**Astronomy & Astrophysics**

# The second data release from the European Pulsar Timing Array

## II. Customised pulsar noise models for spatially correlated gravitational waves

EPTA Collaboration and InPTA Collaboration: J. Antoniadis (Ιωάννης Αντωνιάδης)[1,2], P. Arumugam[3],
S. Arumugam[4], S. Babak[5], M. Bagchi[6,7], A.-S. Bak Nielsen[2,8], C. G. Bassa[9], A. Bathula[10],
A. Berthereau[11,12], M. Bonetti[13,14,15], E. Bortolas[13,14,15], P. R. Brook[16], M. Burgay[17], R. N. Caballero[18],
A. Chalumeau[13,★], D. J. Champion[2], S. Chanlaridis[1], S. Chen[19], I. Cognard[11,12], S. Dandapat[20],
D. Deb[6], S. Desai[21], G. Desvignes[2], N. Dhanda-Batra[22], C. Dwivedi[23], M. Falxa[5,11], R. D. Ferdman[24],
A. Franchini[13,14], J. R. Gair[25], B. Goncharov[26,27], A. Gopakumar[20], E. Graikou[2], J.-M. Grießmeier[11,12],
L. Guillemot[11,12], Y. J. Guo[2], Y. Gupta[28], S. Hisano[29], H. Hu[2], F. Iraci[30,17], D. Izquierdo-Villalba[13,14],
J. Jang[2], J. Jawor[2], G. H. Janssen[9,31], A. Jessner[2], B. C. Joshi[28,3], F. Kareem[32,33], R. Karuppusamy[2],
E. F. Keane[34], M. J. Keith[35,★], D. Kharbanda[21], T. Kikunaga[29], N. Kolhe[36], M. Kramer[2,35],
M. A. Krishnakumar[2,8], K. Lackeos[2], K. J. Lee[3,8,2], K. Liu[2], Y. Liu[37,8], A. G. Lyne[35], J. W. McKee[38,39],
Y. Maan[28], R. A. Main[2], M. B. Mickaliger[35], I. C. Niţu[35], K. Nobleson[40], A. K. Paladi[41],
A. Parthasarathy[2,★], B. B. P. Perera[42], D. Perrodin[17], A. Petiteau[43,5], N. K. Porayko[13,2], A. Possenti[17],
T. Prabu[44], H. Quelquejay Leclere[5], P. Rana[20], A. Samajdar[45], S. A. Sanidas[35], A. Sesana[13,14,15],
G. Shaifullah[13,14,17], J. Singha[3], L. Speri[25], R. Spiewak[35], A. Srivastava[21], B. W. Stappers[35], M. Surnis[46],
S. C. Susarla[47], A. Susobhanan[48], K. Takahashi[49,50], P. Tarafdar[6], G. Theureau[11,12,51], C. Tiburzi[17],
E. van der Wateren[9,31], A. Vecchio[16], V. Venkatraman Krishnan[2], J. P. W. Verbiest[52,8,2],
J. Wang[8,53,54], L. Wang[35], and Z. Wu[37,8]

*(Affiliations can be found after the references)*

**ABSTRACT**

*Aims.* The nanohertz gravitational wave background (GWB) is expected to be an aggregate signal of an ensemble of gravitational waves emitted predominantly by a large population of coalescing supermassive black hole binaries in the centres of merging galaxies. Pulsar timing arrays (PTAs), which are ensembles of extremely stable pulsars at approximately kiloparsec distances precisely monitored for decades, are the most precise experiments capable of detecting this background. However, the subtle imprints that the GWB induces on pulsar timing data are obscured by many sources of noise that occur on various timescales. These must be carefully modelled and mitigated to increase the sensitivity to the background signal.

*Methods.* In this paper, we present a novel technique to estimate the optimal number of frequency coefficients for modelling achromatic and chromatic noise, while selecting the preferred set of noise models to use for each pulsar. We also incorporated a new model to fit for scattering variations in the Bayesian pulsar timing package `temponest`. These customised noise models enable a more robust characterisation of single-pulsar noise. We developed a software package based on `tempo2` to create realistic simulations of European Pulsar Timing Array (EPTA) datasets that allowed us to test the efficacy of our noise modelling algorithms.

*Results.* Using these techniques, we present an in-depth analysis of the noise properties of 25 millisecond pulsars (MSPs) that form the second data release (DR2) of the EPTA and investigate the effect of incorporating low-frequency data from the Indian Pulsar Timing Array collaboration for a common sample of ten MSPs. We used two packages, `enterprise` and `temponest`, to estimate our noise models and compare them with those reported using EPTA DR1. We find that, while in some pulsars we can successfully disentangle chromatic from achromatic noise owing to the wider frequency coverage in DR2, in others the noise models evolve in a much more complicated way. We also find evidence of long-term scattering variations in PSR J1600−3053. Through our simulations, we identify intrinsic biases in our current noise analysis techniques and discuss their effect on GWB searches. The analysis and results discussed in this article directly help to improve the sensitivity to the GWB signal and they are already being used as part of global PTA efforts.

**Key words.** pulsars: general – gravitational waves – methods: statistical

## 1. Introduction

Pulsar timing allows astronomers to track every single rotation of a pulsar. It involves comparing the measured pulse times of arrival (TOAs) with a timing model that contains our current best knowledge of the pulsar's timing properties. Millisecond pulsars (MSPs), which are rapidly rotating neutron stars distributed throughout the Galaxy, pulse with sufficient regularity to function as excellent clocks. The timing residuals, which are differences between the predicted and estimated arrival times, contain imprints of various astrophysical processes, including the subtle signature induced by the stochastic gravitational wave background (GWB).

---

★ Corresponding authors: A. Chalumeau, `aurelien.chalumeau@unimib.it`; M. J. Keith, `michael.keith@manchester.ac.uk`; A. Parthasarathy, `aparthas@mpifr-bonn.mpg.de`

Unlike transient gravitational waves, a GWB will be detected from all directions in the sky. However, a single pulsar is not sufficient to disentangle the contribution of the GWB, as it is affected by many sources of noise with spectral properties similar to those of the GWB. To overcome this, cross-correlations between multiple pulsars are used, as the GWB signal is correlated across multiple sources while the noise (typically intrinsic to the pulsar) is not. Thus, pulsar timing arrays (PTAs; Foster & Backer 1990) extend the concept of pulsar timing to an ensemble of MSPs. However, a GWB is not the only correlated signal amongst an array of pulsars. Imperfections in a common reference clock can cause a monopolar signal (Hobbs et al. 2020), while unmodelled errors related to Solar System ephemerides can induce a dipolar signal (Caballero et al. 2018). The GWB is expected to induce a quadrupolar signal, as described by the Hellings–Downs (H–D) curve (Hellings & Downs 1983). The latter is considered to be the definitive evidence for the detection of a GWB. It is also important to note that measurable deviations from the H–D curve provide critical insight into the origins of gravitational waves (Taylor et al. 2015).

Given the characteristics of the GWB, PTAs need to accumulate decades of data to become sensitive to the signal. Three PTA collaborations were formed in the early 2000s, the European Pulsar Timing Array (EPTA; Desvignes et al. 2016), the North American Nanohertz Observatory for Gravitational Waves (NANOGrav; Demorest et al. 2013), and the Parkes Pulsar Timing Array (PPTA; Manchester et al. 2013). Since the sensitivity of the GWB relies, among other things, on the number of pulsars and the length of the dataset, the three PTA collaborations came together to form the International Pulsar Timing Array (IPTA). The Indian Pulsar Timing Array (InPTA; Tarafdar et al. 2022) has recently also joined the IPTA. The three PTAs, along with the IPTA, have recently detected a common red signal (CRS) that shares many characteristics with the expected GWB (Arzoumanian et al. 2020; Goncharov et al. 2021; Chen et al. 2021; Antoniadis et al. 2022). Although these measurements could be the first sign of the GWB emerging from noise, there was no definitive detection of the H–D correlation. The second data release (DR2) of the EPTA with longer timing baselines and more pulsars was created to better constrain the CRS and potentially detect the H–D curve.

Using ~24 yr of data and six of the best-timed MSPs, EPTA's detection of the CRS had an amplitude of $2.95^{+0.89}_{-0.72} \times 10^{-15}$ for a fixed spectral index ($\gamma$) of 13/3 (cf. Eq. (2)) as expected for a GWB originating from a population of supermassive black hole binaries (Chen et al. 2021). Further investigations of the spectral properties of individual pulsars by Chalumeau et al. (2022) showed a clear need for pulsar-specific noise models and frequency binning. Furthermore, Goncharov et al. (2021, 2022) stated that it is likely that at least part of the CRS can arise from a superposition of independent RN processes in individual MSPs. With simulated datasets, Zic et al. (2022) demonstrated how spurious evidence for CRS arises based on the choice of priors. Thus, robust modelling of the noise within each MSP dataset is critical given the possibility that the CRS may be a subset of the GWB signal.

In addition to radio PTAs, an independent limit on the GWB was placed using gamma ray observations of MSPs (Fermi-LAT Collaboration 2022). This not only provides an independent handle on the properties of the GWB but also, since gamma rays are immune to the effects of the ionised interstellar medium (IISM), it offers the possibility of testing the efficacy of current chromatic models through joint radio and gamma-ray

analysis. Therefore, it is natural for such analyses to be integrated into future IPTA data releases.

In this paper, we focus on the long-term timing noise budget of pulsars in the second data release of the EPTA (hereafter, DR2full). This contains ~25 yr of data for 25 of the most precisely timed MSPs in the EPTA DR1. This new dataset will add significant sensitivity at the lowest Fourier frequencies of the GWB, owing to its long time span, and increased sensitivity at the higher Fourier frequencies due to improved observing cadence and an effective doubling of the observing bandwidth. We also combine low radio frequency pulsar timing data from the InPTA for 10 MSPs that are common to the EPTA DR2 (hereafter, DR2full+), to investigate the effects of the IISM and their impact on the total noise budget. For details on the EPTA and InPTA datasets, observing systems, and pulsar timing parameters, refer to EPTA Collaboration (2023).

The paper is structured as follows. In Sect. 2, we provide an overview of the noise budget, the descriptions of techniques used to estimate optimal Fourier frequency bins, and the selection of favoured noise models. In Sect. 3, we present the estimated noise budget for the EPTA and InPTA datasets and highlight interesting cases. In Sect. 4, we interpret our results, compare them with previous analyses, and discuss implications for GWB searches. In Sect. 5, we discuss the efficacy of our noise models through additional tests, simulations, and cross-validation with independent software packages. In Sect. 6, we present conclusions and provide potential future trajectories to explore, especially by focussing on IPTA datasets. The customised noise models reported in this paper are applied directly to the EPTA's search for spatially correlated GWB signals (EPTA Collaboration & InPTA Collaboration 2023).

## 2. Noise modelling techniques and noise budget

The assumption that TOAs are only composed of a deterministic timing model plus time-uncorrelated instrumental noise is generally untrue in pulsar timing datasets. Stochastic processes, ranging from intrinsic spin noise in the neutron star to chromatic variations resulting from the IISM, contribute additional time-correlated noise in the observed TOAs (Cordes & Shannon 2010). Unless this noise is accounted for, it can bias timing model parameter estimation and reduce the sensitivity to the GWB. The typical method to account for this noise is to solve for both the pulsar timing model and a stochastic Gaussian-process noise model simultaneously (Lentati et al. 2013; van Haasteren & Vallisneri 2014). This is achieved by the use of Bayesian parameter estimation routines to find the posterior probability distribution of the noise model hyperparameters and the pulsar timing model parameters. In practice, significant computation time can be saved by analytically marginalising over most or all of the pulsar timing model parameters. This is achieved by reusing the linearised approximation to the timing model adopted in traditional pulsar timing software such as `tempo2`, which significantly reduces the parameter space to be sampled in the Bayesian analysis. The preparation of the parameters of the pulsar timing model, including the choice of parameters included in the fit, is discussed in EPTA Collaboration (2023).

### 2.1. Description of the noise models

For the work discussed in this paper, we focus primarily on the standard noise models used widely within the IPTA

(Lentati et al. 2013, 2014 and references therein). In brief, we model the excess noise in the pulsar as a combination of Gaussian processes with two main forms, either correlated in time (i.e., 'red' noise) or uncorrelated in time (that is, 'white' noise). Red noise usually dominates on timescales of years to decades and incorporates signals with power spectral densities similar to the GWB. These processes include both chromatic noise, which depends on observing frequency, and achromatic noise, which is independent of observing frequency. We generally consider one achromatic red timing noise and two chromatic processes, variation in the dispersion measure (DM) to the pulsar, and variation in the scattering timescale for the pulsar. White noise reflects unmodelled instrumental errors and intrinsic pulse jitter (Liu et al. 2011, 2012; Parthasarathy et al. 2021) in the arrival times.

### 2.1.1. Time-correlated noise

Pulsar timing noise represents stochastic irregularities in pulsar rotation. Persistent temporally correlated noise that manifests equally across the radio frequency band of the instrument is referred to as spin noise, an achromatic noise process. It is typically modelled as a wide-sense stationary stochastic signal, that is, a process with zero mean and a covariance function that depends only on the absolute time lag between two points. Timing noise is present across the pulsar population and its amplitude seems to vary as a function of the pulsar spin-down rate (Shannon & Cordes 2010). Although the origins of spin noise are not unanimously agreed upon, it is typically well-modelled with a power-law spectrum. Power-law behaviour is expected if spin noise originates due to interactions between the neutron star crust and its superfluid core (Jones 1990), although observations of several canonical pulsars (that is, non-MSPs) show a quasi-periodic behaviour in timing noise properties (Lyne et al. 2010), or spectral turnovers (Parthasarathy et al. 2019) that may warrant additional terms beyond a single power law. However, the relatively small amplitudes of spin noise seen in MSPs is found to be relatively well-modelled by a power-law process (Goncharov et al. 2020).

The IISM introduces a wide range of chromatic noise processes into the TOAs that are dependent on the observing frequency. These include dispersive delays, scintillation, and pulse profile broadening due to multipath propagation (Cordes et al. 2016; Shannon & Cordes 2017; Donner et al. 2020). Dispersive delays can become measurable on timescales of days to weeks and scale with the observing radio frequency as $\nu^{-2}$. DM variations are one of the biggest sources of unmodelled error, and although the wider bandwidths and higher cadence observations of recent PTA datasets significantly improves the ability to model the DM variations and separate them from achromatic signals, the inhomogeneity of the datasets means that it can be difficult to separate chromatic and achromatic noise on the longest timescales. The best approach to account for DM variability depends on the underlying dataset. For EPTA, we currently favour using a stochastic power-law model for the DM variations, which assumes that there is a stationary smooth process that determines the DM. This allows for observations separated in both time and observing frequency to be used for constraining the DM. The theoretical expectation is that DM variations are caused by Kolmogorov turbulence in the IISM, and hence will have a power law index of $\gamma_{DM} = 8/3$ (Foster & Cordes 1990). Alternative models (such as the use of `DMX` parameters; Alam et al. 2021), which typically make independent measurements of DM over discrete time intervals, avoid the assumption of smooth variation and stationary process, but require near-simultaneous observations at a wide range of frequencies to be effective. Additionally,

**Table 1.** Noise model constants and hyperparameter priors.

| Parameter | Red | DM | Scatter |
|---|---|---|---|
| $K_{scale}$ | $12\pi^2$ | $k_{DM}^2$ | $12\pi^2$ |
| $\alpha$ | 0 | 2 | 4 |
| Prior($\log_{10}(A)$) | $\mathcal{U}(-18, -10)$ | $\mathcal{U}(-18, -10)$ | $\mathcal{U}(-18, -10)$ |
| Prior($\gamma$) | $\mathcal{U}(0, 7)$ | $\mathcal{U}(0, 7)$ | $\mathcal{U}(0, 7)$ |

**Notes.** All priors are uniform between the specified bounds. The term $k_{DM} = 2.41 \times 10^{-4}$ cm$^{-3}$ pc MHz$^2$s$^{-1}$ is the DM constant.

there is a variable contribution to the DM from the solar wind. A study of the time-variable solar wind is part of an ongoing EPTA project (Niţu et al., in prep.). For this work, we use a fixed value of 7.9 cm$^{-3}$ (Madison et al. 2019) for all pulsars, except for PSR J1022+1001 where we fit a constant solar wind amplitude[1] (see also EPTA Collaboration 2023).

After DM variations, the second most prominent effect of the interstellar medium at typical observing frequencies is the variation of the pulsar's scattering timescale. This scales as $\nu^{-4}$ and therefore can be separated from DM given enough coverage in observing frequency. Variation in scattering timescale is modelled as a power law in the few cases where it is important for the pulsars in our current dataset.

For each of our time-correlated noise processes, we model the noise of a process with chromatic index $\alpha$ for a TOA at time $t$ and observing frequency $\nu$ ($\nu_{ref}$ is the reference frequency set at 1400 MHz) with a Fourier basis of $N_{coef}$ coefficients as,

$$y(t) = \sum_{j=1}^{N_{coef}} Y_j \left( a_j \cos(j\omega t) + b_j \sin(j\omega t) \right) \left( \frac{\nu}{\nu_{ref}} \right)^{-\alpha}, \quad (1)$$

where $\omega = 2\pi/T_{span}$ for $T_{span}$ typically chosen to be the total observing time span, $a_j$ and $b_j$ are fit parameters with a Gaussian prior $\mathcal{N}(0, 1)$ and $Y_j$ determined by the noise model hyperparameters $A$ and $\gamma$. Specifically, we define $Y_j$ as

$$Y_j = \sqrt{\frac{A^2}{K_{scale}} \frac{s_{yr}^3}{T_{span}} \left( \frac{f_j}{f_{yr}} \right)^{-\gamma}}, \quad (2)$$

where $f_{yr} = 1$ yr$^{-1}$, $s_{yr} = 31557600$ s yr$^{-1}$ converts years to seconds with $T_{span}$ in seconds, and $K_{scale}$ is a scale that can adjust the units of $A$ appropriately for the given noise process. In practise, the parameters $a_j$ and $b_j$ are analytically marginalised following the method described in Lentati et al. (2014) and implemented in `temponest` and `enterprise`.

The choice of constants for each process, red noise, DM noise and scattering variations, and the hyperparameter priors are given in Table 1. Achromatic red noise is scaled so that $A_{red}$ is the equivalent GWB amplitude in yr$^{3/2}$, the DM variations are given in `temponest` units of cm$^{-3}$pc yr$^{3/2}$s$^{-1}$, and the scattering variations are given as the equivalent amplitude of the red noise (in yr$^{3/2}$) at 1400 MHz. These units are chosen to match previous publications and can be used directly in `tempo2`.

### 2.1.2. White noise

Temporally uncorrelated white noise in pulsar timing residuals needs to be modelled to effectively estimate the precision of pulsar timing parameters. For this work, we include the widely

---

[1] The solar wind amplitude is measured at 1 au.

**Table 2.** Parameter priors for the two exponential dips included in the noise models for PSR J1713+0747.

| Parameter | Prior |
|---|---|
| $A$ (s) | $\log_{10}\mathcal{U}(10^{-10}, 10^{-2})$ |
| $\tau$ (days) | $\log_{10}\mathcal{U}(1, 10^{2.5})$ |
| $t_0$ (1st event) [MJD] | $\mathcal{U}(54650, 54850)$ |
| $t_0$ (2nd event) [MJD] | $\mathcal{U}(57490, 57530)$ |
| $\alpha$ (1st event) | 4 (fixed value) |
| $\alpha$ (2nd event) | 1 (fixed value) |

adopted parameters[2] EFAC and EQUAD where the diagonals of the noise covariance matrix are scaled by

$$\sigma_{\text{scaled}}^2 = \text{EFAC}^2 \times \sigma_{\text{original}}^2 + \text{EQUAD}^2, \tag{3}$$

where $\sigma_{\text{original}}$ is the measurement uncertainty of a TOA due to template-fitting errors. These parameters are applied for every observing backend and are tagged with the `-group` flag in the EPTA dataset. We use uniform priors on EFAC (0.1 to 5) and $\log_{10}(\text{EQUAD})$ (−9 to −5).

### 2.1.3. Exponential dips

The exponential dips are signals manifesting as a sudden radio frequency-dependent advance of pulse arrival times and are estimated to impact the measurements of time-correlated signals (Hazboun et al. 2020). Such events have been reported three times for PSR J1713+0747 at MJDs ∼54757(2008) (Coles et al. 2015; Zhu et al. 2015; Desvignes et al. 2016), ∼57510(2016) (Lam et al. 2018; Goncharov et al. 2021; Chalumeau et al. 2022) and ∼59320(2021) (Xu et al. 2021; CHIME/Pulsar Collaboration 2021; Singha et al. 2021). Only the first two events are part of DR2full. This effect can be explained by a sudden drop in the density of the IISM electron column along the line of sight to the pulsar and would induce an effect on the timing residuals with a chromatic index of 2. It could also be caused by a magnetospheric process inducing a profile shape change with a different chromatic index than for a DM event (Shannon et al. 2016; Goncharov et al. 2021).

The model commonly used in `enterprise` describes the exponential dips as a deterministic signal with a waveform $d_{\text{exp}}$ expressed for any observing time $t_i > t_0$ and radio frequency $\nu_k$ as

$$d_{\text{exp}}(t_i, \nu_k ; \theta_{\text{exp}}) = A \left( \frac{\nu_k}{\nu_{\text{ref}}} \right)^{-\alpha} \exp\left( -\frac{t_i - t_0}{\tau} \right), \tag{4}$$

with the reference frequency $\nu_{\text{ref}}$ set at 1.4 GHz and the parameters $\theta_{\text{exp}}$ being the amplitude ($A$; in residual units), the relaxation time ($\tau$; in days), the initial epoch ($t_0$; in MJD) and the chromatic index ($\alpha$). The prior probability distributions are described in Table 2. For the analyses in this work and similarly to Chalumeau et al. (2022), we fix the exponential dips $\alpha$ of the first and second events, respectively, at 4 and 1, chosen from their marginalised posteriors evaluated a priori.

### 2.2. Bayesian inference for pulsar timing noise analysis

We use a Bayesian approach (e.g. Sivia & Skilling 2006) for the parameter estimation and model selection. Given the measured

---

timing residuals $\delta t$, the parameters $\theta_i$ of a chosen model $\mathcal{M}_i$ are considered random variables with a probability density function (parameter posteriors) described with the Bayes theorem as follows.

$$p(\theta_i \mid \delta t, \mathcal{M}_i) = \frac{p(\delta t \mid \theta_i, \mathcal{M}_i)\, p(\theta_i \mid \mathcal{M}_i)}{p(\delta t \mid \mathcal{M}_i)}, \tag{5}$$

where $p(\theta_i \mid \mathcal{M}_i)$, $p(\delta t \mid \theta_i, \mathcal{M}_i)$ and $p(\delta t \mid \mathcal{M}_i)$ are the parameter prior, likelihood and marginal likelihood (or evidence), respectively. The prior distributions for the noise parameters are described in Sect. 2.1 and those for the timing model parameters are shown in EPTA Collaboration (2023).

We use a Gaussian likelihood (van Haasteren & Vallisneri 2015; Chalumeau et al. 2022), where the deterministic signal waveforms are directly reduced to the timing residuals and the stochastic time-correlated and uncorrelated components are included in the time-domain covariance matrix. The evidence corresponds to the marginalised likelihood, that is, the integral of the likelihood over the whole parameter space. This term is particularly useful for model selection, as described below.

Let us now rewrite the Bayes theorem to express the probability density function of a model $\mathcal{M}_i$ given the set of timing residuals $\delta t$

$$p(\mathcal{M}_i \mid \delta t) = \frac{p(\delta t \mid \mathcal{M}_i)\, p(\mathcal{M}_i)}{p(\delta t)}, \tag{6}$$

where $p(\mathcal{M}_i)$ is the prior of model $\mathcal{M}_i$, $p(\delta t \mid \mathcal{M}_i)$ is the evidence that also appears in Eq. (5) and $p(\delta t)$ is the probability of observing the timing residuals marginalised over all models. The odds ratio, that is, the probability ratio between two models $\mathcal{M}_1$ and $\mathcal{M}_2$ given the same data is defined as

$$\frac{p(\mathcal{M}_2 \mid \delta t)}{p(\mathcal{M}_1 \mid \delta t)} = \frac{p(\delta t \mid \mathcal{M}_2)}{p(\delta t \mid \mathcal{M}_1)} \frac{p(\mathcal{M}_2)}{p(\mathcal{M}_1)}. \tag{7}$$

Considering equal prior probability between all models (as in this work), Eq. (7) is reduced to the ratio of evidences, also termed the Bayes factor,

$$\mathcal{B}_{\mathcal{M}_1}^{\mathcal{M}_2} = \frac{p(\delta t \mid \mathcal{M}_2)}{p(\delta t \mid \mathcal{M}_1)}. \tag{8}$$

If $\mathcal{B}_{\mathcal{M}_1}^{\mathcal{M}_2} > 1$, the model $\mathcal{M}_2$ is preferred over $\mathcal{M}_1$ given the measured timing residuals $\delta t$ considered as the data. We use the scale proposed in Kass & Raftery (1995) to interpret and make decisions about the selection of the model and include an additional noise component in a simpler model only if the related Bayes factor is equal to or greater than 150. In this work, most of the Bayes factors are evaluated following a product-space sampling approach (Carlin & Chib 1995; Hee et al. 2015; Taylor et al. 2020), also referred to as the HyperModel method, where an additional hyperparameter is added to switch between two or more models. The Bayes factor between the two models becomes equivalent to the ratio of the number of samples between the models.

### 2.3. Noise modelling packages and samplers

For the single-pulsar noise analysis results reported in this paper, we use two Bayesian pulsar timing analysis packages, `temponest` (Lentati et al. 2014) and `enterprise`[3]

---

[2] However, some PTA collaborations use different notations.

[3] https://gitlab.in2p3.fr/epta/enterprise/-/tree/master/enterprise

(Ellis et al. 2019). Both implement the noise models as described in Sect. 2.1. Although both codes have been tested with a range of samplers, for this work, the sampling for `temponest` is performed using Multinest (Feroz & Hobson 2008) and for `enterprise` is performed using PTMCMCSampler (Ellis & van Haasteren 2017). We have made several optimisations to `temponest`[4], including adding a scattering delay variation model.

## 2.4. Choosing the optimal number of Fourier coefficients

We model the red noise, the DM variations and the scattering variations as stationary Gaussian processes, following the Fourier-sum approach described in van Haasteren & Vallisneri (2014), with a discrete and finite set of sine/cosine basis functions and a power-law power spectral density (PSD). The set of frequencies for each Gaussian process is chosen to be linearly distributed as $1/T_{\rm span}, 2/T_{\rm span}, ..., N_{\rm coef}/T_{\rm span}$. Most PTA analyses in the past used the same number of frequency bins $N_{\rm coef}$ for a given signal for all pulsars. However, different types of covariance expansions have been discussed in van Haasteren & Vallisneri (2015), and methods to select a customised number of frequency bins have been proposed, either from the transitional frequency of a "red noise – white noise" broken power-law PSD (Chen et al. 2021), or from selecting the preferred value from a model selection among chosen sets of frequency bins (Chalumeau et al. 2022). In this work, we propose a novel method to select the number of frequency components for Gaussian processes by setting $N_{\rm coef}$ as a free parameter in the noise model, and thus evaluating a marginalised posterior distribution $p(N_{\rm coef}|\,\delta t)$. This approach extends the method performed in Chalumeau et al. (2022) by enabling tests for any possible value in a selected range of frequencies without having to evaluate a Bayes factor between all possible models. As the prior is the same between two Gaussian processes with a different number of frequency bins, this method is equivalent to a Bayes factor evaluation from a product-space approach.

We decide to set a non-informative prior for the discrete parameter $N_{\rm coef}$. The method implemented consists of including a real hyperparameter $\tilde{N}_{\rm coef}$ with a uniform prior $\mathcal{U}\left([\tilde{N}_{\rm min}, \tilde{N}_{\rm max} + 1[\right)$, and setting the number of frequency bins $N_{\rm coef}$ as the integer floor value $\lfloor \tilde{N}_{\rm coef} \rfloor$, thus ensuring an equal probability between all frequencies inside the prior range. The prior limits for the three stochastic signals considered in this work (red noise, DM, and scattering delay variations) are $\tilde{N}_{\rm min} = 10$ and $\tilde{N}_{\rm max} = 150$. The former is empirically chosen for statistical reasons. The maximum value of 150 allows testing for frequencies up to $\sim 33$ days$^{-1}$ and $\sim 60$ days$^{-1}$ for pulsars with the shortest and longest baselines, respectively, where white noise is expected to dominate over other signals. The value of the upper edge prior for the truncated dataset (DR2new; cf. Sect. 3) is $\tilde{N}_{\rm max} = 100$ to cover frequencies up to 37–33 day$^{-1}$ for all pulsars.

In case of an inconclusive (that is, flat) posterior distribution, we select a minimal value and perform an additional Bayes factor evaluation between the maximum posterior value and the selected minimal value as a performance check. By minimising the number of Fourier modes, we aim at selecting only the low-frequency noise expected to follow a power-law PSD, and reduce the effects of excess noise at high frequencies in the white noise dominated range. Furthermore, this step allows to reduce

---

[4] https://github.com/aparthas3112/Temponest

the computational cost of the likelihood which becomes crucial for a gravitational-wave analysis where the Gaussian process hyperparameters for all the pulsars in the array are simultaneously sampled.

## 2.5. Model selection and parameter estimation methodology

For each pulsar, we perform a Bayesian model selection to evaluate the most favoured combination among a chosen list of time-correlated noise components (cf. Sect. 2.1). The selection between two models is based on a Bayes factor evaluation performed with `enterprise` using the HyperModel method. For these analyses, we marginalise across all timing model parameters and fit the remaining parameters. For PSR J1713+0747, we also include the exponential dips described in Sect. 2.1.3. After obtaining the customised noise models, a full Bayesian analysis is performed with `temponest` that fits all the timing and noise model parameters. The results of these analyses are reported in Sect. 3. The procedure used to obtain the customised noise models is described in the following part of this section.

The six candidate noise models include any combination of achromatic red noise (RN), DM variations (DM), and scattering delay variations (SV). These are NONE (no time-correlated noise), RN, DM, RN+DM, DM+SV and RN+DM+SV (as reported in Tables 3 and 4. In this work, we do not consider models including SV but without DM, as they are physically unlikely at the observed radio frequencies in our dataset. We first evaluate the number of frequency bins that are the most preferred for each candidate noise model following the method described in Sect. 2.4. In the second step, we select the most favoured model among the six candidates with a Bayes factor evaluation. Following the principle of Occam's razor, we select the simplest model (that is, the model with the lowest number of Gaussian process signals) in case of inconclusive results.

After obtaining the final customised models, we perform additional checks to consolidate the robustness of the results described in Sect. 5. This step has been particularly useful in revisiting the noise model for PSR J1022+1011, which exhibited unmodelled achromatic red noise in the low frequencies of the power spectrum. A deeper analysis and a more advanced noise model would likely be required for this pulsar. Therefore, we stick to a standard version of the noise model for this pulsar, including RN and DM with 30 and 100 PSD frequencies, respectively.

# 3. Estimated noise budget for the EPTA DR2 pulsars

Here, we present the results of the noise model selection and parameter estimation for two EPTA datasets (to assist the reader, Table 5 provides a summary of the various datasets used in this paper),

- The first is DR2full, containing data from the EPTA DR1 and the new instrumentation in DR2.
- The second dataset consists only of the EPTA DR2 data from the new instrumentation (starting from ~MJD 55611; hereafter, DR2new).
- The third is DR2full+ which contains the data described as DR2full along with additional low-frequency data from the InPTA collaboration for 10 common MSPs.

In principle, the combined EPTA+InPTA dataset (DR2full+) is much more sensitive for measuring the pulsar noise parameters as well as the GWB, with a total time span of 15–25 yr

**Table 3.** Favoured models listed for the 25 pulsars using both EPTA and InPTA data (for 10 common MSPs; DR2full+) along with estimated values for chromatic and achromatic noise models.

| Pulsar | PTA | Favoured models | Red noise | | | DM noise | | | Time span |
|--------|-----|-----------------|-----------|---|---|----------|---|---|-----------|
| | | | $N_{coef}$ | $A$ | $\gamma$ | $N_{coef}$ | $A$ | $\gamma$ | yr |
| J0030+0451 | EPTA | RN | 10 | $-14.93^{+0.83}_{-1.1}$ | $5.49^{+1.93}_{-1.56}$ | X | X | X | 21.96 |
| J0613−0200 | EPTA+InPTA | RN+DM | 10 | $-14.99^{+0.94}_{-1.24}$ | $5.34^{+2.06}_{-1.6}$ | 129 | $-11.58^{+0.06}_{-0.06}$ | $1.34^{+0.28}_{-0.26}$ | 23.83 |
| J0751+1807 | EPTA+InPTA | DM | X | X | X | 115 | $-11.72^{+0.2}_{-0.2}$ | $2.69^{+0.51}_{-0.49}$ | 25.12 |
| J0900−3144 | EPTA | RN+DM | 135 | $-12.76^{+0.09}_{-0.08}$ | $1.06^{+0.28}_{-0.27}$ | 150 | $-11.94^{+0.67}_{-0.87}$ | $3.89^{+2.12}_{-1.79}$ | 13.64 |
| J1012+5307 | EPTA+InPTA | RN+DM | 149 | $-13.03^{+0.05}_{-0.04}$ | $1.21^{+0.17}_{-0.17}$ | 47 | $-11.95^{+0.11}_{-0.12}$ | $1.74^{+0.39}_{-0.37}$ | 24.61 |
| J1022+1001 | EPTA+InPTA | RN+DM | 30 | $-13.8^{+0.51}_{-0.99}$ | $3.01^{+1.55}_{-0.97}$ | 100 | $-11.46^{+0.09}_{-0.08}$ | $0.14^{+0.26}_{-0.13}$ | 25.37 |
| J1024−0719 | EPTA | DM | X | X | X | 34 | $-11.82^{+0.18}_{-0.21}$ | $2.46^{+0.87}_{-0.66}$ | 23.14 |
| J1455−3330 | EPTA | RN | 49 | $-13.26^{+0.28}_{-0.49}$ | $2.21^{+1.35}_{-1.04}$ | X | X | X | 15.72 |
| J1600−3053 | EPTA+InPTA | RN+DM | 21 | $-14.05^{+0.49}_{-0.89}$ | $2.86^{+1.99}_{-1.24}$ | 148 | $-11.46^{+0.04}_{-0.04}$ | $1.99^{+0.12}_{-0.12}$ | 15.42 |
| J1640+2224 | EPTA | DM | X | X | X | 145 | $-11.66^{+0.14}_{-0.13}$ | $0.48^{+0.49}_{-0.4}$ | 24.44 |
| J1713+0747 | EPTA+InPTA | RN+DM | 12 | $-14.19^{+0.27}_{-0.29}$ | $3.28^{+0.66}_{-0.63}$ | 148 | $-11.86^{+0.05}_{-0.04}$ | $1.59^{+0.19}_{-0.19}$ | 24.5 |
| J1730−2304 | EPTA | DM | X | X | X | 10 | $-11.56^{+0.55}_{-0.57}$ | $2.22^{+1.56}_{-1.45}$ | 16.1 |
| J1738+0333 | EPTA | RN | 11 | $-12.93^{+0.36}_{-0.4}$ | $2.14^{+1.31}_{-1.2}$ | X | X | X | 14.12 |
| J1744−1134 | EPTA+InPTA | RN+DM | 10 | $-14.12^{+0.41}_{-0.72}$ | $3.45^{+1.19}_{-0.75}$ | 150 | $-11.82^{+0.1}_{-0.07}$ | $0.26^{+0.37}_{-0.23}$ | 25.14 |
| J1751−2857 | EPTA | DM | X | X | X | 41 | $-11.08^{+0.22}_{-0.33}$ | $2.13^{+0.99}_{-0.7}$ | 14.69 |
| J1801−1417 | EPTA | DM | X | X | X | 14 | $-10.73^{+0.27}_{-0.26}$ | $1.68^{+1.16}_{-1.06}$ | 13.71 |
| J1804−2717 | EPTA | DM | X | X | X | 38 | $-11.19^{+0.18}_{-0.83}$ | $0.78^{+2.95}_{-0.71}$ | 14.73 |
| J1843−1113 | EPTA | DM | X | X | X | 73 | $-11.03^{+0.08}_{-0.08}$ | $2.07^{+0.36}_{-0.31}$ | 16.8 |
| J1857+0943 | EPTA+InPTA | DM | X | X | X | 11 | $-11.86^{+0.27}_{-0.28}$ | $2.88^{+0.66}_{-0.62}$ | 25.11 |
| J1909−3744 | EPTA+InPTA | RN+DM | 20 | $-14.89^{+0.78}_{-0.85}$ | $4.77^{+1.96}_{-1.79}$ | 150 | $-11.85^{+0.05}_{-0.05}$ | $1.31^{+0.16}_{-0.15}$ | 17.14 |
| J1910+1256 | EPTA | DM | X | X | X | 10 | $-11.71^{+0.66}_{-0.84}$ | $2.98^{+2.38}_{-1.87}$ | 15.21 |
| J1911+1347 | EPTA | DM | X | X | X | 10 | $-11.98^{+0.39}_{-0.47}$ | $3.06^{+1.36}_{-1.06}$ | 14.2 |
| J1918−0642 | EPTA | DM | X | X | X | 138 | $-12.09^{+0.4}_{-0.44}$ | $3.49^{+1.13}_{-1.06}$ | 19.71 |
| J2124−3358 | EPTA+InPTA | DM | X | X | X | 18 | $-11.77^{+0.34}_{-0.39}$ | $2.07^{+1.09}_{-0.98}$ | 17.15 |
| J2322+2057 | EPTA | NONE | X | X | X | X | X | X | 14.68 |

**Notes.** The second column lists the PTAs that contributed to the dataset for each pulsar. The third column reports the favoured model. Columns 4 to 9 report the estimated number of coefficients, amplitude, and slope (medians with 95% confidence intervals) for the achromatic and chromatic noise processes, respectively. The last column reports the total time span for each of the pulsars.

for most pulsars. However, around half of these data are from instruments that were designed prior to the start of the PTA experiment, and in some cases lacked coherent dedispersion or were otherwise limited in time resolution on the fastest pulsars. Furthermore, in some cases the early data do not have good coverage in observing frequency, making it difficult to disentangle chromatic and achromatic noise processes. Hence, although the DR2new is only 10 yr long, it may contain fewer systematics and degeneracies in the noise models. Indeed, as shown in EPTA Collaboration & InPTA Collaboration (2023), DR2new appears to be favourable in sensitivity to the GWB, which may suggest that the noise modelling is insufficient for early EPTA data, or that the characteristics of the pulsar noise or GWB signal are varying with time. Further investigation of the time-stationary properties of noise in each pulsar is briefly discussed in Sect. 4.1 and should be a priority for future IPTA data analysis.

In Table 3, we present the noise budget for 25 MSPs in DR2full+, while Table 4 shows the same results when

applied to DR2new. For each pulsar, we report the properties of the preferred noise model estimated from the method described in Sect. 2.5 and discuss the interpretations of the results in Sect. 4.1.

For the DR2full+ data, we found that RN, DM and RN+DM are favoured for three, thirteen and eight pulsars respectively, and no significant time-correlated noise was evident in PSR J2322+2057. The data for fourteen pulsars do not support the inclusion of achromatic red noise, and this is further investigated in Sect. 5.1. However, including these pulsars in a GWB search analysis is likely still important for the optimisation of the PTA sensitivity as we include the spatial correlations among pulsars. For these pulsars, estimates of DM variations in a multi-pulsar analysis with spatial correlations are largely consistent with single-pulsar results with a few exceptions.

Of the 21 pulsars with significant DM variations, twelve have constrained power laws with a spectral index consistent with the expected $\gamma_{DM} = 8/3$ from Kolmogorov turbulence in IISM. Of

**Table 4.** Same as in Table 3, but for DR2new.

| Pulsar | Favoured models | Red noise | | | DM noise | | | Time span |
|---|---|---|---|---|---|---|---|---|
| | | $N_{coef}$ | $A$ | $\gamma$ | $N_{coef}$ | $A$ | $\gamma$ | yr |
| J0030+0451 | RN | 10 | $-14.72^{+1.26}_{-1.29}$ | $4.6^{+2.27}_{-3.36}$ | X | X | X | 9.78 |
| J0613−0200 | **DM** | X | X | X | 11 | $-11.79^{+0.34}_{-1.0}$ | $1.14^{+3.23}_{-1.08}$ | 10.07 |
| J0751+1807 | DM | X | X | X | 92 | $-11.8^{+0.32}_{-0.82}$ | $2.9^{+2.31}_{-1.04}$ | 10.08 |
| J0900−3144 | RN+DM | 99 | $-12.72^{+0.12}_{-0.13}$ | $1.19^{+0.44}_{-0.4}$ | 100 | $-11.81^{+0.73}_{-1.22}$ | $3.78^{+2.91}_{-2.18}$ | 9.86 |
| J1012+5307 | RN+DM | 92 | $-12.99^{+0.12}_{-0.1}$ | $1.25^{+0.37}_{-0.34}$ | 16 | $-12.66^{+0.78}_{-1.55}$ | $3.93^{+2.86}_{-2.53}$ | 10.08 |
| J1022+1001 | RN+DM | 30 | $-15.42^{+1.9}_{-2.44}$ | $3.79^{+3.04}_{-3.56}$ | 100 | $-11.4^{+0.15}_{-0.17}$ | $0.4^{+0.45}_{-0.38}$ | 10.08 |
| J1024−0719 | DM | X | X | X | 15 | $-11.79^{+0.31}_{-0.41}$ | $2.47^{+1.68}_{-1.28}$ | 10.08 |
| J1455−3330 | RN | 80 | $-13.28^{+0.27}_{-1.74}$ | $1.52^{+4.59}_{-1.37}$ | X | X | X | 9.27 |
| J1600−3053 | **DM+SV** | X | X | X | 97 | $-12.32^{+0.5}_{-0.78}$ | $4.51^{+2.15}_{-1.5}$ | 9.91 |
| J1640+2224 | DM | X | X | X | 47 | $-11.57^{+0.17}_{-0.21}$ | $0.38^{+0.75}_{-0.36}$ | 10.22 |
| J1713+0747 | RN+DM | 10 | $-13.73^{+0.37}_{-0.62}$ | $1.79^{+2.25}_{-1.42}$ | 73 | $-12.17^{+0.13}_{-0.13}$ | $1.36^{+0.51}_{-0.44}$ | 10.08 |
| J1730−2304 | DM | X | X | X | 11 | $-11.91^{+0.49}_{-0.75}$ | $2.57^{+2.6}_{-2.04}$ | 9.93 |
| J1738+0333 | **DM** | X | X | X | 24 | $-11.2^{+0.26}_{-0.57}$ | $1.84^{+2.09}_{-1.09}$ | 9.98 |
| J1744−1134 | **DM** | X | X | X | 87 | $-11.71^{+0.12}_{-0.11}$ | $1.13^{+0.46}_{-0.36}$ | 9.75 |
| J1751−2857 | DM | X | X | X | 25 | $-11.0^{+0.21}_{-0.61}$ | $1.06^{+2.67}_{-0.99}$ | 9.43 |
| J1801−1417 | DM | X | X | X | 11 | $-10.79^{+0.35}_{-0.37}$ | $1.64^{+1.65}_{-1.3}$ | 9.74 |
| J1804−2717 | DM | X | X | X | 23 | $-11.15^{+0.2}_{-1.66}$ | $0.58^{+4.68}_{-0.55}$ | 9.54 |
| J1843−1113 | DM | X | X | X | 95 | $-11.06^{+0.1}_{-0.1}$ | $2.37^{+0.55}_{-0.43}$ | 10.08 |
| J1857+0943 | DM | X | X | X | 20 | $-12.55^{+0.62}_{-0.66}$ | $4.92^{+1.87}_{-1.92}$ | 9.98 |
| J1909−3744 | RN+DM | 14 | $-15.13^{+0.86}_{-0.64}$ | $5.72^{+1.21}_{-2.2}$ | 95 | $-11.89^{+0.12}_{-0.1}$ | $1.63^{+0.36}_{-0.31}$ | 9.04 |
| J1910+1256 | DM | X | X | X | 20 | $-12.11^{+0.75}_{-0.9}$ | $4.47^{+2.31}_{-2.32}$ | 9.91 |
| J1911+1347 | DM | X | X | X | 15 | $-12.14^{+0.49}_{-1.31}$ | $3.5^{+2.71}_{-1.46}$ | 9.92 |
| J1918−0642 | DM | X | X | X | 75 | $-12.34^{+0.74}_{-1.04}$ | $4.24^{+2.44}_{-2.16}$ | 10.08 |
| J2124−3358 | DM | X | X | X | 24 | $-11.47^{+0.22}_{-1.75}$ | $1.23^{+4.17}_{-1.01}$ | 9.68 |
| J2322+2057 | NONE | X | X | X | X | X | X | 9.68 |

**Notes.** Pulsars for which the preferred noise models have changed from Table 3 are highlighted.

**Table 5.** Abbreviations for the different datasets used in this paper.

| Dataset | Referenced as |
|---|---|
| EPTA data with 25 MSPs | DR2full |
| EPTA data with only new backends | DR2new |
| DR2full + InPTA data | DR2full+ |

**Notes.** The InPTA dataset extends the timing baseline of the EPTA dataset from MJD 59385 to 59644. It also provides consistent multi-frequency coverage for a period of a few years, down to the low-end (~300 MHz) of the frequency range whose coverage is very sparse with EPTA data only.

the remaining nine pulsars, seven of them also include an achromatic red noise component in the favoured model, which could potentially impact the measurement of DM variations.

We report a significant measurement of achromatic red noise for eleven pulsars, with six consistent with the predicted spectral index $\gamma_{GWB} = 13/3$ from an idealistic GWB produced from GW-driven circular supermassive black hole binaries (SMBHBs). We find peculiar achromatic red noise for PSRs J0900−3144 and J1012+5307, with flat power-laws (resp. $\gamma_{RN} = 1.06^{+0.28}_{-0.27}$ and $\gamma_{RN} = 1.21^{+0.17}_{-0.17}$), displaying short-term correlated noise, as shown in Fig. 1.

For DR2new, we find significant achromatic noise only for seven pulsars, four of which are consistent with $\gamma_{GWB} = 13/3$. The lower number of pulsars favouring RN as compared to DR2full+ can either be due to the shorter timing baselines which reduce the sensitivity for long-term red noise signals or due to unmodelled noise present in the first half of the dataset. Of the 22 pulsars with significant DM, sixteen are consistent with $\gamma_{DM} = 8/3$. Of the remaining six, three pulsars have RN+DM as the favoured model. PSRs J1640+2224 and J1744−1134 display unusually low spectral indices at $0.38^{+0.75}_{-0.36}$ and $1.13^{+0.46}_{-0.36}$ respectively. Conversely, we report high DM variations spectral index for PSR J1600−3053 at $4.51^{+2.15}_{-1.5}$, for which the favoured noise model is DM+SV. More details on the noise properties for this

pulsar are described in Sect. 4.2. Figures for posterior distributions, time-domain realisations, and relevant tables generated for each pulsar are available on Zenodo[5].

# 4. Interpretation

## 4.1. Comparison of EPTA Noise Models

In this section, we compare the noise models between DR2full, DR2new, and the original models reported for the EPTA DR1 dataset from Caballero et al. (2016). We find that the pulsars largely fall into three categories:

– Category1: consistent noise models in the three datasets: Typically, DR2full improves over DR1 or DR2new, and both DR2full and DR2new do a better job of constraining chromatic noise than DR1.
– Category2: DR2full and DR2new find chromatic noise: Several pulsars in DR1 found only achromatic noise or were unable to distinguish between achromatic and DM noise, but the DR2full and DR2new datasets prefer models with only DM variations.
– Category3: complex noise models: In a handful of pulsars, including several of those with the best TOA precision, we find a more complex relationship between the noise models in the three datasets.

In Fig. 2, we present noise models for four pulsars as an example of belonging to either of the above categories. Below, we present a summary of the noise models in the three datasets for each of the 25 pulsars.

*PSR J0030+0541 (Category 1).* DR2full and DR2new largely agree, both with achromatic noise only. The DR2 posterior is better constrained. DR1 finds the same achromatic noise, but is unable to cleanly distinguish it from the DM noise.

*PSR J0613−0200 (Category 3).* DR2 and DR1 both see achromatic noise of similar amplitude, but the spectral index is unconstrained for DR1, whereas DR2 finds the achromatic noise to have an index >4. DR2new is unable to detect this noise, probably due to the relatively short data (∼10 yr), and only finds DM variations. The posterior for $\gamma_{DM}$ shifts somewhat between the three datasets but are largely consistent and slightly flatter than Kolmogorov.

*PSR J0751+1807 (Category 1).* All three datasets are largely consistent, finding only DM variations with a spectral slope consistent with Kolmogorov.

*PSR J0900−3144 (Category 3).* DR2 and DR2new posteriors are consistent, finding achromatic noise with a very flat spectral shape <2, and DM variations with an unconstrained spectral slope. The achromatic noise is much flatter than expected from classical power-law timing noise and may represent another noise process. The chromatic noise in DR1 data is consistent but unable to discern the achromatic noise.

*PSR J1012+5307 (Category 3).* DR2full has a tightly constrained posterior for DM and achromatic noise. The achromatic index is very flat <2 and appears to be dominated by the high-frequency power above 1/yr. DR2new finds similar achromatic noise, and similar amplitude DM variations, but is barely able to constrain $\gamma_{DM}$. We speculate that this is because the longer dataset of DR2full can rule out very steep DM variations given the absence of steep achromatic noise. DR1 finds achromatic

noise with a marginally steeper spectrum and finds an upper limit on DM variations just below that observed in the DR2full data.

*PSR J1022+1001 (Category 2).* DR2full and DR2new are consistent with each other and show a very flat spectrum DM variation and an achromatic noise with $\gamma_{red} \sim 4$. The flat-spectrum DM variation in this pulsar could likely be due to high-frequency residuals remaining from the solar wind contribution as this pulsar passes close to the ecliptic. A revised interpretation is given in Sect. 5.1, after performing a chromatic index evaluation for each time-correlated component. DR1 showed flat spectrum achromatic noise and little variation in DM. We suspect that this may be a leakage from DM variations due to the limited frequency coverage. We account for the solar wind by fitting for the NE-SW parameter as part of the timing model, which results in an estimated value of $10.9 \pm 0.3$ cm$^{-3}$ relative to the constant value of 7.9 cm$^{-3}$ in the other pulsars.

*PSR J1024−0719 (Category 2).* DR1 showed a marginal detection of DM variations with a steep spectral index. DR2full and DR2new both find only DM variations consistent with Kolmogorov and incompatible with the DR1 value. We find that the DR2full and DR2new models seem more plausible, but they could indicate the presence of smooth DM structures in the early DR1 data.

*PSR J1455−3330 (Category 1).* The three datasets are largely consistent. DR1 cannot distinguish between DM and achromatic noise, but the better frequency coverage of DR2full and DR2new finds only achromatic noise. The DR2full posteriors are more tightly constrained, particularly in $\gamma_{red}$.

*PSR J1600−3053 (Category 3).* DR1 has fairly flat spectrum DM variations, which are split into flat-spectrum scattering delay variations and steep (than Kolmogorov) DM variations in DR2full and DR2new.

*PSR J1640+2224 (Category 2).* DR1 is unable to distinguish between DM and achromatic noise. DR2full and DR2new find only DM variations with an extremely flat spectral index, $\gamma_{DM} < 1$. It is not clear what physical process would give rise to such a flat spectral index for DM variations.

*PSR J1713+0747 (Category 3).* Two DM events (exponential dips) at MJDs ∼54757 and one at MJD ∼57510 are observed for this pulsar in DR2full. The methodologies adopted to model these events are different between DR1 and DR2, but we find that this does not affect the red noise properties. However, from Fig. 2, it is clear that the red noise properties change considerably between DR1, DR2new, and DR2full with the red noise spectral index going from steep to shallow, respectively. We think that this could likely point to either our assumptions of stationarity being incorrect or processes not well-modelled by a power law.

*PSR J1730−2304 (Category 1).* Generally consistent between the three datasets, with DR1 unable to distinguish between DM and achromatic noise. DR2full and DR2new find only DM variations consistent with Kolmogorov.

*PSR J1738+0333 (Category 2).* DR1 does not find any noise. DR2full and DR2new find a flat-spectrum noise process, but the model selection prefers achromatic noise for DR2full and chromatic noise for DR2new. In both cases, the evidence is marginal for any one particular model.
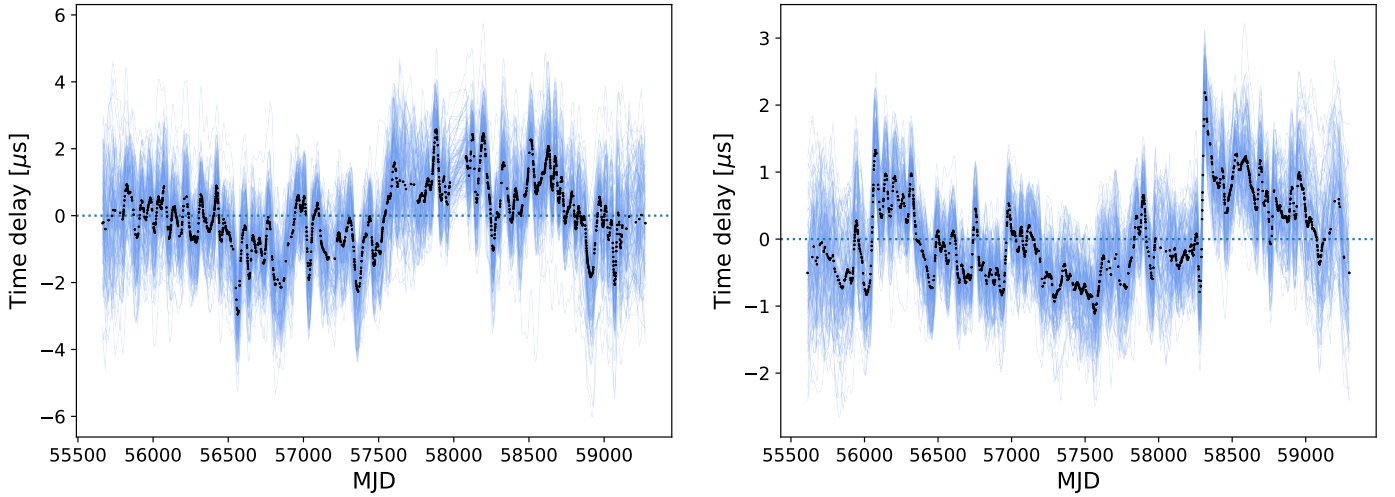
**Fig. 1.** 100 random realisations (blue) and medians for each epoch (black) of the RN time-delay reconstructions for PSRs J0900−3144 (left) and J1012+5307 (right) using the most favoured noise models with the DR2new dataset. The short-timescale stochastic signals seen at a μs level for these two pulsars still have unknown origins.
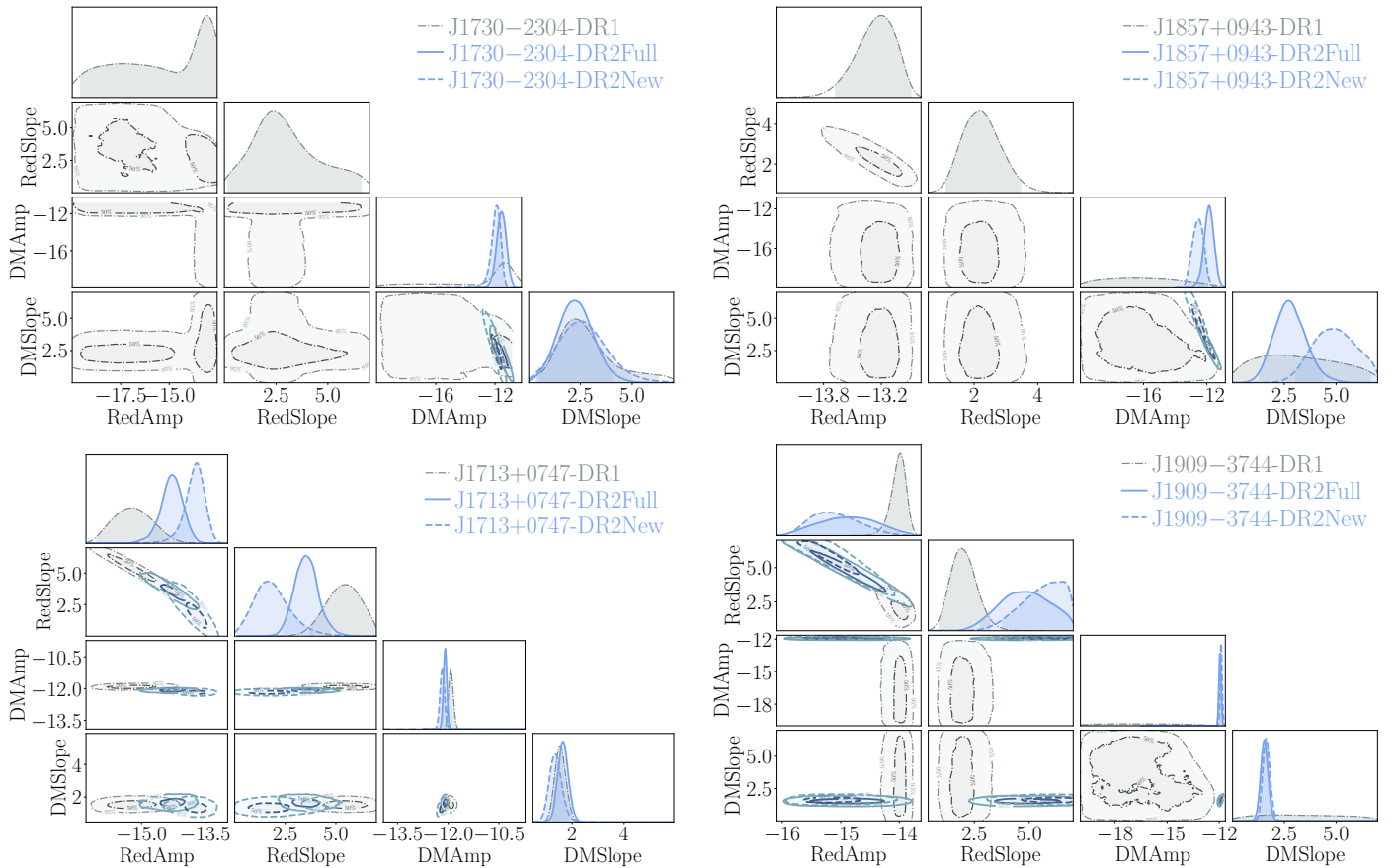


**Fig. 2.** Marginalised posterior distributions of noise hyperparameters obtained with the EPTA DR1, DR2full and DR2new. Top left: PSR J1730−2304 falls in Category 1, where the DR2full/New datasets are able to constrain chromatic noise (DMAmp, DMSlope). Top right: For PSR J1857+0943, belonging to Category 2, it is clear that the DR1 data could not disentangle chromatic from achromatic noise (RedAmp, RedSlope), which is resolved better with the DR2 datasets. Bottom left: PSR J1713+0747 in Category 3, shows clear inconsistencies in the achromatic noise estimates. Bottom right: PSR J1909−3744 shows a much more complicated behaviour as described in the text motivating its inclusion in Category 3.

*PSR J1744−1134 (Category 3).* The three datasets are superficially similar, but paint a confusing picture when taken together. DR1 finds no variations in DM and an achromatic process with $\gamma_{\rm red} \sim 3$. DR2full finds DM variations with $\gamma_{\rm DM} < 1$, amplitude above the upper limit for DM variations from DR1, and achromatic noise of similar amplitude but with a steeper, but largely unconstrained spectral slope, $\gamma_{\rm red} > 2.5$. Indeed, the DR2full posterior for the achromatic noise is bimodal with a component that is largely consistent with DR1 and a component that is much steeper ($\gamma_{\rm red} > 5$). DR2new does not find evidence of achromatic noise, but has consistent DM variations with DR2full. The very flat spectrum DM variations, inconsistent with DR1, suggest that the chromatic noise is not a power-law process but instead dominated by high-frequency variations to which DR1 was insensitive because of the limited instantaneous frequency coverage. The inconsistency of the achromatic noise may indicate there is another underlying process at work in this pulsar beyond a single power-law achromatic noise and DM variations. This pulsar is observed by all members of the IPTA, and hence we suspect that the picture will become clearer when studied with a combined IPTA dataset.

*PSR J1751−2857 (Category 1).* Generally consistent between the three datasets. DR1 only finds upper limits, but DR2full and DR2new find a DM variation consistent with Kolmogorov.

*PSR J1801−1417 (Category 1).* Generally consistent between the three datasets. DR2full and DR2new better constrain the DM variation, and the spectral shape is consistent with Kolmogorov.

*PSR J1804−2717 (Category 1).* Generally consistent between the three datasets. DR2full and DR2new much better constrain the DM variations, and the spectral shape is consistent with Kolmogorov, although both prefer a flatter $\gamma_{\rm DM}$.

*PSR J1843−1113 (Category 1).* Generally consistent between the three datasets. DR2full and DR2new much better constrain the DM variation, and the spectral shape is consistent with Kolmogorov.

*PSR J1857+0943 (Category 2).* DR1 finds achromatic noise that DR2full and DR2new attribute to DM variations, without any achromatic noise. There is some overlap in the DM posteriors for DR1 and DR2full, though we suspect that the detection of achromatic noise in DR1 was incorrect.

*PSR J1909−3744 (Category 3).* DR1 finds achromatic noise that seems to have the same properties as the DM variations in the DR2full/DR2new data. The DM variations in DR2full/DR2new are above the upper limit from DR1. DR1 had minimal frequency coverage for this pulsar, though it is not clear why achromatic noise was preferred if there was a degeneracy. DR2full and DR2new find achromatic noise with a steeper spectral index and at a lower amplitude than the DM variations in the DR1 data. We note that the noise analysis in Liu et al. (2020) which used very similar EPTA data as in this work, reported similar results when the L-band data were divided into sub-bands.

*PSR J1910+1256 (Category 1).* Generally consistent between the three datasets. DR1 only finds upper limits, while DR2full and DR2new find DM variations consistent with Kolmogorov, but $\gamma_{\rm DM}$ is not well constrained.

*PSR J1911+1347 (Category 2).* Generally consistent between the three datasets. DR1 is not able to distinguish between DM and achromatic noise, but DR2full and DR2new find a DM variation consistent with Kolmogorov.

*PSR J1918−0642 (Category 2).* DR1 finds achromatic noise with a preference for large $\gamma_{\rm red}$, and does not find significant DM variations. DR2full does not find evidence for achromatic noise, but instead finds a DM variation consistent with Kolmogorov, although also consistent with a steeper spectrum. DR2new is consistent with DR2full. This is another example of achromatic noise in DR1 being interpreted as chromatic noise when wide-band receivers are used.

*PSR J2124−3358 (Category 2).* Generally consistent between the three datasets. DR1 only finds upper limits, but DR2full and DR2Bew find a variation in DM consistent with Kolmogorov, though preferring a flatter $\gamma_{\rm DM}$.

*PSR J2322+2057.* Does not show evidence of time-correlated noise in any EPTA dataset.

### 4.2. Changes in noise models after the inclusion of the InPTA data

Here we study the impact of including low radio frequency observations from the InPTA on the estimated noise models. The InPTA dataset complements the EPTA data with simultaneous observations at 300–500 MHz and at 1260–1460 MHz between MJDs 58235 and 59496 observed with the upgraded Giant Metrewave Radio Telescope (uGMRT). The frequency coverage at 300–500 MHz is particularly important since EPTA has a limited number of observations at this frequency band. Therefore, the inclusion of the InPTA data is of particular interest in constraining noise due to the IISM such as DM and scattering variations. To allow a quantitative comparison of the posterior noise models before and after the inclusion of InPTA data, we adapt and employ a tension metric package as detailed in Raveri & Doux (2021)[6].

We begin by introducing the probability density for parameter differences between the posteriors of two sets of parameters $\theta_1$ and $\theta_2$ to be $\mathcal{P}(\Delta\theta)$ where $\Delta\theta \equiv \theta_1 - \theta_2$. Employing a joint posterior $\mathcal{P}(\theta_1, \theta_2) \equiv \mathcal{P}(\theta_1, \theta_2 | d_1, d_2)$, we can write $\mathcal{P}(\theta_1, \Delta\theta) = \mathcal{P}(\theta_1, \theta_1 - \Delta\theta)$, where $d_1$ and $d_2$ refer to the two data sets. Integrating the base parameters, we obtain

$$\mathcal{P}(\Delta\theta) = \int_{V_\pi} \mathcal{P}(\theta_1, \theta_1 - \Delta\theta) \, d\theta_1, \tag{9}$$

where $V_\pi$ is the volume of the parameter space where the prior is non-vanishing, also known as the support of the prior. If the two datasets, specified by the parameter $\theta$, are conditionally independent, it is possible to sample the posteriors $\mathcal{P}(\theta_1)$ and $\mathcal{P}(\theta_2)$ separately (Raveri & Doux 2021). In this scenario, the expression for $\mathcal{P}(\Delta\theta)$ becomes

$$\mathcal{P}(\Delta\theta) = \int_{V_\pi} \mathcal{P}(\theta_1) \, \mathcal{P}(\theta_1 - \Delta\theta) \, d\theta_1. \tag{10}$$

Thereafter, we obtain the mean probability of the presence of a parameter shift using the following equation.

$$\Delta = \int_{\mathcal{P}(\Delta\theta) > \mathcal{P}(0)} \mathcal{P}(\Delta\theta) d\Delta\theta, \tag{11}$$
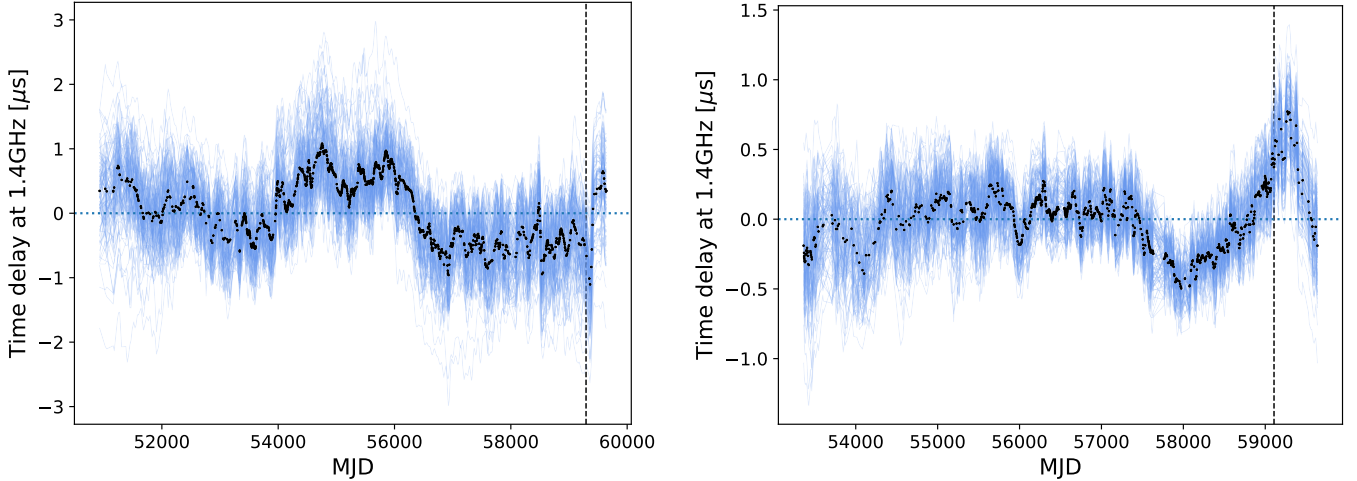
---

**Fig. 3.** 100 random realisations (blue) and medians for each epoch of the time-domain reconstructions of the DM variations delays at 1.4 GHz radio frequency modelled as a Gaussian process for PSRs J0613−0200 (left) and J1909−3744 (right), with the "DR2full+". The black dashed lines display the last epochs of DR2full (that is, EPTA only). The inclusion of InPTA data allows to measure sharp changes at after MJD 59000. The delays are obtained from marginalising the timing model parameters which comprise the DM constant and the first two time derivatives of the dispersion measure (DM1, DM2).

**Table 6.** Estimated tension (Z-score in sigma) between the DR2full and DR2full+ datasets for the red and DM noise models.

| Pulsar | Model | RN-RN | DM-DM |
|--------|-------|-------|-------|
| J0613−0200 | DM+RN | 0.74 | **2.97** |
| J0751+1807 | DM | X | 0.63 |
| J1012+5307 | DM+RN | 0.02 | 0.04 |
| J1022+1001 | DM+RN | 0.08 | 0.52 |
| J1600−3053 | DM | X | **4.64** |
| J1713+0747 | DM+RN | 0.01 | 0.14 |
| J1744−1134 | DM+RN | 0.20 | **2.29** |
| J1857+0943 | DM | X | 0.05 |
| J1909−3744 | DM+RN | 0.05 | **4.39** |
| J2124−3358 | DM | X | 0.84 |

**Notes.** Instances with significant tension are highlighted.

which incorporates the posterior mass above the iso-contour of no shift. We then convert the above $\Delta$ into an effective number of $\sigma$ using the standard normal distribution. Detailed comparisons that arise from the posteriors of DR2full and DR2full+ are available online[7]. In Table 6, we report the estimated tension (in $\sigma$) for the red and DM noise models while dealing with DR2full and DR2full+ datasets. It shows that the 2D posterior distributions of the RN and DM parameters are consistent ($\Delta < 1\sigma$) for all parameters, except for the power law DM variations of the PSRs J0613−0200, J1600−3053, J1744−1134 and J1909−3744. In the following, we discuss possible explanations for these pulsars.

### 4.2.1. PSR J0613−0200

For this pulsar, combining InPTA with the EPTA data yields a lower spectral index and higher amplitude at $f_{yr}$ for the chromatic noise. We observe an interesting sharp jump in the last years of the DM time series after including InPTA data (cf. the left panel of Fig. 3), which might increase the power at high PSD frequencies, and therefore yield to a flatter constrained power law.

---

7 https://github.com/subhajitphy/Posterior_comparisons

### 4.2.2. PSR J1600−3053

The favoured model using DR2full comprises DM and SV, with constraints that are (1) consistent with independent measurements of scattering delays with the Large European Array for Pulsars (LEAP; Main et al. 2023; cf. left panel of Fig. 4), and (2) highly consistent with the expected chromatic indices for both DM and SV as shown in the right panel of Fig. 4. However, the inclusion of InPTA data no longer supports scattering variations, and the favoured model is RN+DM. While the inclusion of SV is still supported after including the L-band data from the InPTA, we find that this discrepancy is happening from including the P-band data (300–500 MHz). Furthermore, the scattering variations in DR2full are unlikely to be related to pulse broadening at low frequencies, as we have no evidence for this in the InPTA P-band data. This discrepancy will be further investigated with data from the IPTA.

### 4.2.3. PSR J1744−1134

As described in Sect. 4.1, this pulsar exhibits a bimodal posterior for achromatic noise and a very flat spectrum for DM variations with the DR2full dataset. The inclusion of the InPTA data measures DM variations with a much lower spectral index and a lower amplitude. This removes the observed bimodality in the achromatic noise and allows a tighter constraint on a single mode as shown in Fig. 5.

### 4.2.4. PSR J1909−3744

The behaviour observed in PSR J1909−3744 appears to be very similar to PSR J0613−0200, in which including the InPTA data produces a lower spectral index and a higher amplitude for the chromatic noise. This is also likely caused by the sharp change in DM variations in the last two years of the dataset (cf. Fig. 3). This feature was already reported in Tarafdar et al. (2022) and is also shown in Curyło et al. (2023) that used, respectively, the first data release of InPTA and the ultra-wide-bandwidth low-frequency data from PPTA. Interestingly, we do not observe any associated impact on the RN posterior distributions.
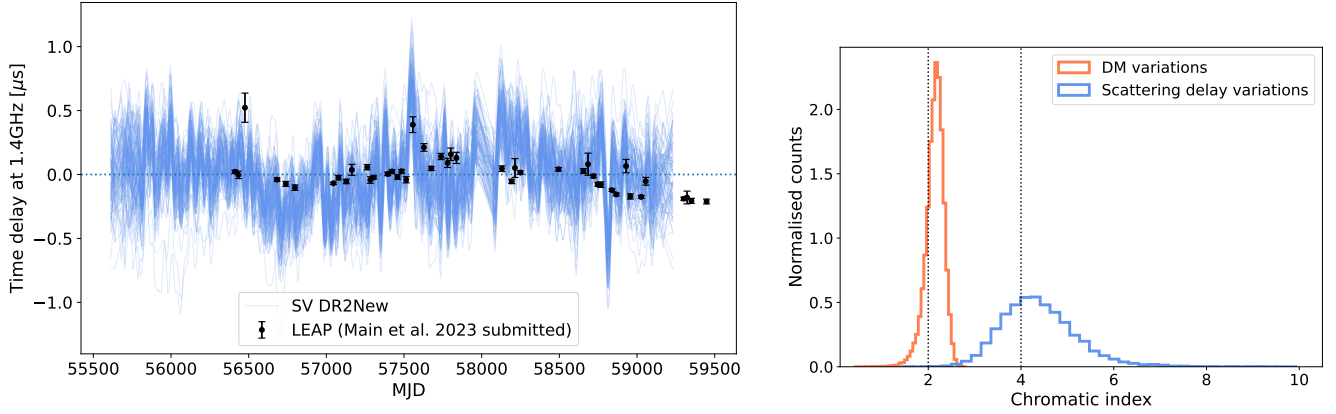
**Fig. 4.** Scattering delay variations measurement for PSR J1600−3053 with DR2new. Left plot: 100 random realisations (blue) of the time-domain reconstructions of the scattering delay variations at 1.4 GHz radio frequency modelled as a Gaussian process with DR2new. Independent scattering delays measured from scintillation analysis with the LEAP data (Main et al. 2023) are shown by the black points. Right plot: chromatic index posterior distributions measured for the DM (orange) and the SV (blue) with DR2new, while fixing the chromatic index of the other component. The favoured model for this dataset includes both DM and SV. The black dotted lines emphasise the expected values for DM and SV, respectively at 2 and 4.
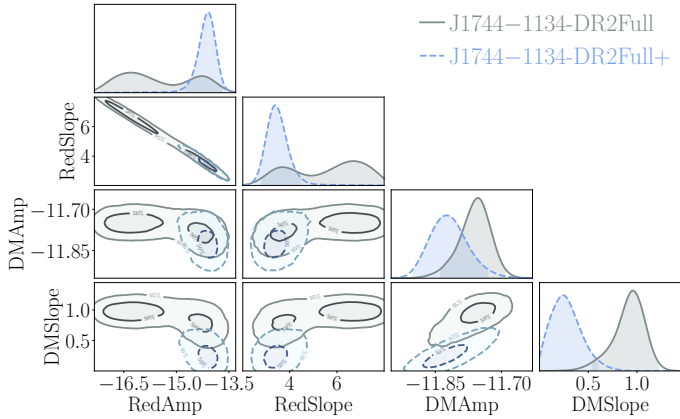


**Fig. 5.** Red and DM noise models for PSR J1744−1134 using DR2full and DR2full+ datasets. The inclusion of InPTA data allows a better constraint on the achromatic noise.

### 4.3. Implications

As expected the much-improved frequency coverage of DR2 has meant that the chromatic noise is much better constrained with the DR2full and DR2new datasets. Surprisingly, in some cases, signals attributed to achromatic processes in DR1 have been attributed to chromatic noise in DR2.

Of particular interest, however, are pulsars such as J1713+0747 and J1744−1134, in which the choice of dataset seems to affect the noise model hyperparameters in a way that is not easily understood. Although these effects are rather subtle, these are two of the best-timed pulsars in the EPTA, and this inconsistency may be an indication that more complex noise models may be needed for the best-timed pulsars and longest datasets. We assume that the noise process is a stationary Gaussian process with a power-law spectral density, and this may suggest that one or more of these assumptions are incorrect. Indeed, it is quite plausible that the pulsar spin noise is not entirely stationary. The IISM is known to have discrete structures that may change the statistics of the DM and scattering variations. In particular, PSR J1713+0747 is observed to undergo 'chromatic

timing events that are neither a stationary process nor fully modelled by achromatic or $\nu^{-2}$ DM variations (Lam et al. 2018). Pulsar timing noise in the population of canonical pulsars also shows behaviour that is either not strictly modelled by a power-law (e.g. quasi-periodic variability; Lyne et al. 2010) or shows discrete changes in spin-down behaviour on long timescales (e.g. Brook et al. 2014; Shaw et al. 2022).

It is worth reiterating that the noise models presented in this work represent our best estimate of the underlying noise using the understanding that we currently have. Further insights are hampered by the fact that it can be hard to disentangle the effect of improvements in the observing systems, especially sensitivity and observing frequency coverage, from changes in the observable pulsar noise properties, either due to time-variable processes or from processes only detectable on long timescales. Combined IPTA datasets bring additional complications, but the combination of improved instantaneous sensitivity and frequency coverage, particularly in the latest instrumentation, will be our best chance to separate instrumental effects from those intrinsic to the pulsars. Furthermore, the direct combination of data can make use of a dropout analysis to identify any telescope- or backend-specific instrumental effects that may mask the real behaviour of the pulsars.

## 5. Noise model validation

In this section, we perform several tests to compare the performance of the customised noise models with more standard models and assess the robustness of the results.

### 5.1. Performance of the customised noise models

As shown by Tables 3 and 4, the favoured noise models include only one component for most pulsars, making them simpler than the standard models commonly used in PTA analyses (RN and DM with 30 and 100 frequency bins, respectively, for all pulsars). We first evaluate the improvements enabled by the model selection process by comparing them with standard noise models, except for PSR J1012+1001, since we use the standard model for this pulsar (cf. Sect. 3). The Bayes factors in favour of the

**Table 7.** Validation and performance statistics for the customised noise models with the "DR2full+" and "DR2new".

| Pulsar | DR2full+ | | | | DR2new | | | |
|---|---|---|---|---|---|---|---|---|
| | $\alpha^{\mathrm{RN}}$ | $\alpha^{\mathrm{DM}}$ | $\mathcal{B}^{\mathrm{cus}}_{\mathrm{std}}$ | $\mathcal{B}^{\mathrm{cus+RN30}}_{\mathrm{cus}}$ | $\alpha^{\mathrm{RN}}$ | $\alpha^{\mathrm{DM}}$ | $\mathcal{B}^{\mathrm{cus}}_{\mathrm{std}}$ | $\mathcal{B}^{\mathrm{cus+RN30}}_{\mathrm{cus}}$ |
| J0030+0451 | $-1.23^{+1.63}_{-1.08}$ | X | 1.8 | 1.0 | $-4.03^{+5.54}_{-0.91}$ | X | 1.6 | 1.1 |
| J0613−0200 | $0.85^{+0.44}_{-0.87}$ | $\mathbf{2.93^{+0.38}_{-0.43}}$ | 79.9 | 1.0 | X | $4.31^{+5.11}_{-4.39}$ | 3.8 | 1.7 |
| J0751+1807 | X | $1.59^{+0.91}_{-0.68}$ | 2.8 | 1.9 | X | $1.34^{+1.26}_{-1.09}$ | 1.0 | 1.1 |
| J0900−3144 | $-0.85^{+1.54}_{-1.55}$ | $1.35^{+7.19}_{-0.64}$ | $\geq 10^3$ | 1.2 | $-0.88^{+1.84}_{-1.77}$ | $4.77^{+4.98}_{-3.88}$ | 287.8 | 1.5 |
| J1012+5307 | $\mathbf{-0.65^{+0.46}_{-0.41}}$ | $2.04^{+0.66}_{-0.46}$ | $\geq 10^3$ | 2.0 | $\mathbf{-0.56^{+0.45}_{-0.43}}$ | $2.10^{+5.16}_{-1.04}$ | $\geq 10^3$ | 1.8 |
| J1022+1001 | $1.02^{+2.28}_{-3.13}$ | $\mathbf{1.54^{+0.41}_{-0.37}}$ | X | 1.0 | $3.04^{+6.34}_{-7.12}$ | $\mathbf{0.28^{+1.48}_{-1.07}}$ | X | 1.6 |
| J1024−0719 | X | $1.42^{+2.18}_{-1.18}$ | 1.2 | 1.2 | X | $1.92^{+2.44}_{-1.53}$ | 3.0 | 1.6 |
| J1455−3330 | $0.58^{+5.10}_{-3.60}$ | X | 1.2 | 1.0 | $1.39^{+6.18}_{-4.71}$ | X | 1.0 | 1.2 |
| J1600−3053 | $1.95^{+0.27}_{-3.21}$ | $2.15^{+0.41}_{-0.30}$ | $\geq 10^3$ | 1.2 | X | $2.16^{+0.31}_{-0.47}$ | $\geq 10^3$ | 1.2 |
| J1640+2224 | X | $2.50^{+5.74}_{-2.34}$ | 2.6 | 1.4 | X | $2.36^{+5.33}_{-2.20}$ | 3.4 | 2.3 |
| J1713+0747 | $\mathbf{-0.85^{+0.63}_{-0.78}}$ | $2.02^{+0.20}_{-0.21}$ | $\geq 10^3$ | 1.2 | $0.25^{+0.74}_{-1.96}$ | $1.76^{+0.47}_{-0.69}$ | $\geq 10^3$ | 1.1 |
| J1730−2304 | X | $2.85^{+2.32}_{-2.04}$ | 6.6 | 1.7 | X | $3.11^{+3.15}_{-2.42}$ | 3.4 | 1.8 |
| J1738+0333 | $0.73^{+2.55}_{-2.29}$ | X | 1.8 | 1.0 | X | $1.53^{+2.65}_{-1.42}$ | 1.3 | 1.0 |
| J1744−1134 | $\mathbf{1.01^{+0.70}_{-0.71}}$ | $1.85^{+1.68}_{-1.28}$ | $\geq 10^3$ | 0.8 | X | $1.85^{+1.68}_{-1.28}$ | 2.4 | 2.0 |
| J1751−2857 | X | $4.60^{+3.87}_{-3.58}$ | 1.9 | 1.3 | X | $5.79^{+3.90}_{-4.66}$ | 3.0 | 1.6 |
| J1801−1417 | X | $3.64^{+2.03}_{-1.73}$ | 27.1 | 1.8 | X | $3.77^{+2.06}_{-1.74}$ | 8.1 | 1.6 |
| J1804−2717 | X | $2.66^{+4.62}_{-2.36}$ | 6.3 | 1.7 | X | $2.99^{+5.87}_{-2.70}$ | 6.2 | 1.9 |
| J1843−1113 | X | $1.66^{+0.97}_{-0.64}$ | 3.3 | 1.9 | X | $1.89^{+0.36}_{-0.37}$ | 2.1 | 2.0 |
| J1857+0943 | X | $2.19^{+0.44}_{-0.57}$ | 8.1 | 2.3 | X | $2.80^{+1.99}_{-1.52}$ | 2.1 | 2.1 |
| J1909−3744 | $0.60^{+0.35}_{-0.87}$ | $\mathbf{2.31^{+0.14}_{-0.14}}$ | $\geq 10^3$ | 1.0 | $0.35^{+0.44}_{-0.76}$ | $\mathbf{3.10^{+0.88}_{-0.74}}$ | 3.0 | 1.1 |
| J1910+1256 | X | $3.75^{+3.31}_{-3.07}$ | 2.2 | 1.8 | X | $3.35^{+2.78}_{-2.57}$ | 1.7 | 1.7 |
| J1911+1347 | X | $3.50^{+1.89}_{-1.57}$ | 2.7 | 2.0 | X | $\mathbf{7.60^{+2.28}_{-4.21}}$ | 1.8 | 1.8 |
| J1918−0642 | X | $2.57^{+1.69}_{-1.74}$ | 2.1 | 2.0 | X | $2.70^{+1.87}_{-1.75}$ | 1.9 | 2.0 |
| J2124−3358 | X | $1.60^{+0.84}_{-1.15}$ | 2.7 | 1.6 | X | $\mathbf{8.12^{+1.80}_{-4.48}}$ | 4.9 | 2.0 |
| J2322+2057 | X | X | 4.1 | 2.0 | X | X | 3.7 | 1.9 |

**Notes.** For each dataset, the two first columns display the medians and 95% confidence intervals of the chromatic index posterior distributions corresponding, respectively, to the RN and DM components. The values displayed in bold emphasise the numbers that are inconsistent with the expected values (0 for RN and 2 for DM). The third columns show the Bayes factor for the customised model against the standard noise model (RN and DM with resp. 30 and 100 frequency bins). The last columns display the Bayes factor evaluations for an additional red noise component with 30 frequency bins with the customised noise models.

customised noise models over the standard models are shown in Col. 3 of Table 7 for each dataset. Following the Kass & Raftery (1995) scale, we find

– weak support ($\mathcal{B}^{\mathrm{cus}}_{\mathrm{stand}} \in [1, 3]$) for 11 and 13 pulsars,
– positive support ($\mathcal{B}^{\mathrm{cus}}_{\mathrm{stand}} \in [3, 20]$) for five and seven pulsars,
– strong support ($\mathcal{B}^{\mathrm{cus}}_{\mathrm{stand}} \in [20, 100]$) for two and zero pulsars,
– very strong support ($\mathcal{B}^{\mathrm{cus}}_{\mathrm{stand}} \geq 100$) for six and four pulsars,

in DR2full+ and DR2new respectively. We notice higher Bayes factor values for DR2full+ compared to DR2new for most pulsars, which can be partially explained by the better performance of the standard noise models to describe higher PSD frequencies for DR2new as its timing baseline is approximately two times shorter than DR2full+. Furthermore, cases with "strong" and "very strong" evidence are obtained for the pulsars that include two time-correlated components (RN+DM, DM+SV), except for PSR J1801−1417 (only DM variations). We also observe a very

significant reduction of Bayes factors for PSR J1909−3744 with the DR2new, even if the preferred model for this pulsar includes both RN and DM for the two datasets.

To assess the validity of the customised noise models, we checked the chromatic index $\alpha$ values used for the time-correlated noise components fixed at 0, 2 and 4 for RN, DM and SV respectively. To do so, we used custom noise models for each pulsar and set $\alpha$ for each noise component as a hyperparameter to estimate its posterior distribution, as performed in recent PTA analyses (e.g. Goncharov et al. 2021; Chalumeau et al. 2022). If the pulsar has two time-correlated components, we perform two analyses to evaluate each $\alpha$ independently while fixing the index of the other component to its relevant value (0 if RN and 2 if DM). We use a uniform prior probability as $\mathcal{U}(-5, 10)$. The medians and 95% credible intervals of the posterior distributions for RN and DM chromatic indices (resp. $\alpha^{\mathrm{RN}}$ and $\alpha^{\mathrm{DM}}$) are shown in

**Table 8.** Tension metrics for the red and DM noise models estimated using `enterprise` and `temponest`.

| Pulsar | Model | RN tension | DM tension |
|--------|-------|-----------|-----------|
| J0030+0451 | RN | 0.02 | X |
| J0613−0200 | DM+RN | 0.40 | 0.07 |
| J0751+1807 | DM | X | 0.07 |
| J0900−3144 | DM+RN | 0.04 | 0.02 |
| J1012+5307 | DM+RN | 0.06 | 0.66 |
| J1022+1001 | DM+RN | 0.01 | 0.23 |
| J1024−0719 | DM | X | 0.16 |
| J1455−3330 | RN | 0.02 | X |
| J1600−3053 | DM | X | 0.03 |
| J1640+2224 | DM | X | 0.16 |
| J1713+0747 | DM+RN | 0.01 | 0.11 |
| J1730−2304 | DM | X | 0.001 |
| J1738+0333 | RN | 0.001 | X |
| J1744−1134 | DM+RN | 0.06 | 0.08 |
| J1751−2857 | DM | X | 0.01 |
| J1801−1417 | DM | X | 0.01 |
| J1804−2717 | DM | X | 0.01 |
| J1857+0943 | DM | X | 0.21 |
| J1909−3744 | DM+RN | 0.03 | 0.04 |
| J1910+1256 | DM | X | 0.001 |
| J1911+1347 | DM | X | 0.001 |
| J1918−0642 | DM | X | 0.02 |
| J2124−3358 | DM | X | 0.01 |
| J2322+2057 | X | X | X |

the first two columns for each dataset in Table 7. In the following, we summarise our findings.

- For RN, we note that the chromatic indices of PSRs J1713+0747 and J1909−3744 are consistent with the expected value of 0 with DR2new, while it is not the case with DR2full. We observe slightly negative values for PSRs J0030+0451, J0900−3144 and J1012+5307 with both datasets.
- For the DM chromatic indices, 18 pulsars are consistent with the expected value for DM variations with both the DR2full+ and the DR2new. However, the high values for PSRs J1911+1347 and J2124−3358 with the DR2new warrant further investigation.
- The only pulsar with a preferred SV model is PSR J1600−3053 with DR2new. The posterior distribution of the SV chromatic index for this pulsar is shown in the right panel of Fig. 4, with a median and 95% credible intervals at $4.31^{+1.88}_{-1.27}$, nicely consistent with the expected value from scattering variations (cf. Sect. 2).
- For PSR J1022+1001, it is interesting that the trend in RN and DM indices swaps between the two datasets (that is, RN is shallower in DR2full+ and steeper in DR2new, while the opposite is the case for DM). To account for this, we stick to the standard noise model for this pulsar, as RN might prioritise the low frequencies, while DM also samples the high frequencies. The estimated indices for the DR2new show that the high-frequency signals are achromatic for this pulsar and the low-frequency components are consistent with the DM variations. While the former could correspond to profile instabilities (Liu et al. 2015; Padmanabh et al. 2020), the latter is likely due to the presence of a variable solar wind (Tiburzi et al. 2021).

For the second validity test, we evaluate a Bayes factor after including an achromatic noise component over the favoured models and therefore check if there is evidence for any remaining red noise. We use 30 frequency components for this additional red noise in order to prioritise the low frequencies. As shown in Col. 5 of Table 7, we only find weak evidence (maximum of 2.3 for PSR J1640+2224) for the inclusion of this component. These results reconfirm the robustness of the customised noise models that do not include an achromatic component.

### 5.2. Consistency among different softwares

In Table 8, we present the tension in $\sigma$ (see Sect. 4.2 for a description of the tension metric) for red and DM noise models estimated using two Bayesian timing packages, `enterprise` and `temponest` for the 25 EPTA MSPs. We find that the noise models for all the pulsars are very consistent.

### 5.3. Marginalisation of the timing model

Preliminary analysis of DR2full using `temponest` showed that the posteriors of the noise model parameters varied significantly depending on whether the timing model was marginalised. This was particularly surprising given that Lentati et al. (2014) demonstrated that neither the analytic marginalisation nor linearisation made a significant difference to any parameters. Further investigation revealed that changes in the implementation of `temponest` in 2017[8] introduced a critical flaw when not marginalising over the timing model whilst marginalising over the arbitrary jumps between instruments; the typical method of operation when not marginalising over the timing model. This error disabled fitting for many jumps, leading to incorrect noise model inferences. Once resolved, we no longer find any discrepancy between noise model parameters when marginalising over the timing model, and find that the linearised timing model is sufficient for further analysis once any highly non-linear binary parameters have been solved.

### 5.4. Simulations

We have implemented a toolkit to generate repeatable simulations of EPTA datasets to validate noise models. Simulations are generated using the `toasim` framework that is distributed with `tempo2`. In brief, the method is to generate a set of 'idealised' TOAs that produce zero residual with respect to a seed pulsar parameter file and then add perturbations for each of the simulated noise processes. For an idealised TOA $t_i$, the final simulated TOA is given by

$$t_{\text{final},i} = t_i + y_{\text{red}}(t_i) + y_{\text{dm}}(t_i) + y_{\text{scat}}(t_i) + y_{\text{white}}(t_i), \tag{12}$$

where each of the perturbations $y$ is discussed in the following sections. In principle, this can only be solved iteratively, but in practise we can neglect second-order corrections as long as the perturbations are small. Therefore, we typically subtract a quadratic from each $y(t)$, which significantly reduces the general amplitude of the perturbations, especially for steep red noise processes. Removing a quadratic in this way does not affect the noise models, but it does mean that the mean pulsar spin frequency and frequency derivative are unchanged from realisation to realisation. This is hence equivalent to fitting for `F0` and `F1` on

---

[8]  commit hash: `e745752`.

each dataset, and small variations in these parameters are largely uninteresting, and in many senses arbitrary for a simulation, this does not affect the overall usefulness of the simulations.

Significant improvements have been implemented to the `toasim` code as part of this work, which were required to ensure that the definitions of the model parameters are consistent between this work and the simulation code. The simulation code has been adapted to be consistent with the `tempo2` and `temponest` definitions of the model parameters, although these are generally trivially related to other definitions.

### 5.4.1. Achromatic red noise

The `toasim` framework provides a method for simulating power-law red-noise processes by means of the inverse Fourier transform. This first computes the noise process over a grid of 4096 equally spaced values over $T_{\rm span}$, extended by a factor of $n = 100$ to avoid periodic boundary effects of the Fourier transform. The spectrum is computed for $N = 1024n$ frequencies with Hermitian symmetry, and Fourier transformed to give a time series:

$$r_k = \sum_{j=-N/2}^{N/2} R_j(a_j + ib_j) \exp\left(2\pi \sqrt{-1}\, jk/N\right), \tag{13}$$

with $a_j$ and $b_j$ being random variables drawn from $\mathcal{N}(0, 1)$. Note that the requirement for Hermitian symmetry means that only 512 independent random values are needed for each of $a_j$ and $b_j$. This is largely analogous to Eq. (1) for the noise modelling except that the Fourier transform includes both positive and negative frequencies and therefore the amplitudes must be scaled down by a factor of two, relative to the noise models in Eq. (2). Therefore, the mean amplitude at each frequency $R_j$ is given by

$$R_j = \frac{1}{2} \sqrt{\frac{A_{\rm red}^2}{12\pi^2} \frac{{\rm s}_{\rm yr}^3}{nT_{\rm span}} \left(\frac{f_j}{f_{\rm yr}}\right)^{-\gamma_{\rm red}}}, \tag{14}$$

where $f_{\rm yr} = 1\,{\rm yr}^{-1}$, and ${\rm s}_{\rm yr} = 31557600\,{\rm s}\,{\rm yr}^{-1}$ converts years to seconds to give a perturbation in seconds for $A_{\rm red}$ in ${\rm yr}^{3/2}$ and $T_{\rm span}$ in s. The factor of $1/2$ corrects from the one-sided PSD used by the noise models to the two-sided PSD needed for the Fourier transform. The final perturbation for the achromatic red noise for TOA $t_i$ is computed by linear interpolation of $r_k$,

$$y_{\rm red}(t_i) = r_k + (t_i - t_k)\frac{r_{k+1} - r_k}{t_{k+1} - t_k}, \tag{15}$$

for $t_k < t_i < t_{k+1}$. This is implemented by the `addRedNoise` `tempo2` plugin.

### 5.4.2. DM variations

Perturbations due to DM variations are similarly computed by the inverse Fourier transform using largely the same method as for the achromatic noise, except that we model a DM time series before converting to a TOA perturbation as a last step. The mean amplitude at each frequency, $D_j$ is given by

$$D_j = \frac{1}{2} \sqrt{A_{\rm DM}^2 \frac{{\rm s}_{\rm yr}^3}{nT_{\rm span}} \left(\frac{f_j}{f_{\rm yr}}\right)^{-\gamma_{\rm DM}}}, \tag{16}$$

with $T_{\rm span}$ in seconds, $A_{\rm DM}$ in `temponest` units of ${\rm cm}^{-3}{\rm pc}\,{\rm yr}^{3/2}{\rm s}^{-1}$.

The final perturbation for a given TOA is then calculated from $d(t)$, a function that linearly interpolates $d_k$ in a similar way to Eq. (15), and

$$y_{\rm DM}(t_i) = \frac{d(t_i)}{\kappa_{\rm DM}\nu_k^2}, \tag{17}$$

where $\nu_k$ is the observing frequency of the TOA and $\kappa_{\rm DM} = 2.41 \times 10^{-4}{\rm cm}^{-3}{\rm pc}\,{\rm MHz}^2{\rm s}^{-1}$ is the DM constant. This is implemented by the `addDmVar` plugin in `tempo2`.

### 5.4.3. Scattering variations

Scattering variations are implemented very similarly to the DM variations. We create a scattering time series $s_k$ using the Fourier transform of $n$ Hermitian complex values given by $S_j(a_j + ib_j)$, where

$$S_j = \frac{1}{2} \sqrt{\frac{A_{\rm scat}^2}{12\pi^2} \frac{{\rm s}_{\rm yr}^3}{nT_{\rm span}} \left(\frac{f_j}{f_{\rm yr}}\right)^{-\gamma_{\rm scat}}}. \tag{18}$$

Then the perturbation associated with the scattering is computed from a linear interpolation function $s(t)$ to give

$$y_{\rm scat}(t_i) = s(t_i)\left(\frac{\nu}{\nu_{\rm ref}}\right)^{-\alpha_{\rm scat}}, \tag{19}$$

where we use $\alpha_{\rm scat} = 4$ for scattering variations, and $\nu_{\rm ref} = 1400\,{\rm MHz}$, chosen to be consistent with the implementation in `temponest` and `enterprise`. This is implemented by the `addChromVar` `tempo2` plugin.

### 5.4.4. White noise

White noise is added to the TOA with uncertainty $\sigma_i$ after applying the EFAC and EQUAD parameters (see Eq. (3)) by drawing $y_{\rm white}(t_i)$ from $\mathcal{N}(0, \sigma_i^2)$.

### 5.4.5. Creating the realisations

The simulation parameters can be given as single values or a uniform prior range from which to draw, so that a different set of parameters is used for each realisation of the simulation. When the parameters are drawn randomly from the prior, we record the sets of parameters used for later comparison with the results.

### 5.5. P–P plots

One tool to validate Bayesian analysis tools is by studying posterior quantiles from simulated datasets (Cook et al. 2006). Our method largely follows that implemented in `bilby` (Ashton et al. 2019), where we simulate a large number of datasets, $j$, with 'true' model parameters $\mathbf{\Theta}_j$ drawn from a prior distribution $p(\mathbf{\Theta})$. We then use our Bayesian software, in this case `temponest`, to generate samples, $\theta_{i,j}$ of the posterior distribution of the parameters given the prior distribution (see Table 9), for each of the simulated datasets, following the same approach as for the real data. We can then, for each parameter $i$, and each simulated dataset $j$, compute the quantile $q_{i,j}$, within the 1-d posterior that the 'true' value of the parameter lies, that is, the weighted fraction of samples for which $\theta_{i,j} < \Theta_{i,j}$. It can be shown that the $q_{i,j}$ should be uniformly distributed between 0 and 1 (Cook et al. 2006). We can visualise this by plotting the cumulative distribution of $q_{i,j}$ over all $j$, yielding a so-called 'P–P

**Table 9.** Uniform prior bounds for the P–P simulations for J1600−3053.

| Parameter | Low | High |
|---|---|---|
| $\log(A_{\rm red})$ | −15.0 | −13.4 |
| $\gamma_{\rm red}$ | 2.0 | 6.0 |
| $\log(A_{\rm dm})$ | −13.0 | −11.0 |
| $\gamma_{\rm dm}$ | 1.6 | 3.2 |
| $\log(A_{\rm scat})$ | −15.0 | −13.2 |
| $\gamma_{\rm scat}$ | 1.1 | 2.2 |
| EFAC | 0.5 | 2.0 |
| $\log_{10}$ EQUAD | −8 | −6.5 |

**Notes.** A separate EFAC and EQUAD is drawn for each `-group` flag.

plot'. In principle, this should follow a straight line, but even with perfect algorithms, there will be deviations caused by the finite number of simulations, which we can estimate with confidence contours taken from the binomial distribution.

It is important that the prior distribution matches between the simulated data and the Bayesian search, particularly for any parameters that may not be constrained on one or both sides and hence be bounded by the prior. This is quite common in the pulsar datasets, as power-law red-noise processes below the detection threshold will have log amplitude unbounded on the low side and the spectral slope may simply return the prior. If the prior does not match the simulation and posterior generation, this will inevitably lead to a perceived error in the posterior computation for such cases. We choose a prior range such that most simulations have noise processes that are largely representative of what we estimated in the real data, and wide enough that the posterior of the real data would not be heavily constrained by the choice of prior.

Results: For simplicity, we first consider the simulation of only achromatic red noise. The P–P plot, shown in Fig. 6a, indicates generally good consistency in the posterior of $\gamma_{\rm red}$, but there is an excess at low confidence intervals for $\log(A_{\rm red})$. The investigation of individual results shows a correlation between the $\log(A_{\rm red})$ confidence interval and the injected $\gamma_{\rm red}$, visualised in Fig. 6b. We believe that the observed correlation suggests that the excess results are driven by cases where $\gamma_{\rm red}$ is large, and hence we suspect that they are cases where the red noise has such a steep spectrum that it is no longer able to be expressed in a Fourier basis with lowest frequency at $1/T_{\rm span}$.

To understand this, we recall that the Fourier-basis Gaussian process models used in this work, and widely used by other similar projects are naturally periodic on $T_{\rm span}$ due to the nature of the Fourier basis. This use of a Fourier basis vastly speeds up the computation compared to directly evaluating the covariance matrix for these large datasets, but it requires that the noise process can be well approximated by a periodic function. It is generally assumed that fitting for the spin frequency (`F0`) and spin frequency derivatives (`F1`) parameters absorbs the low frequencies sufficiently that the periodic model is acceptable, and previous analyses have suggested that the posteriors remain largely unchanged when trying to absorb noise below $1/T_{\rm span}$ even for pulsars with $\gamma_{\rm red} \sim 6$ (Lentati et al. 2016). However, the P–P plot is very sensitive to even small biases, and hence quite likely that the difference may not be noticeable (or have strong Bayesian evidence) for a single analysis. Recently, it has also been seen that Fourier-basis Gaussian processes models periodic

on $T_{\rm span}$ struggle to recover the spin frequency second derivative (`F2`) when $\gamma_{\rm red}$ is greater than ∼4 (Keith & Niṭu 2023), and this implies that fitting for `F0` and `F1` may not be sufficient to allow safe use of the $1/T_{\rm span}$ Fourier basis model when the noise process has a very steep spectral exponent.

In order to test our hypothesis, we repeat the analsis whilst forcing the injected red noise to be periodic on $T_{\rm span}$, with results shown in Fig. 6c. The curve now goes smoothly to zero at 0, and there is no longer any correlation between the $\log(A_{\rm red})$ confidence interval and the injected $\gamma_{\rm red}$. However, the results are still not consistent with a one-to-one relationship (with probability <0.5%), in particular, we find that $\log(A_{\rm red})$ is slightly biased high and $\gamma_{\rm red}$ biased slightly low. We find that a correction of $\gamma_{\rm red}$ by only ∼2% is needed to fully correct the bias, a correction much smaller than the typical uncertainty of the measurement. The slope and amplitude are highly correlated, and indeed scaling $\log(A_{\rm red})$ to a longer time span closer to where the signal can be measured is sufficient to remove the apparent bias here. Our hypothesis is that there is an intrinsic bias for slightly flatter spectral slopes, which then manifests itself as a corresponding bias in the amplitude at $1\,{\rm yr}^{-1}$.

We repeat this with simulations that include achromatic noise red noise, excess white noise, DM variations, and scattering variations. The resulting P–P plots are very similar to the simulation with achromatic red noise only. Figure 6d shows the P–P plot when the noise is periodic on $1/T_{\rm span}$, showing a similarly small but significant bias in the posterior parameters for all three parameters. The similarity between all three components of the noise model to that obtained when only injecting achromatic noise suggests that the bias is due to something fundamental in the Fourier domain Gaussian process modelling of the noise, rather than being driven by the choice of the dataset.

Overall, the P–P plots are encouraging and show that the recovered posteriors are largely correct, but in almost all cases, we find subtle deviations from the correct posterior distribution. In agreement with Sect. 5.2, we find that the analysis does not change significantly if using `enterprise` to compute the noise models rather than `temponest`. The P–P plot methodology is sensitive to very subtle errors in the posterior, including biases much smaller than the uncertainty on the results, so although the results are not perfect, we do not feel that it invalidates the results we present. Nevertheless, we feel that it is important that these tests are repeated on a wider scale across the IPTA, to fully understand the characteristics of the posteriors produced by our noise modelling code and increase confidence in the interpretation of the posterior distributions for the pulsar noise models and the GWB detection parameters, particularly for any pulsars with steep spectrum red noise.

## 6. Conclusions

The noise models presented here are our current best estimates for stochastic noise in the 25-pulsar EPTA DR2 dataset. With the inclusion of wide-band receivers and additional low-frequency data from the InPTA collaboration, all but four pulsars show evidence for chromatic noise, specifically in the form of DM variations. However, several of these pulsars seem to have power-law noise with a much flatter spectrum than expected from DM variations in the turbulent IISM. This may be attributed to additional high-frequency terms, but it may also suggest other processes leaking into the model for DM variations. Compared to the EPTA DR1 and InPTA datasets, we find that the choice of preferred noise model can change, including a couple of cases
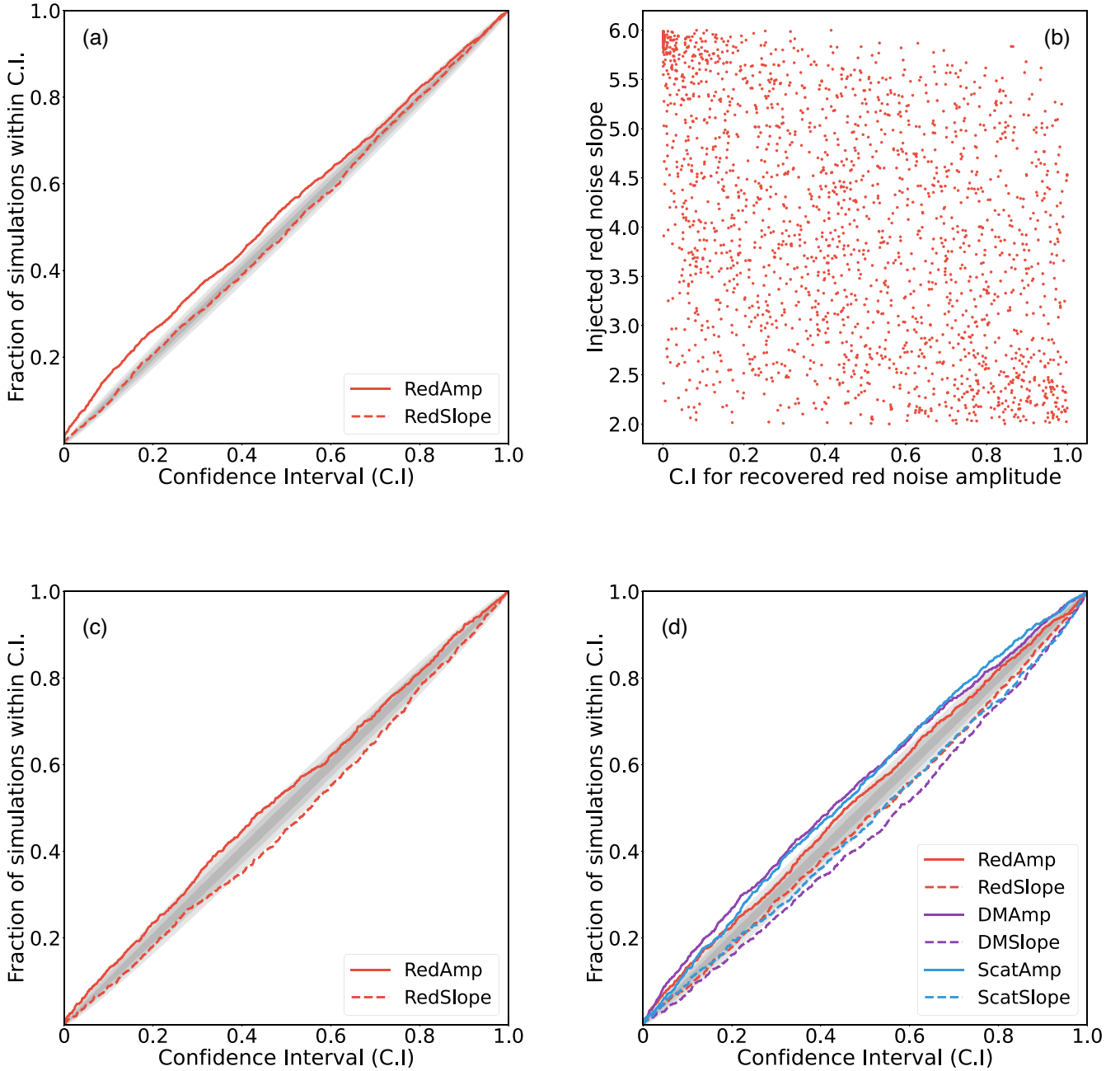
**Fig. 6.** From top left to bottom right: (a) P–P plot for ~2000 simulations with only chromatic noise, injected at much lower than $1/T_{\text{span}}$; (b) the confidence interval of the recovered achromatic amplitude for various levels of injected red noise slope for the simulations is shown in (a); (c) P–P plot for ~1000 simulations with only chromatic noise periodic on $1/T_{\text{span}}$; (d) P–P plot for ~1000 simulations with chromatic, achromatic and scattering variations, periodic on $1/T_{\text{span}}$.

where chromatic noise that was ruled out in EPTA DR1 becomes detected in EPTA DR2.

We feel that this is evidence that the assumptions typically made in noise modelling are not strictly true for all pulsars. Either there are additional processes that need to be considered, such as noise that is not purely modelled by a power law, or the spectral properties of the noise vary over time, or there are systematic instrumental effects that affect particular parts of the dataset. In reality, none of these assumptions are likely completely true, and the improvements in data quality and time span of the EPTA DR2 data particularly highlight these subtle effects.

The fact that the best noise models can depend on the frequency and time coverage of the dataset implies that great care

is needed when comparing the noise models across different PTAs, and hence we suggest that a thorough investigation into the time and frequency stationary of the pulsar noise models, as well as exploration of additional noise terms, is best undertaken within the IPTA framework, which necessarily has better time and frequency coverage than any individual PTA.

We have also tested the estimation of hyperparameters from the noise model in `temponest` and `enterprise` and find that, as might be expected, they are highly consistent in results in real and simulated data. We also demonstrate that previously observed differences in model parameters when marginalising over the timing model parameters in `temponest` were due to a software bug and that marginalising over the timing model

is entirely sufficient when trying to estimate the noise model hyperparameters.

Furthermore, we performed a 'P–P plot' test on our noise model hyperparameter estimation using simulated data and found that there may be two subtle biases in the results. For 'realistic' simulations where there is power at timescales longer than the observing span, a bias in the estimation of the red noise amplitude is observed when the spectral slope of the noise is greater than ~4. We further saw that even if we simulate noise that has the same periodic nature as the noise model, there is a very small bias (~2%) in the recovered spectral slope. Although these tests require large numbers of trials and, hence, large computing costs, we encourage further testing of this nature within the IPTA framework to attempt to better understand the origin and any possible effect of these biases.

It is important to note that although we find several areas of investigation for a better understanding of the pulsar noise processes, the custom noise models presented here model the noise in the EPTA dataset extremely well, and the majority of the pulsar noise models are consistent between all datasets that we have compared. We also demonstrate that our customised noise models improve our overall sensitivity to GWB signals over the 'standard' models for both the complete DR2full+ dataset as well as when focussing only on the new purpose-built EPTA instrumentation in DR2new.

interpretations. APa, AC, MJK equally share the correspondence of the paper. CRediT statement: Conceptualisation: AC, APa, APo, AV, BWS, CT, GMS, GT, JPWV, JWM, KJL, KLi, MJK, MK. Methodology: AC, APa, AV, DJC, GMS, IC, JWM, KJL, KLi, LG, MJK, MK, SB, SChe, VVK. Software: AC, AJ, APa, APe, GD, GMS, KJL, KLi, MJK, RK, SChe, VVK. Validation: AC, APa, BG, GMS, IC, JPWV, JWM, KLi, LG, MJK. Formal Analysis: AC, APa, BG, EvdW, GHJ, GMS, JWM, KLi, MJK. Investigation: AC, APa, APo, BWS, CGB, DJC, DP, GMS, IC, JPWV, JWM, KLi, LG, MBM, MBu, MJK, RK, VVK. Resources: AC, APa, APe, APo, BWS, GHJ, GMS, GT, IC, JPWV, JWM, KJL, KLi, LG, MJK, MK, RK. Data Curation: AC, AJ, APa, BWS, CGB, DJC, DP, EvdW, GHJ, GMS, JA, JWM, KLi, MBM, MJK, MK, NKP, RK, SChe. Writing – Original Draft: AC, APa, GMS, MJK. Writing – Review & Editing: AC, AF, APa, APo, BG, EB, EFK, GMS, GT, JA, JPWV, JWM, KLi, MJK, MK, SChe, VVK. Visualisation: AC, APa, GMS, KLi, MJK. Supervision: AC, APo, ASe, AV, BWS, CGB, DJC, EFK, GHJ, GT, JPWV, KJL, LG, MJK, MK, VVK. Project Administration: AC, APo, ASe, AV, BWS, CGB, CT, GHJ, GMS, GT, JPWV, JWM, LG, MJK, MK. Funding Acquisition: APe, APo, ASe, BWS, GHJ, GT, IC, LG, MJK, MK.

# References

Alam, M. F., Arzoumanian, Z., Baker, P. T., et al. 2021, ApJS, 252, 4
Antoniadis, J., Arzoumanian, Z., Babak, S., et al. 2022, MNRAS, 510, 4873
Arzoumanian, Z., Baker, P. T., Blumer, H., et al. 2020, ApJ, 905, L34
Ashton, G., Hübner, M., Lasky, P. D., et al. 2019, ApJS, 241, 27
Brook, P. R., Karastergiou, A., Buchner, S., et al. 2014, ApJ, 780, L31
Caballero, R. N., Lee, K. J., Lentati, L., et al. 2016, MNRAS, 457, 4421
Caballero, R. N., Guo, Y. J., Lee, K. J., et al. 2018, MNRAS, 481, 5501
Carlin, B. P., & Chib, S. 1995, J. R. Stat. Soc. Ser. B Methodol., 57, 473
Chalumeau, A., Babak, S., Petiteau, A., et al. 2022, MNRAS, 509, 5538
Chen, S., Caballero, R. N., Guo, Y. J., et al. 2021, MNRAS, 508, 4970
CHIME/Pulsar Collaboration (Amiri, M., et al.) 2021, ApJS, 255, 5
Coles, W. A., Kerr, M., Shannon, R. M., et al. 2015, ApJ, 808, 113
Cook, S. R., Gelman, A., & Rubin, D. B. 2006, J. Comput. Graph. Stat., 15, 675
Cordes, J. M., & Shannon, R. M. 2010, ArXiv e-prints [arXiv:1010.3785]
Cordes, J. M., Shannon, R. M., & Stinebring, D. R. 2016, ApJ, 817, 16
Curyło, M., Pennucci, T. T., Bailes, M., et al. 2023, ApJ, 944, 128
Demorest, P. B., Ferdman, R. D., Gonzalez, M. E., et al. 2013, ApJ, 762, 94
Desvignes, G., Caballero, R. N., Lentati, L., et al. 2016, MNRAS, 458, 3341
Donner, J. Y., Verbiest, J. P. W., Tiburzi, C., et al. 2020, A&A, 644, A153
Ellis, J., & van Haasteren, R. 2017, https://zenodo.org/record/1037579
Ellis, J. A., Vallisneri, M., Taylor, S. R., & Baker, P. T. 2019, Astrophysics Source Code Library [record ascl: 1912.015]
EPTA Collaboration (Antoniadis, J., et al.) 2023, A&A, 678, A48 (Paper I)
EPTA Collaboration, & InPTA Collaboration (Antoniadis, J., et al.) 2023, A&A, 678, A50 (Paper III)
Fermi-LAT Collaboration (Ajello, M., et al.) 2022, Science, 376, 521
Feroz, F., & Hobson, M. P. 2008, MNRAS, 384, 449
Foster, R. S., & Backer, D. C. 1990, ApJ, 361, 300
Foster, R. S., & Cordes, J. M. 1990, ApJ, 364, 123
Goncharov, B., Zhu, X.-J., & Thrane, E. 2020, MNRAS, 497, 3264
Goncharov, B., Shannon, R. M., Reardon, D. J., et al. 2021, ApJ, 917, L19
Goncharov, B., Thrane, E., Shannon, R. M., et al. 2022, ApJ, 932, L22
Hazboun, J. S., Simon, J., Taylor, S. R., et al. 2020, ApJ, 890, 108
Hee, W. S., Khor, W. Y., Lim, H. S., & Jafri, M. Z. M. 2015, AIP Conf. Ser., 1657, 040015
Hellings, R. W., & Downs, G. S. 1983, ApJ, 265, L39
Hobbs, G., Guo, L., Caballero, R. N., et al. 2020, MNRAS, 491, 5951
Jones, P. B. 1990, MNRAS, 246, 364
Kass, R. E., & Raftery, A. E. 1995, J. Am. Stat. Assoc., 90, 773
Keith, M. J., & Niţu, I. C. 2023, MNRAS, 523, 4603
Lam, M. T., Ellis, J. A., Grillo, G., et al. 2018, ApJ, 861, 132
Lentati, L., Alexander, P., Hobson, M. P., et al. 2013, Phys. Rev. D, 87, 104021
Lentati, L., Alexander, P., Hobson, M. P., et al. 2014, MNRAS, 437, 3004
Lentati, L., Shannon, R. M., Coles, W. A., et al. 2016, MNRAS, 458, 2161
Liu, K., Verbiest, J. P. W., Kramer, M., et al. 2011, MNRAS, 417, 2916
Liu, K., Keane, E. F., Lee, K. J., et al. 2012, MNRAS, 420, 361
Liu, K., Karuppusamy, R., Lee, K. J., et al. 2015, MNRAS, 449, 1158
Liu, K., Guillemot, L., Istrate, A. G., et al. 2020, MNRAS, 499, 2276
Lyne, A., Hobbs, G., Kramer, M., Stairs, I., & Stappers, B. 2010, Science, 329, 408
Madison, D. R., Cordes, J. M., Arzoumanian, Z., et al. 2019, ApJ, 872, 150
Main, R. A., Antoniadis, J., Chen, S., et al. 2023, MNRAS, 525, 1079
Manchester, R. N., Hobbs, G., Bailes, M., et al. 2013, PASA, 30, e017
Padmanabh, P. V., Barr, E. D., Champion, D. J., et al. 2020, MNRAS, 500, 1178
Parthasarathy, A., Shannon, R. M., Johnston, S., et al. 2019, MNRAS, 489, 3810
Parthasarathy, A., Bailes, M., Shannon, R. M., et al. 2021, MNRAS, 502, 407
Raveri, M., & Doux, C. 2021, Phys. Rev. D, 104, 043504

Shannon, R. M., & Cordes, J. M. 2010, ApJ, 725, 1607
Shannon, R. M., & Cordes, J. M. 2017, MNRAS, 464, 2075
Shannon, R. M., Lentati, L. T., Kerr, M., et al. 2016, ApJ, 828, L1
Shaw, B., Stappers, B. W., Weltevrede, P., et al. 2022, MNRAS, 513, 5861
Singha, J., Surnis, M. P., Joshi, B. C., et al. 2021, MNRAS, 507, L57
Sivia, D. S., & Skilling, J. 2006, Data Analysis – A Bayesian Tutorial, 2nd edn. (Oxford University Press)
Tarafdar, P., Nobleson, K., Rana, P., et al. 2022, PASA, 39, e053
Taylor, S. R., Mingarelli, C. M. F., Gair, J. R., et al. 2015, Phys. Rev. Lett., 115, 041101
Taylor, S. R., van Haasteren, R., & Sesana, A. 2020, Phys. Rev. D, 102, 084039
Tiburzi, C., Shaifullah, G. M., Bassa, C. G., et al. 2021, A&A, 647, A84
van Haasteren, R., & Vallisneri, M. 2014, Phys. Rev. D, 90, 104012
van Haasteren, R., & Vallisneri, M. 2015, MNRAS, 446, 1170
Xu, H., Huang, Y. X., Burgay, M., et al. 2021, ATel, 14642, 1
Zhu, W. W., Stairs, I. H., Demorest, P. B., et al. 2015, ApJ, 809, 41
Zic, A., Hobbs, G., Shannon, R. M., et al. 2022, MNRAS, 516, 410

[1] Institute of Astrophysics, FORTH, N. Plastira 100, 70013, Heraklion, Greece
[2] Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, 53121 Bonn, Germany
[3] Department of Physics, Indian Institute of Technology Roorkee, Roorkee-247667, India
[4] Department of Electrical Engineering, IIT Hyderabad, Kandi, Telangana 502284, India
[5] Université Paris-Cité, CNRS, Astroparticule et Cosmologie, 75013 Paris, France
[6] The Institute of Mathematical Sciences, C. I. T. Campus, Taramani, Chennai 600113, India
[7] Homi Bhabha National Institute, Training School Complex, Anushakti Nagar, Mumbai 400094, India
[8] Fakultät für Physik, Universität Bielefeld, Postfach 100131, 33501 Bielefeld, Germany
[9] ASTRON, Netherlands Institute for Radio Astronomy, Oude Hoogeveensedijk 4, 7991 PD, Dwingeloo, The Netherlands
[10] Department of Physical Sciences, Indian Institute of Science Education and Research, Mohali, Punjab 140306, India
[11] Laboratoire de Physique et Chimie de l'Environnement et de l'Espace, Université d'Orléans / CNRS, 45071 Orléans Cedex 02, France
[12] Observatoire Radioastronomique de Nançay, Observatoire de Paris, Université PSL, Université d'Orléans, CNRS, 18330 Nançay, France
[13] Dipartimento di Fisica "G. Occhialini", Università degli Studi di Milano-Bicocca, Piazza della Scienza 3, 20126 Milano, Italy
[14] INFN, Sezione di Milano-Bicocca, Piazza della Scienza 3, 20126 Milano, Italy
[15] INAF – Osservatorio Astronomico di Brera, via Brera 20, 20121 Milano, Italy
[16] Institute for Gravitational Wave Astronomy and School of Physics and Astronomy, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK
[17] INAF – Osservatorio Astronomico di Cagliari, via della Scienza 5, 09047 Selargius (CA), Italy
[18] Hellenic Open University, School of Science and Technology, 26335 Patras, Greece
[19] Kavli Institute for Astronomy and Astrophysics, Peking University, Beijing 100871, PR China
[20] Department of Astronomy and Astrophysics, Tata Institute of Fundamental Research, Homi Bhabha Road, Navy Nagar, Colaba, Mumbai 400005, India
[21] Department of Physics, IIT Hyderabad, Kandi, Telangana 502284, India
[22] Department of Physics and Astrophysics, University of Delhi, Delhi 110007, India
[23] Department of Earth and Space Sciences, Indian Institute of Space Science and Technology, Valiamala, Thiruvananthapuram, Kerala 695547, India

[24] School of Physics, Faculty of Science, University of East Anglia, Norwich NR4 7TJ, UK

[25] Max Planck Institute for Gravitational Physics (Albert Einstein Institute), Am Muühlenberg 1, 14476 Potsdam, Germany

[26] Gran Sasso Science Institute (GSSI), 67100 L'Aquila, Italy

[27] INFN, Laboratori Nazionali del Gran Sasso, 67100 Assergi, Italy

[28] National Centre for Radio Astrophysics, Pune University Campus, Pune 411007, India

[29] Kumamoto University, Graduate School of Science and Technology, Kumamoto, 860-8555, Japan

[30] Università di Cagliari, Dipartimento di Fisica, S.P. Monserrato-Sestu Km 0,700, 09042 Monserrato (CA), Italy

[31] Department of Astrophysics/IMAPP, Radboud University Nijmegen, PO Box 9010, 6500 GL Nijmegen, The Netherlands

[32] Department of Physical Sciences,Indian Institute of Science Education and Research Kolkata, Mohanpur 741246, India

[33] Center of Excellence in Space Sciences India, Indian Institute of Science Education and Research Kolkata, 741246, India

[34] School of Physics, Trinity College Dublin, College Green, Dublin 2, D02 PN40, Ireland

[35] Jodrell Bank Centre for Astrophysics, Department of Physics and Astronomy, University of Manchester, Manchester M13 9PL, UK

[36] Department of Physics, St. Xavier's College (Autonomous), Mumbai 400001, India

[37] National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, PR China

[38] E.A. Milne Centre for Astrophysics, University of Hull, Cottingham Road, Kingston-upon-Hull HU6 7RX, UK

[39] Centre of Excellence for Data Science, Artificial Intelligence and Modelling (DAIM), University of Hull, Cottingham Road, Kingston-upon-Hull HU6 7RX, UK

[40] Department of Physics, BITS Pilani Hyderabad Campus, Hyderabad 500078, Telangana, India

[41] Joint Astronomy Programme, Indian Institute of Science, Bengaluru, Karnataka 560012, India

[42] Arecibo Observatory, HC3 Box 53995, Arecibo, PR 00612, USA

[43] IRFU, CEA, Université Paris-Saclay, 91191 Gif-sur-Yvette, France

[44] Raman Research Institute India, Bengaluru, Karnataka, 560080, India

[45] Institut für Physik und Astronomie, Universität Potsdam, Haus 28, Karl-Liebknecht-Str. 24/25, 14476, Potsdam, Germany

[46] Department of Physics, IISER Bhopal, Bhopal Bypass Road, Bhauri, Bhopal 462066, Madhya Pradesh, India

[47] Ollscoil na Gaillimhe – University of Galway, University Road, Galway H91 TK33, Ireland

[48] Center for Gravitation, Cosmology, and Astrophysics, University of Wisconsin-Milwaukee, Milwaukee, WI 53211, USA

[49] Division of Natural Science, Faculty of Advanced Science and Technology, Kumamoto University, 2-39-1 Kurokami, Kumamoto 860-8555, Japan

[50] International Research Organization for Advanced Science and Technology, Kumamoto University, 2-39-1 Kurokami, Kumamoto 860-8555, Japan

[51] Laboratoire Univers et Théories LUTh, Observatoire de Paris, Université PSL, CNRS, Université de Paris, 92190 Meudon, France

[52] Florida Space Institute, University of Central Florida, 12354 Research Parkway, Partnership 1 Building, Suite 214, Orlando, FL 32826-0650, USA

[53] Ruhr University Bochum, Faculty of Physics and Astronomy, Astronomical Institute (AIRUB), 44780 Bochum, Germany

[54] Advanced Institute of Natural Sciences, Beijing Normal University, Zhuhai 519087, PR China