Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **537,** 3470–3479 (2025) https://doi.org/10.1093/mnras/staf190
Advance Access publication 2025 February 03

Ensemble noise properties of the European Pulsar Timing Array

Boris Goncharov [1,2]★ and Shubhit Sardana [1,3]

[1]*Max Planck Institute for Gravitational Physics (Albert Einstein Institute), Callinstrasse 38, D-30167 Hannover, Germany*
[2]*Leibniz Universität Hannover, D-30167 Hannover, Germany*
[3]*Department of Physics, IISER Bhopal, Bhauri Bypass Road, Bhopal, 462066, India*

Accepted 2025 January 28. Received 2024 December 31; in original form 2024 September 11

ABSTRACT

The null hypothesis in Pulsar Timing Array (PTA) analyses includes assumptions about ensemble properties of pulsar time-correlated noise. These properties are encoded in prior probabilities for the amplitude and the spectral index of the power-law power spectral density of temporal correlations of the noise. Because multiple realizations of time-correlated noise processes are found in pulsars, these ensemble noise properties could and should be modelled in the full-PTA observations by parametrizing the respective prior distributions using the so-called hyperparameters. This approach is known as the hierarchical Bayesian inference. In this work, we introduce a new procedure for numerical marginalization over hyperparameters. The procedure may be used in searches for nanohertz gravitational waves and other PTA analyses to resolve prior misspecification at negligible computational cost. Furthermore, we infer the distribution of amplitudes and spectral indices of the power spectral density of spin noise and dispersion measure variation noise based on the observation of 25 millisecond pulsars by the European Pulsar Timing Array. Our results may be used for the simulation of realistic noise in PTAs.

Key words: gravitational waves – methods: data analysis – pulsars: general.

1 INTRODUCTION

Pulsar Timing Arrays (PTAs; Foster & Backer 1990) are experiments that monitor pulse arrival times from Galactic millisecond pulsars with a primary goal of detecting nanohertz-frequency gravitational waves (Sazhin 1978; Detweiler 1979). The most promising source of such gravitational waves is the stochastic superposition of inspiralling supermassive binary black holes in the nearby Universe (Rosado, Sesana & Gair 2015). Thanks to space–time metric perturbations from gravitational waves, pulse arrival times experience delays and advances (henceforth, delays). The power spectral density (PSD) of delays $P(f)$ induced by stochastic gravitational waves manifests temporal correlations and corresponds to the background's characteristic strain spectrum $h_c(f)$. The Fourier frequency of timing delays f is exactly the gravitational wave frequency. For the isotropic stochastic gravitational wave background (GWB) from circular binaries where the inspiral is driven by gravitational wave emission alone, $h_c \propto f^{-2/3}$ corresponding to $P(f) \propto f^{-13/3}$, highlighting signal prominence towards lower frequencies or, equivalently, longer observational time-scales (Phinney 2001; Renzini et al. 2022). At low frequencies, PTAs are limited by time-correlated 'red' noise, which is also modelled using a power law. The two basic sources of red noise in PTA data are the dispersion measure (DM) variation noise (Keith et al. 2013) and the spin noise (SN; Shannon & Cordes 2010). DM noise features an additional dependence $P(f) \propto \nu^{-2}$, where ν is a radio frequency. It belongs to a broader class of noise processes called 'chromatic', referring to the dependence of the signal amplitude on radio frequency. Achromatic red noise that is independent of ν is also called SN because it is associated with irregularities in pulsar rotation. A decisive contribution of the stochastic background to PTA data is ultimately determined through the covariance of the signal across pulsar pairs. The covariance follows the Hellings & Downs (1983) function of pulsar angular separation.

Contemporary PTA analyses are performed using Bayesian inference, where a uniform prior probability is assumed for pulsar-specific red noise parameters, namely, the log-10 amplitude, $\lg A$, and the spectral index, γ, of noise PSD. Often the term 'prior' is used to refer to our knowledge of a parameter before evidence is taken into account. However, more precisely, priors are the models of how likely a parameter with a certain value is to be found in a given data realization. A prior choice is sufficient as long as it is consistent with an observation. Uniform priors on PTA noise parameters are designed to be sufficient for single-pulsar noise analyses. However, all pulsars represent different realizations of data with respect to pulsar noise parameters.[1] So, for full-PTA analyses with multiple pulsars, a mismatch between our prior on $(\lg A, \gamma)$ and the observed distribution of these parameters may accumulate across pulsars and lead to biased inference. This case is known as *prior misspecification*.

Implicitly, noise prior misspecification is shown by Hazboun et al. (2020) to bias measurements of the strain amplitude of the GWB in PTAs. The authors have mitigated the bias by allowing the data to choose whether red noise is present or absent in a pulsar. Furthermore, in simulations of Goncharov et al. (2021b) and Zic

★ E-mail: boris.goncharov@aei.mpg.de

[1]While remaining a single realization of data with respect to parameters governing $h_c(f)$ of GWB.

et al. (2022), pulsar SN with a wide range of $(\lg A, \gamma)$ modelled with the standard uniform priors, not representative of ensemble noise properties, yields evidence for a stochastic signal with the same power-law PSD across pulsars (which is not present in the simulated data). Such a common-spectrum process (CP), under the assumption of uniform noise priors, has been reported in real PTA data as a possible precursor to Hellings–Downs correlations of the GWB (NANOGrav Collaboration 2020; Chen et al. 2021; Goncharov et al. 2021b).

Goncharov et al. (2022) regularized incorrect pulsar noise priors by allowing noise amplitudes of the CP of the putative GWB to vary across pulsars. By showing that this variance is consistent with zero, the authors confirmed the consistency of the signal with the GWB. Although the model of Goncharov et al. (2022) provides the working solution and has clear use cases, the ultimate solution to the misspecification of noise priors is to parametrize priors for all relevant noise parameters. Both the statement of the problem and the most general solution are clearly outlined in van Haasteren (2024). The distribution of pulsar noise parameters is inferred simultaneously with a search for Hellings–Downs correlations and other red processes with pulsar-to-pulsar covariance. The technique is shown to mitigate systematic errors introduced by incorrect noise priors.

In this study, we introduce two new methods for modelling ensemble pulsar noise properties in PTA data. However, unlike van Haasteren (2024), (1) one of our methods can only be used to infer ensemble pulsar noise properties and (2) another method can only be used to remove a systematic error from incorrect pulsar noise priors without gaining access to ensemble noise properties. Next, we perform inference of ensemble noise properties of pulsars from the second data release (DR2; EPTA Collaboration 2023) of the European Pulsar Timing Array (EPTA; Kramer & Champion 2013; Desvignes et al. 2016).

The rest of the paper is organized as follows. In Section 2, we outline the data analysis methodology and an overview of sources of noise in PTAs. Our two new methods of modelling ensemble pulsar noise properties are presented in Sections 2.2.1 and 2.2.2. In Section 3, we report on our main results, where each section corresponds to a different model of a distribution of noise parameters $(\lg A, \gamma)$. In Section 4, we test the robustness of our models to circular analysis. Finally, in Section 5, we draw conclusions.

2 METHODOLOGY

In this study, we analyse the second data release (EPTA Collaboration 2023) of the EPTA. In particular, we focus our attention on the 'DR2full', which we will thus refer to as EPTA DR2 or data unless specified otherwise. The data are based on pulse arrival times from a set of 25 millisecond radio pulsars observed over time spans varying between 14 and 25 yr. The data also include pulsar timing models obtained with the least-squared fitting of pulse arrival times to the models (Hobbs, Edwards & Manchester 2006). Timing models describe how pulse arrival times are affected by deterministic properties of individual pulsars such as pulsar spin frequency and derivatives, pulsar position and proper motion in the sky, DM and derivatives, and binary orbital parameters (if applicable). Contributions of other signals and noise processes to data are referred to as 'residuals', implying that they yield a difference between pulse arrival times predicted by the timing model and measured arrival times. Pulse arrival times are referenced to the position of the Solar system barycentre, which is defined based on the DE440 ephemeris (Park et al. 2021).

2.1 Standard PTA data analysis methodology

Contributions to the PTA data beyond the timing model are often determined using Bayesian inference as described below. The likelihood of data $\boldsymbol{\delta t}$ (a vector of the measured pulse times of arrival, ToA) is a Gaussian distribution, which is multivariate with respect to a number of observations,

$$\mathcal{L}(\boldsymbol{\delta t}|\boldsymbol{\theta}) = \frac{\exp\left(-\frac{1}{2}(\boldsymbol{\delta t}-\boldsymbol{\mu})^T\mathbf{C}^{-1}(\boldsymbol{\delta t}-\boldsymbol{\mu})\right)}{\sqrt{\det(2\pi\mathbf{C})}}, \tag{1}$$

where $\boldsymbol{\theta}$ is a vector of parameters of models that describe the data.In other words, $\mathcal{L}(\boldsymbol{\delta t}|\boldsymbol{\theta})$ is the time-domain likelihood. The model prediction for pulse arrival times as a function of time is $\boldsymbol{\mu}(\boldsymbol{\theta})$, and $\mathbf{C}(\boldsymbol{\theta})$ is a covariance matrix that describes stochastic processes. Diagonal elements of $\mathbf{C}$, $\boldsymbol{\sigma}^2$, correspond to temporally uncorrelated 'white' noise. It is modelled as $\boldsymbol{\sigma}^2(\boldsymbol{e}_\mathrm{f}, \boldsymbol{e}_\mathrm{q}) = \boldsymbol{\sigma}^2_\mathrm{ToA}\boldsymbol{e}^2_\mathrm{f} + \boldsymbol{e}^2_\mathrm{q}$, where $\boldsymbol{\sigma}_\mathrm{ToA}$ is the measurement uncertainty on the data provided by the initial timing model fit and $(\boldsymbol{e}_\mathrm{f}, \boldsymbol{e}_\mathrm{q})$ are white noise model parameters, which are chosen to be separate for each telescope backend-receiver combination. Off-diagonal elements of $\mathbf{C}$ describe temporal correlations. For illustrative purposes, it is convenient to represent temporal correlations as components of $\boldsymbol{\mu}$ together with the timing model contributions using a reduced-rank approximation (Lentati et al. 2013; van Haasteren & Vallisneri 2014):

$$\boldsymbol{\mu} = \mathbf{F}\boldsymbol{a} + \mathbf{M}\boldsymbol{\epsilon} + \mathbf{U}\boldsymbol{j} + \boldsymbol{d}(t). \tag{2}$$

Here, $\mathbf{F}$ are the Fourier sine and cosine basis functions and $\boldsymbol{a}$ are Fourier amplitudes of red processes. Timing model contributions are similarly modelled via the design matrix $\mathbf{M}$ and the coefficients $\boldsymbol{\epsilon}$ represent timing model parameters. Whereas a type of white noise corresponding to timing delays $\boldsymbol{j}$ in units (s) that are the same for all ToAs obtained in the same observing epoch ('jitter' noise) are modelled using the basis $\mathbf{U}$. Vector $\boldsymbol{d}(t)$ corresponds to other deterministic signals as a function of pulsar observation time. Deterministic signals include exponential dips associated with pulse shape changes (Goncharov et al. 2021a). Let us denote $\boldsymbol{b} \equiv (\boldsymbol{a}, \boldsymbol{\epsilon}, \boldsymbol{j}) \in \boldsymbol{\theta}$. Bayes theorem is used to obtain a posterior distribution of model parameters $\mathcal{P}(\boldsymbol{\theta}|\boldsymbol{\delta t})$ from the likelihood and the prior $\pi(\boldsymbol{\theta})$:

$$\mathcal{P}(\boldsymbol{\theta}|\boldsymbol{\delta t}) = \frac{\mathcal{L}(\boldsymbol{\delta t}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\mathcal{Z}}, \tag{3}$$

where $\mathcal{Z}$ is the integral of the numerator over $\boldsymbol{\theta}$, it is termed Bayesian evidence.

Hierarchical inference – the approach of parametrizing prior distributions – is already a part of the standard PTA data analysis machinery. It is employed to reduce the parameter space and to make it physically motivated. Instead of Fourier amplitudes $\boldsymbol{a}$ (two per frequency, one per sine term of $\mathbf{F}$ and one per cosine term), it is more convenient to measure power-law parameters (A, γ) of the PSD of red processes:

$$P(f|A,\gamma) = \frac{A}{12\pi^2}\left(\frac{f}{f_\mathrm{yr}}\right)^{-\gamma}. \tag{4}$$

Thus, the prior on $\boldsymbol{a}$ joins single-pulsar likelihoods into a joint posterior:

$$\mathcal{P}(\boldsymbol{\theta}', \boldsymbol{\theta}''|\boldsymbol{\delta t}) = \mathcal{Z}^{-1}\int \mathcal{L}(\boldsymbol{\delta t}|\boldsymbol{b}, \boldsymbol{\theta}')\pi(\boldsymbol{b}|\boldsymbol{\theta}'')\pi(\boldsymbol{\theta}'')\pi(\boldsymbol{\theta}')\,\mathrm{d}\boldsymbol{b}, \tag{5}$$

where $\boldsymbol{\theta}'$ are parameters that are not included in the reduced-rank approximation in equation (2), such that $\boldsymbol{\theta} = (\boldsymbol{\theta}', \boldsymbol{b})$. Values $\boldsymbol{\theta}''$ are referred to as hyperparameters that include (A, γ). The integral in equation (5) is carried out analytically (Lentati et al.

Table 1. Hierarchical Bayesian analysis methods for PTAs. Parameters $\boldsymbol{\theta}$ of a model describing PTA data include parameters of noise and signals such as the GWB. Hyperparameters $\boldsymbol{\Lambda}$ determine the distribution of noise parameters in nature, $\pi(\boldsymbol{\theta}|\boldsymbol{\Lambda})$.

Approach	Deliverable	Pulsar-wise parallelization	Modelling common signals in all pulsars	Primary scope
No modelling of ensemble noise properties, $\boldsymbol{\Lambda} = \varnothing$	$\mathcal{P}(\boldsymbol{\theta}\|\boldsymbol{\delta t})\Big\|_{\boldsymbol{\Lambda}=\varnothing}$	–	Yes	Measurement of model parameters $\boldsymbol{\theta}$, subject to potential biases due to prior misspecification
Goncharov et al. (2022), $\boldsymbol{\Lambda}_{\rm CP} = (\mu_{\lg A_{\rm CP}}, \sigma_{\lg A_{\rm CP}})$	$\mathcal{P}(\boldsymbol{\Lambda}_{\rm CP}\|\boldsymbol{\delta t})$	Yes	Via a prior in individual pulsars	Modelling of the CP (pulsar-to-pulsar autocorrelation of the GWB)
van Haasteren (2024)	$\mathcal{P}(\boldsymbol{\Lambda}, \boldsymbol{\theta}\|\boldsymbol{\delta t})$	No	Yes	Inference of ensemble noise properties, resolving model misspecification *on the fly*
Single-pulsar prior reweighting (equation 8)	$\mathcal{P}(\boldsymbol{\Lambda}\|\boldsymbol{\delta t})$	Yes	Via a prior in individual pulsars	Inference of ensemble noise properties, resolving model misspecification *in the subsequent analysis*
Marginalization over $\boldsymbol{\Lambda}$ (equation 9)	$\mathcal{P}(\boldsymbol{\theta}\|\boldsymbol{\delta t})$	No	Yes	Resolving model misspecification *on the fly*

2013; van Haasteren & Vallisneri 2014). The integral represents *marginalization* over $\boldsymbol{b}$, and it is ultimately performed with ($\mathbf{F}$, $\mathbf{M}$, $\mathbf{U}$) and $\boldsymbol{b}$ modelled in $\mathbf{C}$. For more details, please refer to NANOGrav Collaboration (2016).

Full-PTA analysis with stochastic processes that are correlated between pulsars, such as the GWB, is performed as follows. To model the PSD of each pulsar-specific noise term, which may vary from pulsar to pulsar, we have a vector of power-law parameters $(\boldsymbol{A}, \boldsymbol{\gamma})$, with a pair of noise (hyper)parameters (A, γ) per pulsar. We also have a pair of $(A_{\rm c}, \gamma_{\rm c})$ to model PSD of each 'common' stochastic signal, i.e. applicable to all pulsars. Using notations from equation (2), an '$\boldsymbol{a}$' component of the term $\pi(\boldsymbol{b}|\boldsymbol{\theta}'')$ for arbitrary pulsars (a, b) and frequencies (i, j) becomes

$$\pi_{(a,i),(b,j)}(\boldsymbol{a}|A_a, \gamma_a, A_{\rm c}, \gamma_{\rm c}) = P_{ai}(f_i|A_a, \gamma_a)\delta_{ab}\delta_{ij} + \Gamma_{ab}P_i(f_i|A_{\rm c}, \gamma_{\rm c})\delta_{ij}, \quad (6)$$

where Γ_{ab} is the overlap reduction function such as the Hellings–Downs function. It encodes pulsar-to-pulsar correlations.

2.2 Hierarchical inference of ensemble noise properties

As part of the standard PTA analysis routine described in Section 2.1, priors $\pi(\lg A, \gamma)$ are assumed to be uniform distributions, $\mathcal{U}$. For example, $\pi(\lg A) = \mathcal{U}(-20, -12)$, $\pi(\gamma) = \mathcal{U}(0, 7)$. As we pointed out in Section 1, previous studies suggest that this model may be incorrect. The solution is to propose alternative parametrized models $\pi(\boldsymbol{\theta}|\boldsymbol{\Lambda})$, which now depend on hyperparameters $\boldsymbol{\Lambda}$. Therefore, generalizing $(\lg A, \gamma) \in \boldsymbol{\theta}$, the posterior becomes

$$\mathcal{P}(\boldsymbol{\theta}, \boldsymbol{\Lambda}|\boldsymbol{\delta t}) = \mathcal{Z}^{-1}\mathcal{L}(\boldsymbol{\delta t}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{\Lambda})\pi(\boldsymbol{\Lambda}). \quad (7)$$

The approach of van Haasteren (2024) is to directly evaluate $\mathcal{P}(\boldsymbol{\theta}, \boldsymbol{\Lambda}|\boldsymbol{\delta t})$. In this work, we propose computationally efficient marginalization of this posterior over (a) $\boldsymbol{\theta}$, or (b) $\boldsymbol{\Lambda}$, as described below. A summary of the methods is provided in Table 1.

2.2.1 *Marginalized posterior of hyperparameters*

To obtain $\mathcal{P}(\boldsymbol{\Lambda})$ marginalized over $\boldsymbol{\theta}$, we propose a solution based on prior reweighting, the application of importance sampling to priors. The calculation is done in two steps, both of which manifest one global fit to data. First, one obtains $\mathcal{P}(\boldsymbol{\theta}', \boldsymbol{\theta}''|\boldsymbol{\delta t})$ for every pulsar assuming a fixed prior on noise parameters $\pi(\boldsymbol{\theta}_i^k|\varnothing)$. It is referred to as the proposal distribution. Here, $\varnothing$ means that hyperparameters are fixed as in the standard EPTA analysis (EPTA Collaboration & InPTA Collaboration 2023a). Secondly, one uses the resulting posterior samples and evidence values $\mathcal{Z}_{\varnothing,i}$ to construct a likelihood marginalized over all parameters except $\boldsymbol{\Lambda}$, $\mathcal{L}(\boldsymbol{\delta t}|\boldsymbol{\Lambda})$. It is shown in Thrane & Talbot (2019) that the likelihood from equation (7) can be written in the following form:

$$\mathcal{L}(\boldsymbol{\delta t}|\boldsymbol{\Lambda}) = \prod_i^{N_{\rm psrs}} \frac{\mathcal{Z}_{\varnothing,i}(\boldsymbol{\delta t}_i)}{n_i} \sum_k^{n_i} \frac{\pi(\boldsymbol{\theta}_i^k|\boldsymbol{\Lambda})}{\pi(\boldsymbol{\theta}_i^k|\varnothing)}. \quad (8)$$

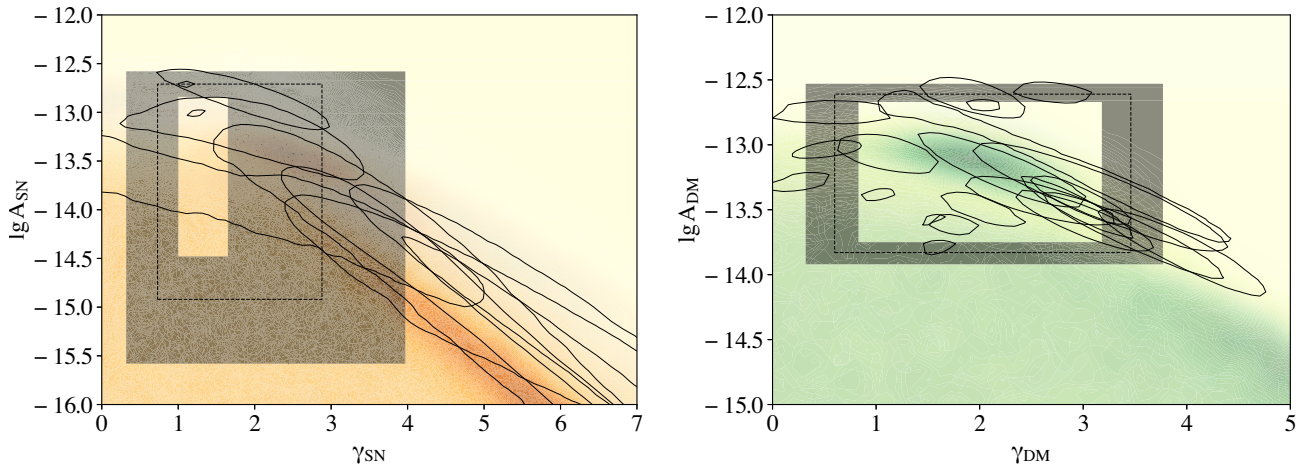
In the above equation, $N_{\rm psrs}$ is the number of pulsars in a PTA, n_i is the number of posterior samples obtained for a given pulsar, and θ_i^k is the kth posterior sample for ith pulsar. Ensemble noise properties of the EPTA presented in this work are obtained based on equation (8). The limitation of this approach is that it is based on expressing the total PTA likelihood as the product of single-pulsar likelihoods. Therefore, the approach is blind to Hellings–Downs correlations or other inter-pulsar (angular) correlations in the timing data. Whereas the presence of temporal correlations with the same $(\lg A, \gamma)$ in all pulsars has to be imposed via a prior in individual pulsars. To model CP (NANOGrav Collaboration 2020; Chen et al. 2021; Goncharov et al. 2021b) associated with the GWB (EPTA Collaboration & InPTA Collaboration 2023b; NANOGrav Collaboration 2023a; Reardon et al. 2023) in our hierarchical likelihood defined by equation (8), we impose an additional spin-noise-like term with $\gamma = 13/3$ in individual pulsar likelihoods. The methodology and implications of this choice are discussed in Appendix A. Improved methods for separating the effect of GWB from pulsar-intrinsic work may be explored in future work.

2.2.2 *Marginalization over hyperparameters*

The posterior from equation (7) marginalized over $\boldsymbol{\Lambda}$ can be written in the following form:

$$\mathcal{P}(\boldsymbol{\theta}|\boldsymbol{\delta t}) = \frac{\mathcal{L}(\boldsymbol{\delta t}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\varnothing)}{\mathcal{Z}} \times \frac{1}{n_{\rm p}} \sum_k^{n_{\rm p}} \frac{\pi(\boldsymbol{\theta}|\boldsymbol{\Lambda}_k)}{\pi(\boldsymbol{\theta}|\varnothing)}, \quad (9)$$

where $\boldsymbol{\Lambda}_k$ are $n_{\rm p}$ samples from the prior. The reader may notice that the resulting posterior in equation (9) is the product ($\times$) of



(a) Spin noise (SN)

(b) Dispersion measure (DM) variation noise

Figure 1. The inferred form of the *uniform* distribution of pulsar red noise parameters. Solid contours correspond to 1σ posterior levels of pulsar noise parameters when analysing single-pulsar data, provided that the noise term is resolved in a pulsar based on model selection from EPTA Collaboration & InPTA Collaboration (2023a) (where inference is performed without the CP term, therefore our solid contours for SN may be different to those in the reference). Background gradient corresponds to the posterior density of pulsars where noise terms were not resolved. Our hierarchical analysis yields hyperparameters that determine how pulsar noise parameters are distributed. The shaded rectangular areas correspond to 1σ credible levels for hyperparameters (**min**, **max**) that govern limits of the uniform distribution based on our hierarchical analysis. Hyperparameter values corresponding to the shaded area and the dashed line represent a row $\mathcal{U}(\boldsymbol{\theta}|\mathbf{min}, \mathbf{max})$ in Table 2.

the standard PTA posterior and the weight factor. A derivation of equation (9) is provided in Appendix. Unlike equation (8), equation (9) allows to simultaneously model Hellings–Downs and other inter-pulsar correlations in the data while accounting for the uncertainty in pulsar noise priors. It is used more extensively in the companion paper (Goncharov et al. 2024), where the likelihood is multiplied by the weight factor from equation (9), and the rest of the analysis is performed in a standard way. Alternatively, the posterior $\mathcal{P}(\boldsymbol{\theta}|\boldsymbol{\delta t})$ can be reweighted into a $\boldsymbol{\Lambda}$-marginalized posterior using rejection sampling.

3 RESULTS

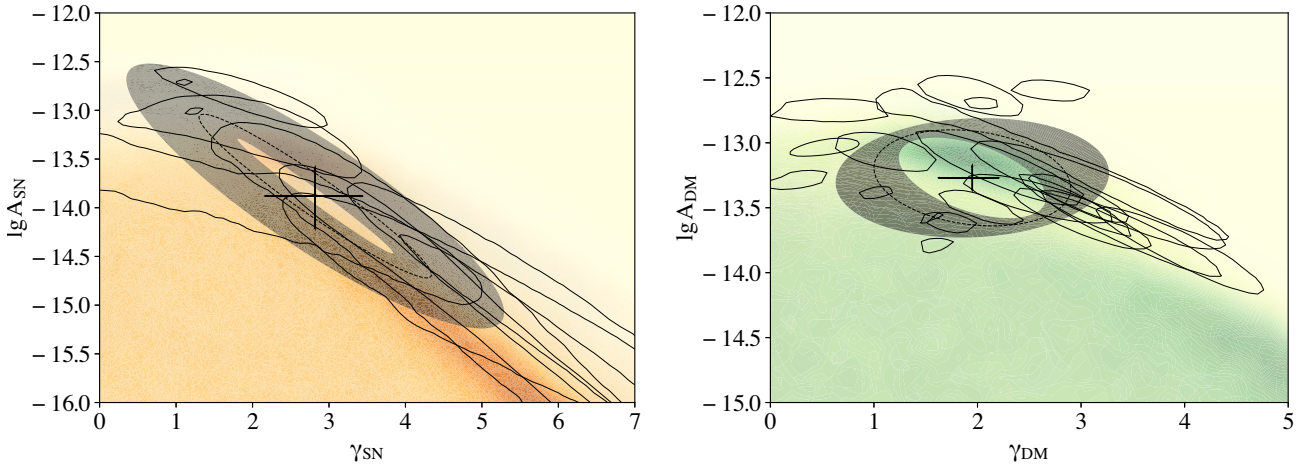
In this work, we infer distributions for parameters governing pulsar SN, $\pi(\lg A_{\rm SN}, \gamma_{\rm SN})$, and DM noise, $\pi(\lg A_{\rm DM}, \gamma_{\rm DM})$, for EPTA DR2. EPTA DR2 contains CP associated with the GWB, such that $\gamma_{\rm CP}$ is consistent with 13/3 (Goncharov et al. 2024). If we do not model this term, it may appear as a 'quasi-common' red noise (as per the terminology of Goncharov et al. 2022). In other words, as a separate cluster in the distribution of SN parameters. Because we are interested in knowing the distribution of *pulsar-intrinsic* parameters, we include an additional red noise term with $\gamma_{\rm c} = 13/3$ to marginalize over a contribution of the common signal. This is done during the analysis of single-pulsar noise corresponding to the first step in Section 2.2. It corresponds to modelling common signal in our data via a prior as per Table 1. Because we are agnostic about $\lg A_{\rm CP}$, the approach imposes a suboptimal uncertainty for measuring $(\lg A, \gamma)$ in pulsars where γ is consistent with $13/3 \approx 4$. Furthermore, to separate contributions of other noise processes such as chromatic noise (including scattering variations) and band- or system-dependent noise, we adopt single-pulsar noise models and optimal numbers of Fourier frequency bins from the EPTA DR2 noise analysis (EPTA Collaboration & InPTA Collaboration 2023a).

3.1 Uniform distribution of noise parameters

We start with the case of *uniform distributions* for noise parameters and we measure hyperparameters that govern uniform prior range. For DM noise, these are $\min(\lg A_{\rm DM})$, $\max(\lg A_{\rm DM})$, $\min(\gamma_{\rm DM})$, and $\max(\gamma_{\rm DM})$. We apply the same principle to SN parameters. Pulsar-specific measurements of $(\lg A, \gamma)$ and the inferred boundaries of the uniform distribution for SN and DM noise parameters – with measurement uncertainties – are shown in Fig. 1. The first important observation is that the data rule out, with high credibility, the standard values of uniform prior boundaries. The data suggest that pulsar noise parameters are distributed in a more narrow range. Posterior density $\mathcal{P}(\lg A, \gamma)$ from pulsars that do not have evidence for the respective noise term according to the EPTA Collaboration & InPTA Collaboration (2023a) is shown in colour. The data from these pulsars have also contributed to the measurement of hyperparameters.

3.2 Normal distribution of noise parameters

The uniform prior model may remain a good approximation, but it is not the most natural choice. Therefore, we further discuss the possibility that pulsar noise parameters are *normally distributed*. Thus, for SN, we introduce $\pi(\lg A_{\rm SN}, \gamma_{\rm SN}|\boldsymbol{\Lambda}_{\mathcal{N}})$ and $\boldsymbol{\Lambda}_{\mathcal{N}} = (\mu_{\lg A_{\rm SN}}, \sigma_{\lg A_{\rm SN}}, \mu_{\gamma_{\rm SN}}, \sigma_{\gamma_{\rm SN}}, \rho_{\rm SN})$. We apply the same principle to DM noise. Here, μ and σ correspond to the mean and the standard deviation of the normal distribution of either $\lg A$ or γ, as subscripts indicate. Parameter $\rho \in [-1, 1]$ is the correlation coefficient between $\lg A$ and γ. It is important to note that we obtained our posterior samples from individual pulsars on step one from Section 2.2 based on uniform priors. So, for step two, there are no posterior samples to recycle from outside of these boundaries. Therefore, we truncate our normal distribution model and apply the same boundaries. The inferred boundaries of the truncated normal distribution for SN and DM noise parameters are shown in Fig. 2. The shaded area corresponds



(a) Spin noise (SN)

(b) Dispersion measure (DM) variation noise

Figure 2. The inferred form of the *normal* distribution of pulsar red noise parameters. The background gradient and the solid contours represent measurements of noise parameters in individual pulsars, as explained in Fig. 1. Our hierarchical analysis yields hyperparameters that determine how pulsar noise parameters are distributed. Shaded ellipses correspond to 1σ credible levels for hyperparameters $(\boldsymbol{\sigma}, \rho)$, which govern the width of the distribution and the tilt, respectively. Arrows correspond to 1σ credible levels for hyperparameters $\boldsymbol{\mu}$, which govern the position of the distribution. Hyperparameter values corresponding to the shaded area and the dashed line represent a row $\mathcal{N}(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\sigma}, \rho)$ in Table 2.

to an uncertainty in $(\boldsymbol{\sigma}, \rho)$ and arrows separately show the uncertainty in $\boldsymbol{\mu}$. The uncertainty in ρ is seen for DM noise as the difference between the tilt of the outer border of the shaded area compared to the inner border. Whereas for SN the value of $\rho \approx -1$ is strongly preferred.

3.3 Distribution of noise parameters as a mixture model

One may also be interested whether the distribution of pulsar noise parameters is more complex. For example, in fig. 1 in Goncharov et al. (2021a), one may notice the clustering of noise parameters in two areas. First, around $\gamma \approx 1$. Secondly, a cluster of pulsars around $\gamma = 13/3$ of the GWB. A similar clustering is observed in the PTA noise analysis by the NANOGrav Collaboration (2023b), where it is explained that pulsars with $\gamma \approx 13/3$ contribute to CP in full-PTA analysis. An analogous figure is shown in Reardon et al. (2023), although the presence of two overdensities is not obvious there. Although we have taken steps to marginalize the contribution of the CP and uncover properties of pulsar-intrinsic noise, it is nevertheless useful to consider the possibility of a mixture model of a Gaussian distribution $\mathcal{N}$ and a uniform distribution $\mathcal{U}$, where a broader uniform distribution may fit potential outliers. The density of such a distribution is then $\nu\mathcal{N} + (1-\nu)\mathcal{U}$, where $\nu \in [0, 1]$ is the contribution of the normal distribution.

3.4 Summary of the results

As part of our measurements of hyperparameters, we have also obtained Bayesian evidence values $\mathcal{Z}$. With this, we perform model selection to determine which of the models fit the observed distribution of pulsar noise parameters best. First, we find that the standard 'static' uniform priors are disfavoured compared to hierarchical uniform priors with natural log Bayes factors of $\gtrsim$100. Therefore, the use of standard uniform priors is not recommended for PTA data analysis. The rest of the results of our hyperparameter estimation and hierarchical model selection are shown in Table 2.

Some ensemble properties of SN are similar to DM noise, whereas some properties are different. For both SN and DM, for all models, the area of the parametrized prior is shrunk as much as possible to achieve the maximum posterior probability (given a fixed area, the narrower the distribution, the higher its maximum), but it also has to be sufficiently wide to be consistent with all the samples of noise amplitudes and spectral indices. For SN, the normal distribution fits the data as well as the uniform distribution. However, the covariance between $\lg A_{\rm SN}$ and $\gamma_{\rm SN}$ plays an important role. In Figs 1(a) and 2(a), one may notice a diagonal trend, which results in finding a covariance parameter $\rho_{\rm SN} \approx -1$. The model with a lack of covariance, $\rho_{\rm SN} = 0$, is disfavoured with $\ln \mathcal{B} = 6$. The covariance between $\lg A_{\rm SN}$ and $\gamma_{\rm SN}$ can be explained by noticing that it follows a line of equal noise power. In contrast, for DM noise, we do not find strong evidence for covariance between $\lg A_{\rm DM}$ and $\gamma_{\rm DM}$. With a lack of covariance, the uniform model is strongly preferred over the Gaussian model with a log Bayes factor of 6. The mixture model is disfavoured by our observations of both SN and DM noise, indicating the lack of outlying noise terms.

Given the posterior biases found in simulated data studies in Appendix A, we repeated all analyses from this section simultaneously with modelling the CP amplitude $\lg A_{\rm CP}$ to be drawn from the truncated normal distribution with hyperparameters $(\mu_{\lg A_{\rm CP}}, \sigma_{\lg A_{\rm CP}})$. This is the approach that mitigated posterior biases in the simulations. SN and DM noise hyperparameters inferred this way are consistent with those in Table 2 at 1σ level, except for the mixture model's $\mu_{\gamma_{\rm SN}} = 1.44^{+0.62}_{-0.84}$. Because the mixture model is disfavoured by the data and the rest of the results are consistent with Table 2, we do not report the results obtained under this alternative model for other hyperparameters.

4 CAVEATS OF THE CIRCULAR ANALYSIS

Circular analysis or 'double dipping' refers to the usage of data to inform on the model that is then applied to analyse the same data. The methodology we propose in our study is not a circular

Table 2. Results of hyperparameter estimation and hierarchical model selection for ensemble pulsar noise properties of the EPTA. DM noise and SN are parametrized by $\boldsymbol{\theta} = (\lg A, \gamma)$. We present the results for the following models of pulsar noise parameter distributions: a uniform distribution $\mathcal{U}$, a normal distribution $\mathcal{N}$ with and without covariance between $\lg A$ and γ, and a mixture model of a normal and a uniform distribution. Columns $\mathcal{P}(\boldsymbol{\Lambda}|\boldsymbol{\delta t})$ contain maximum a posteriori hyperparameter values with 1σ credible levels. Columns $\ln \mathcal{B}$ contain log Bayes factors in favour of a model of interest against the uniform distribution model. The fact that they are negative means that the uniform model is the best fit.

Model, $\pi(\boldsymbol{\theta}\|\boldsymbol{\Lambda})$	Achromatic 'spin' noise $\boldsymbol{\theta} = (A_{\rm SN}, \gamma_{\rm SN})$ $\mathcal{P}(\boldsymbol{\Lambda}_{\rm SN}\|\boldsymbol{\delta t})$		$\ln \mathcal{B}_{\mathcal{U}_{\rm SN}}$	Dispersion measure variation noise $\boldsymbol{\theta} = (A_{\rm DM}, \gamma_{\rm DM})$ $\mathcal{P}(\boldsymbol{\Lambda}_{\rm DM}\|\boldsymbol{\delta t})$		$\ln \mathcal{B}_{\mathcal{U}_{\rm DM}}$
$\mathcal{U}(\boldsymbol{\theta}\|\mathbf{min}, \mathbf{max})$ shown in Fig. 1	$\min(\lg A_{\rm SN})$	$-14.91^{+0.43}_{-0.66}$	–	$\min(\lg A_{\rm DM})$	$-13.83^{+0.08}_{-0.09}$	–
	$\max(\lg A_{\rm SN})$	$-12.71^{+0.13}_{-0.13}$		$\max(\lg A_{\rm DM})$	$-12.61^{+0.08}_{-0.06}$	
	$\min(\gamma_{\rm SN})$	$0.73^{+0.27}_{-0.41}$		$\min(\gamma_{\rm DM})$	$0.60^{+0.23}_{-0.28}$	
	$\max(\gamma_{\rm SN})$	$2.88^{+1.09}_{-1.23}$		$\max(\gamma_{\rm DM})$	$3.46^{+0.31}_{-0.31}$	
$\mathcal{N}(\boldsymbol{\theta}\|\boldsymbol{\mu}, \boldsymbol{\sigma}, \rho)$ shown in Fig. 2	$\mu_{\lg A_{\rm SN}}$	$-13.88^{+0.29}_{-0.33}$	−0.6	$\mu_{\lg A_{\rm DM}}$	$-13.27^{+0.09}_{-0.09}$	−6.0
	$\sigma_{\lg A_{\rm SN}}$	$0.84^{+0.52}_{-0.25}$		$\sigma_{\lg A_{\rm DM}}$	$0.37^{+0.09}_{-0.06}$	
	$\mu_{\gamma_{\rm SN}}$	$2.82^{+0.61}_{-0.65}$		$\mu_{\gamma_{\rm DM}}$	$1.95^{+0.25}_{-0.32}$	
	$\sigma_{\gamma_{\rm SN}}$	$1.52^{+0.95}_{-0.50}$		$\sigma_{\gamma_{\rm DM}}$	$0.95^{+0.37}_{-0.24}$	
	$\rho_{\rm SN}$	$-0.96^{+0.08}_{-0.03}$		$\rho_{\rm DM}$	$-0.21^{+0.29}_{-0.29}$	
$\mathcal{N}(\boldsymbol{\theta}\|\boldsymbol{\mu}, \boldsymbol{\sigma}, \rho = 0)$	$\mu_{\lg A_{\rm SN}}$	$-13.71^{+0.18}_{-0.26}$	−6.5	$\mu_{\lg A_{\rm DM}}$	$-13.27^{+0.08}_{-0.08}$	−6.6
	$\sigma_{\lg A_{\rm SN}}$	$0.54^{+0.21}_{-0.15}$		$\sigma_{\lg A_{\rm DM}}$	$0.35^{+0.07}_{-0.05}$	
	$\mu_{\gamma_{\rm SN}}$	$1.43^{+0.69}_{-0.16}$		$\mu_{\gamma_{\rm DM}}$	$2.01^{+0.19}_{-0.19}$	
	$\sigma_{\gamma_{\rm SN}}$	$0.34^{+0.51}_{-0.26}$		$\sigma_{\gamma_{\rm DM}}$	$0.77^{+0.18}_{-0.15}$	
$\nu\mathcal{N} + (1-\nu)\mathcal{U}$, $0 \leq \nu \leq 1$	$\mu_{\lg A_{\rm SN}}$	$-13.86^{+0.28}_{-0.29}$	−3.7	$\mu_{\lg A_{\rm DM}}$	$-13.28^{+0.09}_{-0.08}$	−9.1
	$\sigma_{\lg A_{\rm SN}}$	$0.78^{+0.43}_{-0.22}$		$\sigma_{\lg A_{\rm DM}}$	$0.37^{+0.08}_{-0.06}$	
	$\mu_{\gamma_{\rm SN}}$	$2.87^{+0.55}_{-0.67}$		$\mu_{\gamma_{\rm DM}}$	$1.94^{+0.26}_{-0.34}$	
	$\sigma_{\gamma_{\rm SN}}$	$1.48^{+0.86}_{-0.44}$		$\sigma_{\gamma_{\rm DM}}$	$0.97^{+0.40}_{-0.24}$	
	$\rho_{\rm SN}$	$-0.97^{+0.07}_{-0.02}$		$\rho_{\rm DM}$	$-0.23^{+0.29}_{-0.27}$	
	$\nu_{\rm SN}$	$0.95^{+0.03}_{-0.07}$		$\nu_{\rm DM}$	$0.96^{+0.03}_{-0.05}$	

analysis; it represents one global fit to PTA data. Despite the fact that importance sampling involves performing parameter estimation on the same data twice, the final result is ultimately independent of the proposal distribution obtained at the first step. Namely, the target distribution is independent of the proposal distribution. However, because circular analysis has been extensively discussed in the PTA community, in this section we take an opportunity to cover this subject from the hierarchical inference standpoint.

Overall, circular analyses may lead to the underestimation of measurement uncertainty and systematic errors. However, some PTA analyses in the past have involved a circular analysis, as pointed out by van Haasteren (2025). Namely, pulsar SN and DM noise terms were included (not included) in the full-PTA analysis based on (a lack of) evidence for these terms in single-pulsar noise model selection. van Haasteren (2025) argues against this approach. Indeed, double dipping is not a proper statistical treatment; it should be abandoned in favour of the proper model selection and model averaging performed in one global fit, as recommended by the author. However, to give a fair overview of the problem, we also show that circular analysis does not always yield incorrect conclusions in analyses of PTA data.

Before we discuss the application of hierarchical inference in the sections below, we would like to point out that it is not clear from van Haasteren (2025) that the aforementioned approach of selecting SN terms for subsequent full-PTA analyses has led to underestimation of measurement uncertainties reported by PTAs. For model averaging, the prior odds between the existence or absence of CP are proposed to be equal in van Haasteren (2025). For the case of including a red noise term to every pulsar, the prior odds will be arbitrary and vary from pulsar to pulsar, which is clear from the equations in appendix in van Haasteren (2025). However, data inform on these odds, so they should become a model (hyper)parameter to avoid prior misspecification, the same problem discussed in this work. These parametrized odds will act as model selection. In the limit where there is no measurable red noise in pulsars and the data inform on it sufficiently well, this will lead to the elimination of the contribution of one of the models. Because different noise models lead to different measurement uncertainties for $(\lg A_{\rm CP}, \gamma_{\rm CP})$, it is possible that the uncertainty introduced by enforcing red noise to every pulsar is suboptimal.

4.1 A comparison between circular analysis, incorrect priors, and a proper analysis

One example of circular analysis in the context of our hierarchical inference is finding best-fitting $\boldsymbol{\Lambda}$ and using it for the gravitational wave search. It also leads to resolving prior misspecification but potentially at the cost of a reduced measurement uncertainty for $(\lg A, \gamma)$ of the GWB. The reduction may come thanks to simply

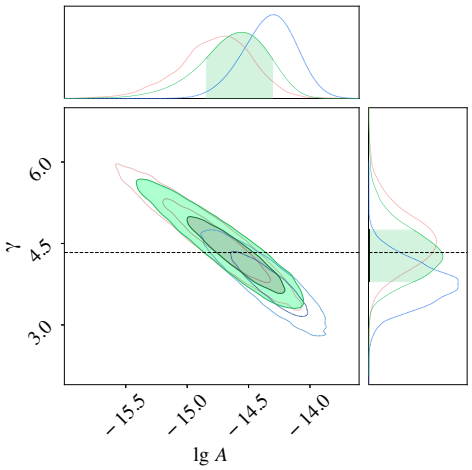


Figure 3. A comparison between circular analysis (dotted contours), incorrect broad uniform priors (dashed contours), and a proper hierarchical analysis (solid contours) of the CP associated with the GWB.

ignoring a part of the intrinsic measurement uncertainty for $\mathbf{\Lambda}$, so a regular noise fluctuation is more likely to render our estimate of $(\lg A, \gamma)$ to be inconsistent with the true value. In Fig. 3, we show three measurements of $(\lg A, \gamma)$ of the CP in the EPTA data. Blue dashed contours correspond to a posterior obtained with the standard uniform SN priors, $\pi(\boldsymbol{\theta}|\mathbf{\Lambda}) = \mathcal{U}(\boldsymbol{\theta}|\varnothing)$. Red dotted contours correspond to $\pi(\boldsymbol{\theta}|\mathbf{\Lambda}) = \mathcal{N}(\boldsymbol{\theta}|\boldsymbol{\mu}_0, \boldsymbol{\sigma}_0, \rho = 0)$, where $(\boldsymbol{\mu}_0, \boldsymbol{\sigma}_0)$ are the values obtained for this model from Table 2. Therefore, the red dotted line corresponds to a circular analysis. Filled green contours correspond to a proper (global fit) measurement obtained based on equation (9); it is discussed in more detail in Goncharov et al. (2024). The measurement uncertainty of $(\lg A, \gamma)$ obtained with circular analysis matches that obtained with a full hierarchical analysis. It is even 5–8 per cent larger, suggesting that in the current setting circular analysis overestimates the intrinsic measurement uncertainty. Note, however, that the circular analysis shifts the maximum a posteriori value of $(\lg A, \gamma)$, so the use of it is still not recommended.

4.2 A toy example: circular analysis of CP

Keeping in mind the caveats above, for the test purposes we also perform the estimation of hyperparameters $\mathbf{\Lambda}$ for the mixture model of two Gaussian distributions where hyperpriors on the second component of the mixture correspond to EPTA measurement of the CP parameters (EPTA Collaboration & InPTA Collaboration 2023b). This exercise remains useful to explore the interplay between pulsar-intrinsic noise and the CP and to understand the performance of the model. For the two-Gaussian mixture model, $\pi(\lg A_{\rm SN}, \gamma_{\rm SN}|\mathbf{\Lambda}_{2\mathcal{N}})$, and $\mathbf{\Lambda}_{2\mathcal{N}} = (\mu^{1,2}_{\lg A_{\rm SN}}, \sigma^{1,2}_{\lg A_{\rm SN}}, \mu^{1,2}_{\gamma_{\rm SN}}, \sigma^{1,2}_{\gamma_{\rm SN}}, \rho^{1,2}_{\rm SN}, \text{and } \nu_{\mathcal{N}})$. Indices (1, 2) refer to each of the two components in a mixture, while ν corresponds to a fraction of the first mixture component in the total prior probability. For the second component to correspond to the CP in the EPTA data, we impose a Gaussian hyperprior determined by the measurement uncertainty on $(\lg A, \gamma)$ reported by the EPTA for our data. We find that the fraction of the second normal component is mostly consistent with one and inconsistent with zero. However, hyperparameters of the second mixture component – mean and (co)variance – are well constrained. Interestingly, we also find a marginal excess posterior density at $(\sigma^2_{\lg A_{\rm SN}}, \sigma^2_{\gamma_{\rm SN}}) = (0, 0)$. This result is related to Goncharov et al. (2022)'s measurement of $\sigma_{\lg A}$ to be consistent with zero. When removing the CP term with $\gamma = 13/3$ from our pulsar-intrinsic noise model, the posterior density increases, as expected. This is shown in Fig. 4.

5 CONCLUSIONS

We performed inference of ensemble properties of SN and DM noise in the 25 yr version of the second data release of the EPTA. Overall, we find that the standard uniform priors used in the previous analysis are too wide and not representative of the observations. However, a parametrized uniform prior distribution is a good fit. We therefore recommend correctly accounting for ensemble pulsar noise properties using equation (9) in future full-PTA analyses to avoid systematic errors (parameter estimation and model selection biases). Our approach of prior reweighting for inferring noise hyperparameters based on equation (8) is well suited for understanding broad ensemble noise properties for simulating data and predicting the PTA noise budget.

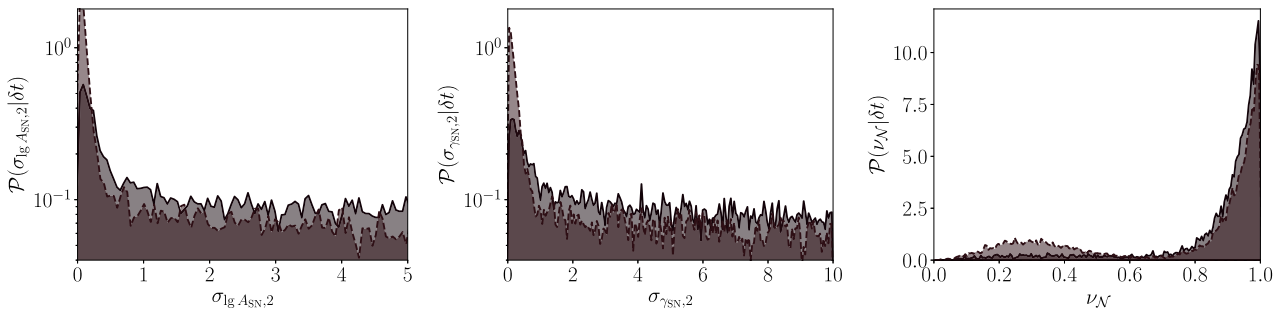


Figure 4. Modelling of the CP as a narrow normally distributed cluster of pulsar noise using data-informed hyperpriors $\pi(\mathbf{\Lambda})$. We assume that pulsar noise parameters $\lg A$ and γ are distributed as a mixture model of two Gaussians, where hyperparameters of the second Gaussian are chosen to match the measurement uncertainty on the CP reported by the EPTA Collaboration & InPTA Collaboration (2023b). Parameters σ with respective indices correspond to standard deviations of the distribution that governs $\lg A$ and γ. Consistency of σ with zero has first been shown in Goncharov et al. (2022) using a more rigorous although less flexible methodology. Parameter $\nu_{\mathcal{N}}$ corresponds to the fraction of the first, wider normal distribution in a model. Solid lines correspond to including the second $\gamma = 13/3$ red noise term to pulsars to account for the presence of the CP, dashed lines correspond to not including it. As expected, the contribution of the CP becomes more visible in the latter case at $\sigma_{\lg A,\gamma} = 0$ and $\nu_{\mathcal{N}} \approx 0.3$.

ACKNOWLEDGEMENTS

We thank Rutger van Haasteren for insightful discussions about hierarchical inference and circular analysis. Some of our calculations were carried out using the OzSTAR Australian national facility (high-performance computing) at Swinburne University of Technology.

DATA AVAILABILITY

The code to reproduce the results of this analysis and the analysis by Goncharov et al. (2022) is available at github.com/bvgoncharov/pta_priors. The data used for our study, EPTA DR2 by the EPTA Collaboration (2023), are available at zenodo. Other data may be provided by the corresponding author upon request. The PTA likelihood is incorporated in ENTERPRISE (Ellis et al. 2020) and posterior sampling is performed using PTMCMCSAMPLER (Ellis & van Haasteren 2017) and DYNESTY (Speagle 2020).

REFERENCES

Chen S. et al., 2021, MNRAS, 508, 4970
Desvignes G. et al., 2016, MNRAS, 458, 3341
Detweiler S., 1979, ApJ, 234, 1100
Ellis J., van Haasteren R., 2017, jellis18/PTMCMCSampler: Official Release, Zenodo. Available at: https://doi.org/10.5281/zenodo.1037579
Ellis J. A., Vallisneri M., Taylor S. R., Baker P. T., 2020, ENTERPRISE: Enhanced Numerical Toolbox Enabling a Robust PulsaR Inference SuitE, Zenodo. Available at: https://doi.org/10.5281/zenodo.4059815
EPTA Collaboration, 2023, A&A, 678, A48
EPTA Collaboration, InPTA Collaboration, 2023a, A&A, 678, A49
EPTA Collaboration, InPTA Collaboration, 2023b, A&A, 678, A50
Foster R. S., Backer D. C., 1990, ApJ, 361, 300
Goncharov B. et al., 2021a, MNRAS, 502, 478
Goncharov B. et al., 2021b, ApJ, 917, L19
Goncharov B. et al., 2022, ApJ, 932, L22
Goncharov B. et al., 2024, preprint (arXiv:2409.03627)
Hazboun J. S., Simon J., Siemens X., Romano J. D., 2020, ApJ, 905, L6
Hellings R. W., Downs G. S., 1983, ApJ, 265, L39
Hobbs G. B., Edwards R. T., Manchester R. N., 2006, MNRAS, 369, 655
Keith M. J. et al., 2013, MNRAS, 429, 2161
Kramer M., Champion D. J., 2013, Class. Quantum Gravity, 30, 224009
Lentati L., Alexander P., Hobson M. P., Taylor S., Gair J., Balan S. T., van Haasteren R., 2013, Phys. Rev. D, 87, 104021
NANOGrav Collaboration, 2016, ApJ, 821, 13
NANOGrav Collaboration, 2020, ApJ, 905, L34
NANOGrav Collaboration, 2023a, ApJ, 951, L8
NANOGrav Collaboration, 2023b, ApJ, 951, L10
Park R. S., Folkner W. M., Williams J. G., Boggs D. H., 2021, AJ, 161, 105
Phinney E. S., 2001, MNRAS, preprint (arXiv:astro-ph/0108028)
Reardon D. J. et al., 2023, ApJ, 951, L6
Renzini A. I., Goncharov B., Jenkins A. C., Meyers P. M., 2022, Galaxies, 10, 34
Rosado P. A., Sesana A., Gair J., 2015, MNRAS, 451, 2417
Sazhin M. V., 1978, SvA, 22, 36
Shannon R. M., Cordes J. M., 2010, ApJ, 725, 1607
Speagle J. S., 2020, MNRAS, 493, 3132
Thrane E., Talbot C., 2019, Publ. Astron. Soc. Aust., 36, e010
van Haasteren R., 2024, ApJS, 273, 23
van Haasteren R., 2025, MNRAS, 537, L1
van Haasteren R., Vallisneri M., 2014, Phys. Rev. D, 90, 104012
Zic A. et al., 2022, MNRAS, 516, 410

Table A1. Parameters of our two simulations.

$\mu_{\lg A}$	$\sigma_{\lg A}$	μ_γ	σ_γ	ρ	$\lg A_{\rm CP}$	$\lg \gamma_{\rm CP}$
−14.5	1.5	4.0	1.5	0.0	−15.0	13/3
−13.8	0.7	2.0	1.0			

APPENDIX A: TESTS WITH THE SIMULATED DATA

We test our inference of hyperparameters based on Section 2.2.1 using two simulated data sets. Both simulations are based on 25 pulsars, timed over 15 yr with the 100 ns precision. The cadence is empirically chosen to visually represent irregular PTA observations. It is based on the Pareto distribution with the shape value of 0.6, scaled to the PTA observation time, with the maximum observation gap of 0.6 of the time span. Simulations contain SN and the CP. SN is drawn from the truncated normal distribution, with hyperparameters given by the first five columns in Table A1. CP represents an additional red noise term in every pulsar, with the same amplitude and spectral index listed in the last two columns in Table A1. Therefore, CP shows the expected spectral properties of the GWB from supermassive black hole binaries.

The simulations are designed to test our ability to correctly model the distribution of SN parameters given the presence of CP. In the first simulated data set, hyperparameters are chosen such that SN amplitudes and spectral indices significantly overlap with those of the CP. Because CP and SN are indistinguishable in the data of single pulsars, the proposal posteriors we construct from such data may show significant differences from the target posteriors we would get in a global fit to all pulsar data. As described in Section 2.2.1, we fix the spectral index of CP to 13/3. Thus, in pulsars where $\gamma_{\rm SN}$ is close to 13/3, proposal posteriors on $A_{\rm CP}$ peak at values of $A_{\rm SN}$, with only the tails extending to the true value of $A_{\rm CP}$. These biases are mitigated when obtaining target posteriors in the global fit, but the residual biases may remain. In the second simulated data, the overlap between CP and SN is less significant. Therefore, the first simulation is considered to be the most difficult case.

First, we perform (hyper)parameter estimation assuming no covariance between SN amplitudes and spectral indices, as simulated ($\rho = 0$). The results for the first simulation are shown in Fig. A1, and for the second simulation in Fig. A2, both as orange contours. The inferred hyperparameters are fully consistent with the simulated values, except for the standard deviation of the SN amplitudes $\sigma_{\lg A}$ for the first simulation. The simulated value of 1.5 lies at the edge of the 1σ credible level in two-dimensional posteriors including $\sigma_{\lg A}$, and just outside of the 1σ level in the one-dimensional posterior for $\sigma_{\lg A}$.

Secondly, we infer hyperparameters assuming that there may be a covariance between SN amplitudes and spectral indices (ρ is a free parameter). The results are shown as green contours in Figs A1 and A2. Although we simulated $\rho = 0$, for both simulations we find ρ to be inconsistent with 0. Instead, the maximum a posteriori ρ is consistent with -1, while the value of 0 is beyond 3σ credibility. Compared to the previous case where we fix $\rho = 0$, a minor tension at 1–2σ level emerges for μ_γ in the second simulation, but another minor tension for $\sigma_{\lg A}$ in the first simulation disappears. We conclude that ρ is most sensitive to posterior biases.

To resolve posterior biases, we extend our model to $A_{\rm CP}$, imposing that it is also drawn from the truncated Gaussian distribution. Thus, we extend both SN and CP priors to become a part of our model: $\pi(\lg A, \gamma|\mu_{\lg A}, \sigma_{\lg A}, \mu_\gamma, \sigma_\gamma, \rho)\pi(\lg A_{\rm CP}|\mu_{\lg A_{\rm CP}}, \sigma_{\lg A_{\rm CP}})$. We im-

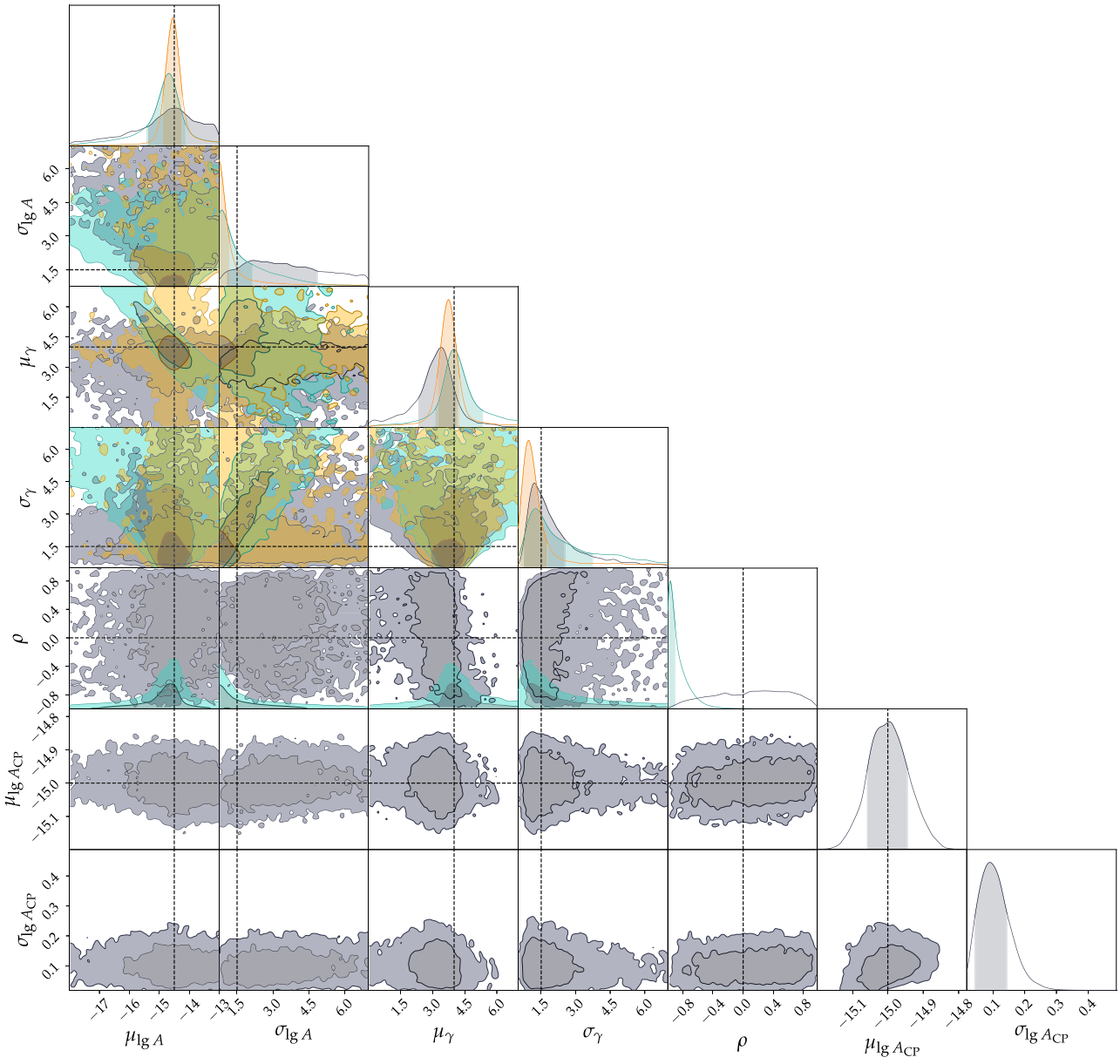


Figure A1. Posteriors on the inferred hyperparameters in the first simulation. Dashed vertical lines correspond to simulated values in Table A1. Yellow contours spanning 4 parameters correspond to $\rho = 0$, green contours spanning 5 parameters correspond to ρ as a free parameter, and black contours spanning all parameters correspond to additionally modelling the distribution of CP amplitudes in pulsars as being drawn from the truncated normal distribution with mean $\mu_{\lg A_{\rm CP}}$ and the standard deviation $\sigma_{\lg A_{\rm CP}}$. Shaded areas correspond to 1σ (1σ and 2σ) credible levels in fully marginalized posteriors (two-dimensional posteriors).

pose a log-uniform prior on σ parameters. The results are shown as black contours in Figs A1 and A2. In this case, we find the inferred hyperparameters to be fully consistent with the simulated values. The values of $\mu_{\lg A_{\rm CP}}$ and $\sigma_{\lg A_{\rm CP}}$ are also in accordance with our expectations. The value $\mu_{\lg A_{\rm CP}}$ is consistent with the simulatedlg $A_{\rm CP} = -15.0$. The value of $\sigma_{\lg A_{\rm CP}}$ peaks at small values; it excludes 0 in the first simulation, despite CP representing an infinitely small range of amplitudes. This is a known effect in importance sampling due to a finite number of the recycled posterior samples (Goncharov et al. 2022).

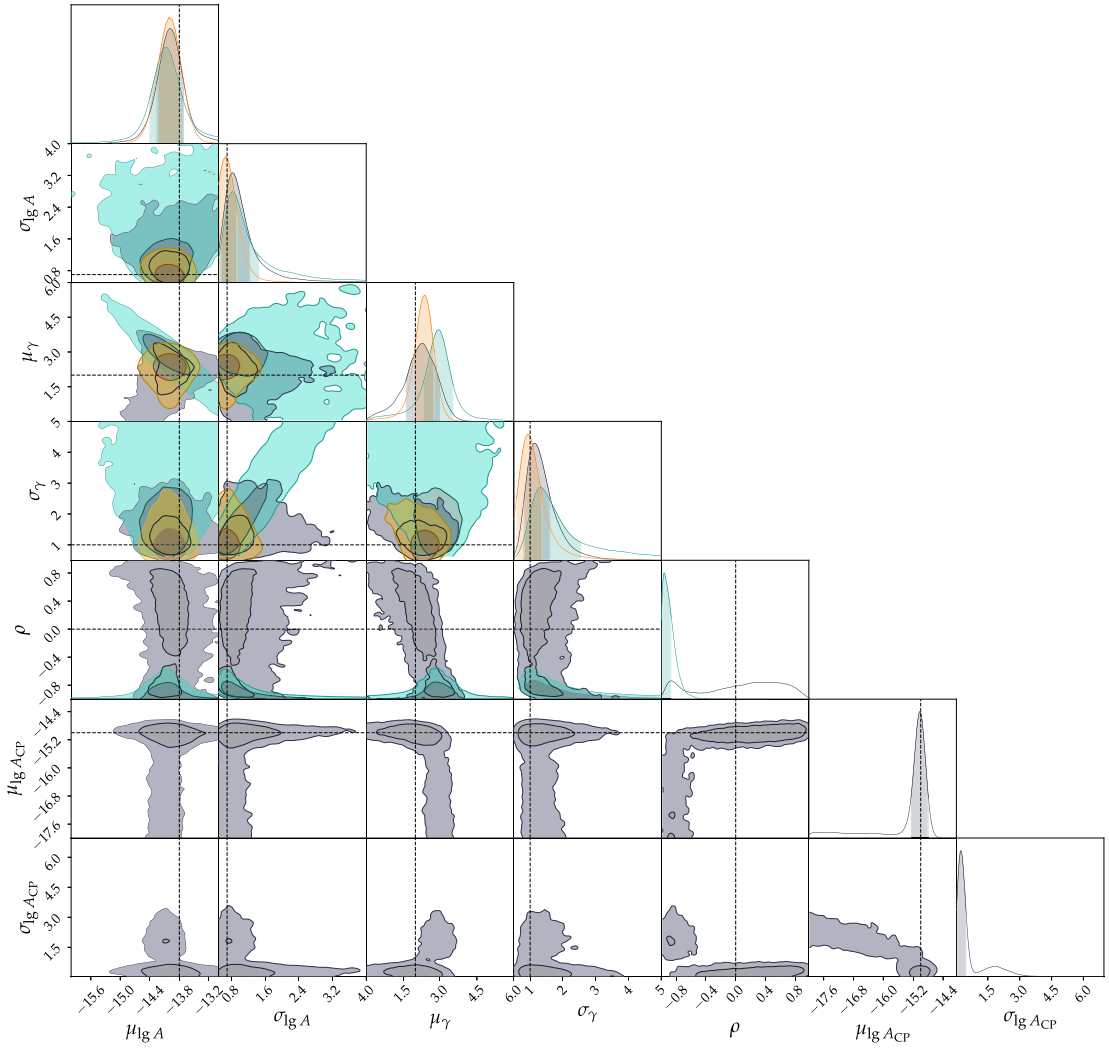


Figure A2. Posteriors on the inferred hyperparameters in the second simulation. Dashed vertical lines correspond to simulated values in Table A1. Colours and contours are explained in Fig. A1.

APPENDIX B: NUMERICAL MARGINALIZATION OVER HYPERPARAMETERS

In this section, we derive equation (9), which describes a hierarchical posterior marginalized over hyperparameters $\mathbf{\Lambda}$. After taking an integral over $\mathbf{\Lambda}$, equation (7) becomes

$$\mathcal{P}(\boldsymbol{\theta}|\boldsymbol{\delta t}) = \mathcal{Z}^{-1}\mathcal{L}(\boldsymbol{\delta t}|\boldsymbol{\theta}) \int \pi(\boldsymbol{\theta}|\mathbf{\Lambda})\pi(\mathbf{\Lambda})\,\mathrm{d}\mathbf{\Lambda}. \tag{B1}$$

Following the formalism of importance sampling, which was used to obtain equation (8), we refer to $\pi(\boldsymbol{\theta}|\mathbf{\Lambda})$ as the target distribution. Similarly, we introduce the proposal distribution, the standard PTA noise prior that does not depend on hyperparameters, $\pi(\boldsymbol{\theta}|\varnothing)$. Next, we multiply equation (B1) by unity and expand the unity on the right-hand side as a proposal distribution divided by itself. Rearranging the multipliers,

$$\mathcal{P}(\boldsymbol{\theta}|\boldsymbol{\delta t}) = \frac{\mathcal{L}(\boldsymbol{\delta t}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\varnothing)}{\mathcal{Z}} \int \frac{\pi(\boldsymbol{\theta}|\mathbf{\Lambda})}{\pi(\boldsymbol{\theta}|\varnothing)}\pi(\mathbf{\Lambda})\,\mathrm{d}\mathbf{\Lambda}. \tag{B2}$$

Next, we use the expression for the expectation value of a probability density $f(x)$ given the known probability density $p(x)$[2]:

$$\langle f(x)\rangle_{p(x)} = \int f(x)p(x)\,\mathrm{d}x \approx \frac{1}{n_\mathrm{s}}\sum_i^{n_\mathrm{s}} f(x_i), \tag{B3}$$

where x_i are n_s samples from $p(x)$. Taking $\pi(\mathbf{\Lambda})$ as $p(x)$, we arrive at equation (9).

[2]Equation (B3) is also used to derive equation (8). There, $p(x)$ is taken to be a posterior distribution instead.

This paper has been typeset from a TEX/LATEX file prepared by the author.