Astronomy
&
Astrophysics

The second data release from the European Pulsar Timing Array

III. Search for gravitational wave signals

EPTA Collaboration and InPTA Collaboration: J. Antoniadis (Ιωάννης Αντωνιάδης)[1,2], P. Arumugam[3],
S. Arumugam[4], S. Babak[5], M. Bagchi[6,7], A.-S. Bak Nielsen[2,8], C. G. Bassa[9], A. Bathula[10],
A. Berthereau[11,12], M. Bonetti[13,14,15], E. Bortolas[13,14,15], P. R. Brook[16], M. Burgay[17], R. N. Caballero[18],
A. Chalumeau[13], D. J. Champion[2], S. Chanlaridis[1], S. Chen[19,*], I. Cognard[11,12], S. Dandapat[20], D. Deb[6],
S. Desai[21], G. Desvignes[2], N. Dhanda-Batra[22], C. Dwivedi[23], M. Falxa[5,11], R. D. Ferdman[24],
A. Franchini[13,14], J. R. Gair[25], B. Goncharov[26,27], A. Gopakumar[20], E. Graikou[2], J.-M. Grießmeier[11,12],
L. Guillemot[11,12], Y. J. Guo[2,*], Y. Gupta[28], S. Hisano[29], H. Hu[2], F. Iraci[30,17], D. Izquierdo-Villalba[13,14],
J. Jang[2], J. Jawor[2], G. H. Janssen[9,31], A. Jessner[2], B. C. Joshi[28,3], F. Kareem[32,33], R. Karuppusamy[2],
E. F. Keane[34], M. J. Keith[35], D. Kharbanda[21], T. Kikunaga[29], N. Kolhe[36], M. Kramer[2,35],
M. A. Krishnakumar[2,8], K. Lackeos[2], K. J. Lee[37,38], K. Liu[2], Y. Liu[38,8], A. G. Lyne[35], J. W. McKee[39,40],
Y. Maan[28], R. A. Main[2], M. B. Mickaliger[35], I. C. Niţu[35], K. Nobleson[41], A. K. Paladi[42], A. Parthasarathy[2],
B. B. P. Perera[43], D. Perrodin[17], A. Petiteau[44,5], N. K. Porayko[13,2], A. Possenti[17], T. Prabu[45],
H. Quelquejay Leclere[5], P. Rana[20], A. Samajdar[46], S. A. Sanidas[35], A. Sesana[13,14,15], G. Shaifullah[13,14,17],
J. Singha[3], L. Speri[25], R. Spiewak[35], A. Srivastava[21], B. W. Stappers[35], M. Surnis[47], S. C. Susarla[48],
A. Susobhanan[49], K. Takahashi[50,51], P. Tarafdar[6], G. Theureau[11,12,52], C. Tiburzi[17], E. van der Wateren[9,31],
A. Vecchio[16], V. Venkatraman Krishnan[2], J. P. W. Verbiest[53,8,2], J. Wang[8,54,55], L. Wang[35], and Z. Wu[38,8]

(Affiliations can be found after the references)

ABSTRACT

We present the results of the search for an isotropic stochastic gravitational wave background (GWB) at nanohertz frequencies using the second data release of the European Pulsar Timing Array (EPTA) for 25 millisecond pulsars and a combination with the first data release of the Indian Pulsar Timing Array (InPTA). A robust GWB detection is conditioned upon resolving the Hellings-Downs angular pattern in the pairwise cross-correlation of the pulsar timing residuals. Additionally, the GWB is expected to yield the same (common) spectrum of temporal correlations across pulsars, which is used as a null hypothesis in the GWB search. Such a common-spectrum process has already been observed in pulsar timing data. We analysed (i) the full 24.7-year EPTA data set, (ii) its 10.3-year subset based on modern observing systems, (iii) the combination of the full data set with the first data release of the InPTA for ten commonly timed millisecond pulsars, and (iv) the combination of the 10.3-year subset with the InPTA data. These combinations allowed us to probe the contributions of instrumental noise and interstellar propagation effects. With the full data set, we find marginal evidence for a GWB, with a Bayes factor of four and a false alarm probability of 4%. With the 10.3-year subset, we report evidence for a GWB, with a Bayes factor of 60 and a false alarm probability of about 0.1% ($\gtrsim 3\sigma$ significance). The addition of the InPTA data yields results that are broadly consistent with the EPTA-only data sets, with the benefit of better noise modelling. Analyses were performed with different data processing pipelines to test the consistency of the results from independent software packages. The latest EPTA data from new generation observing systems show non-negligible evidence for the GWB. At the same time, the inferred spectrum is rather uncertain and in mild tension with the common signal measured in the full data set. However, if the spectral index is fixed at 13/3, the two data sets give a similar amplitude of $(2.5 \pm 0.7) \times 10^{-15}$ at a reference frequency of $1\,\mathrm{yr}^{-1}$. Further investigation of these issues is required for reliable astrophysical interpretations of this signal. By continuing our detection efforts as part of the International Pulsar Timing Array (IPTA), we expect to be able to improve the measurement of spatial correlations and better characterise this signal in the coming years.

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

1. Introduction

The first direct gravitational wave (GW) detection (Abbott et al. 2016) marked the beginning of a new era in the exploration of the Universe. Although terrestrial interferometers such as LIGO, Virgo, and KAGRA are sensitive to GWs at kilohertz frequencies, where stellar mass compact binary mergers leave their imprint, a variety of astrophysical and cosmological phenomena are expected to generate GWs over a much broader frequency spectrum, reaching down to the nanohertz regime and beyond.

Supermassive black holes (SMBHs) are ubiquitous in galaxies (Magorrian et al. 1998; Kormendy & Ho 2013) and there is growing evidence that some of them formed when the Universe was less than a gigayear old (e.g., Wang et al. 2019, 2021). According to the established cold dark matter cosmological scenarios, galaxy formation proceeds in a hierarchical fashion, with small galaxies merging with each other over cosmic history to build progressively larger structures (White & Rees 1978). If

---

these galaxies host SMBHs in their centres, in the aftermath of the merger, SMBH binaries (SMBHBs) will inevitably form (Begelman et al. 1980). Adiabatically inspiralling SMBHBs are anticipated to be the loudest sources of GWs at nanohertz frequencies. The incoherent superposition of their emitted GW signals forms a stochastic GW background (GWB), whose amplitude and spectral index relate to the galactic merger history of the Universe and to the dynamical properties of the emitting binaries (Jaffe & Backer 2003; Sesana et al. 2008; Sesana 2013). Besides SMBHBs, a stochastic GWB can be produced by a number of other physical processes potentially occurring in the early Universe, including non-standard inflationary fields (Guzzetti et al. 2016), first-order phase transitions (Caprini et al. 2010), and cosmological defects such as a network of cosmic strings (Damour & Vilenkin 2000). A comprehensive overview of these phenomena can be found in Caprini & Figueroa (2018).

Currently, this very low-frequency GW regime can only be accessed with pulsar timing arrays (PTAs; Foster & Backer 1990). The technique of pulsar timing relies on the exceptional rotational stability of a particular population of neutron stars, the millisecond pulsars (MSPs). The times of arrival (TOAs) of radio pulses observed at the telescope are measured precisely using maser clocks referenced to the international atomic time. A model, known as a phase-connected timing solution, is then used to account for every rotation of the pulsar for the entire series of TOAs (see Lorimer & Kramer 2004, for a detailed explanation). The pulsar timing technique has allowed high-precision measurements that have led to several significant breakthroughs, including the first indirect detection of GWs through the measured orbital shrinkage of B1913+16 (Taylor & Weisberg 1989). In a PTA, a network of the most stable MSPs is observed regularly and the TOAs are modelled. It is within the small deviations from the model (the residuals) that nanohertz GWs can be searched for.

The idea of using MSPs to detect nanohertz GWs was first proposed by Sazhin (1978) and Detweiler (1979). The distortions in spacetime caused by a GW propagating over the Earth or over a pulsar lead to stochastic advances or delays in TOAs. Astrophysical sources produce GWs that cause larger delays over longer timescales, that is, a temporally correlated (red) signal. However, disentangling the GW signal from other red noise sources, such as variations in the interstellar medium (ISM) or intrinsic pulsar spin noise, with a single pulsar is impossible. Foster & Backer (1990) were the first to suggest a PTA as a method to overcome this problem. Not only would the GW signal result in a common red signal (CRS) in all pulsars, but the signal would be spatially correlated across the sky. This correlation is related to the quadrupolar nature of GWs. Although GWs passing over the individual pulsars would not be correlated, those propagating over the Earth would be. When the degree of correlation for each pair of pulsars is plotted against their angular separation, this results in the Hellings and Downs curve (HD; Hellings & Downs 1983). It is this HD curve that allows us to distinguish the GWB from other potential sources of correlated signal (e.g., the modelling of Earth's motion in the Solar system and local clock instabilities; Tiburzi et al. 2016).

The European Pulsar Timing Array (EPTA; Kramer & Champion 2013) was formed in 2004 to facilitate the detection of GWs. However, it uses pulsar observations taken well before its formal creation, some of which were specifically for PTA-style analysis. The EPTA data are provided by some of the largest radio telescopes in Europe: the *Lovell* Telescope at the Jodrell Bank Observatory, the Nançay decimetric radio telescope, the Westerbork synthesis radio telescope, the Effelsberg

100 m radio telescope, and the Sardinia radio telescope. These telescopes supply independent data sets at a range of observing frequencies (see EPTA Collaboration 2023), but since 2009 they have also worked together as the Large European Array for Pulsars (LEAP), a coherently phased interferometer with an equivalent dish diameter of up to 194 m (Bassa et al. 2016).

The earliest EPTA data date back to 1994, with most pulsars having over 15 yr of data. The bandwidth available and the backends used to record the data have improved over the years. While only some coherently dedispersing backend systems were used initially, considerable upgrades were made around 2005–2010, when most telescopes switched to broadband, coherent dedispersion systems. In addition to offering a wide range of observing frequencies and high observing cadence (with weekly or even shorter spacings between successive observations), multiple telescopes also allow the data sets to be checked against each other, highlighting any local issues. This is crucial for reliability.

The EPTA is a founding member of the International Pulsar Timing Array (IPTA; Verbiest et al. 2009), along with the Parkes Pulsar Timing Array (PPTA) which uses the Parkes telescope (Manchester 2006; Hobbs 2013) and the North American Nanohertz Observatory for Gravitational waves (NANOGrav), which uses data from the Arecibo observatory, Green Bank radio telescope, and Very Large Array (Jenet et al. 2009; Arzoumanian et al. 2015). Recently, the Indian Pulsar Timing Array (InPTA, using the Giant Meterwave Radio Telescope, Joshi et al. 2018) has joined the IPTA as a full member, while the Chinese Pulsar Timing Array (CPTA; using the Five Hundred Meter Spherical radio Telescope; Jiang et al. 2019), and the MeerTIME programme (using the MeerKAT telescope, Bailes et al. 2020) have become observer members.

The first EPTA data release (DR1) was made in 2015 (Desvignes et al. 2016) and was used to place an upper limit on the GWB (Lentati et al. 2015). However, during the analysis of the six best pulsars in the array, a weak CRS was observed in the data. An analysis of the same pulsars in 2021 (Chen et al. 2021) allowed for a direct comparison with the earlier work and clearly showed that not only was the CRS still present in the data, but also that its properties could be significantly better constrained. This CRS is consistent with the findings of the NANOGrav 12.5-yr data set (Arzoumanian et al. 2020), PPTA DR2 (Goncharov et al. 2021b), and IPTA DR2 (Antoniadis et al. 2022). However, six pulsars do not provide enough pairs to sufficiently sample the HD curve, the crucial signature of GWs. To this end, EPTA Data Release 2 (DR2; EPTA Collaboration 2023) has been created using 25 MSPs optimally selected among those timed by the collaboration, following the method described in Speri et al. (2023).

In this paper, we present the results of the search for a stochastic GWB at nanohertz frequencies in the EPTA DR2. A summary of the data set and noise models is given in Sect. 2. For more details, we refer interested readers to the companion papers (EPTA Collaboration & InPTA Collaboration 2023; EPTA Collaboration 2023), respectively. In Sect. 3, we briefly review our analysis methods, which are similar to those used in the six-pulsar analysis presented in Chen et al. (2021). Our main results, including a comparison of the full DR2 data set against a reduced data set that includes only the new generation backend systems, are presented in Sect. 4. In Sect. 5, we discuss the addition of InPTA data and its impact on the GWB search, and draw our conclusions in Sect. 6.

2. Data set and noise models

The DR2 includes observations of 25 pulsars selected from the DR1 source list. These data were collected with six EPTA telescopes, including LEAP. The DR2 data set is a combination of data from DR1 with those recorded with a new generation of data acquisition systems, which offer significantly wider bandwidth and thus greater sensitivity. The DR2 data set offers a variety of time spans for different pulsars, from a minimum of 14 to a maximum of 25 yr. That data set also has a broad observing frequency coverage, starting from (∼300 MHz) and extending up to (∼4 GHz). A subset of DR2, using data for six pulsars, was used for the common red noise process search presented in Chen et al. (2021). Since then, our multi-telescope data have allowed us to detect and correct an issue in the clock corrections applied to the data collected with the 'NUPPI' pulsar backend (Cognard et al. 2013) at the Nançay Radio Telescope. More details of the EPTA DR2 data set and timing analysis results can be found in our data release paper (EPTA Collaboration 2023).

For the analysis presented in this paper, we used two versions of the DR2, the full data set and a truncated version that features data collected with the new generation of pulsar backends only. These were extended by incorporating data from the first InPTA data release (Tarafdar et al. 2022) for an overlapping set of ten pulsars. The InPTA data set was obtained using the upgraded Giant Metrewave Radio Telescope (uGMRT) from MJD 58235 to 59496 covering about 3.5 yr. It complements the EPTA data with simultaneous observations in the 300–500 MHz and 1260–1460 MHz bands and adds about 0.7 yr to the EPTA time span. To summarise, we analyse the following four data sets, additional details of which can be found in the EPTA data release paper (EPTA Collaboration 2023) and the accompanying pulsar noise analysis paper (EPTA Collaboration & InPTA Collaboration 2023):

DR2full. The complete EPTA DR2 data set, covering 24.7 yr of data;

DR2new. A reduced version of the entire data set, including only the last 10.3 yr of data, collected with new generation wide-band backends;

DR2full+. The same as DR2full, but with the addition of InPTA data for 10 pulsars, covering 25.4 yr of data;

DR2new+. The same as DR2new, but with the addition of InPTA data for 10 pulsars, covering 11.0 yr of data.

The new generation backends use improved hardware for the conversion of the electric signals to digital data streams and allow for coherent dedispersion during the observations, whereas previous systems mostly operated with incoherent dedispersion. The increased processing power also enables us to use up to four times the frequency bandwidth as compared to the older, legacy backends.

Before a correlated signal can be searched for, the deterministic properties of individual pulsars need to be modelled. This includes the spin, astrometric, orbital (for binaries), and noise parameters of the pulsar (EPTA Collaboration 2023). Single pulsar noise models for the data sets mentioned above have been obtained from a specific model selection scheme presented in EPTA Collaboration & InPTA Collaboration (2023). For all pulsars, the timing model parameters were analytically marginalised and the white noise parameters EFAC and EQUAD were set at fixed values (cf. Sect. 3). The TOAs are measured by averaging over a frequency range, in which the pulse profile can be considered stable. For the legacy systems, the full bandwidth of about 128 MHz was used. However, for the new generation backends with larger bandwidths, we split the observation into sub-bands, treating each sub-band independently with a template and offset. This method allowes us to reduce the number of TOAs while retaining most of the information from the observations. With at most four TOAs per observation and due to the significantly lowered sensitivity as a result of the sub-banding, we could not measure significant, time-correlated white noise. Thus the ECORR parameter, which describes the presence of such noise, was not included in the analysis. Model selection was applied for other time-correlated signals, allowing for the selection of the most favoured combination among observing-frequency independent red noise (RN), dispersion measure variations (DM), and scattering variations (SV). These correspond to stochastic signals that induce a delay in the timing residuals with a chromatic index k of zero, two, and four, respectively, which characterises the dependence on the observing radio frequency ν^{-k}. Two large events were observed in J1713+0747, one at MJD ∼ 54757 (Coles et al. 2015; Zhu et al. 2015; Desvignes et al. 2016) and one at MJD ∼ 57510 (Lam et al. 2018; Goncharov et al. 2021a; Chalumeau et al. 2022). These were assumed to be caused by sudden changes in the scattering and dispersion variation and modelled as deterministic signals with fixed chromatic indices $k = 4$ and $k = 2$, respectively, as obtained from a Bayesian fit to the data. While both events are spanned by DR2full, only the second event falls within the time-span of DR2new.

Each noise process was modelled as a sum over Fourier components. Following Chalumeau et al. (2022), we did not fix a priori the number of Fourier components of the various processes in the noise analysis. Instead, for all combinations of RN, DM, and SV, we determined the optimal number of Fourier components, that best described the data. We did not consider models that include SV but not DM, as this is not physically motivated. We obtained customised noise models for each pulsar from a Bayes factor (BF) evaluation among the candidate models and performed a final analysis by refitting the timing model parameters simultaneously with these favoured noise models. This enabled further refinement of the timing model parameters and a more reliable evaluation of the white noise parameters, which are subsequently fixed in the GW analyses. The interpretation of custom noise models and further discussion on the robustness of these results are presented in EPTA Collaboration & InPTA Collaboration (2023).

3. Methods

The analysis methods closely follow those of Chen et al. (2021) and references therein. In the following, we summarise the key components of the analysis and provide details of additional analyses included in this work.

The PTA likelihood function for a CRS search is given by (van Haasteren et al. 2009)

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

The post-fit residuals of pulsar I at observation i are denoted as $\widetilde{\delta t}_{I,i}$. The sum of the block diagonal covariance matrices for all pulsars $\tilde{\mathsf{C}}^*$ and the overlap reduction function (ORF) Γ multiplied by the covariance matrix for the correlated common red signal $\mathsf{C}_{\mathrm{CRS}}$ gives $\tilde{\mathsf{C}} = \tilde{\mathsf{C}}^* + \Gamma\mathsf{C}_{\mathrm{CRS}}$[1]. The timing model parameters are analytically marginalised over, see van Haasteren & Levin (2013) and van Haasteren & Vallisneri (2014) for more details.

---

[1] To avoid $\tilde{\mathsf{C}}$ becoming singular, it is regularised as $|\tilde{\mathsf{C}}^*| = |\mathsf{C}^*| \times |\mathsf{M}^{*T}\mathsf{C}^{*-1}\mathsf{M}^*|$, where M^* is the design matrix of all pulsars.

The covariance matrix for each pulsar contains information on the white and red noise components, RN, DM, and SV. Measurement uncertainties on the TOAs can be calibrated with a pair of white noise parameters, EFAC and EQUAD, for each telescope, receiver, and backend combination to modify the initial estimate from the instrument and TOA extraction method

$$\widetilde{\sigma}_{ij}^2 = (\sigma_{ij} \times \mathrm{EFAC})^2 + \mathrm{EQUAD}^2. \tag{2}$$

The red noise power spectra were modelled with a power law

$$S_{\mathrm{RN}} = \frac{A^2}{12\pi^2} \left(\frac{f}{1\,\mathrm{yr}} \right)^{-\gamma} \frac{\mathrm{yr}^3}{T}, \tag{3}$$

representing long-term variations in the ToAs, which are independent of the radio frequency of the observations. This model was used for both, individual pulsars as well as for any putative common red noise.

Propagation of the radio signals through the interstellar medium adds delays, that depend on the frequency of the radio photons. Following Chalumeau et al. (2022) we considered two types of processes; DM variations and scattering of the photons by electrons encountered along the line of sight between the pulsar and Earth. These were also modelled with power laws

$$S_k = \frac{A^2}{12\pi^2} \frac{K}{\nu^k} \left(\frac{f}{1\,\mathrm{yr}} \right)^{-\gamma} \frac{\mathrm{yr}^3}{T}, \tag{4}$$

where K is the DM constant at a reference frequency of 1 MHz, k is the chromatic index of two or four for DM and SV, respectively, and ν is the radio frequency of the propagating photons.

The frequencies f of the Fourier basis were chosen to be $f_n = n/T$ ($n = 1, ..., N$), where T is the time interval between the first and last observations, and N is the number of frequency bins considered. This number was customised for each noise process in each individual pulsar, as described in the companion noise analysis paper (EPTA Collaboration & InPTA Collaboration 2023).

For the CRS, we used two methods to determine the optimal number of frequency bins: 1) we fitted a broken power law to estimate the frequency, where the red noise became dominant over the white noise (Arzoumanian et al. 2020); and 2) we constructed a free-spectrum model, where the power at each frequency was modelled with an independent parameter (Lentati et al. 2013; Arzoumanian et al. 2020).

For the detection of the GWB, the characteristic spatial correlation described by the HD curve

$$\Gamma_{\mathrm{GWB}}(\zeta_{IJ}) = \frac{3}{2} x_{IJ} \ln x_{IJ} - \frac{x_{IJ}}{4} + \frac{1}{2} + \frac{1}{2} \delta x_{IJ}, \tag{5}$$

is the key criterion. Here, ζ_{IJ} is the spatial angular separation between pulsars I and J, $x_{IJ} = [1 - \cos(\zeta_{IJ})]/2$, and $\delta(x_{IJ})$ is the Kronecker delta function.

We employed three types of model to search for generic spatial correlations in the data to compare against the expected HD correlation from a GWB:

Binned correlation function (Taylor & Gair 2013), where we weighted the evidence for the pulsar pair correlations in ten bins of angular separations between pulsars;

Chebyshev polynomial decomposition to the third order (Lentati et al. 2015; Chen et al. 2021)

$$\Gamma(\zeta_{IJ}) \approx c_1 + c_2 y_{IJ} + c_3 (2y_{IJ}^2 - 1) + c_4 (4y_{IJ}^3 - 3y_{IJ}), \tag{6}$$

where $y_{IJ} = (\zeta_{IJ} - \pi/2)/(\pi/2)$ and c_i are the Chebyshev polynomial parameters whose priors are uniform in the range $[-1, 1]$.

The cross-correlation is normalised so that $\Gamma(x) \in [-1, 1]$. This decomposition can be used to compare against constraints from previous EPTA data sets.

Legendre polynomial decomposition to fifth order (Gair et al. 2014)

$$\Gamma(\zeta_{IJ}) \approx \sum_{i=0}^{5} l_i P_i(\cos \zeta_{IJ}), \tag{7}$$

where P_i are the Legendre polynomial functions of order i and l_i are the Legendre polynomial parameters whose priors are uniform in the range $[-1, 1]$. The parameters can be interpreted as the amount of power in the monopole $i = 0$, dipole $i = 1$, quadrupole $i = 2$, and higher modes. A pure GWB would have no monopole or dipole contributions (i.e. $l_0 = l_1 = 0$) in this decomposition with all power at $i \geq 2$.

The off-diagonal elements of the covariance matrix encode the information on cross-pulsar correlated common signals. Apart from the quadrupole HD, we tested for the presence of a monopole (associated with, e.g., clock time errors) and a dipole (associated with, e.g., systematics in the model of the position of the Earth, the Solar system ephemeris, SSE) term. Unless otherwise stated, all analyses were performed with a fixed SSE model, DE440, produced by Park et al. (2021). To check for possible SSE systematics, we performed additional analyses using the BAYESEPHEM model (Arzoumanian et al. 2018; Vallisneri et al. 2020).

Using only the diagonal terms of the covariance matrix allows for fast computational analysis and corresponds to a model without any spatial correlations, which we refer to as common uncorrelated red noise (CURN). It is also possible to use only the cross-terms to search for a common signal in a split-likelihood analysis (Arzoumanian et al. 2020; Antoniadis et al. 2022). If the posterior distribution of the uncorrelated model has a substantial number of samples that are within the support of the correlated model, it is possible to employ the reweighting formalism, which was introduced in Hourihane et al. (2023), to approximate the posterior of the correlated model. The reweighting process involves sampling from the posterior distribution of the uncorrelated model (CURN) and then adjusting the weights of the obtained samples to reflect their likelihood under the correlated model. This technique enables the posterior of a correlated model to be obtained efficiently, the BF between the two models to be obtained, and the quality of the reweighted samples to be quantified.

3.1. Bayesian analysis

We estimated the parameters by evaluating the posterior probability, which is proportional to the likelihood given by Eq. (1) multiplied by the prior. The inverse of the proportionality coefficient is referred to as Bayesian evidence and it is equal to the integral of the likelihood times the prior over the prior range. When searching for a GWB, we fixed the white noise parameters of each pulsar to the maximum likelihood values produced by the single pulsar noise analysis (EPTA Collaboration & InPTA Collaboration 2023) and we simultaneously evaluated the RN, DM, and SV of each individual pulsar and the CRS. The prior ranges adopted for these parameters are given in Table 1. Bayesian evidence was used for model selection, where the BF for one hypothesis over the other is equal to the ratio of the two evidence values corresponding to these hypotheses. The posterior evaluation was done mainly with PTMCMCSAMPLER (Ellis & van Haasteren 2017) with other samplers used for

Table 1. Prior ranges for the parameters of the power laws used in the analysis: amplitude, A, and spectral index, γ.

Parameter	Prior type	Range
$A_{\rm RN}$, $A_{\rm DM}$, $A_{\rm CRS}$	log-Uniform	$[10^{-18}-10^{-10}]$
$\gamma_{\rm RN}$, $\gamma_{\rm DM}$, $\gamma_{\rm CRS}$	Uniform	$[0-7]$

Notes. Subscripts RN, DM, and CRS denote the red noise, DM noise, and common red signal, respectively.

cross-checking: `m3c2` (Falxa et al. 2023), `Eryn` (Karnesis et al. 2023), and `nessai` (Williams et al. 2021).

In this work, we performed model selection in two ways. First, directly, by introducing a hyperparameter that switches between likelihoods corresponding to the two models. The ratio of the fraction of samples using one model to the fraction using the other model is the BF. This method is known as the product-space sampling method (Carlin & Chib 1995; Hee et al. 2016). The second method involves resampling from the CURN model; it is mentioned above and described in detail in Hourihane et al. (2023).

3.2. Frequentist analysis

We also used the frequentist optimal statistic (OS) framework, developed by Anholm et al. (2009), Demorest et al. (2013), and Chamberlin et al. (2015) with the noise marginalisation introduced by Vigeland et al. (2018), to compare against the Bayesian results. The Bayesian output of a CURN analysis was used as the input for the OS. This allowed for high computational efficiency, as the posterior distributions of pulsar noise were directly used to compute the statistics. With the OS we could compute the amount of correlated power for each pulsar pair. Comparing this correlation against different models gives signal-to-noise ratio (S/N) estimates for different types of spatial correlations.

3.3. Software packages

As in Chen et al. (2021), we used both `ENTERPRISE`[2] (Ellis et al. 2020; Taylor et al. 2021) and `FORTYTWO`[3] (Caballero et al. 2016) for the PTA likelihood computation and to cross-check the main results with both pipelines. Some of the more specific analyses were performed only with `ENTERPRISE` for computational cost efficacy since we have demonstrated in Chen et al. (2021) that the two pipelines produce broadly consistent results.

4. Search results on the EPTA data sets `DR2full` and `DR2new`

We first present results for the analysis carried out with EPTA-only versions of the data set, namely `DR2full` and `DR2new`. Results for the EPTA+InPTA data set analysis are presented in Sect. 5.

For simplicity and efficiency, our general analysis setup used the DE440 Solar system ephemeris fit. The starting values for the marginalisation of the timing model were taken from the timing analysis (EPTA Collaboration 2023). Pulsar noise models and observing system white noise parameters were taken from the single pulsar noise analysis (EPTA Collaboration & InPTA Collaboration 2023).

Previous PTA analyses (e.g., Arzoumanian et al. 2020; Antoniadis et al. 2022) have shown the importance of choosing the optimal number of frequencies to model any putative common signal and that most of the power of a common red signal can be found in the lowest frequency bins. We chose the width of the frequency bin to be $1/T$, where T is the time span of the data set. For the power law the lowest 24 (9) frequency bins were used to model the CRS in the `DR2full` (`DR2new`) data set, which correspond to a maximum frequency $f_{\rm max} = 24/24.7 \, {\rm yr}^{-1} (9/10.3 \, {\rm yr}^{-1})$, respectively. We subsequently used these limiting frequencies for the remaining analyses (unless otherwise specified). This choice was verified with a broken power law analysis, which showed that using a larger number of frequency bins did not impact the recovery of the parameters of the CRS.

We show the posterior distributions of the free-spectrum model for the first 50 (20) frequency bins for the `DR2full` (`DR2new`) data set in Fig. 1. The most noticeable difference is in the lowest constrained frequency bin. Extending the `DR2new` best fit power law to lower frequencies would result in a lower CRS in the $1/24.7 \, {\rm yr}^{-1}$ bin, compared to what is measured in `DR2full`. Analogously, fitting a power law to the `DR2full` data set excluding the lowest frequency bin could give constraints that are more consistent with the `DR2new` data set. This could be due to the non-stationary of the processes involved, deviations from the power law model, or some additional unmodelled noise in the full data set. Further investigations are needed to understand better this difference, if, and how it could be mitigated.

4.1. Spectral parameter constraints

The right panel of Fig. 1 shows the posteriors of the recovered parameters for a power law model for a HD correlated process. The recovered power law parameters with `DR2new` are logarithmic amplitude $\log_{10} A = -13.94^{+0.23}_{-0.48}$ and spectral index $\gamma = 2.71^{+1.18}_{-0.73}$ (90% credible regions). The spectral index is shallower than the expected mean value of 13/3 from a population of circular SMBHBs, which still lies within the 3σ credible region (also see Middleton et al. 2021). With `DR2full` the recovered logarithmic amplitude $\log_{10} A = -14.54^{+0.28}_{-0.41}$ and spectral index $\gamma = 4.19^{+0.73}_{-0.63}$ is closer to 13/3. The two data sets give consistent results in the sense that the two posteriors overlap and lie on the same $A - \gamma$ degeneracy line that corresponds to fixing the total HD-correlated power. Therefore, the HD-correlated power measured in `DR2full` and `DR2new` is comparable, although the spectral shape is not well constrained and appears to be different in the two data sets. In support of this statement, when fixing the spectral index to 13/3, the background amplitude inferred from the two data sets is consistent, with a value of $\log_{10} A = -14.61^{+0.11}_{-0.12}$.

Table 2 gives the 90% credible regions of the recovered power law parameters for different analyses and models. Once the data set is fixed, the CRS (CURN or GWB) parameter constraints are consistent between different software packages, namely `ENTERPRISE` and `FORTYTWO`, as well as different models. However, as already indicated by the free-spectrum analysis, there is a systematic difference in the CRS recovery between `DR2full` and `DR2new`.

We quantified the differences between the power law posterior distributions that arise from using different software packages and data sets by adapting the `tensiometer` package, outlined in Raveri & Doux (2021) and summarised in EPTA Collaboration & InPTA Collaboration (2023). This package essentially provides the probability density function of the
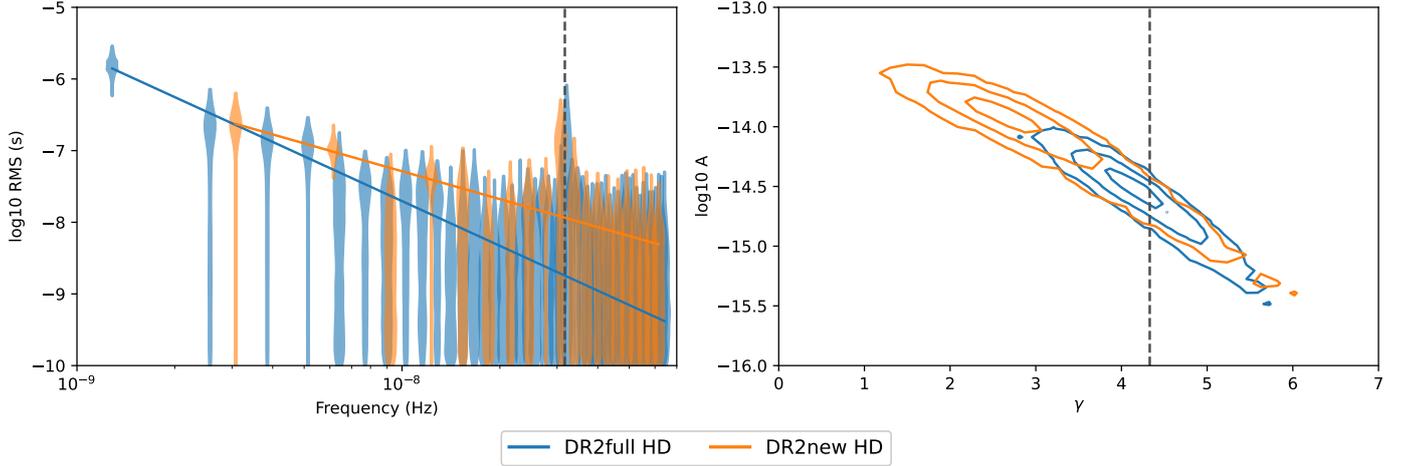
Fig. 1. Spectral properties of a CRS assuming HD correlation. The left panel shows the free spectrum, the independent measurement of common power at each frequency bin, for the two versions of the EPTA-only data set. The right panel shows the 1σ, 2σ, and 3σ contours of the 2D posterior distribution of amplitude and spectral index, when modelling the spectrum with a power law. In both panels, results for `DR2full` are in blue, while those of `DR2new` are in orange. The solid lines in the left panel are the power law best-fits to the GWB (see main text for the parameters of the fit), while the vertical dashed line indicates the position of $f = 1\,\mathrm{yr}^{-1}$. The vertical dashed line in the right panel denotes $\gamma = 13/3$.

Table 2. 90% credible regions for the power law parameters constraints in the different Bayesian analyses with DE440 for both `DR2full` and `DR2new`.

Software + Model	DR2full		DR2new	
	$\log_{10} A_{\mathrm{CRS}}$	γ_{CRS}	$\log_{10} A_{\mathrm{CRS}}$	γ_{CRS}
ENTERPRISE + CURN	$-14.53^{+0.29}_{-0.44}$	$4.13^{+0.80}_{-0.59}$	$-14.00^{+0.28}_{-0.77}$	$2.91^{+1.72}_{-0.87}$
FORTYTWO + CURN	$-14.52^{+0.30}_{-0.40}$	$4.12^{+0.74}_{-0.60}$	$-14.00^{+0.27}_{-0.66}$	$2.91^{+1.51}_{-0.85}$
ENTERPRISE + GWB	$-14.54^{+0.28}_{-0.41}$	$4.19^{+0.73}_{-0.63}$	$-13.94^{+0.23}_{-0.48}$	$2.71^{+1.18}_{-0.71}$
FORTYTWO + GWB	$-14.53^{+0.30}_{-0.40}$	$4.16^{+0.74}_{-0.66}$	$-13.94^{+0.24}_{-0.55}$	$2.71^{+1.30}_{-0.75}$
ENTERPRISE + Binned ORF	$-14.47^{+0.27}_{-0.35}$	$4.10^{+0.64}_{-0.56}$	$-13.89^{+0.22}_{-0.32}$	$2.63^{+0.86}_{-0.71}$
FORTYTWO + Binned ORF	$-14.49^{+0.29}_{-0.39}$	$4.11^{+0.72}_{-0.62}$	$-13.87^{+0.22}_{-0.37}$	$2.58^{+0.98}_{-0.74}$
ENTERPRISE + Chebyshev ORF	$-14.50^{+0.32}_{-0.40}$	$4.17^{+0.73}_{-0.72}$	$-13.87^{+0.22}_{-0.31}$	$2.57^{+0.86}_{-0.76}$
ENTERPRISE + Legendre ORF	$-14.51^{+0.30}_{-0.40}$	$4.19^{+0.74}_{-0.63}$	$-13.89^{+0.23}_{-0.35}$	$2.59^{+0.98}_{-0.72}$

Notes. The analyses included the search for common uncorrelated red noise (CURN), gravitational wave background (GWB), and a common correlated signal with overlap reduction function (ORF) modelled with different methods (binned ORF, Chebyshev ORF, and Legendre ORF).

parameter differences, which can be integrated to obtain the mean probability for the presence of parameter shifts (see Eq. (4) in Raveri & Doux 2021). The resulting probability for a parameter shift can be converted into an effective number of σ using the standard normal distribution. In short, the package produces a score that can be interpreted as 'within how many σ' two distributions are consistent (see also Raveri & Hu 2019, for more details). The results of this analysis in Table 3 indicate that the differences are minimal when comparing posteriors between different analysis software packages (`ENTERPRISE` vs. `FORTYTWO`) regardless of the data set (either `DR2full` or `DR2new`). However, when comparing GWB posteriors between different data sets (`DR2full` vs. `DR2new`), there are tensions of $\sim 1\sigma$ for CURN, $\sim 1.4\sigma$ for HD, and $\sim 1.6\sigma$ for Binned ORF, regardless of the software package used.

Figure 2 shows, in the left panel, the two-dimensional posterior difference distribution between the `ENTERPRISE` and `FORTYTWO` posteriors obtained for the `DR2new` data set, again showing consistency of the results provided by the two independent analysis packages. On the contrary, the corresponding distribution for the difference in the posteriors associated with

`DR2full` and `DR2new`, shown in the right panel, highlights the significant difference between the two data sets, more detailed comparisons can be found online[4].

The parameter constraints from the Bayesian pipelines can be compared with the results of OS estimates. We first fixed the spectral index to $\gamma = 13/3$ and computed the OS amplitude and S/N for a CRS with monopole, dipole, or HD correlation. A summary of our findings is given in Table 4, for the three correlation patterns in the four different data sets. The best-fit amplitudes for the HD correlation from the OS can be compared with the Bayesian value found when slicing the posterior at $\gamma = 13/3$, which is $A^2_{\mathrm{HD}} = 6.0^{+4.0}_{-3.0} \times 10^{-30}$. We notice that this value sits halfway between the OS amplitude estimate for the two data sets, with A^2_{HD} of $2.7^{+3.0}_{-2.5} \times 10^{-30}$ for `DR2full` and $10.0^{+5.1}_{-4.9} \times 10^{-30}$ for `DR2new`. Both estimates overlap with the Bayesian value within their 90% credible region. The median value for the OS S/N estimate for a HD-correlated process increases from 1.3 in `DR2full` to 3.5 for `DR2new`. The A^2_{CRS} and S/N distributions of the correlated processes as estimated by the OS are shown in Fig. 3, which further highlight the HD correlated signal emerging in `DR2new`.

---

[4] https://github.com/subhajitphy/Posterior_comparisons

Table 3. Z-score (in number of σ) produced by the `tensiometer` package, detailed in Raveri & Doux (2021), when comparing posteriors produced by various data sets and software packages.

	Data set comparison DR2new vs. DR2full ENTERPRISE (FORTYTWO)	Software package comparison ENTERPRISE vs. FORTYTWO DR2new (DR2full)
CURN	1.06 (1.15)	0.0063 (0.0274)
GWB	1.50 (1.49)	0.006 (0.0229)
Binned ORF	1.69 (1.68)	0.002 (0.0325)

Notes. The second column compares the posteriors between the `DR2new` and `DR2full` data set while employing `ENTERPRISE` and `FORTYTWO` (in brackets). On the contrary, the third column compares the posteriors given by the `ENTERPRISE` and `FORTYTWO` software packages running on the `DR2new` and `DR2full` data sets (in brackets).
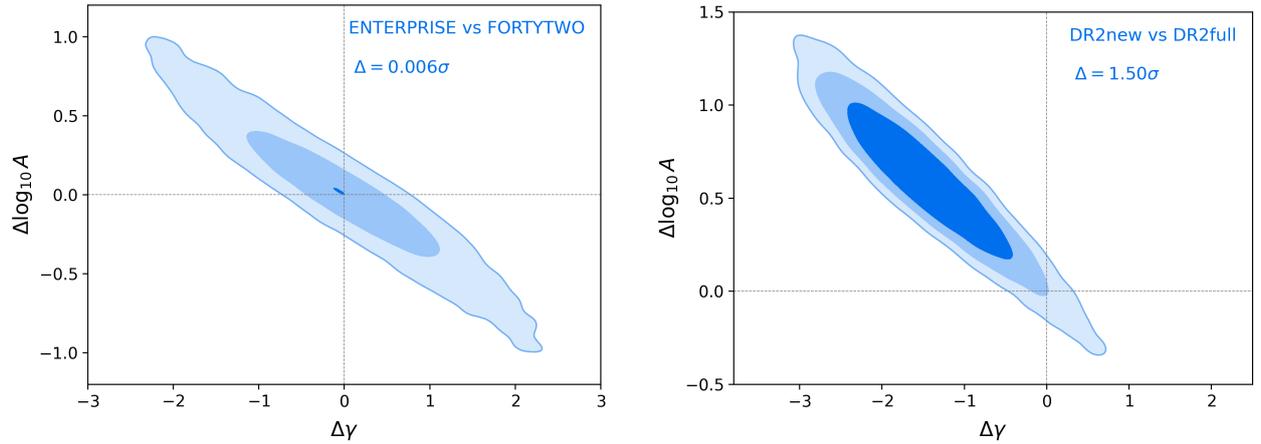
Fig. 2. Difference distributions between two posterior distributions originating from GWB processes. The left panel depicts the difference distribution between `DR2new` and `DR2full` data sets while employing the `ENTERPRISE` package. In comparison, the right panel shows the tension contour between `ENTERPRISE` and `FORTYTWO` software packages when we employ the `DR2new` data set. The plots contain three contours: 1σ, 2σ, and the Δ contours that correspond to the value of the computed tension.

Table 4. Median optimal statistic amplitudes and S/Ns for a single component fit for the monopole, dipole, or HD correlation fixing the spectral index of the common signal to 13/3, where the uncertainties indicate the 90% credible region.

	DR2full	DR2full+	DR2new	DR2new+
A_{MP}^2	$4.1^{+5.0}_{-3.9} \times 10^{-31}$	$3.7^{+5.1}_{-4.1} \times 10^{-31}$	$1.5^{+6.7}_{-4.3} \times 10^{-31}$	$4.0^{+7.0}_{-4.6} \times 10^{-31}$
A_{DP}^2	$-1.9^{+4.6}_{-4.1} \times 10^{-31}$	$-0.8^{+5.1}_{-4.3} \times 10^{-31}$	$0.8^{+9.3}_{-5.8} \times 10^{-31}$	$3.9^{+9.1}_{-6.5} \times 10^{-31}$
A_{HD}^2	$2.7^{+3.0}_{-2.5} \times 10^{-30}$	$2.9^{+2.9}_{-2.4} \times 10^{-30}$	$10.0^{+5.1}_{-4.9} \times 10^{-30}$	$11.0^{+4.6}_{-4.4} \times 10^{-30}$
S/N_{MP}	$1.1^{+1.1}_{-1.0}$	$1.0^{+1.1}_{-1.1}$	$0.3^{+1.4}_{-0.9}$	$0.8^{+1.7}_{-1.0}$
S/N_{DP}	$-0.4^{+0.9}_{-0.8}$	$-0.2^{+1.0}_{-0.9}$	$0.1^{+1.5}_{-0.9}$	$0.6^{+1.5}_{-0.9}$
S/N_{HD}	$1.3^{+1.3}_{-1.2}$	$1.4^{+1.2}_{-1.1}$	$3.5^{+2.4}_{-1.7}$	$4.1^{+2.7}_{-1.7}$

Although we fixed $\gamma = 13/3$ in the previous analysis, the OS can also be computed for a common red process with an arbitrary spectral slope. Figure 4 shows how the OS amplitude and S/N of the `DR2new` data subset change as we varied the spectral index γ of the CRS model. We increased γ from 1.0 to 5.0 in steps of 0.5 and also included $\gamma = 13/3$ to show the expected spectral index of a stationary ensemble of inspiralling SMBHBs. We evaluate the S/N for the monopole, dipole, and HD correlations for each γ. The median of the HD S/N appears to peak around a γ of 2.0, broadly consistent with the shallow posterior found in the Bayesian analysis (cf. the right panel of Fig. 1). The spread of the histograms, however, means that S/N values are self-consistent across the whole range of γ.

4.2. Spatial correlation constraints

After checking for spectral properties, we reconstructed the spatial correlation of the common red signal. The results of the Bayesian search for the correlations with ten binned free parameters and a common red signal power law are shown in Fig. 5. The bins were chosen so that each of them contains 30 pulsar pairs. The grey-shaded histogram represents the distribution of pulsar pairs as a function of angular separation. Since the pulsar distribution is concentrated in the galactic plane, we have more pairs at small angular separations compared to an array of pulsars uniformly distributed across the sky. However, broad coverage of all angles was still achieved with the 25 pulsars chosen by the
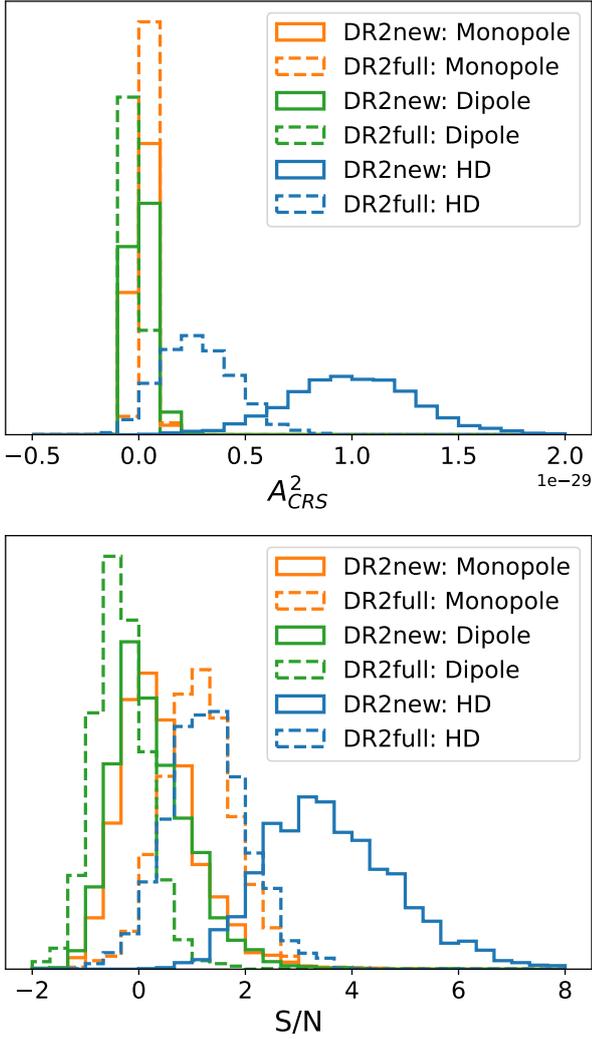
Fig. 3. Amplitude and S/N for a common red signal with $\gamma = 13/3$ for the optimal statistic. The top and bottom panels show the noise-marginalised distributions of the squared amplitude A_{CRS}^2 and S/N, respectively, for a common signal with different correlation patterns, with HD in blue, monopole in orange, and dipole in green. Solid lines are results from `DR2new` and the dashed lines are results from `DR2full`.

Fig. 4. Optimal statistic amplitude (left) and S/N (right) for a range of spectral indices for a common red signal using the `DR2new` data set. Results for the HD model are shown in blue, the dipole model in green, and the monopole model in orange.

ranking procedure of Speri et al. (2023). When comparing the `DR2full` ORF constraints with those of `DR2new`, one can see that the latter appears much more consistent with the expected HD correlation. In particular, the bins around 60, 80, and 135 degrees (i.e. the fifth, sixth, and ninth bins) have more positive correlation coefficients in `DR2full`. These appear to be responsible for the signal in `DR2full` being consistent with a CURN and a monopole, as also implied by the OS amplitude and S/N for a monopole correlation reported in Table 4. In contrast, `DR2new` is very consistent with a HD-correlated process. We also used Chebyshev and Legendre decompositions for the ORF in the Bayesian analysis to find ORF and power law constraints that are consistent with the binned free parameter analysis presented here; see Appendix A.

For comparison, the spatial correlations computed with the OS marginalised over the pulsar noise parameters are shown in Fig. 6, where the correlation coefficients have been obtained by scaling to the median amplitude at fixed $\gamma = 13/3$, as given in Table 4. For each noise realisation, only the median values of

the pulsar pair correlation were used. While Bayesian analysis averaged the correlation within each bin, the OS used each pulsar pair independently and fitted the best correlation across all pairs. For comparison and visual purposes, we chose the same binning to avoid showing 300 individual pulsar pairs. Although the two methods give broadly comparable ORF constraints, several differences can be found. Firstly, the first bin with the pulsar pairs with the closest separations deviates away from the HD and is consistent with no correlation. Secondly, the fourth and seventh bins drop significantly into negative correlations. These dips are most prominent in `DR2full`, while `DR2new` follows the HD curve more closely. Consistent with the Bayesian evaluation, the OS reconstruction also shows prominent positive correlations for
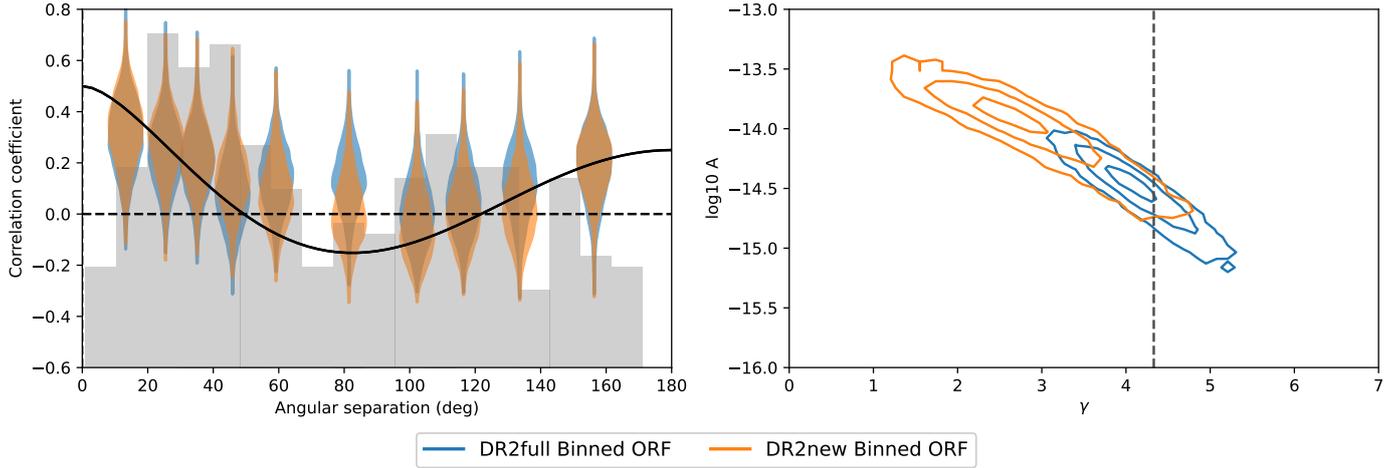
Fig. 5. Binned overlap reduction function. Blue is for `DR2full` while orange is for `DR2new`. The left panel shows violins of the posterior of the correlation coefficients averaged at ten bins of angular separations with 30 pulsar pairs each. The black line is the HD curve based on theoretical expectation of a GWB signal. The grey histogram is the arbitrarily normalised distribution of the number of pulsar pairs at different angular separations. The right panel is the corresponding 2D posterior for the amplitude and spectral index of the common correlated signal, showing 1σ, 2σ, and 3σ contours.
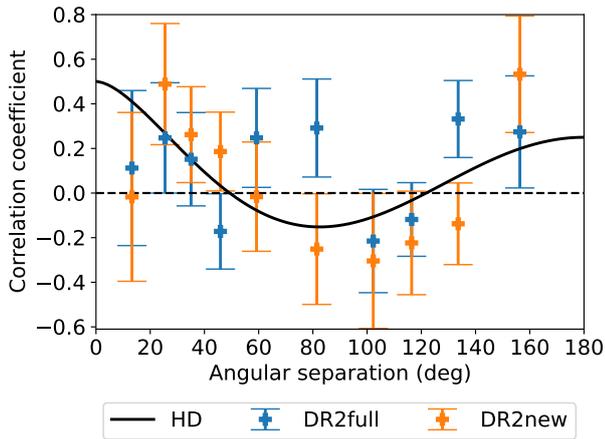
Fig. 6. Constraints on the overlap reduction function from the optimal statistic. Blue and orange points indicate the results for `DR2full` and `DR2new`, respectively. The correlation coefficients for each pair of pulsars are weighted and averaged following the description in Allen & Romano (2023) and grouped in the same way as those in Fig. 5 for comparison. The HD correlation is plotted as a black line for reference.

the fifth, sixth, and ninth bins in the `DR2full` data set, making the overall curve inconsistent with HD.

To quantify how likely the data set is actually showing evidence for a GWB, we compute BFs comparing different spatial correlations: HD correlations that arise from a GWB, monopole correlations that could be produced by clock errors (CLK), and dipole correlations that could be due to SSE systematics (EPH). Firstly, the presence of a common uncorrelated red noise (CURN) is strongly favoured against a model that includes only individual pulsar noises (PSRN). Compared to 3.3 from the 6PSR data set, the log10 BFs are ~5 for `DR2full` and ~3 for `DR2new`. Our main comparison was thus against the PSRN+CURN model, with respect to which we referred all BFs. These are summarised in Table 5. Both the PSRN+CLK and PSRN+EPH models are heavily disfavoured against the PSRN+CURN model in both `DR2full` and `DR2new` (row IDs

3 and 4). The evidence for these two correlations as additional processes to the CURN is inconclusive (row IDs 5 and 6). Conversely, while `DR2full` shows very little evidence for a GWB compared to the CURN, `DR2new` has a significant BF in favour of HD (row ID 2). Since this is a significant result, we have recomputed the BF using several alternative samplers and methods, as described in Sect. 3.1, obtaining BFs of 66, 56, and 62. In this data set, we also find that BFs for models including an additional CLK or EPH or CURN are about a factor of two smaller compared to the model including a GWB alone (row IDs 7, 8, and 9). On the contrary, the analogue BFs for `DR2full` are inconclusive with an indication for an additional monopole process (row ID 8).

These BFs can be compared to the S/Ns from the OS in Table 4 and Fig. 3. The `DR2new` data set yields a median $S/N \approx 3.5$ for the HD correlation, while it is about 1.3 in `DR2full`. These S/N estimates act as semi-independent confirmation of the BFs from Table 5. Consistent with the slightly higher BF for an additional monopole in `DR2full` the S/N is ≈ 1.2. For `DR2new` the S/N for a monopole drops to be consistent with zero. Lastly, no significant signal for a dipole correlation is found in either of the two data sets.

4.3. Significance tests

To quantitatively estimate the significance of the hypothesis that a GWB signal with HD correlation is present in the data, the null hypothesis distribution need to be constructed. Many repetitions of an experiment need to be performed in order to define a strict p-value. This is, unfortunately, not possible for PTAs. Thus, we can only attempt to find a good proxy to estimate the true statistical p-value for the null hypothesis. In the following, we refer to the estimated value from our proxy methods as p-values for simplicity. The respective distributions can be constructed in two different ways, by introducing random phase shifts in the Fourier basis of the common red noise process (Taylor et al. 2017) or by moving the positions of the pulsars in the sky via a random scramble (Cornish & Sampson 2016). The aim of both methods is to effectively destroy the distinctive cross-pulsar correlations, unique to the GWB signal, while retaining the individual pulsar

Table 5. Model selection for different inter-pulsar correlation models for a CRS.

ID	Model	DR2full		DR2full+	DR2new		DR2new+
		ENTERPRISE	FORTYTWO	ENTERPRISE	ENTERPRISE	FORTYTWO	ENTERPRISE
1	PSRN + CURN	–	–	–	–	–	–
2	PSRN + GWB	4	5	4	60	62	65
3	PSRN + CLK	<0.01	<0.01	<0.01	0.2	1.2	0.3
4	PSRN + EPH	<0.01	$\sim 10^{-4}$	<0.01	0.2	0.2	1.3
5	PSRN + CURN + CLK	2	1	2.7	0.8	2	1.6
6	PSRN + CURN + EPH	1	0.1	1	1	1	1.6
7	PSRN + GWB + CURN	3	3	4	27	13	25
8	PSRN + GWB + CLK	5	12	7	28	35	57
9	PSRN + GWB + EPH	3	3	3.6	33	29	43

Notes. We present BFs for different CRS models against the CURN model. We assumed the DE440 SSE fit and used the PSRN+CURN model as the reference model. The model component acronyms are: (i) PSRN = individual Pulsar noise, (ii) CURN = common uncorrelated red noise, (iii) GWB = gravitational wave background 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.

noise characteristics. One should emphasise that both methods should be robust against any mismodelled features in the data set, therefore they, in general, provide more conservative estimates of the significance in comparison to the possibly oversimplified noise simulation bootstrapping.

The distributions of BFs under the null hypothesis (PSRN+CURN) were constructed for DR2full and DR2new using about 200 and 2000 phase shifts, respectively and are displayed in the upper panel of Fig. 7. The DR2full measured BF from Table 5 lies within the 2σ range of the null hypothesis distribution with a p-value of 0.04. The p-value for the BF derived with the DR2new data set reaches a statistically interesting value of 0.0005, which corresponds to the 3σ level of significance ('evidence'). The analysis was performed using both ENTERPRISE and FORTYTWO and shows consistent results between the two software packages. This significance test was repeated for the OS S/N values for the HD correlation and results are shown in the bottom panel of Fig. 7. For DR2full a p-value of 0.07 was found. None of the 10 000 realisations produced a S/N that is comparable to what has been found in DR2new. Therefore, only an upper limit can be set for the p-value < 0.0001, which corresponds to a significance of $>3.5\sigma$.

Figure 8 shows the null distribution obtained with sky scrambles in the OS analysis in the top panel. A matching threshold of 0.2 for any two sky scrambles was imposed to produce about 5000 samples. A large difference particularly in the high S/N tail of the density functions can be found between DR2full and DR2new. The p-value for DR2full of 0.08 is comparable to that obtained with the phase shifts. This could indicate that in the low S/N regime, both methods produce reliable null distributions. In the high S/N regime, however, with DR2new the sky scramble p-value of 0.004 is not consistent with the phase shift method.

The bottom panel of Fig. 8 compares p-values from simulations, theoretical computation and the two methods. A null distribution was generated using a set of realistic simulations resembling the statistical properties of the real DR2new data set and with the injected CURN only. The noise parameters as well as the amplitude and slope of the CURN for the null distribution were taken as random samples from the posteriors of the CURN search with DR2new. Additionally, we compare these p-values with those obtained from the theoretical null distribution described by generalised χ^2 (GX2) distributions and derived in Hazboun et al. (2023). That distribution was computed by fixing the noise parameters to the median values of the posteriors[5].

Both null distributions are compared to the proxy distributions obtained via phase shifting and sky scrambling. For consistency, instead of using the real data set, a target data set was simulated using the same procedure as the simulations for the null distribution, but with GWB injected instead of CURN. The lowest p-value of 0.002 is obtained using phase shifts. The most conservative number is obtained when using all sky scrambles without introducing any threshold or weighting with the p-value of 0.008. At $S/N_{\mathrm{HD}} = 3.5$ the p-values of the simulated (0.006) and theoretical (0.003) null distributions lie between those from phase shifting and sky scrambling (see Fig. 8). We have tested introducing thresholds on the match between the true and scrambled pulsar positions and amongst the scrambles themselves. In general, we found that smaller thresholds lead to a decrease in the proxy null distribution. Similar results are obtained when adding weights to account for the different contributions from each pulsar. From all the above we can conclude that the p-value for the S/N found with DR2new should be ~ 0.004, which was also obtained with sky scrambling at a threshold of 0.2 and no weights. The inconsistencies between the p-values obtained with the real and simulated target data sets can be due to the incompleteness of our simulations (e.g., exponential dips for J1713+0747 are not included and a simpler noise model with a fixed number of frequency components was used).

For an optimal sky scrambling, orthogonality may need to be ensured between different realisations. Additionally, a realistic PTA does not have equally good pulsars, which should be taken into account when assessing the match between different scrambles. This can limit the maximum number of possible sky scrambles or their effectiveness in breaking correlated processes (Di Marco et al. 2023). On the other hand, Hazboun et al. (2023) have shown that sky scrambling can lead to null distributions that are consistent with the theoretical prediction. Further studies are required to determine whether any method can be a good,

---

[5] This difference in how the parameters for the pulsar noise were chosen between the simulated and theoretical null distributions can explain the more conservative p-value for the former. As the randomly drawn parameters for the simulations could allow for more large S/Ns to be found by chance.
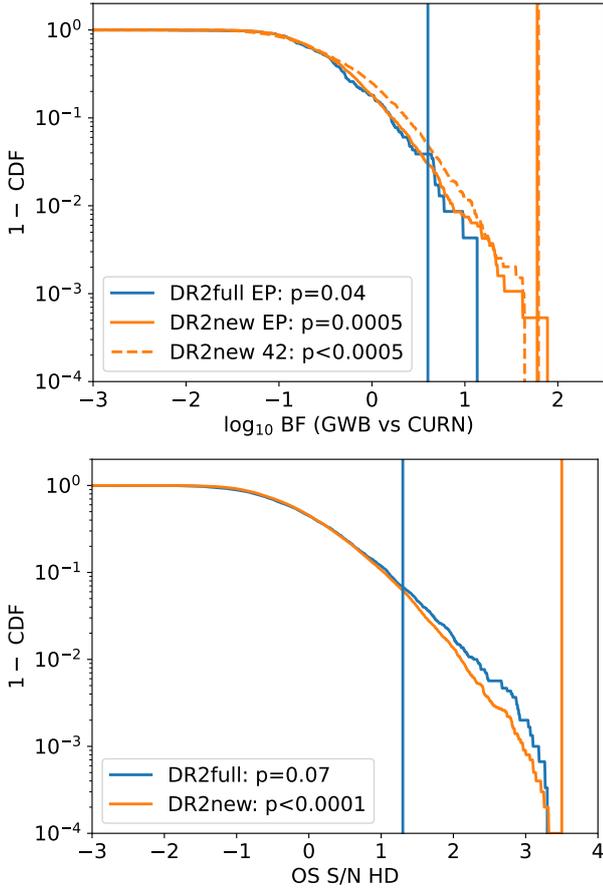
Fig. 7. 1-cumulative density function (CDF) from phase shifts. The top panel is for the BFs for PSRN+GWB versus PSRN+CURN using `ENTERPRISE` (EP) and `FORTYTWO` (42). It should be noted that due to the computational cost, we calculated only a limited number of phase-shifted BFs. This could explain the differences in the 1−CDFs. The OS S/N for the HD correlation with `ENTERPRISE` is shown in the bottom panel. Blue lines are for `DR2full` while orange lines are for `DR2new`. Vertical lines are the measured BF for PSRN+GWB versus PSRN+CURN reported in Table 5 or the OS S/N_{HD} reported in Table 4, respectively. The estimated p-values for each method are given in the legends.

Fig. 8. 1-cumulative density function (CDF) from sky-scrambled optimal statistic HD S/N distributions and a comparison between different methods. In the top panel, the blue histograms are for `DR2full` while the orange histograms are for `DR2new`. Vertical lines are the measured S/N_{HD} values reported in Table 4. The large discrepancy between the two data sets could be an indication that some remaining signal is still present after the scrambling, especially in the strong S/N regime. More checks need to be performed to assess the validity of this method. The bottom panel compares a simulated null distribution in cyan, the generalised χ^2 (GX2) distribution from Hazboun et al. (2023) in purple and the two proxy methods of phase shifting in orange and sky scrambling in green. At the measured value from the `DR2new`, the two methods differ by a factor of a few. The estimated p-values for each method are given in the legends.

4.4. Consistency tests

4.4.1. Comparing the power in the auto-correlation and cross-correlation terms.

For a true GWB, both the auto-correlation and cross-correlation terms should constrain the same process. Since the cross-correlated power is proportional to the square of the expectation value of the ORF, that is Γ^2, which is always $\ll 1$, it is expected that the power in the auto-correlation terms – equivalent to the CURN – is the dominant contributor to the signal. However, we stress that the cross-correlation terms contain the 'smoking gun' that can provide conclusive evidence for the presence of a GWB in the data.

We applied the split likelihood technique (Arzoumanian et al. 2020; Antoniadis et al. 2022) to both `DR2full` and `DR2new`. Figure 9 shows the resulting posterior contours. While the auto-correlation-terms-only analyses recover the CURN with

consistent amplitude and slope for both data sets, noticeable differences can be seen for the power law parameters that are constrained using the cross-correlation terms only assuming HD correlation. For the `DR2full` data set, the cross-correlation terms contain virtually no power and thus only an upper limit is found for a HD-correlated signal. On the contrary, consistent with the much larger evidence for a GWB, the cross-correlation terms in `DR2new` produce a peak in amplitude. Since a long tail still remains, one cannot rule out the possibility of a zero amplitude.

4.4.2. Solar System Ephemeris systematics

The effects of SSE systematics on the PTA GWB search have been studied in Tiburzi et al. (2016) and Guo et al. (2019) and models that can help mitigate the dipolar correlated signal induced by SSE systematics have been added to the tools for PTA analysis. We employed the widely used `BAYESEPHEM` model
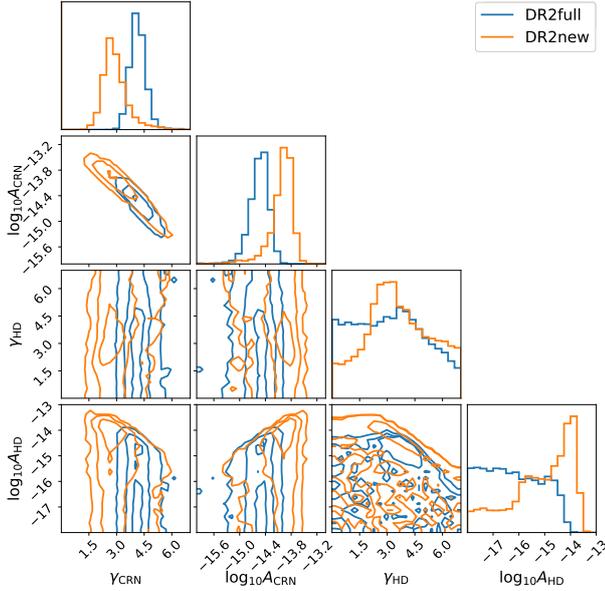
Fig. 9. Power law posterior constraints from a split likelihood analysis using the auto-correlation terms (CURN) and the cross-correlation terms (assuming HD correlation) separately. The left two columns show the recovered power law from the auto-correlation terms and the right two columns show the cross-correlation terms. Blue distributions are for `DR2full` while orange distributions are for `DR2new`.

(Arzoumanian et al. 2018; Vallisneri et al. 2020) on the EPTA-only data sets.

Figure 10 compares the addition of `BAYESEPHEM` to each data set against the use of the DE440 fit alone. Allowing SSE systematics to be present and absorbed by a model is a more conservative approach. In general, the left panels show that certain frequency bins have lower power compared to the DE440 analysis. This in turn broadens the power law posteriors. Comparing the `DR2full` data set against the `DR2new` data set, the longer time span of `DR2full` helps to produce results that are more independent of the SSE fitting used, while the short time span of `DR2new` strongly limits its ability to separate SSE systematics from other common signals. In fact 10.3 yr is close to the Jovian orbital period of about 12 yr, which could lead to a strong degeneracy.

Adding `BAYESEPHEM` also affects the BF for the different signal models. Table 6 shows a selection of the same models as Table 5. The most relevant effect in the search for a GWB is that the BF in favour of the HD correlation in `DR2new` is reduced from about 60 to 17, a factor of about 3. Since `BAYESEPHEM` is known to partially absorb power from the GWB, this reduction follows the expectation, although the exact amount is difficult to predict (Vallisneri et al. 2020).

4.4.3. Distinguishing a common red spectrum from similar individual pulsar noise spectra

It was shown in Goncharov et al. (2021b) that the CURN hypothesis may become strongly favoured over the PSRN hypothesis even when spectra of temporal correlations across pulsars are similar and yet not strictly common, as implied by the CURN model. This is because the standard uniform prior distributions for power law parameters of pulsar-intrinsic red noise do not match the observed distributions of these parameters. The issue was addressed in Goncharov et al. (2022), where

the authors introduced a hierarchical model that governs the distribution of power law amplitudes across pulsars in the CURN. As the distribution width is consistent with zero, this indicates the likely presence of a common-spectrum stochastic process in the data. Dropout analysis (e.g., Arzoumanian et al. 2020) also enables the mitigation of this issue by identifying pulsar outliers. The effect of the range of simulated pulsar noise parameters on the spurious evidence for CURN is demonstrated in Fig. 6 in Zic et al. (2022).

To test each pulsar's consistency with the CURN, we measured both the dropout factors and the distribution of power law amplitudes of the CURN spectrum in the pulsars. The results are shown in Fig. 11. The two top panels show the measurements of $\sigma_{\log_{10} A}$ and $\mu_{\log_{10} A}$, the standard deviation and the mean of the power law amplitude of CURN measured in the pulsars, following Goncharov et al. (2022). As expected, the mean $\mu_{\log_{10} A}$ is consistent with the measurement of $\log_{10} A_{\rm CURN}$. At the same time the standard deviation is consistent with zero, confirming the presence of common temporal correlations in the data. Based on the measured dropout factors, we find that in both data sets only a few pulsars have intrinsic red noise that seems to be inconsistent with the CURN. The majority of pulsars display indifferent behaviour, leaving around five pulsars to contribute most to the constraints of the CURN power law. However, the two data sets have CURNs with very different posterior contours. This can have an impact on the difference between certain pulsars agreeing more or less with the CURN. J1909−0747 and J1744−0744, for example, have very steep intrinsic red noise, thus they are more consistent with the steep CURN from `DR2full` compared to the shallow CURN from `DR2new`. More discussion on the intrinsic pulsar noise properties can be found in EPTA Collaboration & InPTA Collaboration (2023).

4.5. Continuous gravitational wave signal search

In addition to the GWB search, we have also searched our data for the presence of a continuous GW (CGW) signal from an individually resolvable SMBHB in a circular orbit. This subsection presents a preliminary analysis and a detailed investigation (including simulations and significance estimation) is given in Antoniadis et al. (2023b). The main aim of this section is twofold: (i) to look for an alternative explanation of the observed common signal and (ii) to understand how the inclusion of the CGW signal in the model affects the main findings of our analysis.

First, we performed an analysis using the `DR2full` data set. The addition of the CGW signal to the custom pulsar noise and CURN is not informative about the presence of a CGW, with $\rm BF_{\rm CURN}^{\rm CGW+CURN} = 0.5$ (model containing CGW and CURN over the model with CURN only).

Next, we move on to the analysis of the `DR2new` data set. We started by considering a simple model of a CGW source superimposed on PSRN. We found substantial support for the presence of a CGW: the BF of the model CGW+PSRN over the null model, PSRN, is 200 with the Earth term only and 260 if we include the pulsar term in the search. For this search, we used the sampler and method described in Bécsy et al. (2022). The candidate source is localised in sky position and frequency. We have also computed the F_e statistic (a frequentist approach; see Babak & Sesana 2012; Ellis et al. 2012). The results are shown in Fig. 12. The F_e statistic can be marginalised over the individual pulsar noise parameters in a similar manner as the OS. It is shown in blue with the mean value corresponding to $S/N \approx 4.5$. We computed the corresponding p-value ($\approx 0.1\%$) by scrambling the pulsar sky positions, assuming 0.2 match as the orthogonality
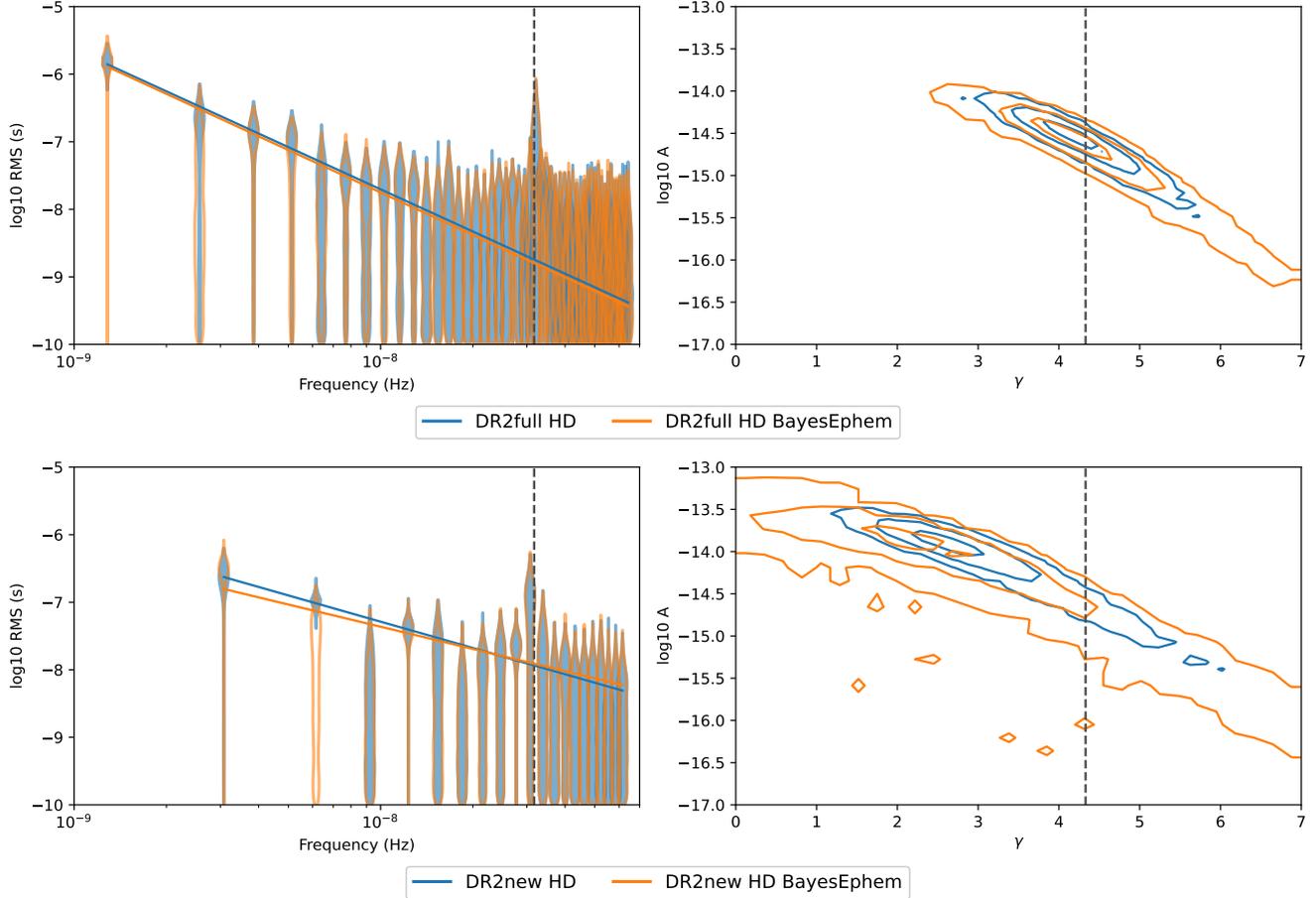
Fig. 10. Spectral properties of a CRS assuming HD correlation in the style of Fig. 1 for the DR2full (top) and DR2new (bottom) comparing the analysis without (blue) and with (orange) BAYESEPHEM.

Table 6. BFs for different CRS models against the CURN model adding BAYESEPHEM to all models with the ENTERPRISE pipeline.

ID	Model	DR2full	DR2new
1	PSRN + CURN	–	–
2	PSRN + GWB	1.5	17
3	PSRN + CLK	0.5	2
4	PSRN + EPH	<0.01	0.4

Notes. The model component acronyms are as in Table 5.

criterion. The F_e-statistic takes only the Earth-term into account and scrambling the pulsar's position destroys the coherence of the CGW, while preserving the noise properties. The F_e of sky scrambles is shown in orange and is compared to the theoretically expected (for Gaussian noise) central χ^2 distribution with four degrees of freedom.

Including the CURN in the model does not change these results significantly: we can still identify the CGW candidate with $\mathrm{BF}_{\mathrm{CURN}}^{\mathrm{CGW+CURN}} = 7$–$20$, with the exact value depending on whether one includes the pulsar term (BF = 7) or only the Earth term, as well as on the sampling method used (BF = $12, 20$). Interestingly, the CURN parameters are much less constrained in the CGW+CURN model. We show partial results of the Bayesian parameter inference in Fig. 13.

Finally, we have considered a model containing a GWB plus a CGW and found that the GWB 'absorbs' most of the CGW,

that is, the CGW becomes poorly constrained. At the same time, our results indicate that we cannot exclude the presence of the CGW, since the BF is not informative ($\mathrm{BF}_{\mathrm{CGW}}^{\mathrm{GWB}} = 1.2$), while the CGW model has a larger parameter space. We note that the identified CGW candidate frequency of approximately 5 nHz is between the two lowest frequency bins that dominate the HD signal, as shown in Fig. 1. The rest of the bins do not contribute much to the GWB signal, due to the relatively short observation span of the DR2new data set and the high level of white noise. To summarise, we find that the observed data is equally well explained by either a GWB or a single CGW. However, given the additional number of parameters for the CGW model and in the absence of additional data, we favour the simpler GWB model. A detailed analysis including extensive simulations can be found in Antoniadis et al. (2023b).

5. Search results on the EPTA+InPTA data sets DR2full+ and DR2new+

Analyses of DR2full+ and DR2new+ data sets including the InPTA data indicate general consistency of the results with the DR2full and DR2new data sets, respectively. We provide a comparison of posteriors for the CURN and GWB between the data sets, with and without InPTA data, in Tables 7 and 8, as well as in Fig. 14; see also Appendix B. While the DR2full+ and DR2full produce very consistent posteriors, a $\sim 0.2\sigma$ difference can be found between DR2new+ and DR2new. The impact of the InPTA data is more noticeable in the shorter data set.
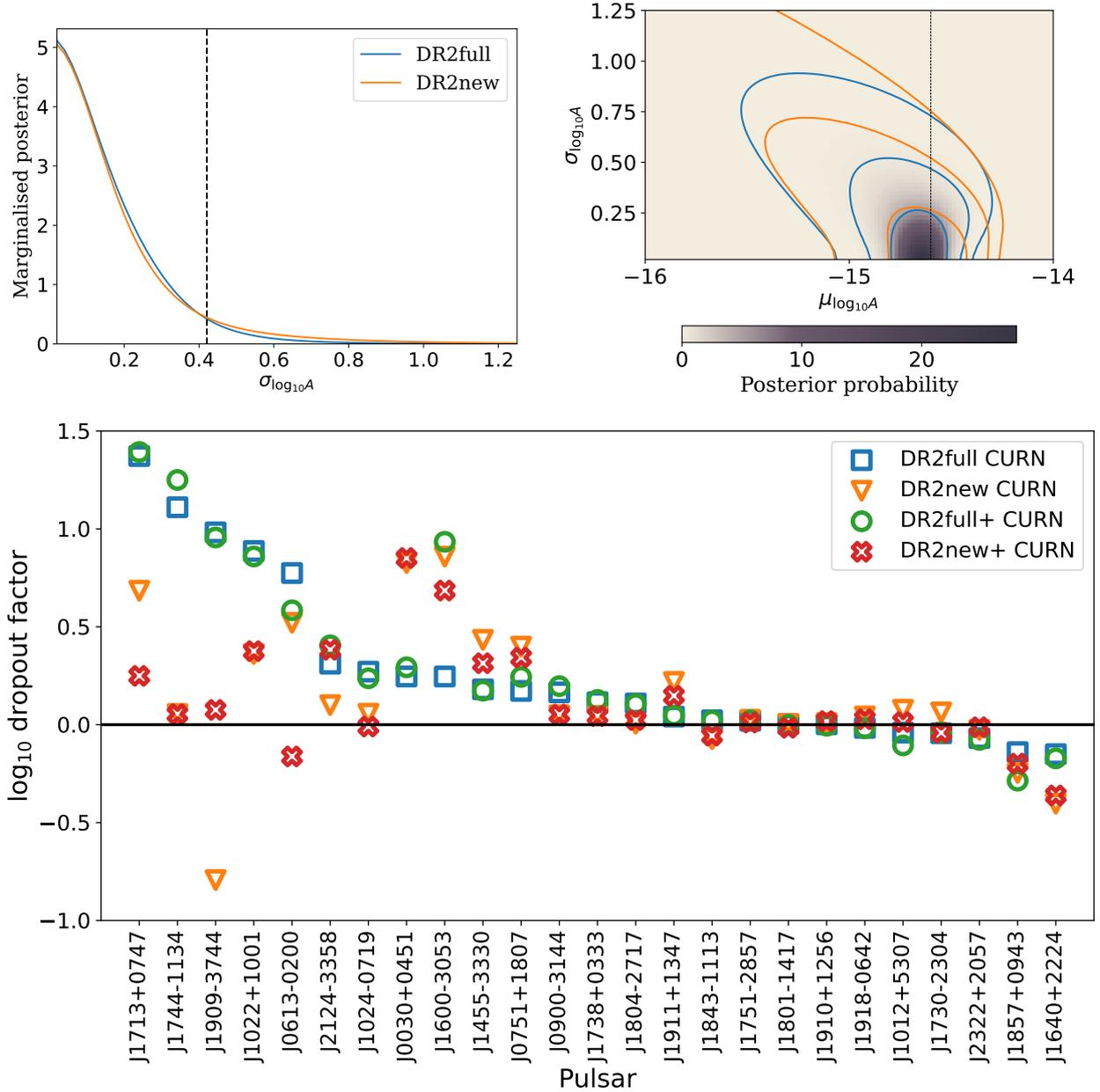
Fig. 11. Tests of the CURN model. The top two panels show that pulsar spectra of temporal correlations attributed to CURN are indeed consistent with representing the same spectrum (Goncharov et al. 2022). The blue lines show the measurement for `DR2full`, whereas the orange lines show the measurement for `DR2new`. The top left panel is the measurement of the standard deviation of the CURN amplitudes across pulsar spectra, $\sigma_{\log_{10} A}$, marginalised over the mean of the CURN amplitudes, $\mu_{\log_{10} A}$, as well as pulsar-intrinsic noise parameters. The dashed vertical line denotes an upper limit at 95% credibility. The top right panel shows both the mean and the standard deviation. The inferred mean value is consistent with the measurement of $\log_{10} A_{\text{CURN}}$ denoted by a dashed vertical line. The bottom panel shows the logarithm of the dropout factors (Arzoumanian et al. 2020) which suggest pulsars' contribution to the CURN model. Positive values represent support for the CURN model, whereas negative values point towards inconsistency of pulsar data with the CURN model. The differences in the dropout factors between the data sets could be due to differences in the recovered CURN or the pulsar red noise spectral properties.

The results for the OS shown in Table 4 are consistent with the EPTA-only data sets. Both `DR2full+` and `DR2full` give similar amplitudes and S/Ns for the three correlation models. An increase in the S/Ns of about 0.5 for monopole, dipole, and HD correlations can be found with the `DR2new+` against the S/Ns in the `DR2new`.

The corresponding BFs are in general agreement with the EPTA-only results (cf. Table 5). The BF between GWB and CURN in the `DR2new+` of 65 is comparable to the 60 from `DR2new`. However, when testing for additional processes, we find significantly larger BFs for PSRN+GWB+CLK (row ID 8) and PSRN+GWB+EPH (row ID 9): 57 versus 28 and 43 versus 33, `DR2new+` versus `DR2new`, respectively. The small BF difference between the models including the CLK or EPH error terms and the GWB-only model further supports the assumption that the signal could be a GWB.

Finally, we investigate the effect of the additional InPTA on the contribution of the individual pulsars to the CURN via the dropout analysis in Fig. 11. As with the previous results, most dropout factors are consistent between the `DR2full+`
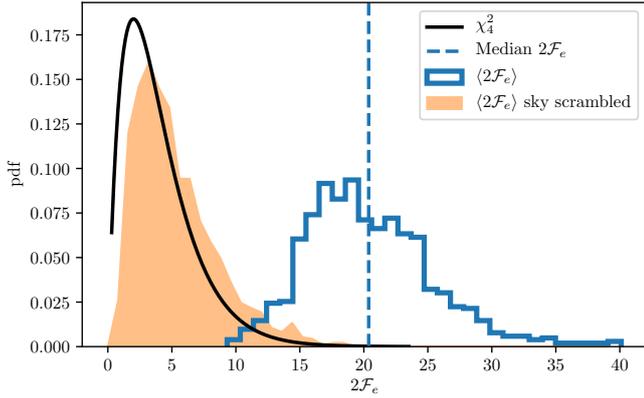
Fig. 12. F_e-statistic for a CGW search marginalised over the noise uncertainties (blue). The F_e-statistic obtained from the analysis of the data with scrambled sky position of pulsars is shown in orange and compared against the theoretical (χ_4^2) distribution in black.
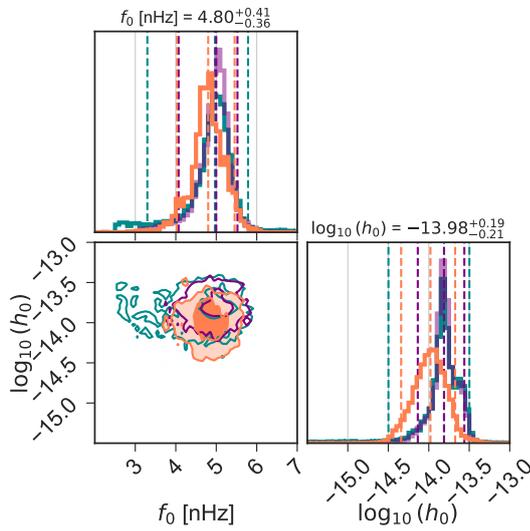
Fig. 13. Inference of the frequency and amplitude of a putative CGW in the CGW+CURN model. Plotted are the 50% and 90% credible regions. Dark-cyan contours are obtained using the Earth term only and ENTERPRISE with PTMCMCSAMPLER, purple contours with Eryn, while coral contours are produced using a sampler described in Bécsy et al. (2022) and also included the pulsar term.

and DR2full data sets. J1600−3053 shows an increase in the dropout factor, possibly due to better single pulsar noise modelling with the addition of the InPTA data. For the new generation EPTA-only and EPTA+InPTA data sets, the differences are more pronounced. Most pulsars have dropout factors that are in agreement. Two pulsars, J1713+0747 and J0613−0200, have reduced dropout factors, whereas J1909−3744 shows an increase.

In summary, the results from the EPTA and InPTA combination are in broad agreement with the EPTA-only data set. The consistency between DR2full+ and DR2full shows the robustness of the results from the full EPTA+InPTA combination. When comparing DR2new+ and DR2new the effects of the InPTA data are more visible with differences in the power law posteriors, increased BFs, and higher OS S/Ns for the GWB, but also for other possible noises. Further investigation is needed to assess and improve the combined sensitivity for GW searches.

6. Discussions and conclusions

The EPTA with its six telescopes and multiple observing systems has collected PTA observations for almost 25 yr and used a new generation of observing systems for the last decade. For the DR2 we have increased the number of pulsars combined with the most recent observations from six to 25. A selection scheme has been applied to find the 25 pulsars that are sufficient to contribute about 95% of the expected sensitivity of the full array with 42 pulsars from DR1 (Speri et al. 2023). Here, we used the optimal timing and noise models obtained in EPTA Collaboration & InPTA Collaboration (2023), EPTA Collaboration (2023) to search for a stochastic GWB. In addition, we combined data for ten common pulsars between InPTA DR1 (Tarafdar et al. 2022) and EPTA DR2 and conducted GWB searches also on those extended EPTA+InPTA data sets.

We present the main results of the GWB search using two versions of the EPTA 25-pulsar data set, the full data set (DR2full) with a time span of 24.7 yr and a new backends-only data set (DR2new) with a time span of 10.3 yr. Both data sets measure a common red signal. By virtue of its length, the full data set yields a better constraint on the spectral properties of the GWB. However, the new backends-only data set provides a better measurement of the cross-correlated power. In the following, we give a summary of the results of our analysis and discuss possible reasons for the discrepancies between the two data sets.

The power law amplitude for a HD-correlated process using DR2full is $\log_{10} A = -14.54^{+0.28}_{-0.41}$, with a spectral index $\gamma = 4.19^{+0.73}_{-0.63}$ that is close to the expected value of 13/3 for a GWB from circular SMBHBs driven by GW emission. With DR2new we obtain a flatter spectrum with $\log_{10} A = -13.94^{+0.23}_{-0.48}$ and $\gamma = 2.71^{+1.18}_{-0.73}$. Fixing the spectral index to 13/3, the amplitudes for the two data sets are consistent, with values around $\log_{10} A = -14.61^{+0.11}_{-0.15}$. This indicates that the two data sets constrain the same amount of power in the GWB, although the detailed spectral shape appears to be different.

The free spectrum analysis provides the possibility to directly compare the results from different data sets in the frequency domain. Ignoring the lowest $1/24.7 \, \mathrm{yr}^{-1}$ frequency bin of DR2full, the remaining bins show good consistency with those of DR2new. Including the lowest frequency bin in the power law fitting may lead to a steepening of the power law. With a shorter data span DR2new probes a different frequency range starting at about two times higher than $1/24.7 \, \mathrm{yr}^{-1}$. Power law fitting could also be affected by the frequency bins just above $10^{-8} \, \mathrm{Hz}$, resulting in a flatter spectrum for DR2new. This could either indicate that a power law model does not provide a good fit to the common signal, or there is additional noise or signal around $10^{-8} \, \mathrm{Hz}$ or $10^{-9} \, \mathrm{Hz}$.

The differences between the Bayesian correlation curves observed in the two data sets are most obvious around angular separations of 60, 80, and 135 degrees. The correlation curve produced by DR2full shows a prominent monopolar-like structure, with a central part shifted upward compared to the HD curve, while the correlation curve produced by DR2new follows the HD correlation much more closely. These results are in agreement with the measured BFs (PSRN+GWB vs. PSRN+CURN), which are four and 60 for DR2full and DR2new, respectively. From the null hypothesis distributions of BFs (PSRN+GWB vs. PSRN+CURN) constructed with phase shifts, we can infer a p-value of 0.04 for DR2full and 0.001 for DR2new. There is essentially no evidence for other correlation patterns or additional common processes in either data set, with
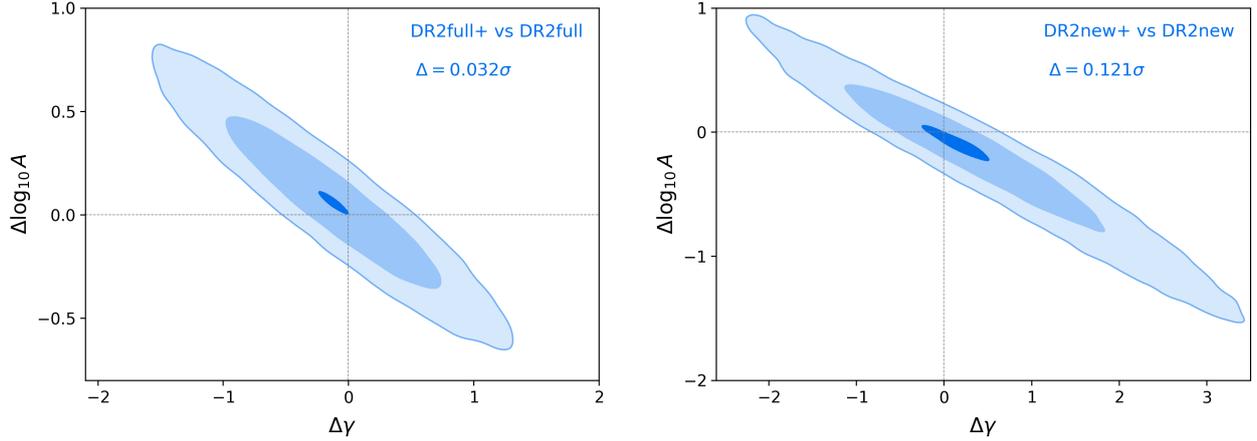
Fig. 14. Difference distributions of posteriors while including the InPTA data. The left panel is associated with posteriors from DR2full+ and DR2full. In contrast, the right panel involves the comparison between DR2new+ and DR2new. Both panels show the tension for the GWB model. These plots provide three contours: 1σ, 2σ, and the Δ contours that represent the computed tension value.

Table 7. 90% credible regions for the constraints of power law parameters in the different Bayesian analyses with DE440 for both DR2full+ and DR2new+, which include the addition of InPTA data.

Software + Model	DR2full+		DR2new+	
	$\log_{10} A_{\mathrm{CRS}}$	γ_{CRS}	$\log_{10} A_{\mathrm{CRS}}$	γ_{CRS}
ENTERPRISE + CURN	$-14.44^{+0.17}_{-0.20}$	$3.98^{+0.40}_{-0.37}$	$-14.30^{+0.33}_{-0.52}$	$3.53^{+1.12}_{-0.84}$
ENTERPRISE + GWB	$-14.48^{+0.18}_{-0.20}$	$4.06^{+0.39}_{-0.40}$	$-14.10^{+0.25}_{-0.44}$	$3.03^{+1.02}_{-0.67}$

perhaps the exception of a tentative hint of an extra monopole in the DR2full data set, which will be the subject of further studies.

Optimal statistic analysis shows even more significant discrepancies between the two data sets. For DR2full the squared amplitude is only 2.7×10^{-30}, which is lower than the amplitude from the Bayesian analysis ($A^2 = 6.02 \times 10^{-30}$). There is a large scatter in the correlation coefficients, giving a similar S/N of around 1.2 for both the HD and the monopole correlations. The p-value for the S/N of the HD correlation is 0.07 from phase shifts and 0.08 from sky scrambles. For DR2new the squared amplitude is 1.0×10^{-29}, which is more consistent with (yet higher than) the Bayesian result. The correlation coefficients also match the HD curve much better. The S/N for the HD correlation increases to 3.5, while the S/N for the monopole correlation drops to almost zero. Sky scrambles give a p-value of 0.002 for HD correlation, while phase shifts yield a p-value < 0.0001. This corresponds to >3σ significance of the GWB signal.

Preliminary analysis including a CGW suggests that its contribution to the observed HD-correlated power cannot be ruled out. The presence of a CGW is not supported in the DR2full data set; its presence is preferred over CURN only in DR2new. The source amplitude and frequency are well-constrained. The candidate is also localised in the sky, but its position and error region depends on whether we include the pulsar term. However, adding a GWB to the analysis absorbs most of the power of the CGW, preventing any strong claim about its actual presence in the data. A more thorough analysis involving the CGW model is presented in Antoniadis et al. (2023b).

The analysis of the combined EPTA DR2 and InPTA DR1 shows broadly consistent results with the EPTA DR2 alone. The power law parameter constraints with DR2full+ show little dif-

Table 8. Tensiometer package based discrepancies between GWB posteriors that arise from DR2full+, DR2full, DR2new+, and DR2new with entries providing the Z-score (in number of σ).

	DR2full+ vs. DR2full	DR2new+ vs. DR2new
CURN	0.037	0.229
GWB	0.032	0.121

ference to those without the InPTA data. For DR2new+, the effect of the additional InPTA data is more pronounced. The power law parameters experience a small shift of 0.17σ towards a steeper spectral index. The BFs and OS S/Ns are also in general agreement with the EPTA-only data sets. Increases in the evidence for additional monopole and dipole correlated signals of $S/N \sim 0.5$ can be found in DR2new+. A larger impact on the shorter data set can be expected, since the InPTA, with three years of time span, is a more significant fraction of DR2new (10.3 yr) compared to DR2full (24.7 yr).

With the high amplitude and large uncertainty in the spectral index, the observed HD correlated signal is broadly consistent with the expectation from a cosmic population of SMBHBs. In particular, as shown by Middleton et al. (2021), the high amplitude of $\log A = -14.61$ inferred when fixing $\gamma = 13/3$ is consistent with the recent discovery of over-massive black holes (e.g., McConnell et al. 2011) and the upward revision of the normalisation of the SMBH-host galaxy relations (see, e.g., McConnell & Ma 2013; Kormendy & Ho 2013). It is not straightforward, however, to construct a self-consistent SMBH and host galaxy cosmic evolution model that results in such a high GWB signal, fulfilling other observational constraints on the SMBH mass function and on the evolution

of the bolometric quasar luminosity function with redshift (Izquierdo-Villalba et al. 2022). A spectrum significantly flatter than $\gamma = 13/3$ can arise for a number of different reasons, including strong coupling with the environment, the predominance of highly eccentric SMBHBs (see e.g., Sesana 2013), or simply by the presence of extra power at high frequencies due to sparse and loud marginally resolvable individual binaries (Middleton et al. 2021). Besides a cosmic population of SMBHBs, the detected signal can be generated by processes occurring in the early Universe (Caprini & Figueroa 2018) as well as specific models of dark matter (Porayko et al. 2018). We plan to investigate the implications of this signal for all these formation scenarios in follow-up papers (Antoniadis et al. 2023a; Smarra et al. 2023).

Our results seem to indicate that `DR2new` provides better constraints on the cross-correlated power than `DR2full`. It would normally be expected that the addition of more data would lead to a more significant detection of a stationary process. There are a few possible factors that could be contributing to this discrepancy, that needs to be investigated in more detail:

Lower quality of the early data, which lacks multi-radio frequency coverage and polarisation calibration, may have allowed for residual unmodelled noise. This can lead to different noises and signals being recovered with the early and new generation observations. Investigating better noise modelling can help to increase the sensitivity and reliability of the early data.

Improper weights for the power law fitting in different frequency bins could introduce bias in the recovery of the spectral properties. In particular, the lowest frequency bins only have contributions from a few pulsars with the longest time spans, but their weights are the highest since the largest amount of power in a common red signal is at the lowest frequencies. Considering a weighting scheme for the different frequency components for the power law fitting could produce an unbiased result.

The presence of excess power at low frequencies can lead to a steepening of the power law. While excess noise at high frequencies can make the spectrum appear shallower. In both cases, noise leaks into the GW signal giving erroneous power law constraints.

Non-stationarity of the pulsar noise or of the putative GW signal can cause the measured spectral properties to be different between the early and late part of a data set, as the properties of the noise and signal could evolve over time.

Some of these differences between pulsars are smoothed out in the `DR2new` data set, as all pulsars have roughly the same time span ~10 yr. This may help to measure the cross-correlations between pulsar pairs more robustly.

An extensive simulation campaign is ongoing to help to better understand the features of our data sets and to build more confidence in the internal consistency of our findings. Verification of the analysis algorithms and their performance on a realistic PTA data set is needed to set our expectations. These can then be compared against the real data set to test the effects that data quality and the noise or signal properties have on the results of the final analysis. Several simulation projects tackling different questions will be published in separate works.

Concurrent efforts from the NANOGrav collaboration on their 15-year data set (Agazie et al. 2023) and the PPTA on their DR3 (Reardon et al. 2023) provide independent results on the search for a gravitational wave background. These will be compared in the IPTA framework to increase our confidence and prepare for the next IPTA data set. Additionally, the CPTA is preparing its first data release and analysis for a GWB signal (Xu et al. 2023).

Moving forward, we plan to add more pulsars timed with the new backends. The EPTA data set is unique in its combination of time span, cadence, number of pulsars, and, when combined with the InPTA data set, in DM monitoring. Indeed, among the pulsars timed by EPTA, there are more than 30 sources observed with new backends for a time span >8 yr displaying RMS < 2 μs, which can add significant value to the EPTA data set. A combination of the resulting data set with NANOGrav 15 yr, PPTA DR3 and MeerKAT DR1 under the aegis of the third data release (DR3) of the IPTA, will produce a data set of unprecedented sensitivity that will help to pin down the nature of the signal presented here and to constrain its properties.

References

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016, Phys. Rev. Lett., 116, 061102

Agazie, G., Anumarlapudi, A., Archibald, A. M., et al. 2023, ApJ, 951, L8

Allen, B., & Romano, J. D. 2023, Phys. Rev. D, 108, 043026

Anholm, M., Ballmer, S., Creighton, J. D. E., Price, L. R., & Siemens, X. 2009, Phys. Rev. D, 79, 084030

Antoniadis, J., Arzoumanian, Z., Babak, S., et al. 2022, MNRAS, 510, 4873

Antoniadis, J., Arumugam, P., Arumugam, S., et al. 2023a, A&A, submitted [arXiv:2306.16227]

Antoniadis, J., Arumugam, P., Arumugam, S., et al. 2023b, ArXiv e-prints [arXiv:2306.16226]

Arzoumanian, Z., Brazier, A., Burke-Spolaor, S., et al. 2015, ApJ, 813, 65

Arzoumanian, Z., Baker, P. T., Brazier, A., et al. 2018, ApJ, 859, 47

Arzoumanian, Z., Baker, P. T., Blumer, H., et al. 2020, ApJ, 905, L34

Babak, S., & Sesana, A. 2012, Phys. Rev. D, 85, 044034

Bailes, M., Jameson, A., Abbate, F., et al. 2020, PASA, 37, e028

Bassa, C. G., Janssen, G. H., Karuppusamy, R., et al. 2016, MNRAS, 456, 2196

Bécsy, B., Cornish, N. J., & Digman, M. C. 2022, Phys. Rev. D, 105, 122003

Begelman, M. C., Blandford, R. D., & Rees, M. J. 1980, Nature, 287, 307

Caballero, R. N., Lee, K. J., Lentati, L., et al. 2016, MNRAS, 457, 4421

Caprini, C., & Figueroa, D. G. 2018, CQG, 35, 163001

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

Carlin, B. P., & Chib, S. 1995, J. R. Stat. Soc. Ser. B-Methodol., 57, 473

Chalumeau, A., Babak, S., Petiteau, A., et al. 2022, MNRAS, 509, 5538

Chamberlin, S. J., Creighton, J. D. E., Siemens, X., et al. 2015, Phys. Rev. D, 91, 044048

Chen, S., Caballero, R. N., Guo, Y. J., et al. 2021, MNRAS, 508, 4970

Cognard, I., Theureau, G., Guillemot, L., et al. 2013, in SF2A-2013: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics, eds. L. Cambresy, F. Martins, E. Nuss, & A. Palacios, 327

Coles, W. A., Kerr, M., Shannon, R. M., et al. 2015, ApJ, 808, 113

Cornish, N. J., & Sampson, L. 2016, Phys. Rev. D, 93, 104047

Damour, T., & Vilenkin, A. 2000, Phys. Rev. Lett., 85, 3761

Demorest, P. B., Ferdman, R. D., Gonzalez, M. E., et al. 2013, ApJ, 762, 94

Desvignes, G., Caballero, R. N., Lentati, L., et al. 2016, MNRAS, 458, 3341

Detweiler, S. 1979, ApJ, 234, 1100

Di Marco, V., Zic, A., Miles, M. T., et al. 2023, ApJ, submitted [arXiv:2305.04464]

Ellis, J., & van Haasteren, R. 2017, https://zenodo.org/record/1037579

Ellis, J. A., Siemens, X., & Creighton, J. D. E. 2012, ApJ, 756, 175

Ellis, J. A., Vallisneri, M., Taylor, S. R., & Baker, P. T. 2020, Astrophysics Source Code Library [record ascl:1912.015]

EPTA Collaboration (Antoniadis, J., et al.) 2023, A&A, 678, A48 (Paper I)

EPTA Collaboration, & InPTA Collaboration (Antoniadis, J., et al.) 2023, A&A, 678, A49 (Paper II)

Falxa, M., Babak, S., & Le Jeune, M. 2023, Phys. Rev. D, 107, 022008

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

Gair, J., Romano, J. D., Taylor, S., & Mingarelli, C. M. F. 2014, Phys. Rev. D, 90, 082001

Goncharov, B., Reardon, D. J., Shannon, R. M., et al. 2021a, MNRAS, 502, 478

Goncharov, B., Shannon, R. M., Reardon, D. J., et al. 2021b, ApJ, 917, L19

Goncharov, B., Thrane, E., Shannon, R. M., et al. 2022, ApJ, 932, L22

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

Guzzetti, M. C., Bartolo, N., Liguori, M., & Matarrese, S. 2016, Nuovo Cimento Rivista Serie, 39, 399

Hazboun, J. S., Meyers, P. M., Romano, J. D., Siemens, X., & Archibald, A. M. 2023, ArXiv e-prints [arXiv:2305.01116]

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, CQG, 30, 224007

Hourihane, S., Meyers, P., Johnson, A., Chatziioannou, K., & Vallisneri, M. 2023, Phys. Rev. D, 107, 084045

Izquierdo-Villalba, D., Sesana, A., Bonoli, S., & Colpi, M. 2022, MNRAS, 509, 3488

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

Jenet, F., Finn, L. S., Lazio, J., et al. 2009, ArXiv e-prints [arXiv:0909.1058]

Jiang, P., Yue, Y., Gan, H., et al. 2019, Sci. China Phys. Mech. Astron., 62, 959502

Joshi, B. C., Arumugasamy, P., Bagchi, M., et al. 2018, J. Astrophys. Astron., 39, 51

Karnesis, N., Katz, M. L., Korsakova, N., Gair, J. R., & Stergioulas, N. 2023, ArXiv e-prints [arXiv:2303.02164]

Kormendy, J., & Ho, L. C. 2013, ARA&A, 51, 511

Kramer, M., & Champion, D. J. 2013, CQG, 30, 224009

Lam, M. T., Ellis, J. A., Grillo, G., et al. 2018, ApJ, 861, 132

Lentati, L., Alexander, P., Hobson, M. P., et al. 2013, Phys. Rev. D, 87, 104021

Lentati, L., Taylor, S. R., Mingarelli, C. M. F., et al. 2015, MNRAS, 453, 2576

Lorimer, D. R., & Kramer, M. 2004, Handbook of Pulsar Astronomy (UK: Cambridge University Press), 4

Magorrian, J., Tremaine, S., Richstone, D., et al. 1998, AJ, 115, 2285

Manchester, R. N. 2006, Chin. J. Astron. Astrophys. Supp., 6, 139

McConnell, N. J., & Ma, C.-P. 2013, ApJ, 764, 184

McConnell, N. J., Ma, C.-P., Gebhardt, K., et al. 2011, Nature, 480, 215

Middleton, H., Sesana, A., Chen, S., et al. 2021, MNRAS, 502, L99

Park, R. S., Folkner, W. M., Williams, J. G., & Boggs, D. H. 2021, AJ, 161, 105

Porayko, N. K., Zhu, X., Levin, Y., et al. 2018, Phys. Rev. D, 98, 102002

Raveri, M., & Doux, C. 2021, Phy. Rev. D, 104, 043504

Raveri, M., & Hu, W. 2019, Phys. Rev. D, 99, 043506

Reardon, D. J., Zic, A., Shannon, R. M., et al. 2023, ApJ, 951, L6

Sazhin, M. V. 1978, Soviet Astron., 22, 36

Sesana, A. 2013, CQG, 30, 224014

Sesana, A., Vecchio, A., & Colacino, C. N. 2008, MNRAS, 390, 192

Smarra, C., Goncharov, B., Barausse, E., et al. 2023, ArXiv e-prints [arXiv:2306.16228]

Speri, L., Porayko, N. K., Falxa, M., et al. 2023, MNRAS, 518, 1802

Tarafdar, P., Nobleson, K., Rana, P., et al. 2022, PASA, 39

Taylor, S. R., & Gair, J. R. 2013, Phys. Rev. D, 88, 084001

Taylor, J. H., & Weisberg, J. M. 1989, ApJ, 345, 434

Taylor, S. R., Lentati, L., Babak, S., et al. 2017, Phys. Rev. D, 95, 042002

Taylor, S. R., Baker, P. T., Hazboun, J. S., Simon, J., & Vigeland, S. J. 2021, enterprise_extensions (GitLab) https://github.com/nanograv/enterprise_extensions

Tiburzi, C., Hobbs, G., Kerr, M., et al. 2016, MNRAS, 455, 4339

Vallisneri, M., Taylor, S. R., Simon, J., 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., Levin, Y., McDonald, P., & Lu, T. 2009, MNRAS, 395, 1005

Verbiest, J. P. W., Bailes, M., Coles, W. A., et al. 2009, MNRAS, 400, 951

Vigeland, S. J., Islo, K., Taylor, S. R., & Ellis, J. A. 2018, Phys. Rev. D, 98, 044003

Wang, F., Yang, J., Fan, X., et al. 2019, ApJ, 884, 30

Wang, F., Yang, J., Fan, X., et al. 2021, ApJ, 907, L1

White, S. D. M., & Rees, M. J. 1978, MNRAS, 183, 341

Williams, M. J., Veitch, J., & Messenger, C. 2021, Phys. Rev. D, 103, 103006

Xu, H., Chen, S., Guo, Y., et al. 2023, Res. Astrono. Astrophys., 23, 075024

Zhu, W. W., Stairs, I. H., Demorest, P. B., et al. 2015, ApJ, 809, 41

Zic, A., Hobbs, G., Shannon, R. M., et al. 2022, MNRAS, 516, 410

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

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

[3] Department of Physics, Indian Institute of Technology Roorkee, Roorkee 247667, India

[4] Department of Electrical Engineering, IIT Hyderabad, Kandi, Telangana 502284, India

[5] Université Paris Cité, CNRS, Astroparticule et Cosmologie, 75013 Paris, France

[6] The Institute of Mathematical Sciences, C. I. T. Campus, Taramani Chennai 600113, India

[7] Homi Bhabha National Institute, Training School Complex, Anushakti Nagar, Mumbai 400094, India

[8] Fakultät für Physik, Universität Bielefeld, Postfach 100131, 33501 Bielefeld, Germany

[9] ASTRON, Netherlands Institute for Radio Astronomy, Oude Hoogeveensedijk 4, 7991 PD Dwingeloo, The Netherlands

[10] Department of Physical Sciences, Indian Institute of Science Education and Research, Mohali, Punjab 140306, India

[11] Laboratoire de Physique et Chimie de l'Environnement et de l'Espace, Université d'Orléans/CNRS, 45071 Orléans Cedex 02, France

[12] Observatoire Radioastronomique de Nançay, Observatoire de Paris, Université PSL, Université d'Orléans, CNRS, 18330 Nançay, France

[13] Dipartimento di Fisica "G. Occhialini", Universitá degli Studi di Milano-Bicocca, Piazza della Scienza 3, 20126 Milano, Italy

[14] INFN, Sezione di Milano-Bicocca, Piazza della Scienza 3, 20126 Milano, Italy

[15] INAF – Osservatorio Astronomico di Brera, via Brera 20, 20121 Milano, Italy

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

[17] INAF – Osservatorio Astronomico di Cagliari, via della Scienza 5, 09047 Selargius (CA), Italy

[18] Hellenic Open University, School of Science and Technology, 26335 Patras, Greece

[19] Kavli Institute for Astronomy and Astrophysics, Peking University, Beijing 100871, PR China

[20] Department of Astronomy and Astrophysics, Tata Institute of Fundamental Research, Homi Bhabha Road, Navy Nagar, Colaba, Mumbai 400005, India

[21] Department of Physics, IIT Hyderabad, Kandi, Telangana 502284, India

[22] Department of Physics and Astrophysics, University of Delhi, Delhi 110007, India

[23] Department of Earth and Space Sciences, Indian Institute of Space Science and Technology, Valiamala, Thiruvananthapuram, Kerala 695547, India

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

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

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

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

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

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

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

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

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

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

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

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

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

[37] Department of Astronomy, School of Physics, Peking University, Beijing 100871, PR China

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Appendix A: Chebyshev and Legendre decomposition for the ORF

In this Appendix, we present the spatial correlation measured among pulsar pairs as reconstructed from: i) a third order Chebishev polynomial decomposition and ii) a fifth order Legendre polynomial decomposition. Results are shown in Figures A.1 and A.2, respectively, which highlight the broad consistency with the average HD correlation expected for a GWB and with the binned Bayesian reconstruction shown in Fig. 5 and discussed in Sect. 4.2.
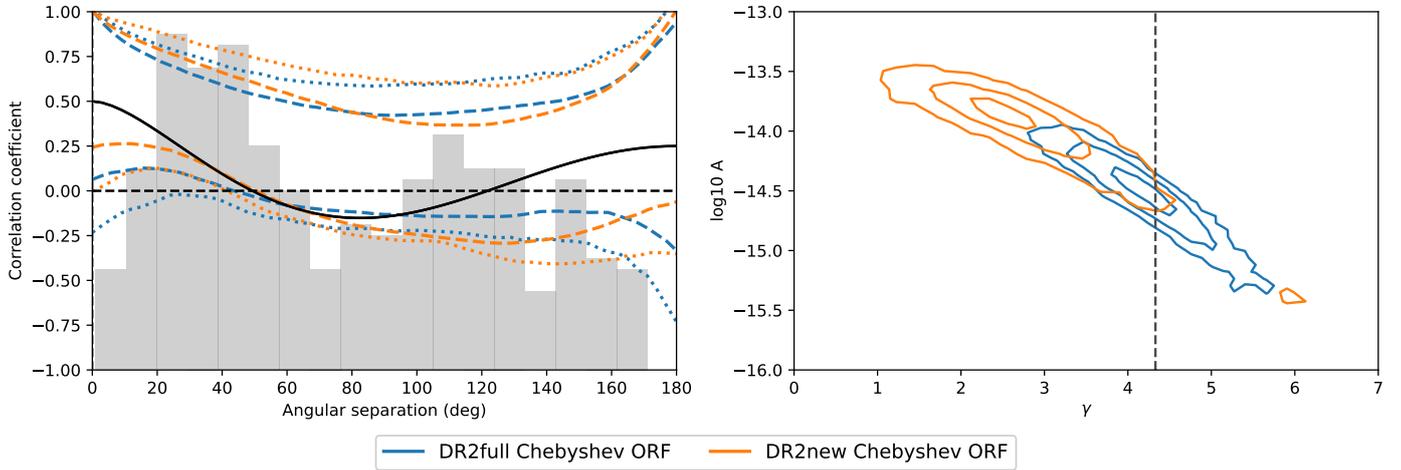

Fig. A.1. Overlap reduction function reconstructed using Chebyshev polynomial. The figure is in the style of Fig. 5, with the difference in the left panel being that the dashed and dotted lines indicate the central 95 and 99.7 % credible regions of the reconstructed spatial correlations.
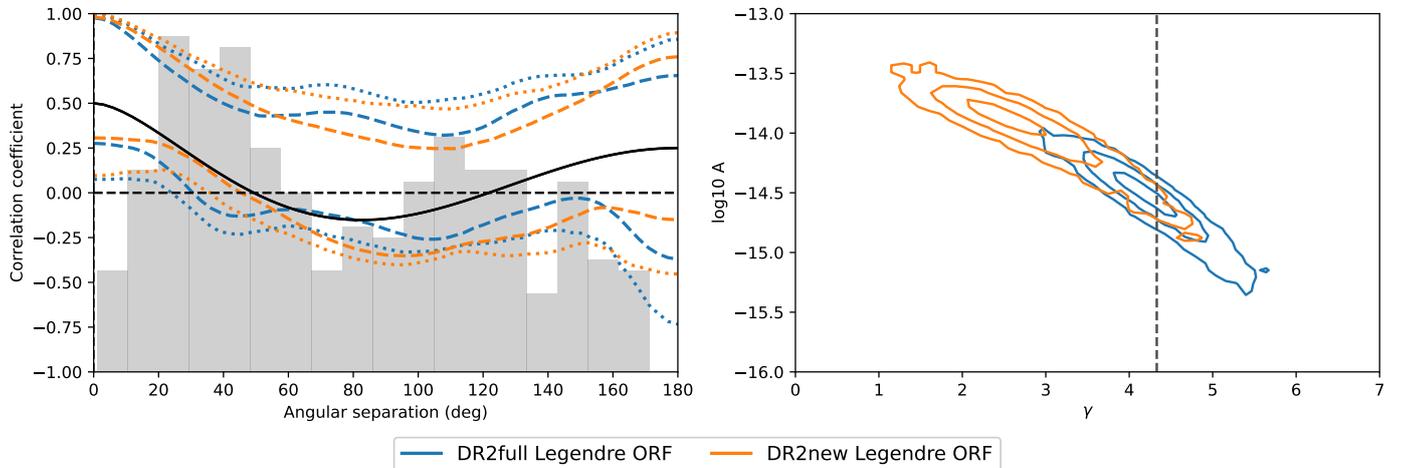

Fig. A.2. Overlap reduction function reconstructed using Legendre polynomial. The figure is in the style of Fig. 5, with the difference in the left panel being that the dashed and dotted lines indicate the central 95 and 99.7 % credible regions of the reconstructed spatial correlations.

Appendix B: Comparison of the four data sets

In this Appendix, we present the constraints for the free spectrum, power law and binned correlations in Figs B.1 and B.2 for the four data sets used in this work: `DR2full`, `DR2new`, `DR2full+` and `DR2new+`.
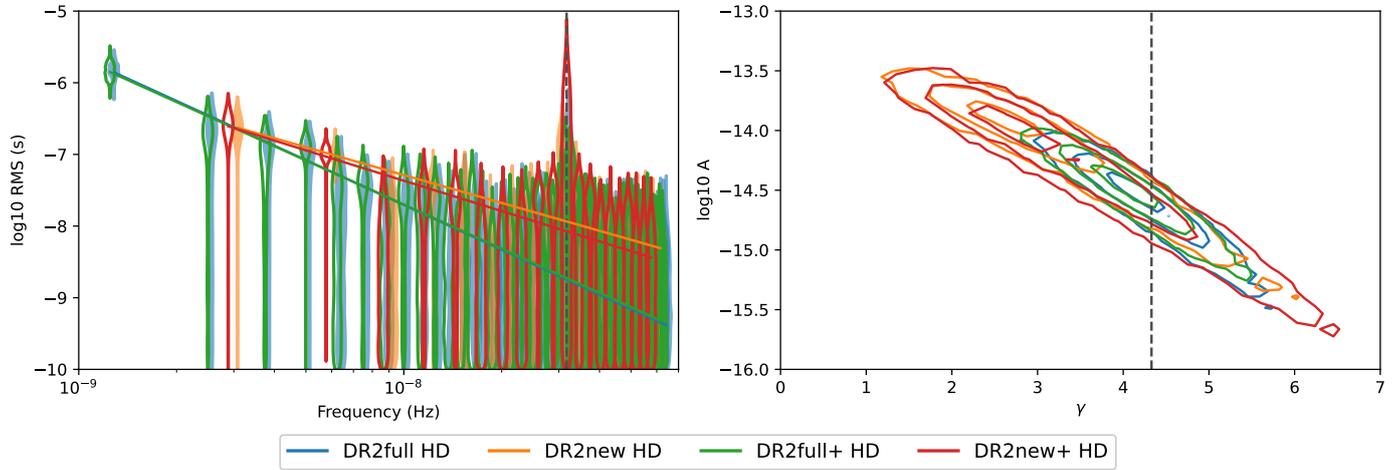

Fig. B.1. Spectral properties of a CRS signal assuming HD correlation in the style of Fig. 1 for all four data sets.
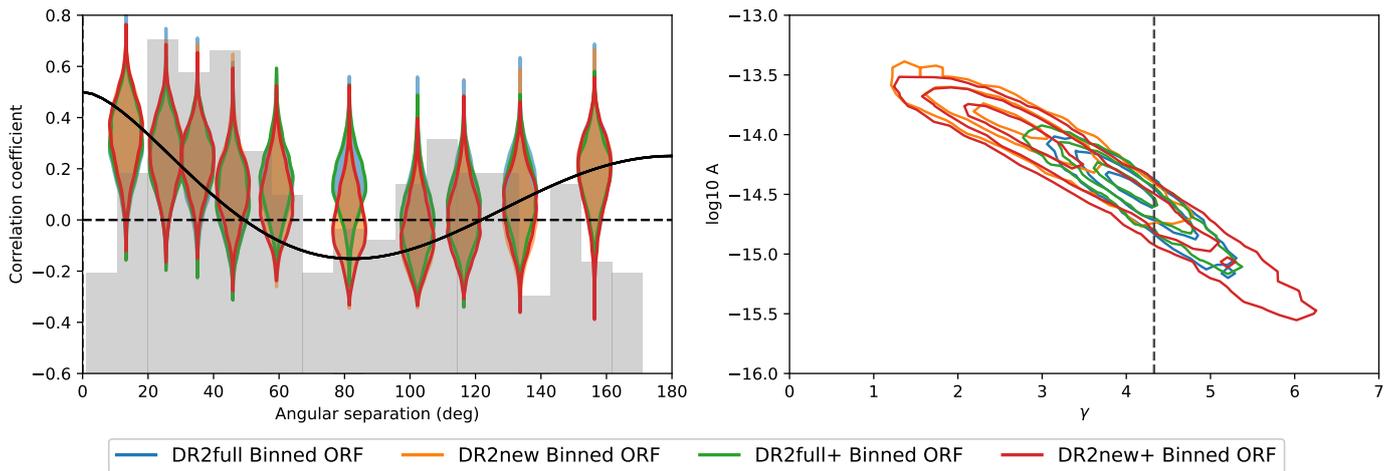
Fig. B.2. Binned overlap reduction function in the style of Fig. 5 for all four data sets.