S. Chen, R.N. Caballero, Y.J. Guo, A. Chalumeau, K. Liu, G. Shaifullah, K.J. Lee, S. Babak, G. Desvignes, A. Parthasarathy, et al.

Common-red-signal analysis with 24-yr high-precision timing of the European Pulsar Timing Array: inferences in the stochastic gravitational-wave background search

S. Chen [1,2]★ R. N. Caballero [3]★ Y. J. Guo,[4] A. Chalumeau,[1,2,5] K. Liu,[4] G. Shaifullah [6,7]
K. J. Lee,[3,4,8] S. Babak,[5,9] G. Desvignes [4,10] A. Parthasarathy [4] H. Hu [4] E. van der Wateren,[11,12]
J. Antoniadis,[4,13,14] A.-S. Bak Nielsen,[4,15] C. G. Bassa,[11] A. Berthereau,[1,2] M. Burgay [16]
D. J. Champion [4] I. Cognard,[1,2] M. Falxa,[5] R. D. Ferdman,[17] P. C. C. Freire [4] J. R. Gair,[18] E. Graikou,[4]
L. Guillemot,[1,2] J. Jang,[4] G. H. Janssen,[11,12] R. Karuppusamy,[4] M. J. Keith,[19] M. Kramer [4,19]
X. J. Liu,[19,20] A. G. Lyne,[19] R. A. Main,[4] J. W. McKee [21] M. B. Mickaliger,[19] B. B. P. Perera [22]
D. Perrodin,[16] A. Petiteau,[5] N. K. Porayko,[4] A. Possenti,[16,23] A. Samajdar,[6] S. A. Sanidas,[19] A. Sesana,[6,7]
L. Speri [18] B. W. Stappers [19] G. Theureau,[1,2,24] C. Tiburzi,[11] A. Vecchio,[25] J. P. W. Verbiest [4,15]
J. Wang,[15] L. Wang [8] and H. Xu[3,8,26]

Affiliations are listed at the end of the paper

ABSTRACT

We present results from the search for a stochastic gravitational-wave background (GWB) as predicted by the theory of General Relativity using six radio millisecond pulsars from the Data Release 2 (DR2) of the European Pulsar Timing Array (EPTA) covering a timespan up to 24 yr. A GWB manifests itself as a long-term low-frequency stochastic signal common to all pulsars, a common red signal (CRS), with the characteristic Hellings-Downs (HD) spatial correlation. Our analysis is performed with two independent pipelines, ENTERPRISE, and TEMPONEST+FORTYTWO, which produce consistent results. A search for a CRS with simultaneous estimation of its spatial correlations yields spectral properties compatible with theoretical GWB predictions, but does not result in the required measurement of the HD correlation, as required for GWB detection. Further Bayesian model comparison between different types of CRSs, including a GWB, finds the most favoured model to be the common uncorrelated red noise described by a power law with $A = 5.13^{+4.20}_{-2.73} \times 10^{-15}$ and $\gamma = 3.78^{+0.69}_{-0.59}$ (95 per cent credible regions). Fixing the spectral index to $\gamma = 13/3$ as expected from the GWB by circular, inspiralling supermassive black hole binaries results in an amplitude of $A = 2.95^{+0.89}_{-0.72} \times 10^{-15}$. We implement three different models, BAYESEPHEM, LINIMOSS, and EPHEMGP, to address possible Solar system ephemeris (SSE) systematics and conclude that our results may only marginally depend on these effects. This work builds on the methods and models from the studies on the EPTA DR1. We show that under the same analysis framework the results remain consistent after the data set extension.

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

1 INTRODUCTION

Radio pulsars, and especially radio millisecond pulsars (MSPs), have been used as astronomical tools to study aspects of fundamental physics with remarkable success, thanks to their exceptional rotational stability. An area of research where MSPs have been particularly useful is gravity (e.g. Taylor 1993; Kramer et al. 2006; Will 2014), especially by employing the 'pulsar timing' technique (e.g. Lorimer & Kramer 2005), which relies on high-precision measurements of the pulses' times-of-arrival (TOAs) being compared to a 'timing model'. The difference between the measured and the model-predicted TOAs is referred to as the 'timing residuals'. Any unmodelled effects will appear in the timing residuals, and the timing model is revised and/or extended accordingly. The timing models have astounding predictive power as it only requires the precise modelling of the pulsar's rotation, orbital motion, and the signal's propagation in space, and not the details of the radiation's physics or emission mechanism. Pulsar timing observations provided the first evidence for the existence of gravitational waves (GWs; Taylor & Weisberg 1982), by confirming that the measured orbital changes of the binary pulsar PSR B1913+16 match those predicted by the theory of General Relativity (GR) due to the system's energy loss through the emission of GWs.

MSPs have been proposed as a tool for the direct detection of GWs at nHz frequencies (Sazhin 1978; Detweiler 1979). The experiment is based on systematically observing an ensemble of MSPs at many

★ E-mail: siyuan.chen@cnrs-orleans.fr (SC); caballero.astro@gmail.com (RNC)

sky positions, a configuration called a 'Pulsar Timing Array' (PTA; Foster & Backer 1990). Various theories of gravity, including GR, predict that the propagation of GWs cause distortions in the space–time metric, with specific polarization modes (e.g. Eardley, Lee & Lightman 1973a; Eardley et al. 1973b; Estabrook & Wahlquist 1975). GWs propagating in the vicinity of the Earth and the pulsars induce spatially correlated variations in the TOAs over the time-scale of several decades. One of the physically motivated sources of GWs in the nHz band are inspiralling supermassive black hole binaries (SMBHBs). The incoherent superposition of a large number of unresolved SMBHB GW signals forms a GW background (GWB) that may be detected with a PTA (Rajagopal & Romani 1995; Jaffe & Backer 2003; Wyithe & Loeb 2003; Sesana et al. 2004). In addition, GWBs from cosmic strings (Kibble 1976; Sanidas, Battye & Stappers 2012) or a cosmological relic GWB from the inflationary era (see Grishchuk 2005) have also been proposed as PTA target signals. As such, PTAs can provide direct observational constraints for large-scale structure and cosmological models.

The basic idea of searching for a GWB with PTAs is to look for this common red signal (CRS) with the characteristic spatial correlation in the array of MSPs (see Hellings & Downs 1983; Lee, Jenet & Price 2008). While PTA data are also being used for other applications, such as probing Solar system planetary parameters (e.g. Caballero et al. 2018) establishing pulsar-based time-scales (e.g. Hobbs et al. 2020) and measuring local clock instabilities (Li et al. 2020), the nHz GW search remains the primary objective of PTA efforts.

After the first experimental efforts to establish a PTA pioneered by Donald Backer and collaborators (e.g. Romani 1988; Foster & Backer 1990; Backer 1995), collaborations with European partners to also utilize the Effelsberg 100-m radio telescope (EFF) and the Nançay Radio Telescope (NRT) started corresponding timing efforts. Together with regular timing efforts with the Lovell Telescope (LT) and the Westerbork Synthesis Radio Telescope (WSRT), this laid the foundation for uninterrupted PTA data sets spanning now up to 24 yr for a number of sources. Inspired by the formation of the Parkes Pulsar Timing Array (PPTA; Manchester 2006; Hobbs 2013), the European Pulsar Timing Array was officially established in January 2006 (Stappers et al. 2006; Janssen et al. 2008) as a collaboration of European radio observatories and research institutions working towards the direct detection of nHz GWs. Apart from the PPTA, the EPTA has also been working alongside the North-American Nanohertz Observatory for Gravitational Waves (NANOGrav; Jenet et al. 2009; Arzoumanian et al. 2015). The three groups collaborate under the International Pulsar Timing Array (IPTA; Verbiest et al. 2016; Perera et al. 2019), with combined data-sets and shared resources and expertise. The advent of further sensitive telescopes able to perform precision pulsar timing observations, especially the Giant Metrewave Radio Telescope (GMRT; Swarup 1990) in India, which recently formally joined the IPTA, and more recently the Five Hundred Meter Spherical Telescope (FAST; Jiang et al. 2019) in China, and the MeerKAT telescope (Camilo et al. 2018) in South Africa with the MeerTIME program (Bailes et al. 2020), adds important capabilities to the IPTA.

With continuous improvements of the hardware used, the current EPTA data set (henceforth Data Release 2; DR2) extends the previous data set (DR1; Desvignes et al. 2016) by up to 7 yr. From early on, coherent-dedispersion data acquisition hardware has been in place at some of the telescopes [e.g. the Effelsberg Berkeley Pulsar Processor (EBPP), the Berkeley Orléans Nançay Instrumentation (BON), or PuMa II at the WSRT, see Section 2]. Additionally, the Large European Array of Pulsars (LEAP; Kramer & Stappers 2010; Bassa et al. 2016a) efforts combine the EPTA telescopes in tied-array

mode to form a 194-m equivalent dish for MSP timing with monthly cadence for the last 8–10 yr. As a result, the EPTA has effectively operated as a 'mini-IPTA', combining data sets from six different telescopes. Consequently, the combination of data sets is complex, but the availability to cross-check the data with multiple overlapping data set helps enormously in identifying and solving instrumental problems.

In the EPTA DR1 GWB upper limit analysis, Lentati et al. (2015) (henceforth LTM15) applied a methodology to probe a number of physically motivated CRSs in the data set and compare their probability against the GWB with the Hellings–Downs (HD) correlation (Hellings & Downs 1983). This methodology was created because there was some modest evidence in this first data combination of a common signal and it was recognized that this would likely require careful examination in a future extended data set. In this work, we analyse the same six pulsars used in EPTA DR1 GW analyses (Lentati et al. 2015; Taylor et al. 2015; Babak et al. 2016; Caballero et al. 2016), and present updated results using the EPTA DR2 on the properties of a CRS in the data. Building upon our DR1 methods we include the same noise terms for the pulsar properties as well as the pulse propagation through the ionized interstellar medium (IISM), which has been the standard first-order analysis approach in pulsar timing for a number of years. While we are working on further optimization in modelling and analysis methods, which could potentially improve our results, conducting the analysis here in the same model framework as in DR1 allows us to get a direct comparison and measure of the improvement achieved with the new data. In order to increase confidence in the results, we make use of multiple independently developed analysis codes, some of which are used for the first time, for the parameter estimation, model selection and modelling of Solar system dynamics. As a first-order examination of the stationarity of the CRS we also analyse the evolution of this signal from DR1 to DR2. This work is the first in a series of papers which will present results from multiple types of analysis using a larger number of pulsars with the updated EPTA DR2 in order to fully establish if the CRS persists and what type of spatial correlation it has.

The rest of the paper is organized as follows: In Section 2, we present the properties of the data used in this study. Section 3 provides an overview of the modelling and analysis framework. Results from the single-pulsar analysis are presented in Section 4. Section 5 shows the results from a search for a CRS with simultaneous analysis of the signal's angular correlations and Bayesian model selection between different possible types of CRSs to gauge which is more supported by the data. The most supported model is further investigated in Section 6, where we investigate effects from choices in modelling the CRS spectrum, the pulsars' contributions to the common signal, the time-stationarity of the CRS and the effects of the SSE. We discuss the results in the framework of the GWB and compare these to results from the literature in Section 7 and discuss our conclusions in Section 8.

2 DATA

The EPTA has continued to monitor approximately 50 pulsars with high observing cadence since the first data release (Desvignes et al. 2016), using the five European radio telescopes both in single-dish and LEAP modes. High-sensitivity TOAs produced from coordinated monthly LEAP observations are included for the first time in the EPTA data set. Current observations utilize the new generation of data recording systems at all telescopes, primarily using the Reconfigurable Open Architecture Computing Hardware (ROACH)

FPGA board developed by the CASPER group,[1] which allows for coherent dedispersion (van Straten & Bailes 2011). At EFF, the observations were conducted mainly at three frequency bands centred at 1347, 2627, and 4850 MHz, and the data were recorded with the 'PSRIX' backend with a bandwidth of 200 MHz (Lazarus et al. 2016). The LT observations were carried out at a central frequency of 1532 MHz, with data recorded using the 'ROACH' pulsar backend with a bandwidth of 400 MHz (Bassa et al. 2016b). At the NRT, observations were performed centered at 1484 and 2539 MHz, and the data were recorded with the 'NUPPI' backend with a bandwidth of 512 MHz (Cognard et al. 2013). At the WSRT, observations were made at 350, 1380, and 2200 MHz, and the data recording was performed with the 'PuMa II' backend (Karuppusamy, Stappers & van Straten 2008). The Sardinia Radio Telescope (SRT) is the latest telescope addition to the EPTA and now effectively participates in EPTA combined data set. The majority of the SRT observations were made as part of LEAP sessions. Due to the small number of the SRT single-telescope observations the corresponding TOAs were not included in this data set. The SRT observations were conducted at 1396 MHz with the data recorded using the 'ROACH1' backend using a bandwidth of 128 MHz. The LEAP observations were performed at 1396 MHz using the same backends but with a recording bandwidth of 128 MHz at each telescope.

The six DR1 priority pulsars used in this paper, are: PSRs J0613−0200, J1012+5307, J1600−3053, J1713+0747, J1744−1134, and J1909−3744. Each pulsar is regularly observed by all the EPTA telescopes, except for J1909−3744 which given its sky location has so far only been monitored by NRT and the SRT; however, as noted above, single-telescope SRT data are not included in the current data set. The data from the new backends collected at each telescope were processed using the PSRCHIVE software package (Hotan, van Straten & Manchester 2004), to carry out calibration and radio-frequency interference mitigation. Then, for each observation epoch, an 'integrated profile' was formed by averaging the data in time and frequency. For data from EFF, WSRT, and LEAP, the frequency averaging was performed over the entire band. For data from LT and NRT, the averaging was done in two and four sub-bands, respectively, to accommodate their larger bandwidth. The TOAs of these profiles were then calculated using the canonical template-matching method (Taylor 1992).

We note that for about 3 yr, observations were made simultaneously both with the older backends (as they appear in DR1) and the new ones with the corresponding TOAs from the new backends.[2] In this work, we replace all those DR1 TOAs. We also exclude single-telescope data from epochs that were used to create the corresponding LEAP TOAs. As seen in Fig. 1, although the timespan extension from DR1 is ∼7 yr, the effective improvement due to the new-generation data is ≳10 yr.

A summary of the combined data set can be found in Table 1. We refer to Lazarus et al. (2016), Bassa et al. (2016a), Perera et al. (2018), Liu et al. (2020), and the forthcoming EPTA DR2 paper for more details of the observations and data from each individual telescope. The full details for the backends, data and TOA extraction for the

DR1 data that are part of this data set can be found in Desvignes et al. (2016).

3 ANALYSIS FRAMEWORK

In this section, we briefly summarize the established mathematical and algorithmic framework used in PTAs to analyse TOAs. While the information described in the section can be found in the literature, we provide this overview as the methods were progressively created in many publications. This section can also serve as a quick reference for the planned follow-up papers. The interested reader can study the provided references for detailed explanations and formula derivations.

The analysis is divided in two main parts: (a) the single-pulsar analysis, which provides the pulsar timing parameters and stochastic noise parameters for each MSP, and (b) the CRS analysis, which describes the methods used in the search for CRSs, including the GWB, in the TOAs of all MSPs and the investigation of their spatial correlations.

3.1 Single-pulsar timing and noise modelling

Our search for common signals between pulsars is preceded by single-pulsar analyses. This process provides pulsar models that comprise of the timing and noise parameters. The former induce deterministic signals, while the latter induce stochastic signals. In the case of GW searches with PTAs, the pulsar noise analysis is equivalent to the characterization of a GW-detector noise and is thus a necessary step before the search analysis, as pulsar noise can correlate with GW signals and reduce the data's sensitivity (see e.g. Caballero et al. 2016). All sources of noise therefore need to be measured in order to be decorrelated from the signal of interest, and to properly evaluate the possibility for the existence of any common signals.

Timing parameters are typically measured progressively as the pulsar data set increases with new data. Least-squares linear fits are very effective in producing phase-connected timing models (i.e. models that account for every pulsar rotation). The timing analysis in all cases was performed using the pulsar timing package TEMPO2 (Hobbs, Edwards & Manchester 2006). It is used to first derive the basic phase-connected timing solution which uses a linear approximation for the pulsar timing model, i.e. assuming that the timing-model parameters may only have small, linear deviations from the true values. In practice, this means that on every update of the model we assume that there are only linear deviations from the pre-fit values of the timing parameters to their post-fit values. The topocentric TOAs (i.e. TOAs at the observatory) were mapped to the Solar system barycentre to form the barycentric TOAs using the Solar system ephemeris (SSE) DE438 from the Jet Propulsion Laboratory (Folkner & Park 2018). The TOAs were referred to the terrestrial time-scale BIPM2019, provided by the Bureau International des Poids et Mesures.[3]

While the least-squares linear approximation timing model is only valid if the timing data are white-noise dominated, pulsar data often contain other types of non-white (i.e. time-correlated) noise processes. In the presence of such correlated noise processes, even for phase-connected timing models, the values and uncertainties of the timing parameters can be biased (e.g. Coles et al. 2011). A full analysis of pulsar timing data with the intention of accurate

---

[1] http://casper.berkeley.edu/

[2] These periods of observations with simultaneous data recordings using older and newer instruments are intentionally carried out at the observatories in order to confirm the good performance of the new instruments during commissioning, and to accurately measure the necessary phase offset between the two data sets, as required when creating the combined data set.

[3] www.bipm.org

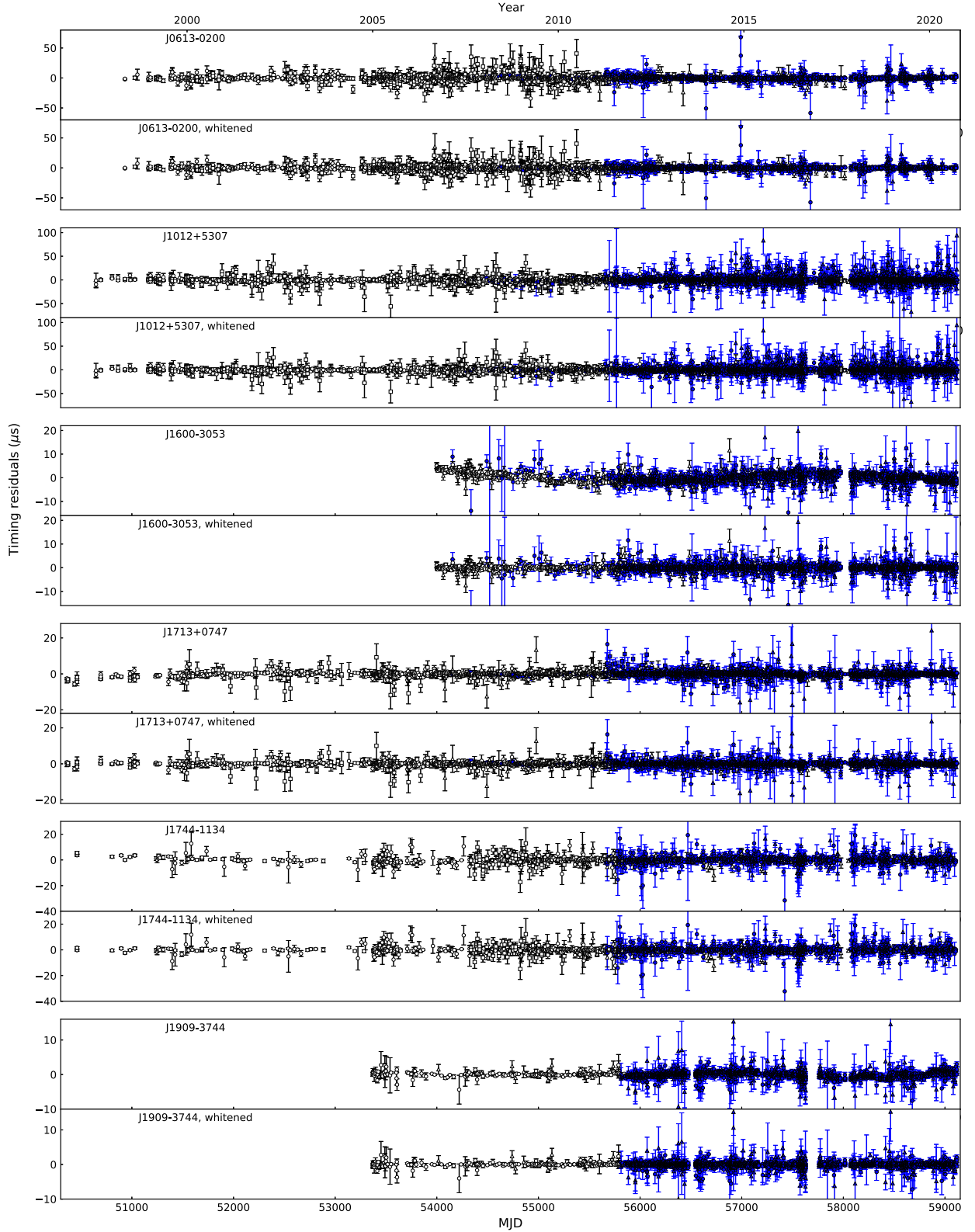**Figure 1.** Timing residuals of the six pulsars used in this paper. For each pulsar, the residuals before and after subtraction of DM and red noise are shown. The squares, circles and triangles represent *P*-band, *L*-band and *S/C*-band observations, respectively (see Table 1 for information on the observing frequency bands). The blue/filled and black/unfilled symbols indicate the new backend data and DR1 TOAs, respectively.

Table 1. Overview of the six-pulsar data set used in this paper's analysis. The columns represent the name of the pulsar, the coverage of frequency band, the number of TOAs, the timespan T of the data set, the median σ_{TOA} in each frequency band, the timing residual rms, the rms after subtraction of DM noise, and the rms after subtraction of both red and DM noise (whitened). The corresponding frequency ranges for the P, L, S, and C bands are 0.3–1.0, 1.0–2.0, 2.0–4.0, and 4.0–8.0 GHz, respectively.

Pulsar name	Band	N_{TOA}	T (yr)	Median σ_{TOA} (μs)				rms (μs)	rms, DM removed (μs)	rms, whitened (μs)
				P	L	S	C			
J0613−0200	L, S, C	3022	22.4	–	1.527	7.113	7.400	1.415	1.281	1.168
J1012+5307	P, L, S, C	5837	23.2	4.800	1.705	7.859	4.756	1.393	1.326	1.233
J1600−3053	L, S	3345	14.0	–	0.475	1.762	–	1.376	0.766	0.439
J1713+0747	P, L, S, C	5052	24.0	1.329	0.308	0.696	0.703	0.415	0.355	0.311
J1744−1134	P, L, S, C	1980	23.7	3.700	0.912	3.185	1.046	0.898	0.736	0.653
J1909−3744	L, S	2817	15.7	–	0.268	0.667	–	0.504	0.424	0.228

pulsar parameter measurements is therefore better achieved via a simultaneous fit of the timing and noise models. In this work we measure the stochastic noise using the linear approximation to the timing model by analytically marginalizing the likelihood over the timing parameters, as we show below. This accounts for possible covariances between the timing and noise parameters and has consequently been used in standard PTA analyses. We note that a more precise timing model can be derived iteratively with TEMPO2 via a generalized least-squares fit, using the covariance matrix of a noise model (see e.g. Coles et al. 2011), such as the maximum-likelihood model from the Bayesian noise analysis; this approach was used to produce the timing residuals presented in Fig 1.

In this study, we use the same pulsar noise models as in the EPTA DR1 analysis (Caballero et al. 2016; Desvignes et al. 2016). The model uses two white-noise parameters to properly re-scale the TOA uncertainties, an achromatic red noise component, and a chromatic dispersion-measure (DM) noise component. We discuss these noise components, in brief.

The noise analysis is conducted using a Bayesian framework, following the same general approach as in the EPTA DR1 (Caballero et al. 2016), and uses two independent analysis codes to cross-check for consistency in the results (see Section 4.1 for details). We present an overview of the mathematical configuration and provide basic details for the calculations of the noise covariance matrices, at the level where the used analysis codes follow the same principles, even though differences can appear in the exact programming implementations of these principles.

For a single-pulsar timing problem, where we have n TOAs, the likelihood function was first introduced in van Haasteren et al. (2009) as

$$L_{PSR} = \frac{e^{-\frac{1}{2}(\delta t_{post})^T \mathbf{C}^{-1}(\delta t_{post})}}{\sqrt{(2\pi)^n |\mathbf{C}|}} . \tag{1}$$

We use the subscript $_{PSR}$ to denote that the function is for the single-pulsar problem, and the ^{T} superscript to denote the matrix transpose. The likelihood function is derived under the assumption of operating in the linear approximation of the timing model, where the initial values of the timing parameters are close to the correct ones and only linear deviations are required when re-fitting the data. In this way, the (post-fit) timing residuals are estimated with the use of the design matrix of timing parameters, $\mathbf{M}$, and the vector ξ of the amplitudes of the timing-parameter signals, as $\delta t_{post} = \delta t_{pre} - \mathbf{M}\xi$, where δt_{pre} are the pre-fit residuals (i.e. prior to the new fit of the timing model). The effects of the additional stochastic parameters are described via the pulsar's total covariance matrix, $\mathbf{C}$. Our noise model assumes three stochastic noise components in the data: white, red, and DM. Each of these noise components is defined by a covariance matrix:

$\mathbf{C}_W$, $\mathbf{C}_R$, and $\mathbf{C}_{DM}$, respectively. The total covariance matrix is the sum of these,

$$\mathbf{C} = \mathbf{C}_W + \mathbf{C}_R + \mathbf{C}_{DM} . \tag{2}$$

van Haasteren et al. (2009) describe a way to analytically marginalize the pulsar timing parameters using the reduced likelihood function

$$L'_{PSR} = \frac{e^{-\frac{1}{2}(\delta t_{pre})^T \mathbf{C}'(\delta t_{pre})}}{\sqrt{(2\pi)^{n-m} \times |\mathbf{C}| \times |\mathbf{M}^T \mathbf{C}^{-1}\mathbf{M}|}} , \tag{3}$$

where $\mathbf{C}' = \mathbf{C}^{-1} - \mathbf{C}^{-1}\mathbf{M}(\mathbf{M}^T\mathbf{C}^{-1}\mathbf{M})^{-1}\mathbf{M}^T\mathbf{C}^{-1}$, and m is the number of timing parameters that are marginalized. In this work, the analytical marginalization over the timing parameters is performed in one of two ways, depending on the computing algorithm used (discussed later). We briefly discuss these methods in chronological order of their appearance in the literature.

One approach currently used for marginalizing the timing model, which we refer to as the 'G-matrix' approach, is a more numerically stable implementation of the initial formulation. It uses a reduced likelihood function defined as (van Haasteren & Levin 2013)

$$L'_{PSR,1} = \frac{e^{-\frac{1}{2}(\delta t_{pre})^T \mathbf{G}(\mathbf{G}^T\mathbf{C}\mathbf{G})^{-1}\mathbf{G}^T(\delta t_{pre})}}{\sqrt{(2\pi)^{n-m} \times |\mathbf{G}^T\mathbf{C}\mathbf{G}|}} , \tag{4}$$

where $\mathbf{G}$ can be derived from the design matrix via a singular value decomposition (see van Haasteren & Levin 2013, for details). As in equation (1), the dimensions of the $\mathbf{C}$ and $\mathbf{G}$ are $n \times n$ and $n \times (n - m)$, respectively.

The second marginalization method follows a different route, and we refer to it as the 'Gaussian-process' approach (e.g. van Haasteren & Vallisneri 2014; Arzoumanian et al. 2016). The post-fit residuals are now expressed by taking into account effects from the stochastic noise components in addition to the timing parameters via the design matrix. Further, the timing model effects are added to the total covariance matrix (equation 2) resulting in $\mathbf{C}'_2 = \mathbf{C}_W + \mathbf{TBT}^T$. The matrix $\mathbf{T}$ includes the design matrix in addition to terms related to the stochastic noise and $\mathbf{B}$ is a matrix used to define the prior distributions of the timing and noise-related parameters (these are frequency-domain parameters that sample the noise, as discussed later in this section, see e.g. equation 10). In this case, the marginalized likelihood becomes

$$L'_{PSR,2} = \frac{e^{-\frac{1}{2}(\delta t_{pre})^T \mathbf{C}'^{-1}_2(\delta t_{pre})}}{\sqrt{(2\pi)^n \times |\mathbf{C}'_2|}} . \tag{5}$$

By setting uniform and infinite priors on the timing parameters via $\mathbf{B}$ and using the Woodbury matrix identity (Woodbury 1950; Hager 1989) on $\mathbf{C}'^{-1}_2$, van Haasteren & Vallisneri (2014) have recovered an equivalent to the marginalization in equation (3) times a constant

from the prior matrix, **B**, that acts only as a renormalization in the Bayesian analysis.

To conclude the basic description of the single-pulsar analysis, we discuss the noise components and their respective covariance matrices. In all cases, the stochastic noise components are modelled as random Gaussian processes. Detailed information on the noise modelling and the likelihood function can be found in van Haasteren & Levin (2013), van Haasteren & Vallisneri (2014), Lentati et al. (2013, 2014a), and references therein, so we only provide an overview here for completeness.

3.1.1 White noise

The white noise is modelled with the so-called EFAC and EQUAD stochastic parameters. For each observing system, defined by the telescope receiver and backend system combination, we apply an (EFAC, EQUAD) pair to re-scale the initial TOA measurement uncertainty, σ, according to equation

$$\hat{\sigma}^2 = (\sigma \cdot \text{EFAC})^2 + \text{EQUAD}^2 . \tag{6}$$

EFAC is a multiplicative factor that accounts for possible errors in the estimation of the formal TOA uncertainty as calculated during the cross-correlation of the pulse profile with a standard template (Taylor 1992). EQUAD is a term added in quadrature to account for additional scatter of the TOAs due to physical effects such as pulsar jitter noise (Ekers & Moffet 1968), an effect that appears to be often relevant in high precision pulsar timing observations (e.g. Liu et al. 2011, 2012; Osłowski et al. 2011; Shannon et al. 2014; Lam et al. 2016, 2019; Parthasarathy et al. 2021).

The TOA uncertainties are re-scaled such that the data conforms to our basic assumption that they are drawn from a Gaussian parent distribution. The case of non-Gaussian uncorrelated TOA noise is discussed theoretically in the literature (see e.g Lentati, Hobson & Alexander 2014b; Vallisneri & van Haasteren 2017), and we note that evidence for non-Gaussianities in TOAs have recently been reported for a small number of MSPs (Goncharov et al. 2021b). Increasing TOA precision may result in the need to take non-Gaussianities into account in the future.

The white-noise covariance matrix, $\mathbf{C}_W$, is a diagonal matrix with the re-scaled variances of the TOAs (equation 6) as its elements, i.e.

$$C_{W,i,j} = \hat{\sigma}_{ij}^2 \delta_{ij} . \tag{7}$$

3.1.2 Red and dispersion-measure noise

The red noise component addresses the pulsar's intrinsic low-frequency noise, achromatic noise, which has mostly been associated with pulsar-spin irregularities (e.g. Cordes & Downs 1985; D'Alessandro et al. 1995). While MSPs have much lower levels of red noise than young pulsars, the effects are measurable, especially for long-term data sets and at this high timing precision (e.g. Verbiest et al. 2009; Caballero et al. 2016; Alam et al. 2021; Goncharov et al. 2021b).

DM is defined as the integrated column density of free electrons in the pulsar's line of sight and can vary over time (You et al. 2007). In principle, it is possible to have DM measurements for each observation if a sufficient range of observing frequencies is covered for each observation. In the case of the EPTA data, however, we are dealing with a very long data set where, due to differing circumstances at our participating telescopes over the last 24 years, there is a significant level of inhomogeneity in the observing frequencies across the data set. We therefore opt to fully

model the temporal DM variations as a combination of deterministic and stochastic processes, as discussed in Lee et al. (2014). The timing model parameters of the DM value at a reference epoch and its first- and second-time derivative give a first-order deterministic approximation. Delays from stochastic turbulences of the IISM add an additional DM noise component.

In the models used in this work, both the red and DM stochastic noise components are described as wide-sense stationary signals with a single power-law spectrum of the form

$$S \propto A^2 f^{-\gamma} , \tag{8}$$

where A is the spectrum's amplitude and γ its spectral index.

The red-noise covariance matrix is calculated via the Woodbury identity (Woodbury 1950; Hager 1989). We consider the Fourier transforms **F** of the time-domain red noise signal t_R, which has been shown to be a well-performing approximation in pulsar timing analysis (Lentati et al. 2013; van Haasteren & Vallisneri 2015). We then have

$$(\mathbf{C}_W + \mathbf{C}_R)^{-1} = \mathbf{C}_W^{-1} - \mathbf{C}_W^{-1}\mathbf{F}[\mathbf{F}^\mathsf{T}\mathbf{C}_W^{-1}\mathbf{F} + \Psi^{-1}]^{-1}\mathbf{F}^\mathsf{T}\mathbf{C}_W^{-1} . \tag{9}$$

In matrix notation, the Fourier transforms are

$$t_R = \mathbf{F}a . \tag{10}$$

The Fourier elements are defined by a limited number of sine-cosine pairs $\{\sin(2\pi tf); \cos(2\pi tf)\}$ with coefficients a and corresponding covariance matrix $\Psi = \langle a_f a_{f'} \rangle$ (where the indices f, f' correspond to the different frequencies).

In contrast to the red noise, the DM noise is chromatic and the model demands that the induced TOA delays are related to the inverse square of their observing radio frequency, ν, following the dispersion law of cold plasma (e.g. Landau & Lifshitz 1960). The covariance matrix of the DM stochastic noise is calculated using the same recipe as in the case of the red noise component, with one change that introduces the dependency of the induced time delay on the observing frequency (Lee et al. 2014). To achieve this in the frequency-domain noise models used here, the sine and cosine Fourier transform elements are multiplied by the corresponding (in time) elements of a vector $\mathbf{F}^{DM}$ with length equal to the number of TOAs and elements (LTM15)

$$F_i^{DM} = 1/K\nu_i^2 , \tag{11}$$

where $K = 2.41 \times 10^{-16}$ (Hz^{-2} cm^{-3} pc s^{-1}) is the dispersion constant and ν_i is the observing frequency of the ith TOA.

3.1.3 J1713+0747 events

The TOAs of PSR J1713+0747 are characterized by two additional structures in the timing residuals in this data set, caused by discrete 'events' affecting the pulsar's signal. They manifest as a rapid decrease in the timing residuals, followed by a recovery occurring on time-scales of the order of ~100 d. The first of these events has been detected by all PTAs (see Demorest et al. 2013; Keith et al. 2013; Desvignes et al. 2016) and has been linked to discrete variations in the DM parameter that are not strictly caused by the turbulence in the IISM. The second one has previously been reported in Lam et al. (2018) and Goncharov et al. (2021b) with some evidence indicating a deviation from the DM nature of the event. These events have been physically interpreted as being caused by plasma lensing effects, due to discrete, underdense IISM structures (Lam et al. 2018). Alternatively, Goncharov et al. (2021b) claim that the second event could be related to a sudden change in the pulsar profile. As these

events are non-stationary, they are not accurately accounted for when modelling the power spectrum of the long-term DM and pulsar red noise.

The first of these events was modelled in the EPTA data using shapelet basis functions (see Desvignes et al. 2016) and we follow the same strategy for both events in this work. A comparison of the single pulsar analyses with and without these two events shows significant changes in the parameters of the DM stochastic component, but mostly unaffected red noise properties. When not including the events, the stochastic DM noise component has a larger amplitude which compensates for the spectral power located in the DM variations induced by the events. However, we do note that quantitative model comparisons significantly support the model with the additional events for PSR J1713+0737. We will address these details, including a comparison with findings from other PTAs (Hazboun et al. 2020; Goncharov et al. 2021b) in an upcoming paper that will focus on more precise noise models for the EPTA DR2 data (Chalumeau et al., in preparation). We have verified with test runs that modelling or not these events separately does not affect the CRS analysis, as the red noise properties of the MSP (which is the noise that can correlate with a GWB or other achromatic CRSs) remain unaffected by including or not the parameters for the additional events. A more detailed discussion can be found in Chalumeau et al. (in preparation). Therefore, we note that for simplicity in this work the GWB search and all analyses regarding the search for common signals discussed in Sections 5 and 6 use the simpler model that does not include these events.

3.2 Gravitational-wave background search with pulsar timing arrays

The GWB is a stochastic signal, which can be parametrized by a power law that describes the GW-frequency dependence of its characteristic strain, h_c, a measure of the space–time deformation that the GWB induces (see e.g. Maggiore 2000; Jenet et al. 2006). This dimensionless strain spectrum is defined as

$$h_c = A_{\mathrm{GWB}} \left(\frac{f}{f_c} \right)^{\alpha} , \tag{12}$$

where f is the GW frequency and f_c is a reference frequency (typically set to 1 yr^{-1}), A_{GWB} is the GWB strain amplitude at reference frequency and α is the spectral index, which varies based on the physical origin of the GWB. Conveniently, we can express the effect of h_c on the observed pulsar timing residuals, via the one-sided power spectral density of the GWB induced residuals as

$$S_{\mathrm{GWB}} = \frac{A_{\mathrm{GWB}}^2}{12\pi^2} \left(\frac{f}{f_c} \right)^{-\gamma_{\mathrm{GWB}}} f_c^{-3} . \tag{13}$$

The spectral indices of the GWB characteristic strain and power spectrum are related as $\gamma_{\mathrm{GWB}} = 3 - 2\alpha$. For a GWB from the cosmic SMBHB population, the value of α depends on the astrophysical details of the inspiralling SMBHBs, such as whether their orbital dynamics are coupled with their stellar and gaseous environments (see e.g. Sesana 2013; Chen, Sesana & Del Pozzo 2017). For the case of circular, GW-driven binaries, the slope is $\alpha = -2/3$ or $\gamma_{\mathrm{GWB}} = 13/3$.

The timing residuals induced by an isotropic GWB are spatially correlated. The GWB-induced residuals comprise of two terms: the pulsar and Earth terms, which are caused by the space–time deformation from the propagation of the GWB at the pulsar and Earth positions, respectively. Only the Earth terms are coherent across all pulsars, leaving the pulsar terms to act as noise. The correlation

coefficients between the timing residuals of all pulsar pairs as a function of the pulsar-pairs angular correlations is called the overlap reduction function (ORF; see Finn, Larson & Romano 2009). Owing to different GW polarization modes or graviton masses, GR and alternative theories of gravity predict different ORFs (see Lee 2013, for an overview). In the context of GR, GWs have a quadrupolar spatial correlation. As pulsars are separated from each other and from the Earth by multiple GW wavelengths (short-wavelength approximation) the ORF is approximated by the HD curve (Hellings & Downs 1983), with terms calculated as

$$\Gamma_{\mathrm{GWB}}(\zeta_{IJ}) = \frac{3}{2} x_{IJ} \ln x_{IJ} - \frac{x_{IJ}}{4} + \frac{1}{2} + \frac{1}{2}\delta x_{IJ} . \tag{14}$$

Here, the I, J indices represent the different pulsars, Γ_{IJ}, is the predicted correlation coefficient per pulsar pair with angular separation ζ_{IJ}, $\delta(x_{IJ})$ is the Kronecker delta, and $x \equiv [1 - \cos(\zeta)]/2$. Note that due to the isotropic nature of the GWB, only the angular separation between pulsar pairs is important, and not the position of each pulsar. The measurement of the ORF curve is central in every method developed for GWB searches and detection, as PTAs can in fact detect different CRSs other than the GWB, as we discuss below. Therefore, no GWB detection claim can be made without a statistically significant measurement of the HD curve. As such, the search for a GWB essentially amounts to searching for a CRS in the PTA data, determining its spectral properties and ORF and comparing the results with the theoretical predictions of a HD correlated signal.

For joint analysis of all pulsar data in search of a CRS like the GWB, the likelihood function can be expressed by generalizing equation (1) as

$$L_{\mathrm{PTA}} \propto \frac{\mathrm{e}^{-\frac{1}{2} \sum_{I,J,i,j} (\widetilde{\delta t}_{I,i})^{\mathsf{T}} \widetilde{\mathbf{C}}^{-1} (\widetilde{\delta t}_{J,j})}}{\sqrt{|\widetilde{\mathbf{C}}|}} . \tag{15}$$

We use the subscript $_{\mathrm{PTA}}$ to indicate that the equation is for the multipulsar case, and we use tilted overlines to indicate the corresponding data and matrices. Therefore, $\widetilde{\delta t}$ is a vector of all pulsar timing residuals concatenated. The total covariance matrix, including the CRS signal is then

$$\widetilde{\mathbf{C}} = \widetilde{\mathbf{C}}^* + \mathbf{C}_{\mathrm{CRS}} . \tag{16}$$

where $\widetilde{\mathbf{C}}^*$ is the block diagonal of all pulsar covariance matrices. For any CRS with ORF $\Gamma_{\mathrm{CRS}}(\zeta_{IJ})$, its covariance matrix elements are calculated as

$$C_{\mathrm{CRS},I,J,i,j} = \Gamma_{\mathrm{CRS}}(\zeta_{IJ}) C_{\mathrm{CRS}}(i, j) , \tag{17}$$

where $C_{\mathrm{CRS}}(i, j)$ is the CRS-induced time-correlation, that is calculated as in the case of pulsar red noise (equation 9). The corresponding reduced likelihood will depend on the method chosen to marginalize over all the timing parameters, as discussed in Section 3.1.

It is also possible to assume the ORF and derive the corresponding spectral properties. As we will see in Section 5.3, these types of analyses can provide a basis for examining how well a given model fits the data with respect to another competing model. This is a strategy followed when the CRS analysis cannot result in a clear measurement on the ORF shape. Here, we briefly discuss the main CRS models other than the GWB that are of particular interest to PTAs and which we examine separately in Section 5.3. We note that these models were also discussed and investigated in the framework of the EPTA DR1 GWB upper limit work presented in LTM15. A general study and discussion of the effects of these types of signals in pulsar timing data and PTA GW searches can be found in Tiburzi et al. (2016).

3.2.1 Common uncorrelated red noise

PTA data may contain a common, but spatially uncorrelated red noise (we denote this with CURN). A CURN reflects the situation where the individual pulsar red noise of all pulsars include some noise component which have very similar spectral properties, either due to common physical origin or by chance. The CURN case therefore is not technically a common term, as it is not induced in every pulsar TOAs from the same extrinsic process. This would, however, mimic a CRS, but the calculated ORF of all pulsar pairs would be expected to be randomly distributed around zero, therefore having no spatial correlations. The CURN is assumed to also be described by a red signal with a power-law spectrum, following equation (8), with a unique amplitude and spectral index, A_{CURN} and γ_{CURN}. The covariance matrix $\mathbf{C}$ calculated as in the case of pulsar red noise in equation (2). In this case, for each pulsar, I, the total red noise covariance matrix becomes $\mathbf{C}'_{\mathrm{R}}(I) = \mathbf{C}_{\mathrm{R}}(I) + \mathbf{C}_{\mathrm{CURN}}$. In turn, this will affect the (block diagonal) multipulsar covariance matrix that we now use as the total covariance matrix $\tilde{\mathbf{C}}$ in the likelihood function (equation 15).

3.2.2 Clock-error monopolar signal

One source of possible correlated CRS in pulsar TOAs is an error in the terrestrial time-scale to which we refer all TOAs (we denote this with CLK). For example, as noted in Section 2, the TOAs in this work were referred to the BIPM2019 time-scale. As all TOAs are referred to the same time-scale, any imperfections in this time-scale will affect the TOAs of all pulsars in a fully correlated way. This effect has been extensively discussed in the literature and it is exploited in order to create pulsar-based time-scales that can be complementary to atomic time-scales (e.g. Guinot & Petit 1991; Hobbs et al. 2012, 2020).

The CLK signal in this work is also modelled as a stochastic red signal with a power-law spectrum as in previous studies (Lentati et al. 2015; Caballero et al. 2016; Hobbs et al. 2020) with parameters A_{CLK} and γ_{CLK}. It has a monopolar correlation, with ORF terms

$$\Gamma_{\mathrm{CLK}}(\zeta_{IJ}) = 1, \ \forall(I,J). \tag{18}$$

3.2.3 Ephemeris-error dipolar signal

Possible errors in the planetary masses and orbital elements in the used SSE can induce TOA delays in the PTA pulsars with dipolar spatial correlations (we denote this with EPH). These signals reflect the oscillation of the calculated Solar system barycentre position with respect to the true position. As in the case a of clock-error signal, this can also be exploited in order to provide upper limits or measurements of planetary parameters using PTA data (see e.g. Champion et al. 2010; Caballero et al. 2018; Guo et al. 2019). Unlike the other signals discussed in this Section, the SSE signals are deterministic signals associated with planetary orbits. The ORF is defined as (see e.g. Tiburzi et al. 2016, Appendix A),

$$\Gamma_{\mathrm{EPH}}(\zeta_{IJ}) \propto \cos(\zeta_{IJ}). \tag{19}$$

3.3 Bayesian analysis

3.3.1 Parameter estimation

We now remind the reader of the basics of parameter estimation using Bayesian inference, as this is central to this paper's work. For in-depth descriptions of these methods, the interested reader can find information in technical textbooks, such as Gregory (2005). The same parameter estimation methods apply both to the single-pulsar and the common-signal analyses.

Bayes' theorem is the central equation of the analysis and is expressed as

$$Pr(\Theta|\mathcal{D}, H) = \frac{Pr(\mathcal{D}|\Theta, H)Pr(\Theta|H)}{Pr(\mathcal{D}|H)} \Leftrightarrow P(\Theta) = \frac{L(\Theta)p(\Theta)}{Z}. \tag{20}$$

In this equation, which we have written in the extended and a simplified version, $\mathcal{D}$ denotes our data, H denotes the hypothesis (i.e. the model), and Θ denotes the model parameters. $Pr(\Theta|\mathcal{D}, H)$, is the probability of the parameter, given the data and the model and is the posterior probability distribution. This is the distribution we are interested in estimating and we denote henceforth as $P(\Theta)$. $Pr(\mathcal{D}|\Theta, H)$ is the likelihood function, henceforth denoted as $L(\theta)$.

Bayesian inference inherently relies on the assumption of what our prior information (or belief) is for the probability distribution of the parameters. This information is encoded in $Pr(\Theta|H)$, the probability of the parameter given the hypothesis/model, which is known as the prior probability distribution (or simply the prior) and henceforth denoted as $p(\Theta)$. The inclusion of priors is integral in Bayesian inference and the choice of prior distributions has a central effect on the outcome of the inference, unless we are in the high signal-to-noise (S/N) regime, at which point the posterior distribution becomes insensitive to the choice of prior.

The final term in Bayes' theorem, $Pr(\mathcal{D}|H)$, is the probability of the data given the model. This is known as the evidence or marginal likelihood, henceforth denoted as Z. Note that this term is independent of the model parameters, Θ. Therefore, it is not necessary for parameter estimation, but it is used for comparing the relative probabilities between models (and will be implemented in work presented in Section 5.3).

In Bayesian inference we define the likelihood function and the prior distribution, and use the data to update the posterior distribution of the model parameters. For most practical problems, this requires Monte Carlo (MC) sampling algorithms. For the single-pulsar noise parameter estimation discussed in this paper, we only use pre-determined models as discussed in Section 3.1, therefore we are not concerned with accurate estimations of the evidence, Z. Finding a more optimal noise model for EPTA pulsar data will be the topic of a separate paper (Chalumeau et al., in preparation). Parameter estimation therefore is performed by sampling from the unnormalized posterior distribution and can be expressed as

$$P(\Theta) \propto L(\Theta)p(\Theta). \tag{21}$$

3.3.2 Model selection with Bayes factors

In Section 5.3, we will be comparing models of different CRSs in order to evaluate the relative ability of the models to sufficiently describe the data. This is described by the posterior odds ratio of the two analyses. We express this via the Bayes factor, $\mathcal{B}$ as

$$R = \frac{Z_1}{Z_0}\frac{p_1}{p_0} = \mathcal{B}_{10}\frac{p_1}{p_0}. \tag{22}$$

We use the numerical subscripts to denote the model, where 0 denotes the *null hypothesis*, which is the simpler of the two models in the case of nested models. We see that in contrast to the case of Bayesian parameter estimation, for model selection we need a reliable calculation of the Bayesian evidence, Z, which is the likelihood integrated over the prior distribution of dimensionality

n, i.e.

$$Z = \int L(\Theta) p(\Theta) \, \mathrm{d}^n \Theta. \tag{23}$$

The Bayes factor can be calculated directly from the evidences of the two models Z_1 and Z_0. The ratio of the samples in each model is equal to $\mathcal{B}$, under the assumption that the prior probabilities for the two compared models are equal. Alternatively, both models can be compared directly against each other in a hypermodel structure, which allows the sampler to decide which of them is more likely (Hee et al. 2016).

In this context, the value of $\mathcal{B}$ is used as a metric for the model selection problem. The difficulty lies in the numerical interpretation of the Bayes factor and empirical guidelines exist based on studies of multiple problems. In other words, the $\mathcal{B}$ values are not calibrated on a scale to provide a detection significance for this problem and can fluctuate between different analysis runs. In this paper, we will comment on our results according to the interpretation categories listed in tables from Kass & Raftery (1995), noting that in any case these singular $\mathcal{B}$ values only serve the purpose of discussing evidence in favour or not of one model with respect to another. As noted in Kass & Raftery (1995): 'these categories are not a calibration of the Bayes factor but rather a rough descriptive statement about standards of evidence in scientific investigation'.

4 SINGLE-PULSAR ANALYSIS AND RESULTS

In this section, we discuss the algorithms and settings used for the single-pulsar analysis and present the results.

4.1 Analysis algorithms

The two algorithms used for pulsar noise analysis are ENTERPRISE (Ellis et al. 2020) and TEMPONEST (Lentati et al. 2014a). Both packages use Bayesian inference but have been developed independently. They use different approaches to marginalize the timing-model parameters and additionally, each code uses different Monte Carlo samplers implementing different sampling methods.

The first analysis was performed using ENTERPRISE with PTM-CMCSAMPLER (Ellis & van Haasteren 2017), a parallel-tempering Markov chain Monte Carlo sampler, that has been designed for a high dimensional parameter space. To verify the analysis, we repeated the ENTERPRISE analysis with a different (nested) sampler, DYNESTY (Speagle 2020). Since the results are consistent, only those obtained with PTMCMCSAMPLER are reported. ENTERPRISE employs the 'Gaussian-process' approach to the marginalization of the timing parameters (equation 5).

The second analysis was performed using TEMPONEST, a Bayesian pulsar analysis package used as a TEMPO2 plugin which uses a nested-sampling approach (Skilling 2004). Depending on the problem's dimensionality we either use MULTINEST (Feroz, Hobson & Bridges 2009) or POLYCHORD (Handley, Hobson & Lasenby 2015) as the sampler, since the latter is more efficient only in cases with a large number of dimensions. Following suggestions discussed in Lentati et al. (2016) and also based on results from new tests, we used POLYCHORD in all cases where the number of parameters was greater than 55. We mark the cases where POLYCHORD was used in the tables reporting the results. TEMPONEST uses the 'G-matrix'approach for the marginalization of the timing parameters (equation 4).

Table 2. The prior ranges for the analyses using single power-law spectra models with parameters amplitude, A, and spectral index, γ. We use the subscripts RN, DM, and CRS to denote the red noise, DM noise and CRS for any type of common noise. We also include the pulsar white-noise parameters EFAC and EQUAD. Uniform and log-Uniform refer to flat priors and priors that are flat in the $\log_{10}$ space, respectively.

Parameter	Prior type	Range
A_{RN}, A_{DM}, A_{CRS}	log-Uniform	$[10^{-18}$–$10^{-10}]$
$\gamma_{RN}, \gamma_{DM}, \gamma_{CRS}, \delta_{CRS}$	Uniform	$[0$–$7]$
EFACs	Uniform	$[0.1$–$5]$
EQUADs	log-Uniform	$[10^{-9}$–$10^{-5}]$

4.2 Analysis settings and results

We first present some details regarding the noise analysis settings, which were set to be common between the two analyses.

(i) In principle, the optimal numbers of frequency components of $\mathbf{F}$ from equation (10) to describe the red noise and DM noise stochastic power-laws can be a free parameter to be estimated (see Chalumeau et al., in preparation). For this work we choose conservative numbers based on tests to ensure stable constraints on the power-law parameters. We use identical numbers for all six MSPs, the lowest 30 frequency bins for the red noise and 100 for the DM noise. Frequency binning is linear at frequencies N/T, with $N = 1, 2, 3...n$, such that n/T is the highest Fourier frequency of the TOA time series.

(ii) The prior distributions of power-law amplitudes and EQUADs are uniform in log-space, which we refer to as log-uniform priors. These types of priors are argued to function as good approximations of non-informative priors for scale-invariant parameters (Gregory 2005). The spectral indices and EFACs have uniform distributions in linear space, which we refer to as uniform distributions. Based on tests performed, we have decided on a given set of prior ranges for all parameters that we concluded to be adequate for this data set. The prior types and ranges are overviewed in Table 2.

We conducted the noise analysis with ENTERPRISE and TEMPONEST for the EPTA DR2 data set. The results are overviewed in Table 3. The two analysis codes have produced nearly identical results.[4]

5 GRAVITATIONAL-WAVE BACKGROUND SEARCH WITH EPTA DATA

In this section, we present the analysis and results for a search for a GWB in the EPTA data. We first conduct a search for a power-law spectrum CRS with simultaneous estimation of the ORF, as the most general method to look for the presence of a GWB in the data. Subsequently, we perform spectral parameter estimation for the physically motivated CRS cases discussed in Section 3.2, using their pre-defined ORFs, and employ Bayesian model selection to determine which of these models is better supported by the data.

---

[4]We note that the same analyses were performed on DR1, showing consistent results. This is important in both increasing our confidence in the estimated values of the noise parameters, as well as in the validity and quality of the TOAs.

Table 3. Results of single-pulsar noise analysis for ENTERPRISE (EP) and TEMPONEST (TN). The table shows the median values from the 1D marginalized posterior distribution. The uncertainties are calculated such that 95 per cent of the area under the 1D marginalized posterior distribution of the parameter is symmetrically distributed around the median. TEMPONEST analysis for the pulsars noted with * was performed using the POLYCHORD sampler.

| Pulsar | $\log_{10}A_{RN}$ | | γ_{RN} | | $\log_{10}A_{DM}$ | | γ_{DM} | |
	EP	TN	EP	TN	EP	TN	EP	TN
J0613−0200	$-14.93^{+1.11}_{-1.1}$	$-14.94^{+1.05}_{-1.0}$	$5.09^{+1.77}_{-2.08}$	$5.12^{+1.68}_{-1.93}$	$-12.0^{+0.33}_{-0.67}$	$-12.02^{+0.31}_{-0.49}$	$2.55^{+1.68}_{-1.07}$	$2.59^{+1.28}_{-0.95}$
J1012+5307*	$-13.13^{+0.17}_{-0.17}$	$-13.13^{+0.19}_{-0.2}$	$1.68^{+0.71}_{-0.7}$	$1.68^{+0.84}_{-0.81}$	$-11.72^{+0.1}_{-0.11}$	$-11.73^{+0.12}_{-0.12}$	$1.21^{+0.45}_{-0.39}$	$1.2^{+0.48}_{-0.42}$
J1600−3053	$-14.02^{+0.51}_{-1.08}$	$-14.01^{+0.52}_{-1.13}$	$3.46^{+2.45}_{-1.49}$	$3.45^{+2.4}_{-1.43}$	$-11.46^{+0.08}_{-0.07}$	$-11.46^{+0.07}_{-0.07}$	$2.14^{+0.26}_{-0.23}$	$2.15^{+0.24}_{-0.22}$
J1713+0747*	$-14.18^{+0.35}_{-0.44}$	$-14.17^{+0.41}_{-0.52}$	$3.37^{+1.14}_{-0.98}$	$3.35^{+1.38}_{-1.13}$	$-11.84^{+0.09}_{-0.09}$	$-11.84^{+0.1}_{-0.09}$	$1.44^{+0.43}_{-0.41}$	$1.44^{+0.51}_{-0.46}$
J1744−1134	$-15.23^{+1.28}_{-1.17}$	$-15.41^{+1.19}_{-0.9}$	$5.34^{+1.58}_{-2.14}$	$5.59^{+1.27}_{-1.9}$	$-11.69^{+0.1}_{-0.11}$	$-11.7^{+0.08}_{-0.11}$	$1.24^{+0.5}_{-0.5}$	$1.24^{+0.36}_{-0.35}$
J1909−3744	$-14.57^{+0.6}_{-0.77}$	$-14.57^{+0.63}_{-0.78}$	$4.47^{+1.98}_{-1.55}$	$4.46^{+1.99}_{-1.59}$	$-12.06^{+0.11}_{-0.15}$	$-12.06^{+0.11}_{-0.16}$	$2.03^{+0.58}_{-0.44}$	$2.03^{+0.63}_{-0.43}$

5.1 Algorithms and analysis settings

For the search of a GWB and other CRSs in the EPTA data, we once more use two independently developed algorithms for evaluating the likelihood and with independent Monte Carlo (MC) samplers in order to increase our confidence in the results by having a cross-check for possible bugs and analysis mistakes.

The first code employed is ENTERPRISE[5] with PTMCMCSAMPLER,[6] allowing us to perform a fully integrated pulsar noise and correlated signals analysis within the same analysis suite. We also use ENTERPRISE_EXTENSIONS[7] (Taylor et al. 2021) adding helpful functionality to ENTERPRISE. As in the case of the single-pulsar noise analysis, ENTERPRISE uses the 'Gaussian-process' approach to perform the marginalization of the timing model.

The second code used is FORTYTWO, which accepts pulsar noise analysis parameters from TEMPONEST results as input for further analysis. Individual modules of FORTYTWO have previously been used in various publications as they were being developed (Lentati et al. 2015; Caballero et al. 2016; Caballero et al. 2018; Guo et al. 2019; Hobbs et al. 2020) and a unified version is planned to be released in the future. For MC sampling, FORTYTWO uses either PYMULTINEST[8] or POLYCHORDLITE,[9] PYTHON implementations of MULTINEST and POLYCHORD, or its own Metropolis–Hastings sampler, depending on the analysis. To remain consistent with TEMPONEST FORTYTWO implements the 'G-matrix' timing parameter marginalization method. The only exception to this is the analysis in Section 6.5 where the timing model additionally includes SSE planetary parameters, in which case FORTYTWO also employs the 'Gaussian-process' approach enabling the algorithm to run significantly faster.

In all analyses for common signals in the six pulsar, their red-noise and DM-noise parameters are simultaneously sampled with the CRS. There is a significant probability for these pulsar parameters to correlate with the CRS, especially if the latter is not detected in the high S/N regime, either due to its weakness or lack of sufficient pulsar pairs to disentangle common from non-common signals. In the single-pulsar analysis, we model the white noise with two parameters per observing system. Keeping such a configuration in these CRS analyses would result in a currently unmanageable number

of parameters. One approach is to fix the EFAC and EQUAD for each observing system and use a 'global EFAC' parameter per pulsar that acts as a global multiplication factor to regulate each pulsar's white noise level. This has been shown to be a good strategy during the EPTA DR1 GWB analysis, as shown also in LTM15, where in all cases the global EFAC was found to be very consistent with unity. This means that the white-noise estimation during single-pulsar analysis is very robust. In the analyses presented in this work, we have verified the global EFAC values to be ∼1 once again, allowing us to fix the pulsar white-noise parameters from the single-pulsar analysis without significant loss of accuracy in our parameter estimations.

The general setup of the correlated search is as follows:

(i) For CRS models, we use the lowest 30 frequency bins to describe the single power-law spectrum. The CRS frequency bins are determined by a time grid set by the total timespan of the combined TOA data set. As we also sample the pulsar red and DM noise parameters simultaneously, we maintain the same frequency binning for each pulsar as in the single-pulsar analysis, i.e. the lowest 30 and 100 of each pulsar, for red and DM noise, respectively. As such, although all pulsar red noise components and the CRS component use 30 frequency bins, they all correspond to different frequencies.

(ii) The prior distribution of the CRS parameters are noted in Table 2.

(iii) The majority of the analyses use the DE438 SSE model, without any modification. Exceptions to this are discussed in Section 6.5.

5.2 Common-red-signal search and overlap-reduction-function estimation

We conduct a general search for a CRS including an estimation of the ORF using both the ENTERPRISE and FORTYTWO packages. We sampled the CRS spectral properties using the single power-law spectrum with amplitude and spectral index A_{crs} and γ_{crs}, respectively. The priors are found in Table 2. The spatial correlation curve of the CRS is modelled using a Chebyshev polynomial. We have introduced this method in LTM15, and we simply rewrite the formalism here for easier reading. We again use four Chebyshev coefficients (c_i, $i \in [1, 4]$), therefore the correlation curve is approximated by

$$\Gamma(x) \approx c_1 + c_2 x + c_3(2x^2 - 1) + c_4(4x^3 - 3x), \qquad (24)$$

where $x = (\zeta_{IJ} - \pi/2)/(\pi/2)$. Chebyshev polynomial priors are flat in the range [−1,1]. The analysis limits the resulting cross-correlations $\Gamma(x) \in [-1,1]$. In LTM15, we have showed that this approach approximates very well the direct individual measurements of pulsar-pair cross-correlations. To confidently transition from our DR1 results, for this analysis we use FORTYTWO with the Metropolis–

[5]https://github.com/nanograv/enterprise
[6]https://github.com/jellis18/PTMCMCSampler
[7]https://github.com/nanograv/enterprise_extensions
[8]https://johannesbuchner.github.io/PyMultiNest/
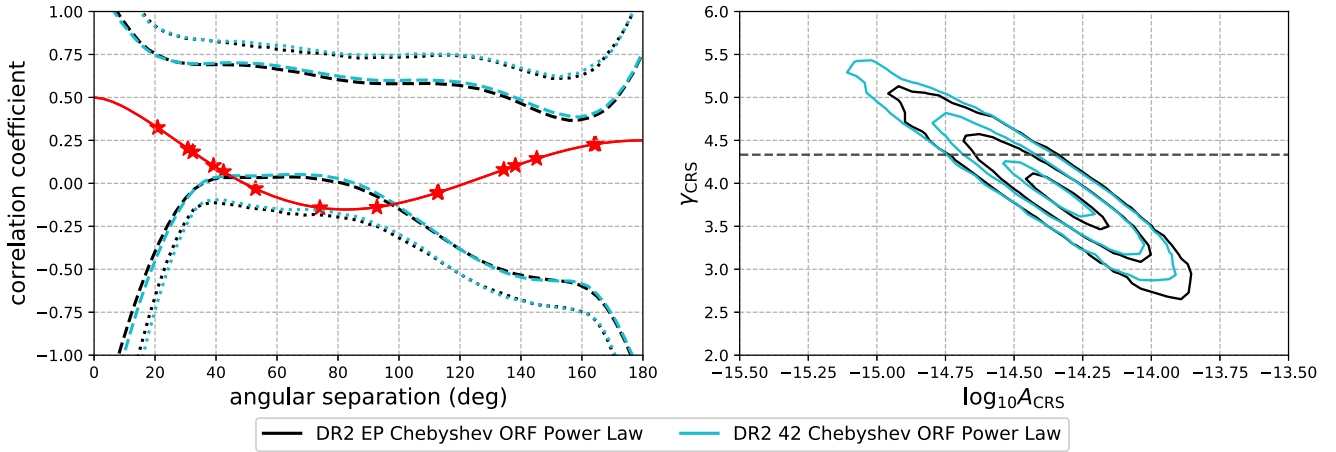[9]https://github.com/PolyChord/PolyChordLite

Figure 2. ENTERPRISE (EP) and FORTYTWO (42) results from a search for a CRS modelled with a single power-law spectrum, with simultaneous sampling of the ORF. The ORF is approximated with a 4th-order Chebyshev polynomial. Left: The posterior distribution for the angular correlation curve of the CRS, shown as boundaries of the credible regions. The dashed and dotted lines denote the 95 and 99.7 per cent credible regions, respectively. The theoretical HD curve is overplotted for comparison as a solid red line. The red stars denote the angular separations of the pulsars used in this study. Right: 2D posterior distribution of the spectral parameters for the single power-law CRS model with the dashed line indicating the expected $\gamma = 13/3$ from a GWB from SMBHBs.

Hastings sampler as it was done in LTM15 and compare the results with those of the ENTERPRISE analysis.

Fig. 2 shows the results of this general CRS search. The analyses clearly recover a common signal, with the two pipelines providing very consistent results. The left-hand panel shows the estimation of the ORF curve and the right-hand panel shows the posterior distributions for the spectral parameters. The 1D posterior distributions of the spectral parameters are $\log_{10} A_{CRS} = -14.32^{+0.31}_{-0.39}$ and $\gamma_{CRS} = 3.83^{+0.82}_{-0.72}$, where we denote the median and uncertainties at the 95 per cent credible region (see also Table 5). The ORF figure shows the 95 and 99.7 per cent credible regions. The results are compatible with the EPTA DR1 GWB analysis from LTM15. The boundary encompasses the probability for ORF to be the HD curve, however, other possibilities remain. We therefore examine in the next section the level of support the data provides to the physically motivated CRS signals discussed in Section 3.2.

5.3 Bayesian model selection for common red signals

The base model to which we make the model comparisons is one where the pulsar TOAs only have independent, uncorrelated individual pulsar noise (we denote this as PSRN), without any measurable commonality in the spectral properties of the different pulsars. We compare this base model to models that add only one CRS, namely either CURN, GWB, CLK, and EPH, as well as models which add two CRSs, i.e. a combination of CURN with one of the remaining three ORFs. The models are listed in Table 4. Given the uncertainty on the ORF this analysis cannot be expected to be fully conclusive, but can provide indications on whether some of these CRS models are more supported by the present data set.

We carried out calculations of Bayes factors with both the ENTERPRISE and the FORTYTWO packages. For ENTERPRISE the Bayes factors were obtained through a hypermodel structure comparing two models against each other. With FORTYTWO Bayes factors were calculated using the global logarithmic evidence for each analysis, which is calculated by PYMULTINEST using the 'Importance Nested Sampling' option. The two analyses give similar results.

Table 4 shows a summary of the Bayes factors for the different models. According to the criteria from Kass & Raftery (1995),

the addition of either of the CURN, GWB, or EPH signals to the base PSRN signal is decisively favoured with a $\log_{10}$ Bayes factor ($\log_{10}$BF) > 2. The strongest Bayes factor is for the CURN model, although the evidence for the GWB is only lower by $\log_{10}$BF ≈ 0.4. This difference provides only a marginal advantage to the CURN, barely disfavouring the GWB signal. The EPH model, however, is clearly less favoured with an $\log_{10}$BF difference to CURN or the GWB of order ~ 1, which is a substantial difference. We will examine the case of EPH in more detail in Section 6.5.2. In contrast to the three models discussed above, the monopolar correlation is only mildly favoured with respect to the PSRN base model.

Since the CURN model has the strongest evidence of the models with a single CRS, we can compare it against models which include another additional common process. The idea is to test whether there may be evidence for several physically motivated common processes coexisting in the data. In general, none of the three spatially correlated processes add substantial evidence to the single CURN. The ability to distinguish between different spatial correlations could be improved by using more than 6 pulsars in the analysis. We thus plan to expand the analysis to include a larger number of MSPs in the future.

The $\log_{10}$BFs values obtained with ENTERPRISE and FORTYTWO are very similar and show the same trends. The difference in the exact values are within the estimated uncertainties by MULTINEST, which are typically of the order ~ 0.2 at the 1σ level, suggesting consistency between the two results. MULTINEST calculates the uncertainty on the evidence using the relative entropy of the full sequence of samples (see Skilling 2004; Feroz & Hobson 2008), a computationally efficient method that does not require multiple runs to estimate the variance of the calculated evidence.

In the case of ENTERPRISE, the uncertainties can be estimated from the number of jumps between the different models (Cornish & Littenberg 2015) or by randomly selecting different sections of the MC chain to get a distribution of Bayes factors (Efron & Tibshirani 1994). From one hypermodel analysis run the calculated 1σ uncertainty is ~ 0.03 for both methods. However, the Bayes factor values are observed to fluctuate at levels about five times higher than the formal uncertainties when running the same analysis multiple times. Therefore, further investigations are warranted.

Table 4. Results from model selection analysis, in logarithmic (base 10) Bayes factors ($\log_{10}$BF), for different CRS models, and with fixed SSE (DE438). The model-components acronyms are: (i) PSRN = individual Pulsar noise only, (ii) CURN = common uncorrelated red noise, (iii) GWB = isotropic GWB with quadrupolar, HD, angular correlation, (iv) CLK = common signal with monopolar spatial correlation, as expected from a clock error, (v) EPH = common signal with dipolar spatial correlation, as expected from SSE errors. PSRN has no $\log_{10}$BF values as it serves as the base model. See Sections 3.2–3.2.3 for the discussion on these models.

| | | $\log_{10}$BF | |
ID	Model	ENTERPRISE	FORTYTWO
0	PSRN	–	–
1	PSRN + CURN	3.8	3.6
2	PSRN + GWB	3.4	3.2
3	PSRN + CLK	0.6	0.8
4	PSRN + EPH	2.1	2.1
5	PSRN + CURN + GWB	3.6	3.7
6	PSRN + CURN + CLK	3.7	3.4
7	PSRN + CURN + EPH	3.7	3.4

6 DETAILED ANALYSIS OF THE COMMON-UNCORRELATED RED NOISE

Since the CURN is the favourable model without sufficient evidence to justify the additional inclusion of another physically motivated CRS such as a GWB, we proceed with investigating the CURN more closely. In principle, the same tests can be performed in the case where we would be dealing with a GWB. In this section, we will specifically examine: (i) the frequency spectrum of the signal, (ii) the effect of the choice of frequency bins when modelling the signal with power-law spectra, (iii) the stationarity of the signal and consistency between DR1 and DR2 inferences, (iv) the consistency of individual pulsar noise with the common signal, and (v) the possible effects from SSE inaccuracies in the CRS analysis.

The power-law spectral parameters inferred from all the different CURN-related analyses can be found in Table 5, along with Jenson–Shannon divergence calculations that compare the results.

6.1 Individual-frequency modelling vs power-law spectrum

In order to further investigate our results, we use the case of the CURN to investigate the CRS spectrum modelling. We therefore proceed to also perform the analysis with an alternative approach to the power-law spectrum model, where the power of each individual CRS-spectrum frequency bin is sampled independently. This approach has been employed in LTM15 and Arzoumanian et al. (2020), and was first discussed in Lentati et al. (2013). We refer to this as the

'free spectrum' analysis. We conducted the analysis employing both ENTERPRISE and FORTYTWO, which provided fully consistent results. We note that for the FORTYTWO analyses, we used the PY-MULTINEST sampler when implementing the power-law spectrum model and POLYCHORDLITE for free-spectrum analyses due to the problem's high dimensionality. The full posterior distributions can be found in the Appendix A1. A comparison with the single-pulsar noise analysis from Table 3 shows the absorbtion of the pulsar red noises into the CURN, while the DM noises remain relatively consistent. Fig. 3 shows the main CURN results of these analyses. The left-hand panel of Fig. 3 shows the power of the CURN at each frequency, the free spectrum, with the straight lines indicating the median values of A and γ of the posterior distributions from the power-law spectral analyses with ENTERPRISE and FORTYTWO respectively. The full 2D posterior contours for the power-law parameters are shown on the right-hand panel. Having confirmed the agreement of the two algorithms, we will be using ENTERPRISE in the rest of the work in this section, except Section 6.5.

The free spectrum figure in general has two features. At high frequencies the power is white-noise dominated and can thus be modelled with a flat horizontal line. The presence of red noise becomes obvious at the lowest frequency bins and appears to be dominant for about 10 frequency bins.

The posterior distributions on the parameters of the power-law model of the CRS are found to be $\log_{10} A = -14.29^{+0.26}_{-0.33}$ and $\gamma = 3.78^{+0.69}_{-0.59}$ (95 per cent credible regions), as seen in the right-hand panel of Fig. 3 and Table 5.

6.2 Choice of the number of power-law frequency bins

The simple power-law model can be modified to include a smooth transition from red to white noise. This allows us to use the data to determine how many frequency bins are needed to optimally sample the low frequency common red noise and which higher frequencies are likely to be white noise dominated. We can replace equation (13) with a broken power law (Arzoumanian et al. 2020)

$$S_{\text{CRS}} = \frac{A_{\text{CRS}}^2}{12\pi^2} \left(\frac{f}{f_c} \right)^{\gamma_{\text{CRS}}} \left(1 + \left(\frac{f}{f_b} \right)^{1/\kappa} \right)^{\kappa(\gamma_{\text{CRS}} - \delta_{\text{CRS}})}, \quad (25)$$

where γ_{CRS} and δ_{CRS} are the spectral indices in the low and high frequency regimes, respectively, f_b is the bend frequency, where the common signal transitions from the red-noise to the white-noise dominated regime. The transition smoothness is determined by κ, which has been fixed at 0.1 for this analysis, but could be sampled over. In order to test which frequencies contribute to the red noise we

Table 5. 95 per cent constraints on the power-law (PL) parameters for the different analyses discussed in Sections 5 and 6 with the Jenson–Shannon divergence computed relative to the ENTERPRISE fixed SSE run.

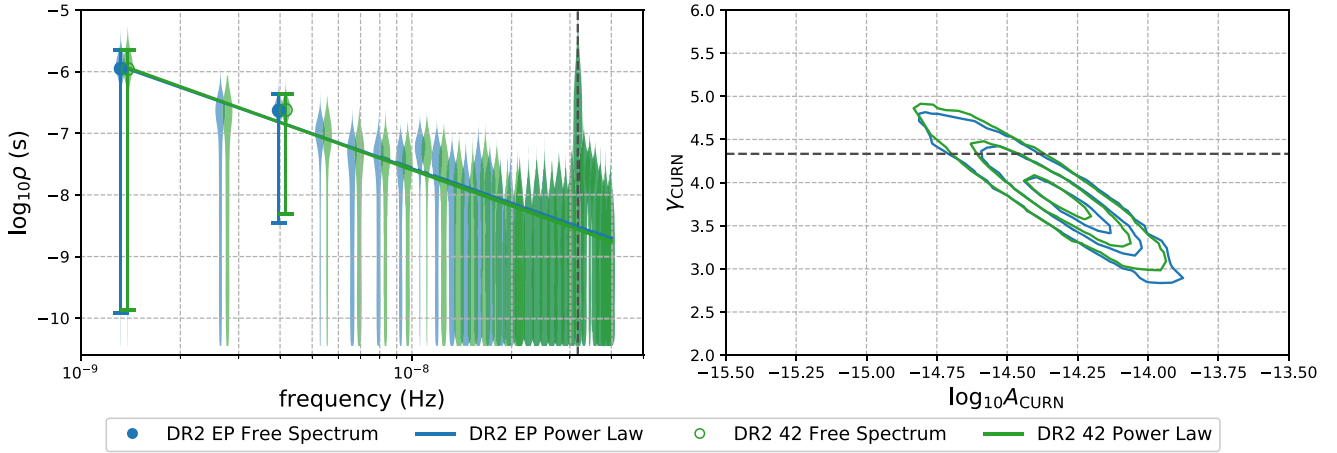| Algorithm + Model	$\log_{10} A_{\text{CRS}}$	J–S div.	γ_{CRS}	J–S div.
ENTERPRISE + DE438 PL	$-14.29^{+0.26}_{-0.33}$	0	$3.78^{+0.69}_{-0.59}$	0
FORTYTWO + DE438 PL	$-14.33^{+0.27}_{-0.31}$	0.00904	$3.87^{+0.67}_{-0.60}$	0.009 42
ENTERPRISE + DE438 CRS PL	$-14.32^{+0.31}_{-0.39}$	0.00833	$3.83^{+0.82}_{-0.72}$	0.008 06
FORTYTWO + DE438 CRS PL	$-14.39^{+0.33}_{-0.43}$	0.04521	$3.97^{+0.89}_{-0.74}$	0.039 29
ENTERPRISE + BAYESEPHEM PL	$-14.32^{+0.30}_{-0.37}$	0.00604	$3.70^{+0.78}_{-0.80}$	0.015 86
ENTERPRISE + EPHEMGP PL	$-14.42^{+0.32}_{-0.41}$	0.07145	$3.91^{+0.85}_{-0.83}$	0.024 82
FORTYTWO + LINIMOSS PL	$-14.41^{+0.35}_{-0.54}$	0.06118	$3.91^{+1.06}_{-0.87}$	0.036 56
ENTERPRISE + DE438 broken PL	$-14.24^{+0.31}_{-0.37}$	0.02293	$3.67^{+0.76}_{-0.71}$	0.020 97

Figure 3. Results from CURN analysis using ENTERPRISE (EP) and FORTYTWO (42), both for the free-spectrum (left-hand panel) and the single power-law (right-hand panel) analyses. The left plot shows the posterior distribution for the amplitude of the power at each frequency bin using violin plots. Where the inference provides good measurement of the power, we denote the median with a circle and the 95 per cent uncertainties. We consider the measurement good, if more than 95 per cent of the posterior probability lies above the lowest 6.25 per cent of the prior. As the results of the two algorithms are almost identical, we slightly shift the 42 distributions of the lowest 15 frequency bins for easier visual comparison. The right plot shows the 2D posterior distribution for the CURN power-law amplitude and spectral index with the dashed line indicating the expected $\gamma = 13/3$ from a GWB from SMBHBs. The two analysis pipelines have produced consistent results.

need to probe the high frequency regime beyond the 30th frequency bin. Thus, a log-uniform prior for $\log_{10} f_b \in [-9, -6]$ is set.

We find that the most likely bend frequency is around the 20th frequency bin $20/T \approx 2.5 \times 10^{-8}$ Hz, below which most of the power is concentrated. To be more conservative we opted to use the conventional 30 frequencies for single power-law models, knowing that we are accounting for the majority of significant frequency bins. Additionally, we have verified that the power-law spectrum parameters are consistent between the single power-law and the low-frequency end of the broken power-law model, as can be seen in Fig. 4.

6.3 Consistency between individual pulsars and the common signal

In order to find out how much each pulsar's red noise is consistent with the CURN or in other terms how much a given pulsar contributes to the CURN we employ the 'dropout' method (e.g. Aggarwal et al. 2019; Arzoumanian et al. 2020). The dropout factor is defined as

$$\text{dropout}_k = \frac{p_k(\text{CURN})}{p_k(\text{no CURN})} \times$$
$$\int \frac{p(\theta_{\text{CURN}}|d_{\text{no }k}) p(\theta_{\text{CURN}}|d_k)}{p(\theta_{\text{CURN}})} d\theta_{\text{CURN}}, \quad (26)$$

where $\theta_{\text{CURN}} = (A_{\text{CURN}}, \gamma_{\text{CURN}})$ is a vector denoting the amplitude and spectral index of the CURN, $p_k(\text{CURN})$, and $p_k(\text{no CURN})$ are the probabilities that pulsar k supports or rejects the CURN. The terms $p(\theta_{\text{CURN}}|d_{\text{no }k})$, $p(\theta_{\text{CURN}}|d_k)$, and $p(\theta_{\text{CURN}})$ are the probabilities of a CURN with certain spectral properties without pulsar k, from pulsar k alone and the overall probability, respectively. The spectral index γ_{CURN} can be fixed to 13/3 for simplicity.[10] The integral then becomes a function of only the amplitude A_{CURN}. In summary, the dropout factor is a measure of how much a given pulsar k supports the existence of the common signal. It can also be seen as a consistency

factor comparing the intrinsic red noise of pulsar k against the CURN constrained by the other pulsars. We can use the Bayesian framework from Section 5.3 to estimate the dropout factors by computing how much each pulsar favours the CURN in a model comparison with the full data set.

The values of the dropout factor for each pulsar for a CURN with freely varying parameters (circles) and fixed $\gamma_{\text{CURN}} = 13/3$ (squares) can be seen in Fig. 5. A dropout factor of about 1 (or 0 in the logarithmic scale, as in Fig. 5) indicates that the pulsar is indifferent to the CURN. A large dropout factor indicates strong commonality between the pulsar and the CURN. Five out of the six pulsars are in 'support' of the CURN in DR2. PSR J1012+5307 is noticeably not and will be further discussed in Section 7.

Differences between the dropout factors for the same pulsar and data set depending on whether γ_{CURN} is fixed or not can be seen in Fig. 5. As the dropout factor can be viewed as a consistency factor, we can compare the intrinsic red noise of a pulsar against the overall constraints on the CURN. For example from Table 3 we find the median red noise $\gamma_{\text{RN}} \sim 3.4$ for PSR J1713+0747, while the overall CURN has $\gamma_{\text{CURN}} \sim 3.8$ (see Fig. 3). One can expect the single pulsar red noise of PSR J1713+0747 to be slightly more consistent with the varied CURN posterior than a distribution fixed at 13/3, thus giving a slightly larger dropout factor when varying γ_{CURN}. These differences become more pronounced using DR1, as the constraints on the CURN are tighter in DR2, such that the slice at $\gamma_{\text{CURN}} = 13/3$ is more representative of the recovered 2D CURN posterior with DR2 in contrast to DR1.

6.4 Consistency with DR1

In this work we have added a substantial amount of more precise data to the DR1 data. Therefore, we investigate whether the CURN properties are consistent between DR1 and this new, extended data set. If the CURN is stationary, the analysis of the two data sets ought to produce consistent results, where we should get better constraints with the added data. This is indeed an important test in the framework of searching for a stochastic GWB, as the signal is theoretically

---

[10]This also allows for an easier decoupling between the pulsar intrinsic red noise and the CURN.
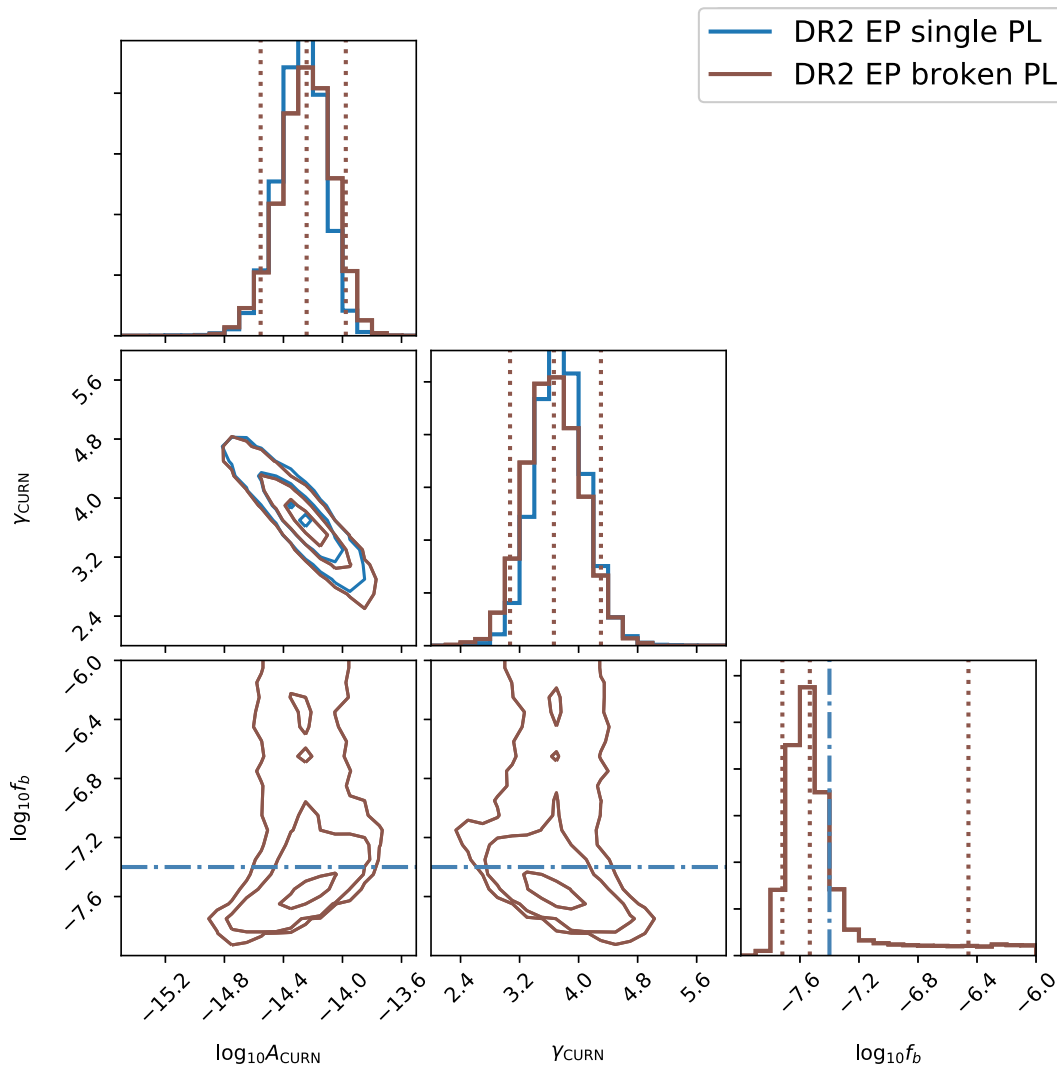
Figure 4. 2D posterior distributions comparing the `ENTERPRISE` CURN search with the single and the broken power-law model with the additional 'bend frequency' (f_b) parameter. The blue dash–dotted line indicates the 30th frequency bin, the highest sampled frequency of the single power-law model. The dotted lines show the median and central 90 per cent credible regions of the 1D marginalized parameter posteriors. The broken power-law analysis suggests a f_b that corresponds to 20 frequency bins. The results of the two analyses are completely compatible, suggesting that the single power-law model with 30 frequency bins describes the data reasonably well.

expected to be stationary. As DR1 is a very well studied data set, it is straightforward to confidently make this investigation. We repeated the single-pulsar analysis for DR1 as with DR2, using the same SSE (DE438) and terrestrial time-standard (BIPM2019) in order to have an appropriate comparison. We performed the DR1 CURN power-law analysis using 22 frequency bins, as this was found to be adequate in LTM15. Again, we used both `ENTERPRISE` and `TEMPONEST` for the single-pulsar noise analysis and cross-checked the CURN analysis with `ENTERPRISE` and `FORTYTWO`. As the result with both codes are compatible, we use the `ENTERPRISE` results here to make the comparisons of the DR2 and the DR1 subset.

Fig. 6 shows how the common signal has evolved from EPTA DR1 to DR2, using the posterior distributions of the single power-laws and free-spectra parameters. One can see that DR2 provides a much more constrained probability distribution of the power-law parameters. While the DR1 data set shows a CURN centred around $\gamma = 2.83^{+2.14}_{-1.96}$ and $\log_{10} A = -13.96^{+0.34}_{-1.41}$ (95 per cent credible region), there is considerable uncertainty in the parameter space beyond the 95 per cent credible region. The additional data from the

DR2 data set constrain the spectral index closer to the expected value of $\gamma = 13/3$ from a GWB by SMBHBs. The amplitude has decreased, also more in line with more probable theoretical expectations (e.g. Chen, Sesana & Conselice 2019; Middleton et al. 2021). The DR2 free spectrum on the left of Fig. 6 also seems to be extending the DR1 free spectrum. In DR1, about four of the lowest frequencies support the existence of a CURN. The median DR2 power law also passes through the DR1 free spectrum power distributions.

While the timespan extension has contributed to the improvement of the CURN analysis, we note that this also appears to be to a large degree the result of the much better multifrequency coverage of the newly added data. This resulted in very significantly improved constraints of the pulsars' DM parameter spaces and decorrelation of said DM parameters from the pulsar red noise parameters. This is in contrast to DR1, where the DM and red noise parameters were significantly correlated for multiple pulsars, adding uncertainty to the pulsar red noise parameters that would subsequently result in similar uncertainties of common red signals. We can see how much pulsars have improved in their ability to contribute to the
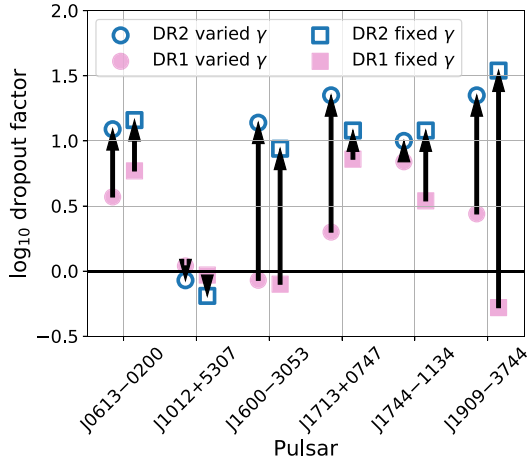
Figure 5. Dropout factors for both DR1 and DR2 with varied and fixed $\gamma_{\mathrm{CURN}} = 13/3$ spectral index shown with circles and squares, respectively. The number of pulsars contributing the CURN detection has increased from three to five with the DR2 data extension. Only PSR J1012+5307 seems indifferent to the CURN (see Section 7 for a discussion).

recovered CURN, by examining the changes in the dropout factors for each pulsar, as presented in Fig. 5. PSR J1909−3744 is the most prominent example of the achieved improvement, as it has moved from having the smallest contribution to the largest. This pulsar has the highest TOA precision, however in DR1 it only had a time-span of 9.38 yr (in contrast to 15.7 yr in DR2) and had highly correlated red and DM noise parameters. The decorrelation of red and DM noise components is achieved thanks to the wide bandwidth of NUPPI (as mentioned in Section 2, for this MSP we only use NRT data). Four other MSPs have increased their dropout factors, supporting the stationarity assumption of the CURN.

We finally examine if the extension of the data set from DR1 to DR2 creates any unexpected differences in the Bayes factors between the different models examined in Section 3.3.2. For the CURN case, and using the DE438 SSE, the $\log_{10}$BF has increased from ≈ 1.2 to ≈ 3.7, further supporting the stationarity assumption, and strongly suggesting that the signal, irrespective of its origin and interpretation, is not a statistical fluctuation. We finally note that despite increased

Bayes factors for the different CRS signals in DR2 by comparison to DR1, the difference in the evidence between CURN and the GWB, has not drastically change from DR1 (see LTM15), thus still not allowing to support the finding of a GWB or other spatially correlated signal. This is most likely due to only using six pulsars in both cases, which does not offer the necessary sampling of the angular separations. We also note that the clock-error signal remains the least favourable physically motivated CRS. This is expected from the posterior distribution of the ORF in Fig. 2, which is consistently away from 1 across the pulsar angular separations axis. The full comparison of Bayes factors between DR1 and DR2 can be found in Table A2.

6.5 Addressing possible common red signals from Solar system ephemeris systematics

Previous studies (e.g. Tiburzi et al. 2016; Guo et al. 2019; Vallisneri et al. 2020) have shown that the SSE modelling plays an important role in the search for common signals with PTA data. We therefore investigate the degree by which SSE inaccuracies affect the CURN parameter estimation, and whether modelling possible SSE-induced signals affects the CRS model selection results. In this study we apply three independently developed algorithms that introduce modelling of the SSE uncertainties into the CRS search. This lays the ground-work for a robust and cross-checked mitigation of the SSE effects in future GWB searches. All three algorithms assume that the SSE parameters are close to the correct ones and as such investigate linear deviations from their values. The algorithms differ in the method used to derive the induced TOA delays by SSE parameter inaccuracies and the SSE used as reference.

The first method applies the BAYESEPHEM model (Vallisneri et al. 2020), which has previously been used in studies estimating upper limits for the GWB and examining common signals (Arzoumanian et al. 2018, 2020). This algorithm is based on a physical model that accounts for induced TOA delays due to linear deviations in planetary masses, rotation rate about the ecliptic pole, and planetary average orbital elements, resulting in a quasi-Keplerian model for the orbit. Allowing these parameters to vary with reference to the SSE DE436 (Folkner & Park 2018) create what we refer to as variational partials. The linear combination of partials that minimize
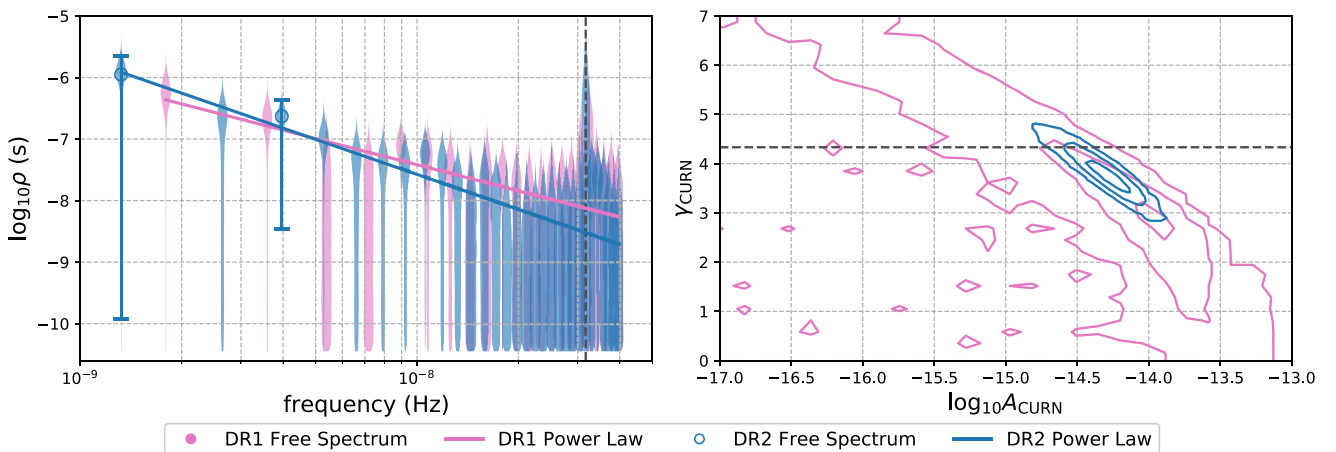

Figure 6. Comparison of the CURN recovered signals with DR2 and DR1 in the same style as Fig. 3, both using free-spectrum and power-law analyses with ENTERPRISE. The CURN signal properties are in agreement with the expected detection evolution of a stationary red signal when extending the timespan. The improvement is largely due to the significant increase in data quality with the DR2 extension.
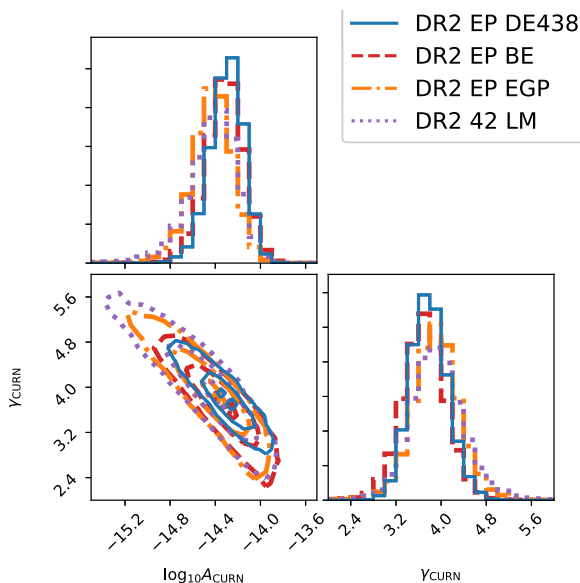
Figure 7. Posterior distributions of CURN power-law parameters using ENTERPRISE with (i) fixed DE438 SSE, (ii) BAYESEPHEM and (iii) EPHEMGP and FORTYTWO with LINIMOSS. BAYESEPHEM and EPHEMGP sample the mass and orbital parameters of Jupiter and Saturn using partials derived from DE436 and INPOP19a. The LINIMOSS analysis analytically marginalizes these parameters together with the pulsar timing models in a fully dynamical analysis with the PMOE SSE.

the differences in the orbit with respect to DE436 (differences well below detectability by TOA precision of real data) define the values for the planetary parameters that are used as the BAYESEPHEM initial values. In this work, BAYESEPHEM includes terms accounting for the masses of Jupiter, Saturn, Uranus, and Neptune, the rotation rate around the ecliptic pole and orbital elements for Jupiter as well as for Saturn, since, the EPTA DR2 is approaching 25 yr of timespan. Each of these terms is linearly perturbed around the initial values adding a delay to the TOAs. The overall linear delay (calculated by projecting the partials on the TOAs) is treated as a deterministic signal. The SSE model parameters are MC sampled together with the CRS and pulsar-noise parameters in the Bayesian framework.

In addition to BAYESEPHEM, in this work we also use two other algorithms to control for the effects of SSE errors, namely EPHEMGP and LINIMOSS. For both, this is the first time they have been used in the context of any CRS analysis and this work provides a first comparison of their performances against BAYESEPHEM with real data. While EPHEMGP is similar to BAYESEPHEM, it implements a different model from independent SSE information. LINIMOSS, however, is different from both models, as it has access to the full equations and input parameters of a published dynamical SSE model.

EPHEMGP (Chalumeau et al., in preparation), describes SSE uncertainties as a Gaussian process. The SSE design matrix is based on the partials derived using the INPOP19a SSE fit (Fienga et al. 2019), which are mapped onto the TOAs via the induced delays. In this work, we use EPHEMGP paired with the ENTERPRISE package and fit for the parameters of Jupiter and Saturn orbital elements in the Bayesian search of the CURN. The modified residuals correspond to the deviations in the initial orbital elements of Jupiter and Saturn and mimic the time-varying uncertainties in the position of the Solar system barycentre. Similarly to BAYESEPHEM, we have

sampled EPHEMGP parameters together with CURN and pulsar noise parameters. The addition of this analysis allowed for a cross-check of the BAYESEPHEM results, within the same analysis framework of likelihood estimation and MC sampler.

LINIMOSS (Guo et al. 2019) is integrated with FORTYTWO, and is a fully dynamical model of the major Solar system bodies based on the PMOE SSE (Li & Ni 2003; Li et al. 2008b). This ephemeris was used to optimize the orbit (Li et al. 2008a) of the space-based GW observatory 'Laser Interferometer Space Antenna' (LISA; Amaro-Seoane et al. 2017). LINIMOSS is built by making modifications in PMOE to match the initial conditions used in the DE435 SSE. In Guo et al. (2019) it has been demonstrated that LINIMOSS/PMOE is compatible with the DE435 SSE, with differences well below our data precision, and can therefore be confidently used for pulsar timing. The design matrix for the planetary parameters is directly derived by first linearly perturbing the planetary masses and orbital elements and numerically re-integrating the SSE; TOAs predicted using the original SSE are then fitted with the modified SSE, and these SSE-induced TOA delays are added as deterministic signals in the pulsar design matrices. In this study we use LINIMOSS to analytically marginalize the SSE mass and orbital parameters for Jupiter and Saturn together with the rest of the timing models during the search for common signals.

6.5.1 Effects on the parameter estimation of the common signal

We first focus on the CURN parameter estimation and produce their posterior distributions with the three different algorithms. The analyses with BAYESEPHEM and EPHEMGP perform MC sampling on the SSE parameters and priors are therefore carefully defined. Both methods determine the prior range phenomenologically by allowing the parameters to vary enough to cover differences between various SSE models, as well as keeping the resulting residuals of the pulsar TOAs below a certain threshold to stay within the linear regime. For BAYESEPHEM the delays are limited to about μs level, while EPHEMGP allows for delays from SSE systematics up to about 100 μs. Both use uniform priors for the orbital elements, BAYESEPHEM uses Gaussian priors for the planetary masses, while they are held fixed in this analysis with EPHEMGP, after confirming that no mass-error signals could be detected by the pulsar data.

LINIMOSS analysis performs analytical marginalization of planetary masses and orbital elements together with the pulsar timing model using uniform infinite priors, making this the analysis with the wider priors. We note that this analytical marginalization of the SSE parameters used here is not the only way it is possible to use LINIMOSS, as in principle we may marginalize with specified prior types and ranges, or also MC sample these parameters together with the CRS parameters during the Bayesian inference process. In this study, we use the full analytical marginalization as a complementary analysis to the MC sampling used by BAYESEPHEM and EPHEMGP, and will present the details of general LINIMOSS use in GWB searches separately (Guo et al., in preparation). This has certain limitations in the model selection process, but also serves as a useful check to see whether our analyses produce the expected results, as we discuss below in Sections 6.5.1 and 6.5.2.

Fig. 7 shows the comparison of the ENTERPRISE results without any SSE fitting, and with the use of BAYESEPHEM, EPHEMGP, and LINIMOSS (while the LINIMOSS analysis uses FORTYTWO, note that the fixed DE438 distributions from both ENTERPRISE and FORTYTWO are nearly identical, see Fig. 3, and therefore it is

Table 6. Results from model selection analysis, in logarithmic (base 10) Bayes factors ($\log_{10}$BF), for different CRS models, using ENTERPRISE and fitting SSE parameters with BAYESEPHEM. The model IDs and acronyms are the same as in Table 4. PSRN has no $\log_{10}$BF values as it serves as the base model.

ID	Model	ENTERPRISE+BAYESEPHEM
0	PSRN	–
1	PSRN + CURN	2.9
2	PSRN + GWB	2.7
3	PSRN + CLK	2.3
4	PSRN + EPH	1.0

Table 7. Results from model selection analysis, in logarithmic (base 10) Bayes factors ($\log_{10}$BF), for different CRS models, using FORTYTWO and fitting SSE parameters with LINIMOSS. The model IDs and acronyms conventions are the same as in Table 4. PSRN has no $\log_{10}$BF values as it serves as the base model. Note that these results are not directly comparable to those in Table 6 as explained in the main text.

ID	Model	FORTYTWO+LINIMOSS
0	PSRN	–
1	PSRN + CURN	1.3
2	PSRN + GWB	1.3
3	PSRN + CLK	1.2
4	PSRN + EPH	0.0

sufficient to only show one fixed DE438 result). We can see that all three methods show consistent posterior distributions. The inclusion of the SSE models slightly increases the uncertainties in the recovered parameters, while still keeping them highly confined. As expected, the contours become progressively broader than the DE438 contour, as the allowed prior increases from BAYESEPHEM to EPHEMGP, to the full marginalization in LINIMOSS.

6.5.2 Effects on the model selection for common red signals

The next step is to examine how including the SSE parameters in the analysis affects the results for the model selection, as discussed in Section 3.3.2. We first examine this in the framework of ENTERPRISE+BAYESEPHEM. The BAYESEPHEM model acts as an additional common signal to all pulsars in the array with a dipolar nature. As such, it is possible to do a Bayesian model comparison between a CURN analysis (or indeed any of the models listed in Table 4) with a fixed SSE and one using BAYESEPHEM. We find only a small $\log_{10}$BF ~ 0.4 in favour of the addition of BAYESEPHEM to the PSRN model, while for the CURN model the addition of BAYESEPHEM is disfavoured by approximately the same number. Although the addition of BAYESEPHEM models the data better than pulsar noise can alone, a simpler CURN provides an equivalently good fit to the TOAs and is therefore the more preferred model. The small $\log_{10}$BFs indicate that the TOAs are not strongly dependent on any (possible) SSE-parameter inaccuracies.

The $\log_{10}$BFs when including BAYESEPHEM in the analysis for the same models as in Section 3.3.2 can be found Table 6. As we have seen that models with two CRS components in addition to PSRN are not an improvement to models with only one CRS added, we only focus on the latter models. The evidence for CURN, GWB remain very significant even if we add uncertainty modelling into the SSE. However, the $\log_{10}$BF for the EPH CRS is significantly lower. This is an expected result as BAYESEPHEM is designed to take care of the possible systematics in the SSE encoded in the Solar system barycentre. The drop in significance of the (PSRN+EPH) model corresponds to the absorption of dipolar correlations by BAYESEPHEM. At the same time, the CLK model now has a much stronger support. This suggests that after modelling the SSE parameters the ORF may become flatter and more similar to a monopolar signal, and that indeed the use of only six pulsars makes it difficult to distinguish between different angular correlation shapes.

In addition to the model selection process using BAYESEPHEM, we also perform an analogous analysis using LINIMOSS. It differs from the BAYESEPHEM analysis in that we cannot discern if a model selection would prefer a fixed SSE or the inclusion of SSE fitting with LINIMOSS. This is because the analytical marginalization of

the SSE parameters implies the use of improper (un-normalizable) priors that makes the model with and without the SSE fit not directly comparable. Nevertheless, we may still perform Bayesian model selection, if all models in question are affected by the marginalization over the improper prior. We therefore additionally compare the same models reported in the Table 6, but now using LINIMOSS and analytical marginalization (infinite priors) over the planetary parameters. Table 7 overviews these results. There are two important observations one can immediately make. First, the model (PSRN+EPH) is significantly disfavoured by comparison to other CRS models as LINIMOSS fully absorbs any dipolar correlations related to SSE signal in the data with the analysical marginalization. This is an expected and highly desired result. The second observation is that the Bayes factors for the CURN and GWB are reduced more than in the case of BAYESEPHEM, and in addition, the evidences for the CRS to be a CURN, a GWB, or a CLK signal are equalized, following a similar trend as with BAYESEPHEM, but more prominently. These differences from the BAYESEPHEM results are most likely because the full marginalization of the SSE parameters, while only using six pulsars in the analysis, can lead to indiscriminate absorbtion of other types of CRS, correlated or not. Tests using simulated data have confirmed this scenario. The equal evidences in CURN, a GWB, or a CLK signals after fitting the SSE with LINIMOSS also supports that the data with six pulsars cannot distinguish the nature of any remaining CRS after fitting for the SSE. A more detailed analysis using LINIMOSS with more precise techniques in order to fully interpret the results, will be published separately (Guo et al., in preparation).

We also plan to compute Bayes factors and perform similar investigations with EPHEMGP in a future publication (see Chalumeau et al., in preparation). As we have access to the full design matrix of INPOP19a, more planet masses and orbits can be added to the EPHEMGP model used in this work. The effects of a properly chosen prior range will also be tested.

The general conclusion is that overall the SSE analytical marginalization is very good at absorbtion of dipolar signals but is less safe with respect to leaving a true GWB signal in the data unabsorbed; this is in agreement with findings by Tiburzi et al. (2016). As such, this type of analysis is a good basis for a very conservative GWB search or strain upper limit. On the other hand, the approach of sampling the SSE parameters requires careful prior choices as it may leave some dipolar signal unmodelled, potentially affecting the measurement of parameters and detection significance of a true GWB signal. These issues can be reduced significantly with more pulsars to better sample the angular separations, which highly motivates the work for the preparation of the upcoming full EPTA DR2 data set.
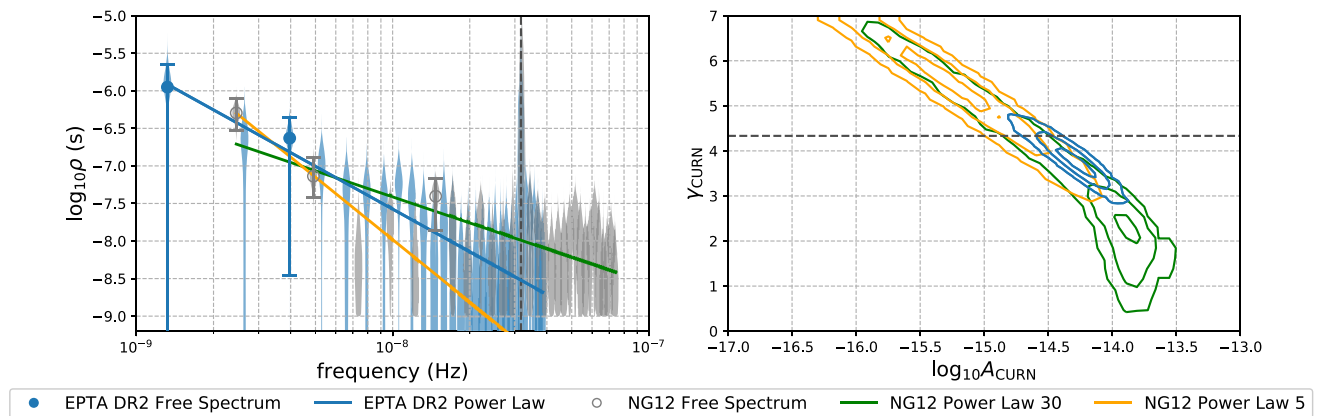
Figure 8. CURN search free spectrum and power-law recovery comparison between EPTA DR2 and NG12 with DE438 (Arzoumanian et al. 2020) in the same style as Fig. 3.
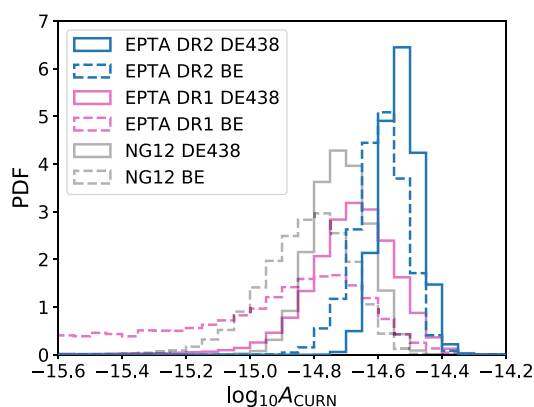

Figure 9. Comparison of the CURN recovered signals with EPTA DR2, DR1 and NG12 (Arzoumanian et al. 2020), using fixed DE438 and `BAYESEPHEM` with `ENTERPRISE` and fixed $\gamma_{\mathrm{CURN}} = 13/3$.

7 DISCUSSION OF RESULTS AND COMPARISON WITH LITERATURE

A CURN process has been reported by independent data sets and their analysis by other PTA collaborations (Arzoumanian et al. 2020; Goncharov et al. 2021a). We will briefly compare the results from this work against the constraints on the power-law parameters of the CURN and the model selection Bayes factors from Arzoumanian et al. (2020). A more detailed comparison between the inferences from the different regional PTA groups will be presented in a relevant upcoming IPTA paper (Antoniadis et al., in preparation).

Fig. 8 compares the fixed DE438 CURN analysis from the EPTA DR2 against the NANOGrav 12.5 yr data set. The left-hand panel shows a broad agreement in the power distributions by frequency. As the EPTA DR2 is almost twice as long as the NANOGrav data, it provides a tighter constraint on the CURN power-law spectral slope, and consequently amplitude. In addition, this result is largely insensitive to the choice of the number of frequency bins used in the analysis, see Section 6.2 and Fig 4, in contrast to NANOGrav's recovered spectral properties.

We can also compare the amplitude of a CURN at fixed $\gamma = 13/3$, which corresponds to a GWB from massive black hole binaries. The EPTA DR2 results give an amplitude of $A_{\mathrm{CURN}} = 2.95^{+0.89}_{-0.72} \times 10^{-15}$ (95 per cent credible region). Fig. 9 shows a comprehensive compar-

ison of the relevant CURN amplitude posterior distributions when using the DE438 SSE and when fitting SSE planetary parameters with `BAYESEPHEM`, for EPTA DR2 and DR1 and the NANOGrav 12.5-yr results from Arzoumanian et al. (2020). We note that the EPTA results are obtained with an upper prior bound of $\log_{10}A = -10$ and simultaneously fit for the pulsar DM stochastic noise, whereas NANOGrav uses a $\log_{10}A = -14$ upper prior bound and observationally measured piecewise DMX values to model the pulsar DM variation over time (e.g. Arzoumanian et al. 2015; Jones et al. 2017). As such, this comparison can only be indicative.

In this paper, we have compared power-law models for common red signals (CRSs) with different overlap reduction functions (ORFs) while retaining the spectral index as a free parameter. We obtain a $\log_{10}$BFs of ≈3.7 for PSRN+CURN versus PSRN (DE438). In order to make a more direct comparison with the NANOGrav model selection results, we have repeated our analysis by fixing the spectral index $\gamma = 13/3$ and matching the NANOGrav prior on the common signal amplitude. In this case, we find $\log_{10}$ BF $\gtrsim 4$ compared to the NANOGrav result of 4.5 (DE438) with an estimated uncertainty of 0.9. We can therefore conclude that the EPTA DR2 and NANOGrav 12.5-yr results are in general agreement.

Our EPTA DR2 model comparison results appear to be less affected by SSE uncertainties than the NANOGrav 12.5-yr results. A possible reason is that the NANOGrav data set has a maximum timespan of roughly half a year longer than the Jovian orbital period, while the EPTA data covers close to two Jovian orbits and is a few years short of Saturn's orbit. As such, the EPTA data may be more effective in distinguishing signals induced by errors in Jupiter's orbit from pulsar noise. Another possibility, however, is that the present EPTA data set does not efficiently recover the dipolar correlations with `BAYESEPHEM` due to only using six pulsars, with sparse coverage of the pulsars angular separations space. The addition of more pulsars in the EPTA array will provide more information on this.

The commonality of the single pulsar red noises, which could be interpreted as a CRS, has been investigated by the PPTA (Goncharov et al. 2021a). Though not unexpected, the PPTA has used simulated data to unambiguously demonstrate that individual pulsar noises can be recovered by the analysis code as a CURN, if the spectral properties are similar. This is something that must be considered carefully, as MSPs possibly may have common underlying mechanisms that produce intrinsic stochastic noise (see e.g Jones 1990; Shannon & Cordes 2010; Melatos & Link 2014). In Fig. 10, we
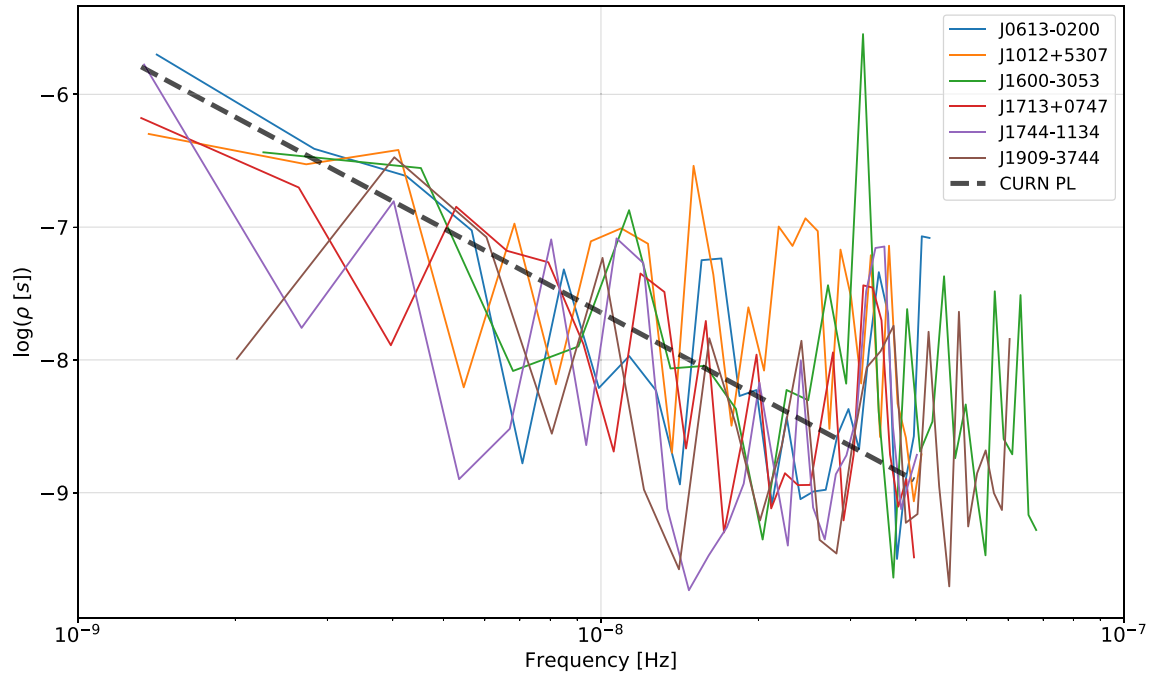
Figure 10. Single pulsar noise spectra (Maximum aposteriori values from `ENTERPRISE` noise analysis) for each of the six pulsars, compared to the CURN power law (median values) shown with the dashed line.

show the power spectra for each of the 6 EPTA pulsars using the maximum a posteriori values of the SPNA runs from `ENTERPRISE`. One can see a broad agreement of all pulsars with the CURN, thus strengthening this CURN as a common noise floor. However, PSR J1909−3744's red noise power is poorly constrained at the lowest frequency, consequently, plotting a point estimate can give the impression of a dip below the CURN. Looking at PSR J1012+5307, the red noise is clearly present in this pulsar and its shape is consistent with other pulsars. However, the level of white noise is higher and, as a result, its slope appears to be lower and thus inconsistent with the CURN. This is a possible cause of results of dropout analysis observed for this pulsar in Fig. 5.

The noise properties of the six pulsars used in this study are analysed in greater detail and the results will be published separately in an upcoming paper (Chalumeau et al., in preparation). In a similar fashion that Lentati et al. (2016) have examined the noise properties of the IPTA DR1 and more recently Goncharov et al. (2021b) of the PPTA DR2, the EPTA is optimizing the pulsar noise models via Bayesian model selection. One important aspect will be on the chromaticity, i.e. radio-frequency dependence, of the noise. Apart from DM, scattering variation can also introduce a significant noise term to the TOAs (Main et al. 2020). Other additional noise components such as band noise (i.e. noise confined in a specific observing-frequency band) and system noise (i.e. noise attributed to one specific observing system, therefore not being intrinsic pulsar noise) will also be investigated. Finding system noise requires overlapping data by multiple observing systems in the same frequency band. The addition of new-generation-backened data are now allowing better such investigations than DR1, where only some basic investigation could be applied (see Caballero et al. 2016) and further work had to be completed on the IPTA DR1 (Lentati et al. 2016). Despite the DR2 improvement, The IPTA remains the best framework to identify possible system noise components.

8 CONCLUSIONS

The EPTA has collected and analysed observations for six MSPs using five radio telescopes for a timespan of up to 24 yr. Data are collected both from single-telescope and phased-array (LEAP) observations, in a wide range of radio frequencies. The DR1 has been published in Desvignes et al. (2016) and analysed to place an uppper limit on the isotropic stochastic GWB strain amplitude (LTM15); the corresponding amplitude limit at the 95 per cent credible region for circular, GW-driven SMBHB was 3×10^{-15} on the GWB. Initial analysis with the DR1 on the correlated search had shown the possible presence of a CRS amongst the six MSPs. In this work we re-analyse the same six MSPs using the extended DR2 data and find increased and strong evidence for the CRS, the properties of which are now very well measured and remain consistent with the DR1 results.

We have determined the pulsar noise properties with two separate pipelines with fully consistent results. A simultaneous search for the spectral properties and spatial correlation of a common red signal detects a well-constrained contour for the power-law spectral parameters with $\log_{10} A = -14.32^{+0.31}_{-0.39}$ and $\gamma = 3.83^{+0.82}_{-0.72}$ (95 per cent credible regions), but no conclusive measurement on the spatial correlation. Thus, we employed Bayesian model selection to compare different physically motivated spatial correlations in their ability to fit the data. In the case where the SSE is fixed to DE438, we find strong evidence for a CURN in addition to intrinsic, individual pulsar red noise with a $\log_{10} \mathrm{BF} \sim 3.7$ for varying spectral index. This is a significant increase from the $\log_{10} \mathrm{BF} \sim 1$ in DR1. The model selection for CRSs with different ORFs shows very little evidence for a monopolar and some evidence for a dipolar signal. The HD correlation is slightly disfavoured compared to the CURN, but more favoured than the dipolar correlation. None the less, we cannot rule out that the CRS is due to SSE systematics with this current results. We further analysed the data sets using three independent SSE-error mitigation models: `BAYESEPHEM`, `EPHEMGP`, and `LINIMOSS`.

While we concluded that our results are not significantly affected by SSE inaccuracies, we have confirmed that including SSE modelling reduces the evidence for a dipolar CRS in the data, and confirmed that the evidence reduction is stronger with wider priors on the values of planetary mass and orbital parameters. The effect of the SSE fitting in the general CURN search is a slight increase in the confidence intervals of the amplitude and spectral index of the power-law model. Longer timespan, more precise TOAs and more pulsars can help to diminish the effects of SSE systematics in the CURN recovery.

Assuming a single power-law model for the CURN, the recovered parameters constraints from the EPTA DR2 are $\log_{10} A = -14.29^{+0.26}_{-0.33}$ and $\gamma = 3.78^{+0.69}_{-0.59}$ (95 per cent credible regions). This power-law spectrum model uses the 30 lowest frequency bins. An analysis with a 'broken power law' spectrum has indicated the optimal number of frequency bins for the red part of the spectrum is ~20. However, the corresponding 2D posterior contours are consistent with those of the single power-law analysis, confirming the robustness our results to the exact number of frequency bins.

We have measured the contributions of the individual pulsars to the CURN using the dropout method. This produces a factor of consistency between the red noise of a given pulsar with the CURN. We find support for five out of six pulsars contributing to the CURN. This is an increase from three pulsars in DR1. The overall dropout consistency also increases significantly, particularly for J1600−3053 and J1909−3744, where DR2 adds the most data relative to DR1.

The nominal amplitude A_{GWB} for a power law at fixed $\gamma = 13/3$ at $1 \, yr^{-1}$ has been found to be $2.95^{+0.89}_{-0.72} \times 10^{-15}$. While there is a notable difference in the reported amplitude median values compared to the NG12 amplitude of 1.9×10^{-15} and PPTA DR2 amplitude of 2.2×10^{-15}, an overlap at the 2–3σ-level remains; in particular the significantly large NG12 parameter space for the spectral properties of the CURN cautions in drawing conclusions only from median values. Differences at this point, nevertheless, could have a multitude of reasons. One particular point could be in the covariance and modelling of the pulsar noise, which could leak into the CURN. Further investigations and cross-comparisons between the different data sets will be done in the IPTA framework.

If the full amplitude of 3×10^{-15} of the EPTA DR2 is due to the GWB, this would necessitate various strong astrophysical effects (see e.g. Sesana 2013; Kelley et al. 2017; Siwek, Kelley & Hernquist 2020). Following Middleton et al. (2021), the two main observables are the black hole binary mass and the merger time-scale. The hypothetical amplitude of 3×10^{-15} is close to the upper bound of possible values. In order to achieve such a strong GWB emission, black hole binaries have to be very heavy and the overall merger time-scale needs to be short. This would place very stringent limits on current massive black hole formation and evolution models. Alternatively, a number of other GW emission mechanisms can produce such an amplitude, such as cosmic strings (see e.g. Ölmez, Mandic & Siemens 2010; Sanidas, Battye & Stappers 2012), primordial black holes (see e.g. Grishchuk 2005; Lasky et al. 2016), phase transitions (see e.g. Caprini, Durrer & Siemens 2010; Kobakhidze et al. 2017) and numerous other sources.

This paper is the first of a series introducing the next generation of the EPTA data set DR2. In this work, we have focused on the six MSPs that have been analysed previously with DR1. Despite the strong evidence for a CRS, we have no measurement of the characteristic HD spatial correlation of the GWB. This is expected given the small number of pulsars used. The EPTA is preparing the expansion data set to ~25 pulsars, carefully selected based on their data quality, noise properties and sky positions to maximize our sensitivity to a GWB with HD ORF. The EPTA is also working on analysing the expanded DR2 data set more carefully regarding all aspects of the analysis, including, but not limited to, the pulsar timing and noise analysis, chromatic noise mitigation, Solar system dynamics modelling and more consistency checks on any CRS signal.

ACKNOWLEDGEMENTS

DATA AVAILABILITY

The timing data used in this article shall be shared on reasonable request to the corresponding authors.

REFERENCES

Aggarwal K. et al., 2019, ApJ, 880, 116

Alam M. F. et al., 2021, ApJS, 252, 4

Amaro-Seoane P. et al., 2017, preprint (arXiv:1702.00786)

Arzoumanian Z. et al., 2015, ApJ, 813, 65

Arzoumanian Z. et al., 2016, ApJ, 821, 13

Arzoumanian Z. et al., 2018, ApJ, 859, 47

Arzoumanian Z. et al., 2020, ApJ, 905, L34

Babak S. et al., 2016, MNRAS, 455, 1665

Backer D. C., 1995, JA&A, 16, 165

Bailes M. et al., 2020, PASA, 37, e028

Bassa C. G. et al., 2016a, MNRAS, 456, 2196

Bassa C. G. et al., 2016b, MNRAS, 460, 2207

Caballero R. N. et al., 2016, MNRAS, 457, 4421

Caballero R. N. et al., 2018, MNRAS, 481, 5501

Camilo F. et al., 2018, ApJ, 856, 180

Caprini C., Durrer R., Siemens X., 2010, Phys. Rev. D, 82, 063511

Champion D. J. et al., 2010, ApJ, 720, L201

Chen S., Sesana A., Del Pozzo W., 2017, MNRAS, 470, 1738

Chen S., Sesana A., Conselice C. J., 2019, MNRAS, 488, 401

Cognard I., Theureau G., Guillemot L., Liu K., Lassus A., Desvignes G., 2013, in Cambresy L., Martins F., Nuss E., Palacios A., eds, SF2A-2013: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics. Société Francaise d'Astronomie et d'Astrophysique, Paris, p. 327

Coles W., Hobbs G., Champion D. J., Manchester R. N., Verbiest J. P. W., 2011, MNRAS, 418, 561

Cordes J. M., Downs G. S., 1985, ApJS, 59, 343

Cornish N. J., Littenberg T. B., 2015, Class. Quantum Gravity, 32, 135012

D'Alessandro F., McCulloch P. M., Hamilton P. A., Deshpande A. A., 1995, MNRAS, 277, 1033

Demorest P. B. et al., 2013, ApJ, 762, 94

Desvignes G. et al., 2016, MNRAS, 458, 3341

Detweiler S., 1979, ApJ, 234, 1100

Eardley D. M., Lee D. L., Lightman A. P., 1973a, Phys. Rev. D, 8, 3308

Eardley D. M., Lee D. L., Lightman A. P., Wagoner R. V., Will C. M., 1973b, Phys. Rev. Lett., 30, 884

Efron B., Tibshirani R. J., 1994, An Introduction to the Bootstrap. CRC Press, Boca Raton

Ekers R. D., Moffet A. T., 1968, Nature, 220, 756

Ellis J., van Haasteren R., 2017, jellis18/PTMCMCSampler: Official Release, available at https://zenodo.org/record/1037579#.YXQLhhpBzIU

Ellis J. A., Vallisneri M., Taylor S. R., Baker P. T., 2020, ENTERPRISE: Enhanced Numerical Toolbox Enabling a Robust PulsaR Inference SuitE, available at https://zenodo.org/record/4059815#.YXQLuBpBzIU

Estabrook F. B., Wahlquist H. D., 1975, Gen. Relat. Gravit., 6, 439

Feroz F., Hobson M. P., 2008, MNRAS, 384, 449

Feroz F., Hobson M. P., Bridges M., 2009, MNRAS, 398, 1601

Fienga A., Deram P., Viswanathan V., Di Ruscio A., Bernus L., Durante D., Gastineau M., Laskar J., 2019, Notes Scientifiques et Techniques de l'Institut de Mecanique Celeste. Institut de mécanique céleste et de calcul des éphémérides, Paris, p. 109

Finn L. S., Larson S. L., Romano J. D., 2009, Phys. Rev. D, 79, 062003

Folkner W. M., Park R. S., 2018, IOM 392R-18-004

Foster R. S., Backer D. C., 1990, ApJ, 361, 300

Goncharov B. et al., 2021a, ApJ, 917, L19

Goncharov B. et al., 2021b, MNRAS, 502, 478

Gregory P. C., 2005, Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with 'Mathematica' Support. Cambridge Univ. Press, Cambridge

Grishchuk L. P., 2005, Phys. Usp., 48, 1235

Guinot B., Petit G., 1991, A&A, 248, 292

Guo Y. J., Li G. Y., Lee K. J., Caballero R. N., 2019, MNRAS, 489, 5573

Hager W. W., 1989, SIAM Rev., 31, 221

Handley W. J., Hobson M. P., Lasenby A. N., 2015, MNRAS, 453, 4384

Hazboun J. S. et al., 2020, ApJ, 890, 108

Hee S., Handley W. J., Hobson M. P., Lasenby A. N., 2016, MNRAS, 455, 2461

Hellings R. W., Downs G. S., 1983, ApJ, 265, L39

Hobbs G., 2013, Class. Quantum Gravity, 30, 224007

Hobbs G. B., Edwards R. T., Manchester R. N., 2006, MNRAS, 369, 655

Hobbs G. et al., 2012, MNRAS, 427, 2780

Hobbs G. et al., 2020, MNRAS, 491, 5951

Hotan A. W., van Straten W., Manchester R. N., 2004, PASA, 21, 302

Jaffe A. H., Backer D. C., 2003, ApJ, 583, 616

Janssen G. H., Stappers B. W., Kramer M., Purver M., Jessner A., Cognard I., 2008, in Bassa C., Wang Z., Cumming A., Kaspi V. M., eds, AIP Conf. Ser. Vol. 983, 40 Years of Pulsars: Millisecond Pulsars, Magnetars and More. Astron. Soc. Pac., San Francisco, p. 633

Jenet F. A. et al., 2006, ApJ, 653, 1571

Jenet F. et al., 2009, preprint (arXiv:0909.1058)

Jiang P. et al., 2019, Sci. China Phys. Mech. Astron., 62, 959502

Jones P. B., 1990, MNRAS, 246, 364

Jones M. L. et al., 2017, ApJ, 841, 125

Karuppusamy R., Stappers B., van Straten W., 2008, PASP, 120, 191

Kass R. E., Raftery A. E., 1995, J. Amer. Stat. Assoc., 90, 773

Keith M. J. et al., 2013, MNRAS, 429, 2161

Kelley L. Z., Blecha L., Hernquist L., Sesana A., Taylor S. R., 2017, MNRAS, 471, 4508

Kibble T. W. B., 1976, J. Phys. A Math. Gen., 9, 1387

Kobakhidze A., Lagger C., Manning A., Yue J., 2017, Eur. Phys. J. C, 77, 570

Kramer M., Stappers B., 2010, in van Leeuwen J., ed., ISKAF2010 Science Meeting. p. 34

Kramer M. et al., 2006, Science, 314, 97

Lam M. T. et al., 2016, ApJ, 819, 155

Lam M. T. et al., 2018, ApJ, 861, 132

Lam M. T. et al., 2019, ApJ, 872, 193

Landau L. D., Lifshitz E. M., 1960, Electrodynamics of Continuous Media. Pergamon Press, Oxford, UK

Lasky P. D. et al., 2016, Phys. Rev. X, 6, 011035

Lazarus P., Karuppusamy R., Graikou E., Caballero R. N., Champion D. J., Lee K. J., Verbiest J. P. W., Kramer M., 2016, MNRAS, 458, 868

Lee K. J., 2013, Class. Quantum Gravity, 30, 224016

Lee K. J., Jenet F. A., Price R. H., 2008, ApJ, 685, 1304

Lee K. J. et al., 2014, MNRAS, 441, 2831

Lentati L., Alexander P., Hobson M. P., Taylor S., Gair J., Balan S. T., van Haasteren R., 2013, Phys. Rev. D, 87, 104021

Lentati L., Alexander P., Hobson M. P., Feroz F., van Haasteren R., Lee K. J., Shannon R. M., 2014a, MNRAS, 437, 3004

Lentati L., Hobson M. P., Alexander P., 2014b, MNRAS, 444, 3863

Lentati L. et al., 2015, MNRAS, 453, 2576 (LTM15)

Lentati L. et al., 2016, MNRAS, 458, 2161

Li G. Y., Ni W. T., 2003, Publications of Purple Mountain Observatory, 22, 3

Li G. Y. et al., 2008a, Int. J. Mod. Phys. D, 17, 1021

Li G. Y., Zhao H. B., Xia Y., Zeng F., Luo Y. J., 2008b, in Jin W. J., Platais I., Perryman M. A. C., eds, IAU Sym. 248, A Giant Step: from Milli- to Micro-arcsecond Astrometry. Kluwer, Dordrecht, p. 560

Li Z., Lee K., Caballero R. N., Xu Y., Hao L., Wang M., Wang J., 2020, Sci. China Phys. Mech. Astron., 63, 219512

Liu K., Verbiest J. P. W., Kramer M., Stappers B. W., van Straten W., Cordes J. M., 2011, MNRAS, 417, 2916

Liu K., Keane E. F., Lee K. J., Kramer M., Cordes J. M., Purver M. B., 2012, MNRAS, 420, 361

Liu K. et al., 2020, MNRAS, 499, 2276
Lorimer D. R., Kramer M., 2005, Handbook of Pulsar Astronomy. Cambridge University Press, Cambridge
Maggiore M., 2000, Phys. Rep., 331, 283
Main R. A. et al., 2020, MNRAS, 499, 1468
Manchester R. N., 2006, IAU Joint Discussion, Highlights of Astronomy. Cambridge Univ. Press, Cambridge, Vol. 26, p. 16
Melatos A., Link B., 2014, MNRAS, 437, 21
Middleton H., Sesana A., Chen S., Vecchio A., Del Pozzo W., Rosado P. A., 2021, MNRAS, 502, L99
Ölmez S., Mandic V., Siemens X., 2010, Phys. Rev. D, 81, 104028
Osłowski S., van Straten W., Hobbs G. B., Bailes M., Demorest P., 2011, MNRAS, 418, 1258
Parthasarathy A. et al., 2021, MNRAS, 502, 407
Perera B. B. P. et al., 2018, MNRAS, 478, 218
Perera B. B. P. et al., 2019, MNRAS, 490, 4666
Rajagopal M., Romani R. W., 1995, ApJ, 446, 543
Romani R., 1988, Phys. Lett. B, 215, 477
Sanidas S. A., Battye R. A., Stappers B. W., 2012, Phys. Rev. D, 85, 122003
Sazhin M. V., 1978, Sov. Astron., 22, 36
Sesana A., 2013, MNRAS, 433, L1
Sesana A., Haardt F., Madau P., Volonteri M., 2004, ApJ, 611, 623
Shannon R. M., Cordes J. M., 2010, ApJ, 725, 1607
Shannon R. M. et al., 2014, MNRAS, 443, 1463
Siwek M. S., Kelley L. Z., Hernquist L., 2020, MNRAS, 498, 537
Skilling J., 2004, in Fischer R., Preuss R., von Toussaint U., eds, AIP Conf. Ser. Vol. 735, Bayesian Inference And Maximum Entropy Methods in Science and Engineering. Astron. Soc. Pac., San Francisco, p. 395
Speagle J. S., 2020, MNRAS, 493, 3132
Stappers B. W., Kramer M., Lyne A. G., D'Amico N., Jessner A., 2006, Chin. J. Astron. Astrophys. Suppl., 6, 298
Swarup G., 1990, Indian J. Radio Space Phys., 19, 493
Taylor S. R. et al., 2015, Phys. Rev. Lett., 115, 041101
Taylor J. H., 1992, Phil. Trans. R. Soc. Lond. Ser. A, 341, 117
Taylor J. H., 1993, Class. Quantum Gravity, 10, S167
Taylor J. H., Weisberg J. M., 1982, ApJ, 253, 908
Taylor S. R., Baker P. T., Hazboun J. S., Simon J., Vigeland S. J., 2021, enterprise_extensions, available at https://github.com/nanograv/enterprise_extensions
Tiburzi C. et al., 2016, MNRAS, 455, 4339
Vallisneri M., van Haasteren R., 2017, MNRAS, 466, 4954
Vallisneri M. et al., 2020, ApJ, 893, 112
van Haasteren R., Levin Y., 2013, MNRAS, 428, 1147
van Haasteren R., Vallisneri M., 2014, Phys. Rev. D, 90, 104012
van Haasteren R., Vallisneri M., 2015, MNRAS, 446, 1170
van Haasteren R., Levin Y., McDonald P., Lu T., 2009, MNRAS, 395, 1005
van Straten W., Bailes M., 2011, Publ. Astron. Soc. Aust., 28, 1
Verbiest J. P. W. et al., 2009, MNRAS, 400, 951
Verbiest J. P. W. et al., 2016, MNRAS, 458, 1267
Will C. M., 2014, Liv. Rev. Relat., 17, 4
Woodbury M. A., 1950, Memorandum Report. Princeton University, Princeton, NJ, p. 42
Wyithe J. S. B., Loeb A., 2003, ApJ, 590, 691
You X. P. et al., 2007, MNRAS, 378, 493

APPENDIX A: SUPPLEMENTARY MATERIAL

Table A1. Results of single-pulsar noise analysis for ENTERPRISE (EP) and TEMPONEST (TN) for DR1 in the same style as Table 3.

Pulsar	$\log_{10}A_{RN}$ EP	TN	γ_{RN} EP	TN	$\log_{10}A_{DM}$ EP	TN	γ_{DM} EP	TN
J0613−0200	$-14.44^{+1.3}_{-1.81}$	$-14.34^{+1.13}_{-1.49}$	$4.11^{+2.69}_{-3.12}$	$4.03^{+2.64}_{-2.57}$	$-11.6^{+0.14}_{-0.21}$	$-11.6^{+0.12}_{-0.15}$	$1.1^{+0.87}_{-0.66}$	$1.11^{+0.64}_{-0.53}$
J1012+5307	$-12.99^{+0.17}_{-0.2}$	$-12.99^{+0.14}_{-0.16}$	$1.48^{+0.86}_{-0.68}$	$1.47^{+0.64}_{-0.57}$	$-15.22^{+3.24}_{-2.65}$	$-15.41^{+3.17}_{-2.37}$	$2.8^{+3.93}_{-2.67}$	$3.03^{+3.48}_{-2.77}$
J1600−3053	$-13.27^{+0.2}_{-4.38}$	$-13.29^{+0.21}_{-4.45}$	$1.38^{+4.75}_{-1.05}$	$1.44^{+4.91}_{-1.12}$	$-13.93^{+2.48}_{-3.86}$	$-13.52^{+2.08}_{-4.24}$	$2.0^{+4.65}_{-1.76}$	$1.92^{+4.68}_{-1.65}$
J1713+0747	$-15.01^{+1.2}_{-1.91}$	$-15.03^{+0.97}_{-0.92}$	$4.89^{+1.97}_{-3.36}$	$5.09^{+1.69}_{-2.05}$	$-11.69^{+0.11}_{-0.11}$	$-11.7^{+0.08}_{-0.08}$	$1.28^{+0.56}_{-0.53}$	$1.22^{+0.44}_{-0.4}$
J1744−1134	$-14.21^{+0.8}_{-3.38}$	$-14.0^{+0.56}_{-3.0}$	$3.03^{+3.61}_{-2.47}$	$3.01^{+3.24}_{-1.78}$	$-11.79^{+0.19}_{-0.22}$	$-11.83^{+0.2}_{-0.2}$	$0.64^{+0.88}_{-0.54}$	$0.58^{+0.93}_{-0.54}$
J1909−3744	$-13.9^{+0.21}_{-3.61}$	$-13.9^{+0.21}_{-3.62}$	$1.91^{+3.67}_{-1.36}$	$1.91^{+3.7}_{-1.35}$	$-14.87^{+2.79}_{-2.98}$	$-14.94^{+2.86}_{-2.91}$	$2.55^{+4.14}_{-2.36}$	$2.58^{+4.12}_{-2.39}$

Table A2. $\log_{10}BF$ comparison between DR1 and DR2 in the same style as Table 4.

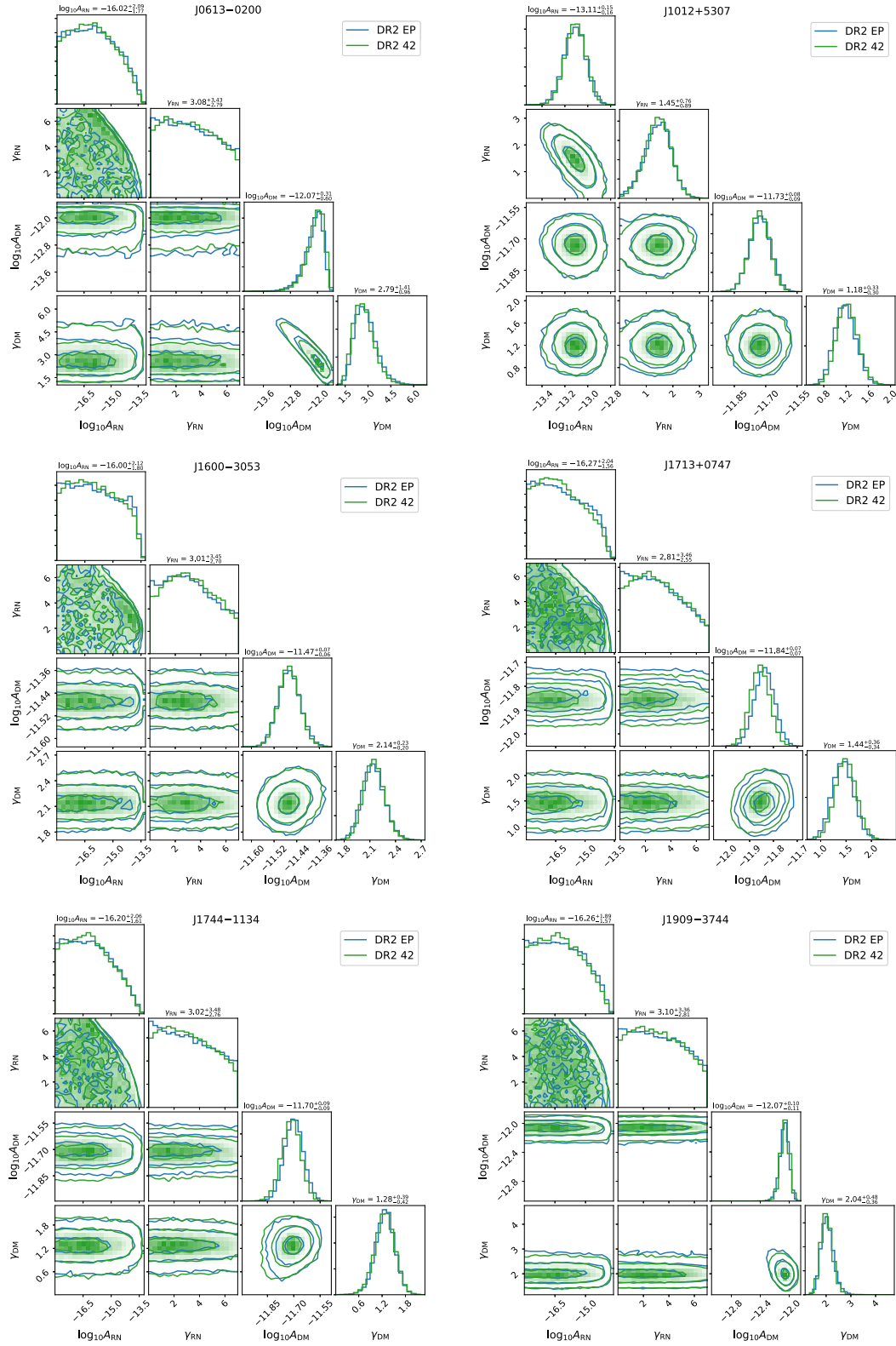| Model ID	Model	DR1 DE438 ENTERPRISE	FORTYTWO	DR2 DE438 ENTERPRISE	FORTYTWO
0	PSR	−	−	−	−
1	PSR + CURN	1.3	1.2	3.8	3.6
2	PSR + GWB	0.9	0.8	3.4	3.2
3	PSR + CLK	−0.3	−0.3	0.6	0.8
4	PSR + EPH	0.6	0.3	2.1	2.2
5	PSR + CURN + GWB	1.2	0.5	3.6	3.7
6	PSR + CURN + CLK	1.0	0.4	3.7	3.4
7	PSR + CURN + EPH	1.1	0.5	3.7	3.4

Figure A1. Comparison of the parameter posterior distributions from the CURN search between ENTERPRISE (EP) and FORTYTWO (42) split by pulsar noise parameters. The numbers indicate the 95 per cent credible regions from the EP analysis.

[1]*Laboratoire de Physique et Chimie de l'Environnement et de l'Espace LPC2E UMR7328, CNRS, Université d'Orléans, F-45071 Orléans, France*

[2]*Station de Radioastronomie de Nançay, Observatoire de Paris, PSL University, CNRS, Université d'Orléans, F-18330 Nançay, France*

[3]*Kavli Institute for Astronomy and Astrophysics, Peking University, Beijing 100871, P. R. China*

[4]*Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, D-53121 Bonn, Germany*

[5]*CNRS, Astroparticule et Cosmologie, Université de Paris, F-75013 Paris, France*

[6]*Dipartimento di Fisica 'G. Occhialini', Università degli Studi di Milano-Bicocca, Piazza della Scienza 3, I-20126 Milano, Italy*

[7]*INFN, Sezione di Milano-Bicocca, Piazza della Scienza 3, I-20126 Milano, Italy*

[8]*National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, P. R. China*

[9]*Moscow Institute of Physics and Technology, Dolgoprudny, Moscow 141700, Russia*

[10]*LESIA, Observatoire de Paris, Université PSL, CNRS, Sorbonne Université, Université de Paris, 5 place Jules Janssen, F-92195 Meudon, France*

[11]*ASTRON, Netherlands Institute for Radio Astronomy, Oude Hoogeveensedijk 4, NL-7991 PD Dwingeloo, the Netherlands*

[12]*Department of Astrophysics/IMAPP, Radboud University Nijmegen, P.O. Box 9010, NL-6500 GL Nijmegen, the Netherlands*

[13]*Institute of Astrophysics, FORTH, N. Plastira 100, 70013 Heraklion, Greece*

[14]*Argelander Institut für Astronomie, Auf dem Hügel 71, D-53117 Bonn, Germany*

[15]*Fakultät für Physik, Universität Bielefeld, Postfach 100131, D-33501 Bielefeld, Germany*

[16]*INAF – Osservatorio Astronomico di Cagliari, via della Scienza 5, I-09047 Selargius, CA, Italy*

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

[18]*Max-Planck-Institut für Gravitationsphysik (Albert Einstein Institut), Am Mühlenberg 1, D-14476 Golm, Germany*

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

[20]*Laboratory of Gravitational Waves and Cosmology, Advanced Institute of Natural Sciences, Beijing Normal University at Zhuhai, Zhuhai 519087, P. R. China*

[21]*Canadian Institute for Theoretical Astrophysics, University of Toronto, 60 St. George Street, Toronto, ON M5S 3H8, Canada*

[22]*Arecibo Observatory, HC3 Box 53995, Arecibo, PR 00612, USA*

[23]*Dipartimento di Fisica, Università di Cagliari, S.P. Monserrato-Sestu Km 0, 700, I-09042 Monserrato, CA, Italy*

[24]*Laboratoire Univers et Théories LUTh, Observatoire de Paris, Université PSL, CNRS, Université de Paris, F-92190 Meudon, France*

[25]*Institute for Gravitational Wave Astronomy and School of Physics and Astronomy, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK*

[26]*Department of Astronomy, Peking University, Beijing 100871, P. R. China*

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