pattern function, $\hat{p}_a$ the unit vector pointing at the position of pulsar a , $\hat{k}$ the direction of propagation of the wave, $s_A(t)$

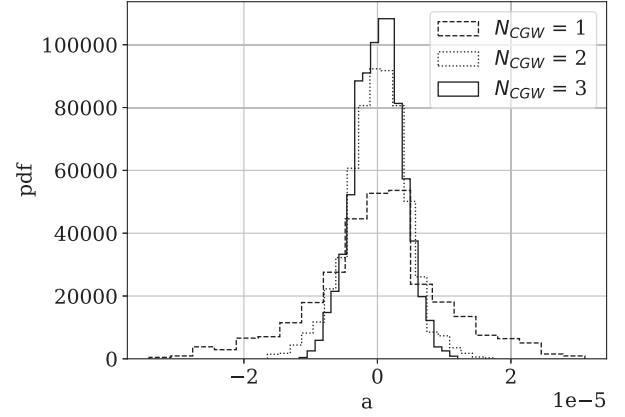


FIG. 6. Histogram of the FFT coefficients calculated for each pulsar timing residual at frequency $3/(10 \text{ yr})$ where a CRN and three CGW sources were injected in sequence. We gradually added the three CGWs with amplitudes of $\log_{10} h = -14.5$, -14.65 and -15 corresponding to SNRs of, respectively 37, 17, 7. The value N_{CGW} marks the total number of CGWs injected in the data. By adding more CGWs the tails of the distribution get progressively damped.

the circular binary waveform for each polarization modes $+$ and $\times$ that is a function of the frequency f_{gw} of the wave, the polarization angle and the inclination of the binary, and τ_a the relative time delay between Earth time and pulsar time (for more details about this type of signal and its full mathematical expression, see [37–40]).

We inject the same GWB as in the previous section with characteristic strain $h_c = 10^{-15} (f/f_{\text{yr}})^{-2/3}$. We inject in the data one, two or three CGWs with a frequency of $3/(10 \text{ yr})$ according to Eq. (22) and check for Gaussianity of the common noise at this frequency. The SNR of the sources with respect to the GWB is calculated in time domain as

$$\text{SNR}^2 = \sum_{a=1}^{N_p} s_a^T C^{-1} s_a, \quad (23)$$

where $C(t - t') = \sigma^2 \mathcal{I}_{N_t} + \sum_n S(f_n) \Delta f \cos[2\pi f_n(t - t')]$.

To illustrate the effect of single sources in the statistic of the Fourier coefficients, in Fig. 6, we show the distribution of the Fourier coefficients at frequency $3/(10 \text{ yr})$ obtained with the fast Fourier transform (FFT) of simulated pulsar timing data where a GWB and some CGW sources are injected. Each pulsar FFT gives two values (real and imaginary part at the considered frequency) and we plot a histogram of all combined coefficients across pulsars. We see that a single CGW produces a heavy-tailed distribution, that becomes progressively more Gaussian when adding few more CGWs.

The fact that adding a single source introduces non-Gaussianities is illustrated in Fig. 7, where we show the added higher-order moments as a function of the SNR of

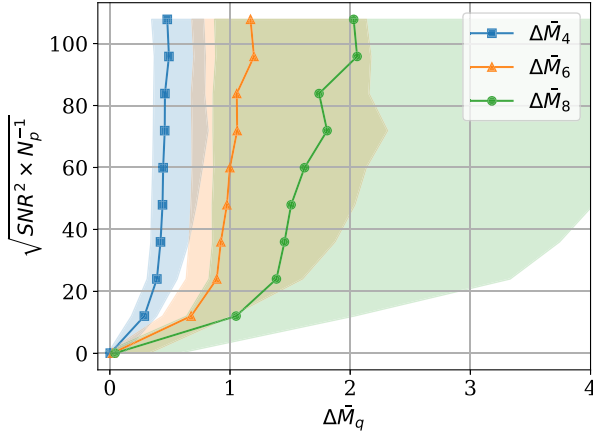


FIG. 7. Excess moments as a function of the averaged SNR of the source for N_p pulsars with 10 years of observation, 100- ns noise level and a GWB signal with $h_c = 10^{-15} (f/f_{\text{yr}})^{-2/3}$. Since $f_{\text{gw}} = 3/(10 \text{ yr})$, the source produces three cycles in the data. This plot shows the median excess moments estimated from a 100 realization of the CGW parameters. The shaded colored areas show the 16th and 84th percentiles of the realizations. Only the amplitude of the CGW was fixed to control the SNR.

that source, directly calculated from the histograms of the FFT coefficients for different CGW parameter realizations as shown in Fig. 6. In particular, we show the relative excess moments with respect to a Gaussian as defined in Eq. (7). It is clear that the deviations from Gaussianity grow with the SNR. This has to do with the pulsar response to GWs $F^{+, \times}$ introduced in Eq. (22). When one CGW is present in the data, it is going to appear very bright only in some pulsars that respond well to it, usually the pulsars close to the sky location of the source [41]. This is what causes the heavier tails in the distribution of the Fourier coefficients in Fig. 6 and puts more power in the higher-order moments in Fig. 7. As more CGW sources are present and isotropically distributed in the sky, all the pulsars will equally respond to a signal. Because the signals they see have random phases, for a large number of CGWs, the central limit theorem tells us that the Fourier coefficients will be asymptotically Gaussian distributed and spatially correlated following the HD correlation pattern [8]. Then, there is an intricate relation between the anisotropy of an astrophysical GWB and its non-Gaussianity.

We test our model on the scenario where only one CGW is injected with an amplitude $\log_{10} h = -14.25$ giving $\text{SNR} = 50$.³ In Fig. 8 we show the posterior distribution

³The SNR is high because it scales as $\sqrt{N_p}$ and we simulate $N_p = 100$ equally good pulsars. Even though a high SNR should yield more informative posterior distributions, it does not systematically mean that the injected excess moments $\Delta \tilde{M}_q$ are very large, since the latter depend on the statistics of Fourier coefficients in individual pulsars. To that extent, $\Delta \tilde{M}_q$ should scale as the average of the SNRs calculated for individual pulsars.

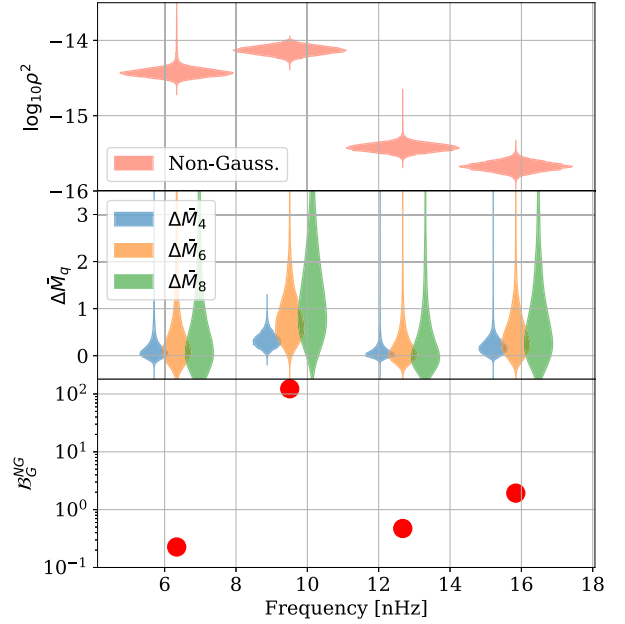


FIG. 8. Recovered free spectrum, excess moments and Bayes factor for the non-Gaussian model at the 2nd, 3rd, 4th, 5th harmonics of $f = 1/10 \text{ yr}$ for the simulated data containing a CGW at frequency $3/(10 \text{ yr})$. Top: violin plot of the estimated free spectrum; the $3/(10 \text{ yr})$ shows the excess power due to the presence of the CGW [42]. Middle: violin plot of the excess moment posterior distributions; the uncertainties are large but we see that the frequency $3/(10 \text{ yr})$ is well constrained to a higher value than other bins, showing that the model catches the excess of power in higher moments introduced by the CGW. Bottom: the Bayes factor $\mathcal{B}_G^{NG}$ at each frequency shows that the model favors non-Gaussianity where the CGW is present.

for higher-order moments calculated with Eqs. (18) and (7). We notice that the second bin of the free spectrum in the top panel shows an excess of power due to the presence of the CGW [42]. The middle panel shows that the posterior distributions for higher-order moments are consistent with zero, except where the CGW is present and the Bayes factor favors the non-Gaussian model over the Gaussian model. Still, the error bars are quite large, and the model does not seem to fully catch the characteristics of the underlying distribution. This is because a mixture of two Gaussians is a limited model that will saturate in the amount of information it can capture. Finding the optimal number of Gaussians in the GMM will be essential for future developments.

In Fig. 9 we show the recovered posterior distribution for the non-Gaussianity parameters when we include or not the CGW as a deterministic signal in the model used in the recovery. We choose a realization of the data where the CGW introduced significant non-Gaussianities, again modeled as the mixture of 2 central Gaussians controlled with parameters α and $\log_{10} c$. The introduction of the deterministic CGW signal in the model removes the non-Gaussian features and destroys their significance. For the model without CGW, we find posterior median values of

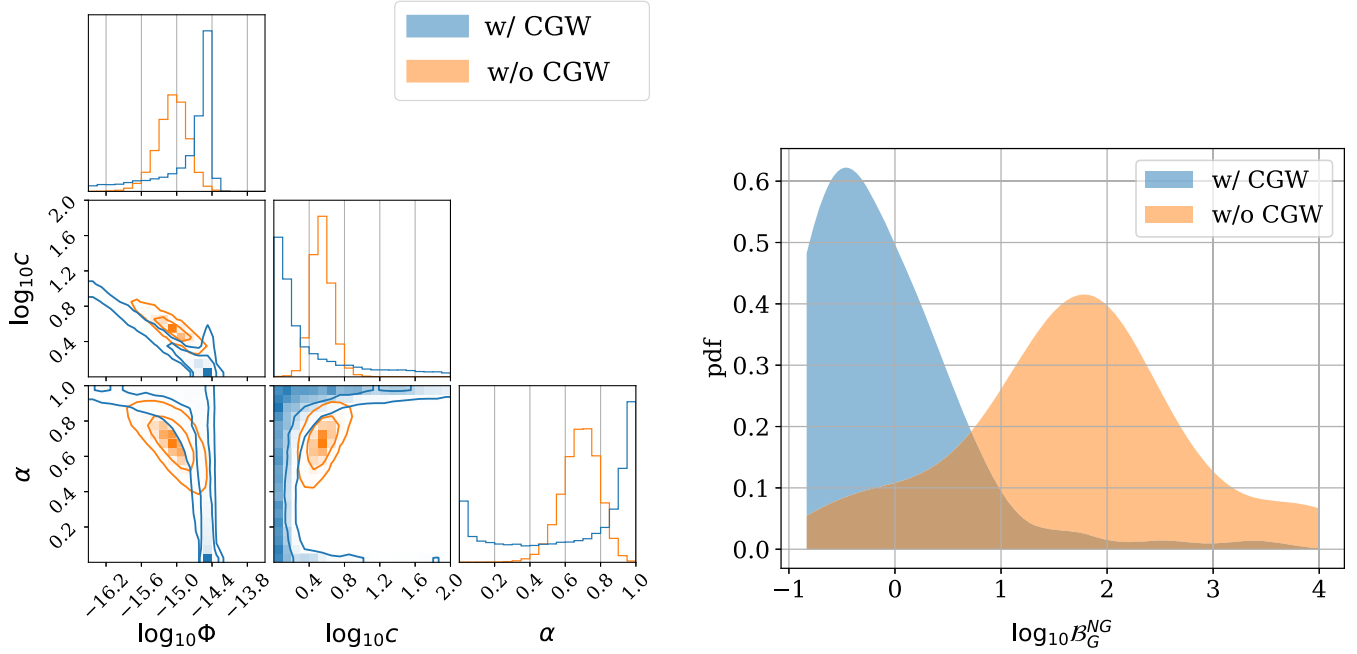


FIG. 9. Left: corner plot of the non-Gaussianity parameters obtained at frequency $3/(10 \text{ yr})$ for a model with CGW as a deterministic signal (blue histogram) and without the CGW (orange histogram). The shaded areas show the 1 and 2σ credible regions. While the orange histogram is well constrained around $\log_{10} c \approx 0.5$ and $\alpha \approx 0.7$, the blue one is bimodal and consistent with $\alpha = 0$ or $\alpha = 1$ and $\log_{10} c = 0$, showing that the model favors only one Gaussian in the mixture. Right: distribution of the recovered Bayes factors B_G^{NG} for 100 simulated datasets containing Gaussian noise with a single source at frequency $3/(10 \text{ yr})$. The significance of non-Gaussianities in the data is high when the CGW is not accounted for in the model (orange histogram), but the inclusion of the CGW as a deterministic signal in the model destroys this significance (blue histogram).

$\alpha = 0.7$ and $\log_{10} c = 0.5$ while the inclusion of the deterministic CGW gives posteriors consistent with $\alpha = 0$, or $\alpha = 1$ and $\log_{10} c = 0$, corresponding to the solutions where only one Gaussian of the mixture is active (i.e., Gaussian behavior is favored). Moreover, the recovered Bayes factor drops from $B_G^{NG} \approx 100$ to $B_G^{NG} \approx 0.5$, showing that the significance is greatly impacted by the subtraction of the CGW waveform in the timing residuals (see Fig. 9). This fact could be used to assess the significance of the presence of individual sources in the data by measuring how including the correct CGW signal in the model correctly yields a Gaussian likelihood.

D. Realistic astrophysical background

Finally, we generate a realistic astrophysical GWB by estimating the individual CGW emissions of a population of SMBHBs. The injection pipeline uses a simulated catalog of SMBHBs providing their chirp mass, frequency, distance, inclination, polarization and sky location. Their contributions are added one by one using Eq. (22) in a 100-pulsar array with 10 years of observation and only white noise at a level of 10^{-7} s . The catalogs that are used are the same as in [42] to produce the simulations referred to as dataset injGWB03 and dataset injGWB03 + injCGW5. Both contain the same realization of a GWB signal but the second has a CGW emitting at 5 nHz,

resolved on top of the GWB (for more details, see section 2.1.2 of [42]). The datasets are analyzed using the non-Gaussian model constructed from a GMM of 2 Gaussians controlled by the parameters α and c . We estimate Bayes factors with hypermodeling [32] or, using the fact that one can turn off non-Gaussianities by setting $\alpha = 0$, with the Savage-Dickey density ratio as [33]

$$B_G^{NG} = \frac{\Pi_\alpha(\alpha = 0)}{p_\alpha(\alpha = 0)}, \quad (24)$$

where $\Pi_\alpha(\alpha = 0) = 1$ is the prior probability distribution for parameter α evaluated at 0, and $p_\alpha(\alpha = 0)$ is the posterior probability distribution for α evaluated at 0.

In Fig. 10 we show the recovered free spectrum, excess kurtosis and Bayes factor B_G^{NG} for the first 8 harmonics of $1/(10 \text{ yr})$. We display only the excess kurtosis because the higher-order moments have very broad posteriors that are not very informative. The free spectra between the two simulations differ from the injection of the CGW at 5 nHz. Most of the injected power by the CGW is present in the first two frequency bins and appears to slightly leak at higher frequencies.⁴ The most striking difference is when

⁴This is due to the finite time of observation that transforms sinusoids in time domain into sinc functions in frequency domain, causing leakage between neighboring frequencies [43].

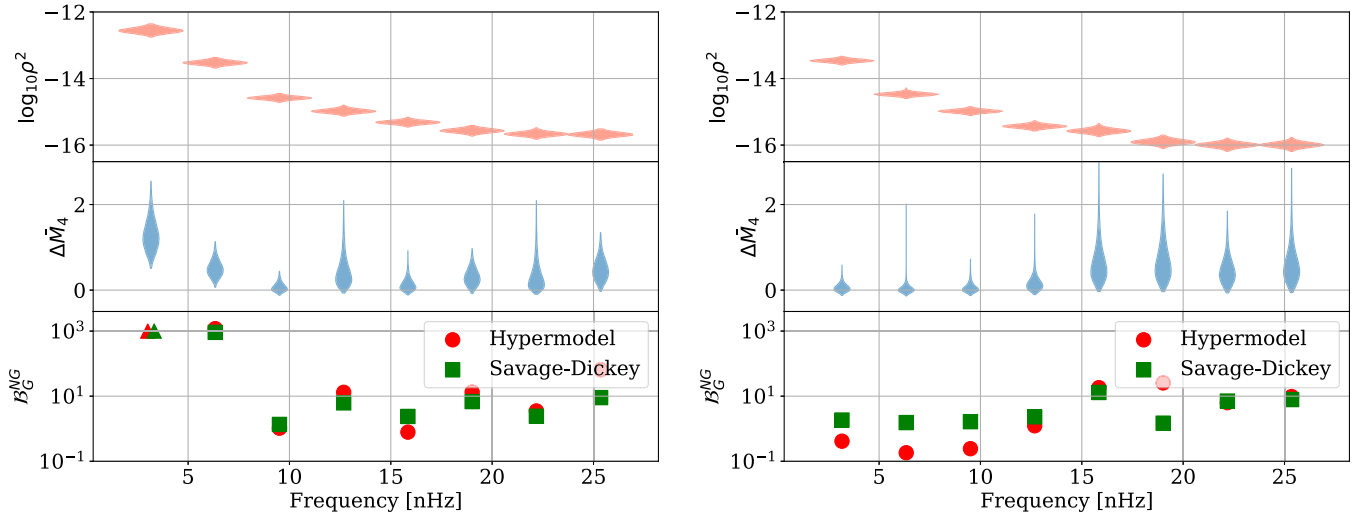


FIG. 10. Posterior distributions for the free spectrum ρ^2 , excess kurtosis $\Delta\bar{M}_4$ and Bayes factor $\mathcal{B}_G^{NG}$ as a function of frequency inferred from a simulated PTA with a realistic GWB signal (left) with a resolved CGW signal injected at 5 nHz (right) without any CGW signal injected. The Bayes factors obtained with the two methods are roughly consistent. The triangles for $\mathcal{B}_G^{NG}$ in the first bin of the left panel indicate that the recovered $\mathcal{B}_G^{NG}$ is far greater than 10^3 and difficult to precisely estimate numerically. Both method yielded very high values ranging between 10^6 and 10^{12} .

looking at the $\Delta\bar{M}_4$ posteriors where the 5-nHz CGW injects a significant excess kurtosis in the first two frequencies, and the Bayes factor favoring the presence of non-Gaussianities is high ($\mathcal{B}_G^{NG} > 10^3$), as expected from what is discussed in the previous section. The simulation without CGW does not show significant signs of non-Gaussianities, especially at low frequencies where $\Delta\bar{M}_4$ is well constrained around zero and $\mathcal{B}_G^{NG} < 1$. This can be explained by the higher number of contributing GW sources at low frequencies, producing a more isotropic and Gaussian GWB signal. Both $\Delta\bar{M}_4$ and $\mathcal{B}_G^{NG}$ seem to slightly increase for higher frequencies, as the number of contributing GW sources decreases, which might enhance the non-Gaussianities in the spectrum. The detection of mild non-Gaussian behavior in PTA might require a larger number of pulsars to increase our sensitivity to the statistic of the spectrum.

IV. CONCLUSION

In this work, we introduced a method to model non-Gaussian features in PTA data by using GMMs as priors for the Fourier coefficients of stochastic processes. This approach enables analytical marginalization of the likelihood, making it computationally efficient and directly compatible with existing PTA analysis pipelines. We validated the method through simulations involving both single pulsar noise and CRN processes, demonstrating that the model can robustly identify departures from Gaussianity using Bayesian model selection. The Bayes factor increases significantly in the presence of excess kurtosis, confirming the sensitivity of the approach to

higher-order statistical structure. Furthermore, we applied the method to a more astrophysically motivated scenario involving a superposition of a Gaussian GWB and a few CGW sources. In this case, we showed that non-Gaussian features become detectable when a small number of loud sources dominate the signal, and that these features are suppressed once the deterministic CGWs are correctly modeled and subtracted. When a realistic GWB is generated from the superposition of the individual CGW emissions of a simulated population of SMBHBs, the non-Gaussianities are harder to detect and might require very large arrays of pulsars to increase their detectability. This is illustrated in Fig. 12 of Appendix E where we show the Bayes factor $\mathcal{B}_G^{NG}$ for different sizes of array and varying excess kurtosis.

One of the key outcomes of this study is the demonstration that posterior estimates of higher-order central moments, particularly the fourth moment (kurtosis), can provide complementary information beyond the PSD, which is typically inferred under the assumption of Gaussianity. By constructing posterior distributions for the higher-order moments from the Fourier coefficients, the model can capture statistical features such as heavier tails or asymmetries in the signal that are not accounted for in standard analyses. This opens up the possibility of using spectral refitting techniques that explicitly include these moments as informative observables [23,25]. In particular, the excess moments may carry physical information about the population of GW sources, for example, indicating the presence of a few dominant binaries within an otherwise Gaussian background. Incorporating this information could help refine the characterization of the astrophysical origin

of the signal and improve constraints on the source population models, especially in the intermediate regime where the number of contributing sources is not large enough for the central limit theorem to fully apply.

Although we focused here on mixtures of two centered Gaussian distributions and considered spatially uncorrelated physical processes for simplicity and computational tractability, the framework can be readily extended. Future work could explore nonsymmetric and multimodal mixtures to capture a broader class of non-Gaussian behaviors, including skewness. Additionally, methods such as reversible-jump MCMC could be employed to dynamically infer the number and properties of the mixture components. The need to implement a version of this analysis accounting for the interfrequency spectral correlations due to the finite time of observation will be essential for the future, as highlighted in [18]. From a modeling perspective, the inclusion of higher-order interpulsar correlation structures (4-point correlator) through Gram-Charlier or Edgeworth expansions has not yet been implemented due to its increased computational cost. Finally, the integration of these techniques into full Bayesian PTA analysis pipelines

may provide a novel diagnostic to disentangle the statistical properties of astrophysical signals, potentially offering a route to discriminate between different GW source populations and to detect outliers or individual resolvable binaries within a stochastic background.

ACKNOWLEDGMENTS

We thank Irene Ferranti for providing the realistic population catalogs. We acknowledge support from the European Union’s H2020 ERC Advanced Grant “PINGU” (Grant No. 101142079) and from the PRIN 2022 grant “GRAPE” (Grant No. CUP H53D2300091 0006).

DATA AVAILABILITY

The data are not publicly available. The data are available from the authors upon reasonable request.

APPENDIX A: MOMENTS OF A GAUSSIAN MIXTURE

A mixture of 2 Gaussians can be expressed using Eq. (4) as

$$p(x) = (1 - \alpha)\varphi(x, 0, \sqrt{\Phi}) + \alpha\varphi(x, \mu_0, \sqrt{c\Phi}), \quad (\text{A1})$$

with mean $\int xp(x)dx = \alpha\mu_0$. Its central moments are given by

$$\begin{aligned} M_n &= \int (x - \alpha\mu_0)^n p(x) dx \\ &= (1 - \alpha) \int (x - \alpha\mu_0)^n \varphi(x, 0, \sqrt{\Phi}) dx + \alpha \int (x - \alpha\mu_0)^n \varphi(x, \mu_0, \sqrt{c\Phi}) dx \\ &= (1 - \alpha) \int \tilde{x}^n \varphi(\tilde{x}, -\alpha\mu_0, \sqrt{\Phi}) d\tilde{x} + \alpha \int \tilde{x}^n \varphi(\tilde{x}, (1 - \alpha)\mu_0, \sqrt{c\Phi}) d\tilde{x} \\ &= (1 - \alpha)\mathcal{M}_n(-\alpha\mu_0, \sqrt{\Phi}) + \alpha\mathcal{M}_n((1 - \alpha)\mu_0, \sqrt{c\Phi}), \end{aligned} \quad (\text{A2})$$

with $\tilde{x} = x - \alpha\mu_0$ and $\mathcal{M}_n(m, \sigma)$ the raw (noncentral) moments of a normal distribution with mean m and standard deviation σ that are well known and can be found in the literature. The first four M_n are given in Table I.

APPENDIX B: MARGINALIZED GAUSSIAN LIKELIHOOD

Consider a dataset δt containing uncorrelated Gaussian white noise characterized by a diagonal noise matrix N . We want to model a signal $s = Fa$ decomposed on a basis F (typically a discrete Fourier basis), characterized by a vector of coefficients a . The coefficients a follow a Gaussian distribution with covariance Φ . The likelihood for this model can be written as

$$\mathcal{L}(\delta t|a, \Phi, \mu) = \frac{\exp\{-\frac{1}{2}[\delta t - Fa]^\top N^{-1}[\delta t - Fa]\}}{|2\pi N|^{1/2}} \times \frac{\exp\{-\frac{1}{2}[a - \mu]^\top \Phi^{-1}[a - \mu]\}}{|2\pi \Phi|^{1/2}}, \quad (\text{B1})$$

where on the left we have the Gaussian likelihood for uncorrelated white noise and on the right, the prior probability distribution for coefficients a . Developing this expression, we get

$$\mathcal{L}(\delta t|a, \Phi, \mu) = \frac{\exp\{-\frac{1}{2}\delta t^\top N^{-1}\delta t - \frac{1}{2}a^\top [F^\top N^{-1}F + \Phi^{-1}]a + a^\top [F^\top N^{-1}\delta t + \Phi^{-1}\mu]\}}{|2\pi N|^{1/2}|2\pi \Phi|^{1/2}} \times \exp\left\{-\frac{1}{2}\mu^\top \Phi^{-1}\mu\right\}. \quad (\text{B2})$$

Defining $\Sigma = F^\top N^{-1} F + \Phi^{-1}$ and completing the squares in the numerator we get

$$\begin{aligned} & -\frac{1}{2} a^\top \Sigma a + a^\top [F^\top N^{-1} \delta t + \Phi^{-1} \mu] \\ & = -\frac{1}{2} [a - \Sigma^{-1} F^\top N^{-1} \delta t - \Sigma^{-1} \Phi^{-1} \mu]^\top \Sigma [a - \Sigma^{-1} F^\top N^{-1} \delta t - \Sigma^{-1} \Phi^{-1} \mu] \\ & \quad + \frac{1}{2} [\mu^\top \Phi^{-1} \Sigma^{-1} + \delta t^\top N^{-1} F \Sigma^{-1}] \Sigma [\Sigma^{-1} F^\top N^{-1} \delta t + \Sigma^{-1} \Phi^{-1} \mu]. \end{aligned} \quad (\text{B3})$$

Defining $\hat{a} = \Sigma^{-1} F^\top N^{-1} \delta t + \Sigma^{-1} \Phi^{-1} \mu$ and rearranging the terms we get

$$\begin{aligned} \mathcal{L}(\delta t | a, \Phi, \mu) &= \frac{\exp \left\{ -\frac{1}{2} \delta t^\top [N^{-1} - N^{-1} F \Sigma^{-1} F^\top N^{-1}] \delta t \right\}}{|2\pi N|^{1/2} |2\pi \Phi|^{1/2}} \times \exp \left\{ -\frac{1}{2} [a - \hat{a}]^\top \Sigma [a - \hat{a}] \right\} \\ & \times \exp \left\{ -\frac{1}{2} \mu^\top [\Phi^{-1} - \Phi^{-1} \Sigma^{-1} \Phi^{-1}] \mu + \mu^\top \Phi^{-1} \Sigma^{-1} F^\top N^{-1} \delta t \right\}, \end{aligned} \quad (\text{B4})$$

where we recognize the Woodbury matrix inversion $[N + F^\top \Phi F]^{-1} = N^{-1} - N^{-1} F \Sigma^{-1} F^\top N^{-1}$. Marginalizing over the coefficients a gives⁵

$$\begin{aligned} \mathcal{L}(\delta t | \Phi, \mu) &= \int_{-\infty}^{+\infty} da \mathcal{L}(\delta t | a, \Phi, \mu) \\ &= \frac{\exp \left\{ -\frac{1}{2} \delta t^\top [N + F^\top \Phi F]^{-1} \delta t \right\}}{|2\pi N|^{1/2} |\Phi|^{1/2} |\Sigma|^{1/2}} \\ & \times \exp \left\{ -\frac{1}{2} \mu^\top [\Phi^{-1} - \Phi^{-1} \Sigma^{-1} \Phi^{-1}] \mu \right. \\ & \quad \left. + \mu^\top \Phi^{-1} \Sigma^{-1} F^\top N^{-1} \delta t \right\}, \end{aligned} \quad (\text{B5})$$

where in the denominator we have the Woodbury determinant $|N + F^\top \Phi F| = |N| |\Phi| |\Sigma|$. Then, the left-hand side corresponds to a Gaussian likelihood centered on zero with covariance $C = N + F^\top \Phi F$. This expression generalizes the marginalized likelihood presented in [14] for noncentral Gaussian priors. For $\mu = 0$, they are identical.

When considering an array of N_p pulsars, the analytical expression of the likelihood is identical, except that now, the data δt and basis F are a concatenation of the timing residuals δt_I and basis F_I of each pulsar I . Additionally, when considering spatially correlated noise between pulsars the Gaussian prior on Fourier coefficients a must incorporate these correlations between a_I of different pulsars and the Φ is a nondiagonal matrix. As a consequence, the likelihood cannot be factorized in a product of individual pulsar likelihood and the covariance C has a block structure of size $N_p \times N_p$ where each block is

⁵The integral over a is a standard Gaussian integral, here producing a term equal to $|2\pi \Sigma^{-1}|^{1/2}$ in the numerator. The 2π factor simplifies with $|2\pi \Phi|$ in the denominator because Σ and Φ have the same rank.

$C_{IJ} = N_I \delta_{IJ} + F_I^\top \Phi_{IJ} F_J$. For HD correlated noise, $\Phi_{IJ} \propto \chi_{IJ}^{(2)}$. The inversion of the covariance $C^{-1} = [N + F^\top \Phi F]^{-1}$ then becomes computationally expensive.

APPENDIX C: GRAM-CHARLIER A EXPANSION FOR KURTOSIS

We can add the contributions of higher-order statistics by expanding the prior probability distribution $\Phi(\vec{a}_n)$ in terms of the Gram-Charlier A series (or similarly using the Edgeworth series [35]). Then, the non-Gaussian distribution is approximated from the Gaussian prior in Eq. (8) as

$$p(\vec{a}_n) \approx \Pi(\vec{a}_n) \times \left[1 + \frac{1}{4!} \sum_{IJKL} \kappa_{IJKL} H_{IJKL}(\vec{a}_n) \right], \quad (\text{C1})$$

where $\kappa_{IJKL} \propto \chi_{IJKL}^{(4)} h_n^4$ is the 4th cumulant (or excess kurtosis), h_n^4 the level of excess kurtosis at frequency n , and H_{IJKL} the multivariate Hermite polynomials expressed in terms of the $\vec{a}_n$.

This expression has to be normalized and is only valid for small excess kurtosis because this type of expansion behaves poorly for large h_n^4 and might even produce unphysical negative probability density values [44]. Still, it is possible to analytically marginalize the likelihood with respect to the coefficients $\vec{a}_n$ using Isserlis's theorem for different combinations of the elements of $\vec{a}_n$. We have

$$p(\vec{a}_n) \approx \mathcal{C}^{-1} \times \Pi(\vec{a}_n) \times \left[1 + \frac{1}{4!} \sum_{IJKL} \kappa_{IJKL} H_{IJKL}(\vec{a}_n) \right], \quad (\text{C2})$$

with $\mathcal{C}$ a normalization coefficient and

$$\Pi(\vec{a}_n) = \frac{\exp\{-\frac{1}{2} \sum_{IJ} a_{I,n} [\chi_{IJ} \Phi_n]^{-1} a_{J,n}\}}{\sqrt{\det 2\pi \chi \Phi_n}}, \quad (\text{C3})$$

where χ is the matrix of 2-point correlation coefficients $\chi_{IJ}^{(2)}$.

The multivariate 4th-order Hermite polynomials H_{IJKL} are given by the derivatives of the distribution $\Pi(\vec{a}_n)$ with respect to the components a_I of $\vec{a}_n$ as [45,46]

$$\begin{aligned} H_{IJKL}(\vec{a}) &= (-1)^4 \Pi(\vec{a})^{-1} \frac{\partial^4}{\partial a_I \partial a_J \partial a_K \partial a_L} \Pi(\vec{a}) \\ &= \Phi^{-4} \left[\left(\sum_i \chi_{iI}^{-1} a_i \right) \left(\sum_j \chi_{jJ}^{-1} a_j \right) \left(\sum_k \chi_{kK}^{-1} a_k \right) \left(\sum_l \chi_{lL}^{-1} a_l \right) \right] \\ &\quad - \Phi^{-3} \left[\chi_{IJ}^{-1} \left(\sum_k \chi_{kK}^{-1} a_k \right) \left(\sum_l \chi_{lL}^{-1} a_l \right) + \chi_{IK}^{-1} \left(\sum_j \chi_{jJ}^{-1} a_j \right) \left(\sum_l \chi_{lL}^{-1} a_l \right) + \chi_{IL}^{-1} \left(\sum_j \chi_{jJ}^{-1} a_j \right) \left(\sum_k \chi_{kK}^{-1} a_k \right) \right. \\ &\quad \left. - \chi_{JK}^{-1} \left(\sum_i \chi_{iI}^{-1} a_i \right) \left(\sum_l \chi_{lL}^{-1} a_l \right) + \chi_{JL}^{-1} \left(\sum_i \chi_{iI}^{-1} a_i \right) \left(\sum_k \chi_{kK}^{-1} a_k \right) + \chi_{KL}^{-1} \left(\sum_i \chi_{iI}^{-1} a_i \right) \left(\sum_j \chi_{jJ}^{-1} a_j \right) \right] \\ &\quad + \Phi^{-2} [\chi_{IJ}^{-1} \chi_{KL}^{-1} + \chi_{IK}^{-1} \chi_{JL}^{-1} + \chi_{IL}^{-1} \chi_{JK}^{-1}], \end{aligned} \quad (\text{C4})$$

where we have dropped the index n for better visibility.

Then, using the rearranged form of the Gaussian likelihood $\mathcal{L}(\delta t|a, \Phi, \mu = 0)$ in Eq. (B4), the likelihood $\mathcal{L}(\delta t|\Phi, \kappa)$ marginalized with respect to coefficients a is given by

$$\begin{aligned} \mathcal{L}(\delta t|\Phi, \kappa) &= \int_{-\infty}^{+\infty} da \mathcal{L}(\delta t|a, \Phi, \kappa, \mu = 0) \\ &= C^{-1} \times \mathcal{L}(\delta t|\Phi, \kappa = 0) \int da \frac{\exp\{-\frac{1}{2} [a - \hat{a}]^\top \Sigma [a - \hat{a}]\}}{|2\pi \Sigma^{-1}|} \left[1 + \frac{1}{4!} \sum_{IJKL} \kappa_{IJKL} H_{IJKL}(\vec{a}_n) \right] \\ &= C^{-1} \times \mathcal{L}(\delta t|\Phi, \kappa = 0) \left[1 + \frac{1}{24} \sum_{IJKL} \kappa_{IJKL} E_\Sigma \{H_{IJKL}\} \right], \end{aligned} \quad (\text{C5})$$

with $E_\Sigma \{H_{IJKL}\}$ denoting the expected value of the Hermite polynomials for $a \sim \mathcal{N}(\hat{a}, \Sigma)$ that can be obtained using Isserlis's theorem:

$$\begin{aligned} E_\Sigma \{H_{IJKL}\} &= \Phi^{-4} \sum_{ijkl} \chi_{iI}^{-1} \chi_{jJ}^{-1} \chi_{kK}^{-1} \chi_{lL}^{-1} E_\Sigma \{a_i a_j a_k a_l\} - \Phi^{-3} \left[\chi_{IJ}^{-1} \sum_{kl} \chi_{kK}^{-1} \chi_{lL}^{-1} E_\Sigma \{a_k a_l\} + \text{other permutations} \right] \\ &\quad + \Phi^{-2} [\chi_{IJ}^{-1} \chi_{KL}^{-1} + \text{other permutations}], \end{aligned} \quad (\text{C6})$$

where

$$\begin{aligned} E_\Sigma \{a_i a_j a_k a_l\} &= \hat{a}_i \hat{a}_j \hat{a}_k \hat{a}_l + \hat{a}_i \hat{a}_j \Sigma_{kl}^{-1} + \hat{a}_i \hat{a}_k \Sigma_{jl}^{-1} + \hat{a}_i \hat{a}_l \Sigma_{jk}^{-1} + \hat{a}_j \hat{a}_k \Sigma_{il}^{-1} + \hat{a}_j \hat{a}_l \Sigma_{ik}^{-1} + \hat{a}_k \hat{a}_l \Sigma_{ij}^{-1} \\ &\quad + \Sigma_{ij}^{-1} \Sigma_{kl}^{-1} + \Sigma_{il}^{-1} \Sigma_{jk}^{-1} + \Sigma_{ik}^{-1} \Sigma_{jl}^{-1} \end{aligned} \quad (\text{C7})$$

and

$$E_\Sigma \{a_k a_l\} = \Sigma_{kl}^{-1} + \hat{a}_k \hat{a}_l. \quad (\text{C8})$$

The calculation of each component Σ_{IJ}^{-1} requires knowing the inverse of matrix Σ , which is an expensive task. In general, this inverse is implicitly calculated using Cholesky linear system solver to compute $\Sigma^{-1} \delta t$, thus greatly optimizing the computation time. For now, we found no way of optimizing the calculation of Eq. (C6).

Finally, using Eq. (C6), we can calculate the normalization coefficient $\mathcal{C}$ for $a \sim \mathcal{N}(0, \chi^{(2)} \Phi_n)$, that is, following the distribution $\Pi(\vec{a}_n)$,

$$\begin{aligned}
C = 1 + \frac{1}{24} \Phi_n^{-2} \sum_{IJKL} \kappa_{IJKL} & \left[\sum_{ijkl} \chi_{il}^{-1} \chi_{jj}^{-1} \chi_{kk}^{-1} \chi_{ll}^{-1} (\chi_{ij} \chi_{kl} + \chi_{il} \chi_{kj} + \chi_{ik} \chi_{jl}) \right. \\
& - \chi_{IJ}^{-1} \left(\sum_{kl} \chi_{kk}^{-1} \chi_{ll}^{-1} \chi_{kl} \right) - \chi_{IK}^{-1} \left(\sum_{jl} \chi_{jj}^{-1} \chi_{ll}^{-1} \chi_{jl} \right) - \chi_{IL}^{-1} \left(\sum_{jk} \chi_{jj}^{-1} \chi_{kk}^{-1} \chi_{jk} \right) \\
& - \chi_{JK}^{-1} \left(\sum_{il} \chi_{il}^{-1} \chi_{ll}^{-1} \chi_{il} \right) - \chi_{JL}^{-1} \left(\sum_{ik} \chi_{il}^{-1} \chi_{kk}^{-1} \chi_{ik} \right) - \chi_{KL}^{-1} \left(\sum_{ij} \chi_{il}^{-1} \chi_{jj}^{-1} \chi_{ij} \right) \\
& \left. + \chi_{IJ}^{-1} \chi_{KL}^{-1} + \chi_{IK}^{-1} \chi_{JL}^{-1} + \chi_{IL}^{-1} \chi_{JK}^{-1} \right]. \tag{C9}
\end{aligned}$$

In the case of Edgeworth or Gram-Charlier A expansion, the marginalization with respect to coefficients a yields a quite unpractical expression of the likelihood. Especially for a large array of pulsars, the number of terms in the sum becomes extremely large. If such a method was to be used for PTA data analysis, sampling a from the unmarginalized likelihood might be more efficient and practical.

APPENDIX D: PP PLOT

To validate the simulations, we need to ensure that the data generation and the fitting tool are calibrated. Since the true parameter values θ_0 are known in each simulation, we compute their posterior quantiles $q(\theta_0)$ from the 1d marginalized posterior distributions of each parameter. According to [47], if the model and fitting are correct,

these quantiles should follow a Uniform(0, 1) distribution. The PP plot in Fig. 11 seems to show that we indeed recover a uniform distribution within error bars.

APPENDIX E: PERFORMANCE WITH NUMBER OF PULSARS N_p

In this appendix, we show the dependence of the recovered Bayes factor $\mathcal{B}_G^{NG}$ on the number of pulsars in the array N_p , justifying our choice of 100 pulsars. The comparison is obtained for the same realization of non-Gaussian noise, only varying the number of pulsars. The conditions of the simulation are the same as those presented in Sec. III for different N_p .

Since this plot is obtained for one realization, it does not show any variance as in Fig. 3.

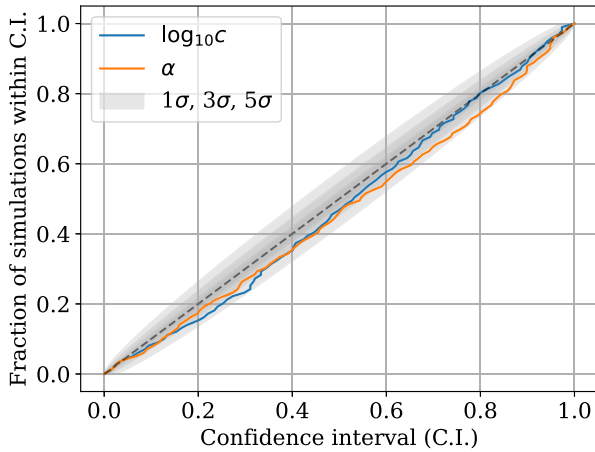


FIG. 11. PP plot comparing the fraction of simulations within confidence interval produced from 100 simulations for the non-Gaussianity parameters c and α . The gray shaded regions are the 1-, 3- and 5- σ errors.

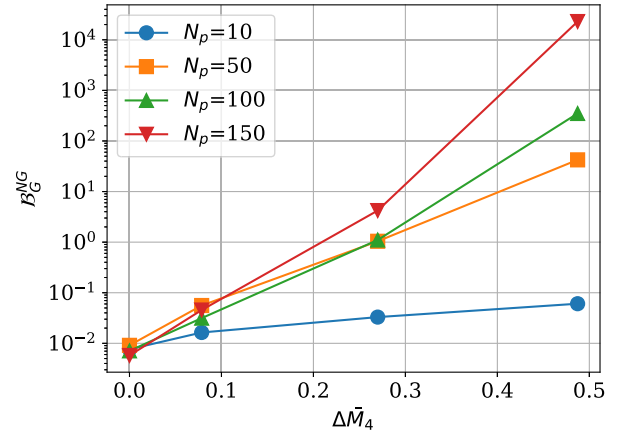


FIG. 12. Bayes factor $\mathcal{B}_G^{NG}$ between the non-Gaussian and Gaussian models as a function of the injected relative excess kurtosis $\Delta \bar{M}_4$ calculated using Eq. (7) for one realization of the noise, with varying number of pulsar N_p .

- [1] M. V. Sazhin, Opportunities for detecting ultralong gravitational waves, *Sov. Astron.* **22**, 36 (1978), <https://ui.adsabs.harvard.edu/abs/1978AZh....55...65S/abstract>.
- [2] S. Detweiler, Pulsar timing measurements and the search for gravitational waves, *Astrophys. J.* **234**, 1100 (1979).
- [3] R. S. Foster and D. C. Backer, Constructing a pulsar timing array, *Astrophys. J.* **361**, 300 (1990).
- [4] A. Sesana, A. Vecchio, and C. N. Colacino, The stochastic gravitational-wave background from massive black hole binary systems: Implications for observations with pulsar timing arrays, *Mon. Not. R. Astron. Soc.* **390**, 192 (2008).
- [5] A. Sesana, Insights into the astrophysics of supermassive black hole binaries from pulsar timing observations, *Classical Quantum Gravity* **30**, 224014 (2013).
- [6] C. Caprini and D. G. Figueroa, Cosmological backgrounds of gravitational waves, *Classical Quantum Gravity* **35**, 163001 (2018).
- [7] R. W. Hellings and G. S. Downs, Upper limits on the isotropic gravitational radiation background from pulsar timing analysis, *Astrophys. J. Lett.* **265**, L39 (1983).
- [8] B. Allen, Variance of the Hellings-Downs correlation, *Phys. Rev. D* **107**, 043018 (2023).
- [9] J. Antoniadis *et al.*, The second data release from the European pulsar timing array: III. Search for gravitational wave signals, *Astron. Astrophys.* **678**, A50 (2023).
- [10] G. Agazie *et al.*, The nanograv 15 yr data set: Evidence for a gravitational-wave background, *Astrophys. J. Lett.* **951**, L8 (2023).
- [11] D. J. Reardon *et al.*, Search for an isotropic gravitational-wave background with the parkes pulsar timing array, *Astrophys. J. Lett.* **951**, L6 (2023).
- [12] M. T. Miles *et al.*, The meerkat pulsar timing array: The first search for gravitational waves with the meerkat radio telescope, *Mon. Not. R. Astron. Soc.* **536**, 1489 (2024).
- [13] H. Xu *et al.*, Searching for the nano-hertz stochastic gravitational wave background with the Chinese pulsar timing array data release I, *Res. Astron. Astrophys.* **23**, 075024 (2023).
- [14] R. van Haasteren and M. Vallisneri, New advances in the gaussian-process approach to pulsar-timing data analysis, *Phys. Rev. D* **90**, 104012 (2014).
- [15] N. Laal *et al.*, Deep neural emulation of the supermassive black-hole binary population, *Astrophys. J.* **982**, 55 (2025).
- [16] L. Lentati, M. P. Hobson, and P. Alexander, Bayesian estimation of non-Gaussianity in pulsar timing analysis, *Mon. Not. R. Astron. Soc.* **444**, 3863 (2014).
- [17] U. von Toussaint, Bayesian inference in physics, *Rev. Mod. Phys.* **83**, 943 (2011).
- [18] M. Crisostomi *et al.*, Beyond diagonal approximations: Improved covariance modeling for pulsar timing array data analysis, [arXiv:2506.13866](https://arxiv.org/abs/2506.13866).
- [19] G. Agazie *et al.*, The nanograv 15 yr data set: Bayesian limits on gravitational waves from individual supermassive black hole binaries, *Astrophys. J. Lett.* **951**, L50 (2023).
- [20] N. Seto, Non-Gaussianity analysis of a gravitational wave background made by short-duration burst signals, *Phys. Rev. D* **80**, 043003 (2009).
- [21] B. Allen, D. Agarwal, J. D. Romano, and S. Valtolina, Source anisotropies and pulsar timing arrays, *Phys. Rev. D* **110**, 123507 (2024).
- [22] X. Xue, Z. Pan, and L. Dai, Non-Gaussian statistics of nanohertz stochastic gravitational waves, *Phys. Rev. D* **111**, 043022 (2025).
- [23] W. G. Lamb and S. R. Taylor, Spectral variance in a stochastic gravitational-wave background from a binary population, *Astrophys. J. Lett.* **971**, L10 (2024).
- [24] N. Laal, S. R. Taylor, R. van Haasteren, W. G. Lamb, and X. Siemens, Solving the PTA data analysis problem with a global Gibbs scheme, *Phys. Rev. D* **111**, 063067 (2025).
- [25] W. G. Lamb, S. R. Taylor, and R. van Haasteren, Rapid refitting techniques for Bayesian spectral characterization of the gravitational wave background using pulsar timing arrays, *Phys. Rev. D* **108**, 103019 (2023).
- [26] A. Mitridate *et al.*, Ptarcade, [arXiv:2306.16377](https://arxiv.org/abs/2306.16377).
- [27] H. Quelquejay Leclerc *et al.*, Practical approaches to analyzing PTA data: Cosmic strings with six pulsars, *Phys. Rev. D* **108**, 123527 (2023).
- [28] J. A. Ellis, M. Vallisneri, S. R. Taylor, and P. T. Baker, Enterprise: Enhanced numerical toolbox enabling a robust pulsar inference suite, Zenodo, [10.5281/zenodo.4059815](https://zenodo.org/record/4059815) (2020).
- [29] S. Valtolina and R. van Haasteren, Regularizing the pulsar timing array likelihood: A path towards Fourier space, *Phys. Rev. D* **112**, 043046 (2025).
- [30] EPTA Collaboration and InPTA Collaboration, The second data release from the European pulsar timing array—IV. Implications for massive black holes, dark matter, and the early universe, *Astron. Astrophys.* **685**, A94 (2024).
- [31] B. P. Carlin and S. Chib, Bayesian model choice via Markov Chain Monte Carlo methods, *J. R. Stat. Soc. Ser. B* **57**, 473 (1995).
- [32] S. Hee *et al.*, Bayesian model selection without evidences: Application to the dark energy equation-of-state, *Mon. Not. R. Astron. Soc.* **455**, 2461 (2015).
- [33] Z. Arzoumanian *et al.*, The nanograv 11 year data set: Pulsar-timing constraints on the stochastic gravitational-wave background, *Astrophys. J.* **859**, 47 (2018).
- [34] J. S. Hazboun *et al.*, The nanograv 11 yr data set: Evolution of gravitational-wave background statistics, *Astrophys. J.* **890**, 108 (2020).
- [35] É. Racine and C. Cutler, Gaussianity of LISA's confusion backgrounds, *Phys. Rev. D* **76**, 124033 (2007).
- [36] A. Sesana, A. Vecchio, and M. Volonteri, Gravitational waves from resolvable massive black hole binary systems and observations with Pulsar Timing Arrays, *Mon. Not. R. Astron. Soc.* **394**, 2255 (2009).
- [37] S. Babak *et al.*, European pulsar timing array limits on continuous gravitational waves from individual supermassive black hole binaries, *Mon. Not. R. Astron. Soc.* **455**, 1665 (2015).
- [38] J. Antoniadis *et al.*, The second data release from the European pulsar timing array: V. Search for continuous gravitational wave signals, *Astron. Astrophys.* **690**, A118 (2024).
- [39] G. Agazie *et al.*, The nanograv 15 yr data set: Bayesian limits on gravitational waves from individual supermassive black hole binaries, *Astrophys. J. Lett.* **951**, L50 (2023).

- [40] M. Falxa *et al.*, Searching for continuous gravitational waves in the second data release of the international pulsar timing array, *Mon. Not. R. Astron. Soc.* **521**, 5077 (2023).
- [41] L. Boyle and U.-L. Pen, Pulsar timing arrays as imaging gravitational wave telescopes: Angular resolution and source (de)confusion, *Phys. Rev. D* **86**, 124028 (2012).
- [42] I. Ferranti, G. Shaifullah, A. Chalumeau, and A. Sesana, Separating deterministic and stochastic gravitational wave signals in realistic pulsar timing array datasets, *Astron. Astrophys.* **694**, A194 (2025).
- [43] J. S. Hazboun, J. D. Romano, and T. L. Smith, Realistic sensitivity curves for pulsar timing arrays, *Phys. Rev. D* **100**, 104028 (2019).
- [44] C. R. Contaldi, P. G. Ferreira, J. Magueijo, and K. M. Górski, A Bayesian estimate of the skewness of the cosmic microwave background, *Astrophys. J.* **534**, 25 (2000).
- [45] F.-S. Kitaura, Non-Gaussian gravitational clustering field statistics, *Mon. Not. R. Astron. Soc.* **420**, 2737 (2012).
- [46] S. Berkowitz and F. J. Garner, The calculation of multidimensional Hermite polynomials and gram-charlier coefficients, *Math. Comput.* **24**, 537 (1970).
- [47] S. Cook *et al.*, Validation of software for Bayesian models using posterior quantiles, *J. Comput. Graph. Stat.* **15**, 675 (2006).