The noise properties of 42 millisecond pulsars from the European Pulsar Timing Array and their impact on gravitational-wave searches

R. N. Caballero, K. J. Lee, L. Lentati, Grégory Desvignes, D. J. Champion, J. P. W. Verbiest, G. H. Janssen, B. W. Stappers, M. Kramer, P. Lazarus, et al.

MNRAS **000**, 1–?? (2016) Preprint 21 January 2016 Compiled using MNRAS LATEX style file v3.0

The noise properties of 42 millisecond pulsars from the European Pulsar Timing Array and their impact on gravitational wave searches

R. N. Caballero,[1⋆] K. J. Lee[2,1], L. Lentati[3], G. Desvignes[1], D. J. Champion[1], J. P. W. Verbiest[4,1], G. H. Janssen[5,6], B. W. Stappers[6], M. Kramer[1,6], P. Lazarus[1], A. Possenti[7], C. Tiburzi[4,1], D. Perrodin[7], S. Osłowski[4,1], S. Babak[8], C. G. Bassa[5], P. Brem[8], M. Burgay[7], I. Cognard[9,10], J. R. Gair[11], E. Graikou[1], L. Guillemot[9,10], J. W. T. Hessels[5,12], R. Karuppusamy[1], A. Lassus[1], K. Liu[1], J. McKee[6], C. M. F. Mingarelli[13,1], A. Petiteau[14], M. B. Purver[6], P. A. Rosado[15,16], S. Sanidas[12,6], A. Sesana[17,8], G. Shaifullah[4,1], R. Smits[5], S. R. Taylor[18], G. Theureau[9,10], R. van Haasteren[18] and A. Vecchio[17]

[1] *Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, 53121 Bonn, Germany*
[2] *Kavli institute for Astronomy and Astrophysics, Peking University, Beijing 100871,P.R.China*
[3] *Astrophysics Group, Cavendish Laboratory, JJ Thomson Avenue, Cambridge, CB3 0HE, UK*
[4] *Fakultät für Physik, Universität Bielefeld, Postfach 100131, 33501 Bielefeld, Germany*
[5] *ASTRON, the Netherlands Institute for Radio Astronomy, Postbus 2, 7990 AA, Dwingeloo, The Netherlands*
[6] *Jodrell Bank Centre for Astrophysics, School of Physics and Astronomy, The University of Manchester, Manchester M13 9PL,UK*
[7] *INAF - Osservatorio Astronomico di Cagliari, via della Scienza 5, I-09047 Selargius (CA), Italy*
[8] *Max-Planck-Institut für Gravitationsphysik, Albert Einstein Institut, Am Mühlenberg 1, 14476 Golm, Germany*
[9] *Laboratoire de Physique et Chimie de l'Environnement et de l'Espace LPC2E CNRS-Université d'Orléans, F-45071, Orléans, France*
[10] *Station de radioastronomie de Nançay, Observatoire de Paris, CNRS/INSU F-18330 Nançay, France*
[11] *School of Mathematics, University of Edinburgh, King's Buildings, Edinburgh EH9 3JZ, UK*
[12] *Anton Pannekoek Institute for Astronomy, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands*
[13] *TAPIR, California Institute of Technology MC 350-17, Pasadena, California 91125, USA*
[14] *Université Paris-Diderot-Paris7 APC - UFR de Physique, Batiment Condorcet, 10 rue Alice Domont et Léonie Duquet 75205 PARIS CEDEX 13, France*
[15] *Centre for Astrophysics & Supercomputing, Swinburne University of Technology, PO Box 218, Hawthorn VIC 3122, Australia*
[16] *Max-Planck-Institut für Gravitationsphysik, Albert-Einstein-Institut, Callinstraße 38, 30167, Hanover, Germany*
[17] *School of Physics and Astronomy, The University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK*
[18] *Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91106, USA*
[19] *Laboratoire Univers et Théories LUTh, Observatoire de Paris, CNRS/INSU, Université Paris Diderot, 5 place Jules Janssen, 92190 Meudon, France*

21 January 2016

ABSTRACT

The sensitivity of Pulsar Timing Arrays to gravitational waves depends on the noise present in the individual pulsar timing data. Noise may be either intrinsic or extrinsic to the pulsar. Intrinsic sources of noise will include rotational instabilities, for example. Extrinsic sources of noise include contributions from physical processes which are not sufficiently well modelled, for example, dispersion and scattering effects, analysis errors and instrumental instabilities. We present the results from a noise analysis for 42 millisecond pulsars (MSPs) observed with the European Pulsar Timing Array. For characterising the low-frequency, stochastic and achromatic noise component, or "timing noise", we employ two methods, based on Bayesian and frequentist statistics. For 25 MSPs, we achieve statistically significant measurements of their timing noise parameters and find that the two methods give consistent results. For the remaining 17 MSPs, we place upper limits on the timing noise amplitude at the 95% confidence level. We additionally place an upper limit on the contribution to the pulsar noise budget from errors in the reference terrestrial time standards (below 1%), and we find evidence for a noise component which is present only in the data of one of the four used telescopes. Finally, we estimate that the timing noise of individual pulsars reduces the sensitivity of this data set to an isotropic, stochastic GW background by a factor of > 9.1 and by a factor of > 2.3 for continuous GWs from resolvable, inspiralling supermassive black-hole binaries with circular orbits.

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

1 INTRODUCTION

Over the past decades, pulsar astronomy has been instrumental in the experimental tests of general relativity (GR) and alternative theories of gravity. Some of the most notable highlights from this research field include the first evidence of the existence of gravitational waves (GWs) (Taylor & Weisberg 1989), the most precise tests of GR (Kramer et al. 2006b), as well as tests of alternative theories of gravity, such as tensor-scalar gravity, in the quasi-stationary, strong-field regime (see e.g. Freire et al. 2012; Shao et al. 2013). These results rely on the pulsar timing technique (e.g. Lorimer & Kramer 2005), which fits the precisely recorded times-of-arrival (TOAs) of the pulses with a model of the pulsar's rotational, astrometric and orbital parameters, as well as signal propagation delays induced by the ionised interstellar medium between the pulsar and Earth. The differences between the observed TOAs and those predicted by the model are called the timing residuals and contain the effects of any unmodelled physical or instrumental processes.

One of the applications of pulsar timing is the possibility of direct detection of GWs via the precise timing of an ensemble of pulsars, commonly referred to as a Pulsar Timing Array (PTA; Foster & Backer 1990). The expected effects of GW propagation on the TOAs were first examined by Sazhin (1978). Later, the idea of using a PTA for unambiguous direct detection of low-frequency (nHz regime) GWs based on the predicted cross-correlation of the residuals of pulsars in various sky positions was proposed by Hellings & Downs (1983). Subsequent work has identified the potential of modern timing data for detecting nHz GWs and formulated the detection methodologies (e.g. Jenet et al. 2004, 2005; Sanidas et al. 2012).

PTAs are sensitive to the stochastic GW background (GWB) resulting from the incoherent superposition of the GW signals from the cosmic population of unresolved inspiralling supermassive black-hole binaries (SMBHBs) (e.g. Rajagopal & Romani 1995), continuous GWs (CGWs) from individual, resolvable SMBHB systems (e.g. Estabrook & Wahlquist 1975), the GWB created from the decaying loops of a cosmic string network that may have formed in the early Universe (e.g. Kibble 1976), a cosmological relic GWB from the Universe's inflationary era (e.g. Grishchuk 2005) and the memory term (long-term change in the GW's amplitude) from GW bursts from SMBHB mergers (e.g. Favata 2009). Prior to the detection, upper limits on the GW amplitudes can impose limits on the properties of the cosmic SMBHB population (e.g Shannon et al. 2015), and rule out the presence of nearby SMBHBs proposed by independent observations (Jenet et al. 2004). In the era of GW astronomy, PTAs using future, hyper-sensitive telescopes will also be able to test theories of gravity in the radiative regime. The GW polarisation modes predicted by GR or alternative theories result in different spatial cross-correlations of the pulsar timing residuals (e.g. Chamberlin & Siemens 2012). These cross-correlations can be further modified if the graviton is not massless as predicted by GR (e.g. Lee 2013).

The pursuit of GW detection using pulsar timing is coordinated by three consortia; the European Pulsar Timing Array (EPTA; Kramer & Champion 2013) in Europe, the North-American Nanohertz Observatory for Gravitational Waves (NANOGrav; McLaughlin 2013) in North America and the Parkes Pulsar Timing Array (PPTA; Hobbs 2013) in Australia. The PTAs employ in total eight large single-dish radio telescopes. The EPTA uses five telescopes, namely the Effelsberg Radio Telescope, the Nançay Radio Telescope, the Lovell Telescope, the Westerbork Radio Synthesis Telescope and the Sardinia Radio Telescope. NANOGrav uses two telescopes, the Green Bank Telescope and the Arecibo Radio Telescope, while the PPTA uses the Parkes Radio Telescope. The three consortia co-operate under the International Pulsar Timing Array (IPTA) consortium, maximising the observing efficiency and data set sensitivity.

The sensitivity of a given PTA is mainly limited by the uncertainties of the TOA measurements, the number of observations and the data time-span, the number of pulsars, their sky distribution and the presence of low-frequency noise in the data (see e.g. Lee et al. 2012; Siemens et al. 2013). While improvements in the instrumentation, increase of the allocated telescope time to PTAs and discoveries of new pulsars can address the first three factors, low-frequency noise needs to be characterised and understood on a pulsar-by-pulsar basis.

A number of methods have been developed to mitigate the dominant sources of noise in pulsar timing. Temporal variations in the dispersion measure (DM; integrated free electron density of the interstellar medium) along the line of sight to the pulsar is a primary source of low-frequency stochastic noise. DM time delays, however, depend on the observing frequency, ν, as $t_{DM} \propto DM\nu^{-2}$, and therefore DM variations can be, to a large degree, corrected using multi-frequency data, (e.g. Keith et al. 2013; Lee et al. 2014). Improper calibration of the gain of the two receiver feeds or cross-coupling between the two feeds can potentially lead to distortions of the total intensity profiles. These instrumental artefacts will introduce additional non-stationary noise components in the timing residuals (van Straten & Bailes 2003; van Straten 2006). By performing standard calibration observations during every observing run, we can minimise the presence of such noise in the data (e.g. Britton 2000). By comparing the noise properties of the same pulsars using overlapping data from from different telescopes, uncorrected noise from instrumental instabilities can potentially be identified (Lentati et al., submitted).

Unfortunately, pulsar timing data also exhibit some levels of "timing noise" (TN), low-frequency, stochastic, achromatic noise, the physical origin of which is unknown and, as such, cannot be mitigated. TN is primarily thought to be caused by pulsar rotational instabilities from various mechanisms. One approach is to consider simultaneous random walks and discrete jumps (caused, e.g., by micro-glitches) in the pulsar's spin frequency and the spin-down rate (e.g. Cordes & Downs 1985; D'Alessandro et al. 1995; Shannon & Cordes 2010). Based on observational evidence, it is also suggested that TN can result from accumulated periodic and quasi-periodic changes in the spin-down rates due to magnetospheric state switching (Kramer et al. 2006a; Lyne et al. 2010). In addition, intrinsic noise has also been proposed to be the result of undetected (and therefore unmodelled) bodies in orbit, such as asteroid belts (Shannon et al. 2013) or planetary-mass objects in long, decadal orbits (Thorsett et al. 1999). Clearly, the measured TN in pulsar timing data can be a superposition of noise intrinsic to the pulsar, and any of the above non-intrinsic noise which is not properly

mitigated, e.g. noise by DM variations not properly corrected due to the lack of sufficient multi-frequency data.

While young pulsars show large amounts of low-frequency noise, millisecond pulsars (MSPs), typically show very low levels of such noise (Verbiest et al. 2009). It is theorised that MSPs have spun-up to the observed ms-order rotational periods via mass transfer from their companions during the system's evolution (Alpar et al. 1982). Their highly stable rotations, short periods and absence of significant temporal changes in their pulse profile shapes (see e.g. Shao et al. 2013) make them excellent celestial clocks which can be timed to sub-100 ns precision over decades. MSPs are therefore the observed sources for GW-detection experiments, and indeed for all high-precision pulsar timing applications.

Despite their demonstrated rotational stability, some MSPs show significant amounts of TN. While their TN is considerably weaker than that of non-recycled pulsars, it can be significant enough to hinder GW detection. PSR B1937+21 (J1939+2134), the first ever discovered MSP, is a notable example of an MSP with strong TN (Kaspi et al. 1994; Shannon et al. 2013). Other MSPs show more moderate noise levels, comparable to the predicted strength of the targeted GWs signals (e.g. PSR J1713+0747; see Zhu et al. 2015). The characterisation of TN is therefore of central importance in high-precision pulsar timing applications.

The measured TN will also contain signals from spatially correlated low-frequency noise (e.g. Tiburzi et al. 2016). Primary examples are the long sought-after stochastic GWB, the signal caused by errors in the reference terrestrial time standards (see e.g. Hobbs et al. 2012) and errors in the solar system ephemeris (see Champion et al. 2010). These signals can be distinguished by the spatial cross-correlations they induce on the timing residuals. The GWB induces a quadrupole signature (see Section 7.1). Errors in the terrestrial time standards produce a fully correlated signal in all pulsars (see Section 6) while errors in the solar system ephemeris can potentially produce a superposition of dipolar correlations between pulsars, each produced by the error in the predicted location of a solar system body. PTAs allow such correlated signals to be recovered or put upper limits on their power.

Different methods have been proposed and employed to characterise the statistical properties of TN in pulsar data and to perform pulsar timing analysis in the presence of correlated noise. These cover techniques based on frequentist (Matsakis et al. 1997; Coles et al. 2011) and Bayesian statistics (e.g. van Haasteren et al. 2009; Lentati et al. 2014), both in the time- and frequency-domain. As part of the efforts to detect GWs, an increasing number of algorithms are being used by the various PTAs to determine the TN properties of MSPs, motivating work to examine the possible biases inherent to different methods. In this context, we perform characterisation of the TN using two established methods based on different statistical analyses, Bayesian and frequentist, and make a comparison of their performance and results. We subsequently use the measured TN properties to search for the presence of TN unique to specific observing systems, place an upper limit on the contribution of clock errors to the measured noise and investigate the impact of the TN on the data set's sensitivity to GWs.

This paper is organised as follows. In Section 2, we describe the data we use. In Section 3, we present the methods used to calculate the noise parameters. The results from both methods are presented in Section 4. In Section 5, we check for TN present only in individual data subsets and continue to investigate systematics by making a search for a correlated clock error signal in Section 6. In Section 7, we evaluate the effects of the TN present in our data on their sensitivity to GWBs and CGWs and finally discuss our conclusions in Section 8.

2 THE EPTA LEGACY TIMING DATASET

We use the EPTA Legacy data set that is presented in Desvignes et al., (submitted; henceforth D15). The data set is composed of data recorded with four EPTA radio telescopes: The Effelsberg Radio Telescope (EFF) in Germany, the Nançay Radio Telescope (NRT) in France, the Westerbork Synthesis Radio Telescope (WSRT) in the Netherlands and the Lovell Telescope (JBO) in the United Kingdom. The data-recording systems (backends) used are the Effelsberg-Berkeley Pulsar Processor (EBPP), the Berkeley-Orléans-Nançay (BON), the Pulsar Machine I (PuMaI) and the Digital Filterbank (DFB) respectively. A more detailed description of the instruments and data reduction techniques can be found in D15, where the timing solutions of the pulsars are also presented.

The data set includes TOAs from 42 MSPs. Their key properties are summarised in Table 1. We identify observing systems as unique combinations of telescope, backend and central observing frequency (receiver). In total, the data set has 18 distinct systems. The EBPP L-band[1] data have the longest time-span, with a maximum of 18 years, starting from October 1996, divided into two observing systems, due to a change in the receiver in 2009. For most of the sources with EBPP data, all other instruments started recording from 2007 onwards, dividing our longest pulsar data sets into two subsets: the first, with single-telescope, single-frequency data and the second, with multi-telescope, multi-frequency data. The lack of multi-frequency data in the first half of the data set makes direct measurements and corrections of the DM variations impossible. It is however possible to extrapolate the signal measured in the second epoch to the first, under the assumption that the DM variations signal is stationary (see Lee et al. 2014). This is performed using the Bayesian analysis methods described in Section 3.2. For a number of MSPs (e.g. PSR J1713+0747, PSR J1012+5307), multi-telescope coverage begins in 1999 with PuMaI data, which contain good quality low-frequency data, allowing direct measurements of the DM variations almost throughout the data set. We note that four MSPs (see Table 1) suffer from a gap in the Effelsberg L-band data for the period between April 1999 and October 2005. The gap is due to changes in the observing priorities.

[1] 1 to 2 GHz range in centre frequency

Table 1. General characteristics of the EPTA Legacy data set. For each pulsar we note the total time-span, T, the ranges of the observing frequencies, ν, the number of observing systems and the number of TOAs. Sources marked with a star suffer from a gap of ∼6 years (1999-2005) in the Effelsberg 1410 MHz data.

PSR J-Name	T (yrs)	ν range (MHz)	number of systems	number of TOAs
J0030+0451⋆	15.1	1345-2678	7	907
J0034−0534	13.5	323-1628	6	276
J0218+4232	17.6	323-2683	13	1196
J0610−2100	6.9	1365-1632	3	1034
J0613−0200	16.1	323-2636	14	1369
J0621+1002	11.8	323-2635	10	673
J0751+1807	17.6	1352-2695	9	796
J0900−3144	6.9	1365-2303	5	875
J1012+5307	16.8	323-2636	15	1459
J1022+1001	17.5	323-2634	10	908
J1024−0719⋆	17.3	1346-2628	9	561
J1455−3330	9.2	1367-1698	3	524
J1600−3053	7.6	1366-2298	4	531
J1640+2224	17.3	1335-2636	8	595
J1643−1224	17.3	1353-2639	11	759
J1713+0747	17.7	820-2637	14	1188
J1721−2457	12.7	1335-1698	4	150
J1730−2304⋆	16.7	1352-2629	8	268
J1738+0333	7.3	1366-1630	3	318
J1744−1134	17.3	323-2634	9	536
J1751−2857	8.3	1397-1631	3	144
J1801−1417	7.1	1395-1697	3	126
J1802−2124	7.2	1366-2048	4	522
J1804−2717	8.1	1374-1698	3	116
J1843−1113	10.1	1335-1629	5	224
J1853+1303	8.4	1397-1698	3	101
J1857+0943	17.3	1335-2632	9	444
J1909−3744	9.4	1367-2681	3	425
J1910+1256	8.5	1366-1630	3	112
J1911−1114	8.8	1397-1630	4	130
J1911+1347	7.5	1365-1698	3	140
J1918−0642	12.8	1372-1630	6	278
J1939+2134	24.1	820-2278	12	3172
J1955+2908	8.1	1395-1629	4	157
J2010−1323	7.4	1381-2298	5	390
J2019+2425	9.1	1365-1629	3	130
J2033+1734	7.9	1367-1631	4	194
J2124−3358	9.4	1365-2298	5	544
J2145−0750	17.5	323-2683	12	800
J2229+2643	8.2	1355-2637	6	316
J2317+1439⋆	17.3	1352-2637	8	555
J2322+2057	7.9	1395-1698	4	229

3 METHODS FOR ESTIMATING NOISE PROPERTIES

For the estimation of the noise properties, we use two different methods. The first method follows a Bayesian approach, in the time-frequency domain and is described in Lentati et al. (2014). The second method uses frequentist statistics based on power-spectral estimation of the residuals and using algorithms described in Section 3.3, which are an extension of those introduced in Coles et al. (2011). We first discuss the noise model components, which we use for both approaches, and then present the details of each method used.

3.1 Noise Modelling

We form the timing residuals using the pulsar timing analysis package `Tempo2` (Hobbs et al. 2006), which iteratively performs a weighted least-squares (wLS) fit of the model to the TOAs until the reduced chi-squared of the residuals is minimised. Timing models are gradually improved over many years by incorporating more data. These solutions will often result in timing residuals scattered beyond what would be expected based on their formal uncertainties, due to the absence, at this point, of the stochastic signals in the model. These signals are in general divided into the time-correlated and uncorrelated components.

The uncorrelated (white-noise) components correct the uncertainties of the timing residuals. The formal uncertainties of the TOAs are derived by the cross-correlation of the recorded integrated pulse profile with a reference template, which is constructed using the best available observations. These uncertainties are correct if the recorded profiles are characterised solely by (white) radiometer noise and the profile template precisely represents the intrinsic shape of the integrated profile. However, possible presence of un-excised radio frequency interference (RFI), temporal variations in the pulse profile, artefacts in the profiles from instrumental instabilities or imperfect profile templates can lead to errors in the uncertainty estimations (e.g. Liu et al. 2011). It is therefore common practice to include a multiplicative correction factor called EFAC. We also add a correction term quadratically to the formal uncertainty to account for additional scatter in the TOAs caused by statistically independent physical processes, such as pulse phase jitter noise (e.g. Shannon et al. 2014). This term is commonly referred to as EQUAD. We do not investigate the physical origin of the noise included in the EQUADs. This requires a more detailed analysis of the white noise; for example, jitter noise is dependent on the integration time of the observation and this needs to be properly taken into consideration if one wants the EQUAD number to describe an underlying physical process.

We include one EFAC and one EQUAD term per observing system to mathematically model the uncorrelated noise from all possible processes. The white-noise correction factors should be such that the data satisfy the central assumption of pulsar timing, that they are drawn from a random Gaussian process. In other words, when subtracting the waveforms (induced residuals) of all calculated stochastic signals from the residuals, their uncertainties should be such that the residuals are white and the timing solution has a reduced chi-squared of unity. The original TOA uncertainty, σ, EFAC (f), EQUAD (q) and corrected uncertainy, $\hat{\sigma}$, are related[2] as:

$$\hat{\sigma}^2 = (\sigma \cdot f)^2 + q^2 \quad (1)$$

We include two stationary time-correlated noise components, namely the chromatic low-frequency noise from DM variations and the achromatic TN. Previous studies (e.g. Shannon & Cordes 2010; Coles et al. 2011) have shown that

[2] This definition is not unique. `Tempo2` by default defines the correction as $\hat{\sigma}^2 = f^2 \cdot (\sigma^2 + q^2)$

the low-frequency power spectra of pulsar timing residuals can be adequately modelled with single power-laws for the majority of MSPs. This does not mean that the TN is necessarily a pure power-law, but rather that this functional form is sufficient to describe the data, given the measurement precision. We examined whether deviations from the single power-law model are supported by the data using the Bayesian analysis method. In particular, we performed the noise analysis with two additional models for the TN spectrum: (i) a model that allows the power of individual frequency bins to vary independently from the power law model and (ii) a model that includes the power-law and an additional sinusoid signal of varying frequency, amplitude and phase. We evaluated the results using the Bayes factor, i.e. the ratio of the Bayesian evidence of two competing models (see also Section 3.2). A common interpretation of the Bayes factor is given by Kass & Raftery (1995), based on which we required a value equal or greater than 3 to justify the addition of any extra model parameter. This was not the case for any of the models we compared to the simple single power-law model.

In this work, we have followed the single power-law formalism for both analysis methods in order to facilitate their comparison and the comparison of the measured TN parameters with those usually used as GW stochastic parameters in the PTA literature. For isotropic GW signals (see Section 7) one of the most important properties is the characteristic strain spectrum, $h_c(f)$, of the GWB on the one-sided power spectrum of the induced timing residuals. For most models of interest, this can be written as a power-law function of the GW frequency (e.g. Jenet et al. 2005), f as:

$$h_c(f) = A\Big(\frac{f}{f_r}\Big)^{\alpha} \tag{2}$$

where A is the (dimensionless) amplitude of the wave, α is the spectral index[3] and f_r is the reference frequency, typically set to 1 yr^{-1}. The one-sided power spectral density of the signal is then given by:

$$S(f) = \frac{A^2}{12\pi^2}\Big(\frac{f}{f_r}\Big)^{-\gamma} \tag{3}$$

where the power spectrum and strain spectral indices are related as $\gamma \equiv 3 - 2\alpha$. This is the functional form we use to model the TN. We set a cut-off at frequency 1/T, where T is the time-span of the data. The cut-off arises naturally due to the absorption of the lowest frequencies by the fitting of the pulsar's spin and spin-down in the timing model. It has been shown (Lee et al. 2012; van Haasteren & Levin 2013) that if the spectral index is $\gamma \lesssim 7$ (which is the case for all MSPs in this paper), the cut-off at frequency 1/T is sufficient.

The DM variations have been mitigated using first- and second-order DM derivatives in the timing model (which are first- and second-order polynomials) and additionally a power law equivalent to Eq. (3). The DM derivatives absorb any DM variation signal with frequencies below the cut-off frequency, in the same way the spin and spin-down do for the achromatic TN (Lee et al. 2014). The observing frequency dependence of the DM variations signal is measured in the time-domain via the (multi-frequency) timing residuals, as we show in Section 3.2. The choice of a power-law spectrum for the DM variations is motivated by the fact that, across a wide spatial frequency range, the electron density fluctuation spectrum usually follows a power-law (Armstrong et al. 1995).

[3] We define the index positive, but note that in the literature it is sometimes defined as a negative number

3.2 Noise Parameter Estimation Using Bayesian Inference

The first Bayesian investigation of the GWB detectability with PTAs was performed by van Haasteren et al. (2009). The algorithms were later applied on EPTA data to derive the EPTA GWB upper limit (van Haasteren et al. 2011). In that analysis, the timing noise parameters of the MSPs were simultaneously estimated with the GWB parameters. Further work on Bayesian analysis methods for pulsar timing provided more algorithms, both in time- and time-frequency-domains, to characterise the properties of timing noise and DM variations and to perform robust pulsar timing analysis in the presence of correlated noise (e.g. van Haasteren et al. 2011; Lentati et al. 2013; Lee et al. 2014).

Bayes' theorem, which is the central equation for these analysis methods, states that:

$$Pr(\Theta) = \frac{L(\Theta)\pi(\Theta)}{Z}, \tag{4}$$

where Θ is the model's parameters, $Pr(\Theta)$ is the posterior probability distribution (PPD) of the parameters (probability distribution of the parameters given the model and the data), $\pi(\Theta)$ is the prior probability distribution (pPD) of the parameters (the initial hypothesis of the probability distribution of the parameters for a given model), $L(\Theta)$ is the likelihood function (which gives the probability that the data are described by a given model) and Z is the Bayesian evidence. Z is only a normalising factor independent of Θ and can therefore be ignored when one is interested only in parameter estimation, such that $Pr(\Theta) \propto L(\Theta)\pi(\Theta)$. On the other hand, when one is interested in model selection, the ratio of the evidence between two different models, $\mathcal{R}$, known as the Bayes factor, is used. The probability, $\mathcal{P}$, of a model compared to another, can the expressed (Kass & Raftery 1995) as:

$$\mathcal{P} = \frac{\mathcal{R}}{1+\mathcal{R}} \tag{5}$$

The various Bayesian analysis algorithms are distinguished by the mathematical description of the model parameters and the computational methods used to sample the unnormalised PPD.

Lentati et al. (2014) introduced `TempoNest`, a Bayesian software package for the analysis of pulsar timing data, available to use as a `Tempo2` plug-in. The timing solution and the additional stochastic parameters such as EFACs, EQUADs, DM variations and the TN (referred to as "excess red noise") can be determined simultaneously. `TempoNest`

uses the Bayesian inference tool MultiNest (Feroz & Hobson 2008) to explore this joint parameter space, whilst using `Tempo2` as an established means of evaluating the timing model at each point in that space. For the PPD sampling, `TempoNest` uses the nested sampling Monte-Carlo method (Skilling 2004).

We perform a joint analysis for the timing model and the stochastic parameters. Both the TN and the DM variations are modelled as Gaussian stochastic signals with power-law spectra as described by Eq. (3). `TempoNest` employs the time-frequency analysis described in Lentati et al. (2013). The TN waveform is expressed as $\mathbf{t}_{\rm TN} = \mathbf{F}_{\rm TN}\mathbf{a}$, where $\mathbf{F}_{\rm TN}$ is the Fourier transform with components $\mathsf{F} = \sin(2\pi f) + \cos(2\pi f)$ and corresponding coefficients, $\mathbf{a}$, which are free parameters. Here, and henceforth, we use boldface characters in equations to denote matrices. The Fourier frequencies take values $f = n/\mathrm{T}$, with n integers ranging from 1 up to the value necessary to sample frequencies as high as $1/14$ days^{-1}. The covariance matrix of the TN is then described by the following equation (see Lentati et al. 2015):

$$\mathbf{C}_{\rm TN} = \mathbf{C}_{\rm w}^{-1} - \mathbf{C}_{\rm w}^{-1}\mathbf{F}_{\rm TN}\left[(\mathbf{F}_{\rm TN})^{\mathbf{T}}\mathbf{C}_{\rm w}^{-1}\mathbf{F}_{\rm TN} + (\Psi^{\rm SN})^{-1}\right]^{-1} (\mathbf{F}_{\rm TN})^{\mathbf{T}}\mathbf{C}_{\rm w}^{-1}. \quad (6)$$

Here, $\Psi = \langle \mathrm{a}_i \mathrm{a}_j \rangle$, is the covariance matrix of the Fourier coefficients and $\mathbf{C}_{\rm w}$ is the covariance matrix of the white noise component, a diagonal matrix with the main diagonal populated by the residual uncertainties squared, $\hat{\sigma}^2$ (as in Eq. 1). The superscript $\mathbf{T}$ denotes the transpose of the matrix.

The covariance matrix for the DM variations, $\mathbf{C}_{\rm DM}$, is equivalent to Eq. (6), but including an observing frequency dependence. This is achieved by replacing $\mathbf{F}$ components with $\mathsf{F}^{DM}_{ij} = \mathsf{F}_{ij} D_i D_j$, where the i,j indices denote the residual numbers, $D_i = \frac{1}{K\nu_i}$, ν_i is the observing frequency of the TOA, typically set as the central frequency of the observing band, and K=2.41×10^{-16} Hz^{-2}cm^{-3}pc s^{-1}, is the dispersion constant.

The likelihood function is the probability that the data (TOAs), noted as $\boldsymbol{t}$, are fully described by the timing model signal, $\boldsymbol{\tau}(\boldsymbol{\epsilon})$, with parameters $\boldsymbol{\epsilon}$ and the stochastic noise. The latter is encoded in the residuals' total covariance matrix,

$$\mathbf{C} = \mathbf{C}_{\rm w} + \mathbf{C}_{\rm DM} + \mathbf{C}_{\rm TN}. \quad (7)$$

Following van Haasteren et al. (2009), and noting that the difference $\mathbf{t} - \boldsymbol{\tau}(\boldsymbol{\epsilon})$ gives the timing residuals vector, we can write the likelihood function as:

$$L = \frac{1}{\sqrt{(2\pi)^n |\mathbf{C}|}} e^{-\frac{1}{2}(\mathbf{t}-\boldsymbol{\tau}(\boldsymbol{\epsilon}))^{\mathbf{T}}\mathbf{C}^{-1}(\mathbf{t}-\boldsymbol{\tau}(\boldsymbol{\epsilon}))}. \quad (8)$$

After the noise properties are estimated, we produce the TN waveforms, which can be estimated from the data using the maximum likelihood (ML) value of its statistical model parameters, A and γ. As shown in Lee et al. (2014), the ML waveform, $\mathbf{t}_{\rm TN}$, and its uncertainties, $\sigma_{\rm TN}$, are optimally estimated as

$$\mathbf{t}_{\rm TN} = \mathbf{C}_{\rm TN}\mathbf{C}^{-1}\mathbf{t}\ , \quad (9)$$

with uncertainties estimated as:

$$\sigma_{\rm TN} = \mathbf{C}_{\rm TN} - \mathbf{C}_{\rm TN}\mathbf{C}^{-1}\mathbf{C}_{\rm TN}. \quad (10)$$

The uncertainties are estimated as the standard deviation of the estimator. However, as noted in Lee et al. (2014), since the components of TN waveforms are correlated, their interpretation in terms of uncertainties is meaningless, since this is only valid under the assumption that the noise is uncorrelated. The uncertainties can therefore only be used as an indication of the variance of each point.

We have performed the Bayesian inference analysis twice using different combination of pPDs. The pPDs on the timing parameters are always uniform, centred around the value from the wLS fit of the timing model by `Tempo2` with a range of 10 to 20 times their 1-σ `Tempo2` uncertainties. This range was chosen after testing verified that is sufficient for all timing parameters PPDs to converge. For the noise parameters, the ranges are from 0 to 7 for spectral indices, −20 to 8 for the logarithm of the amplitudes, −10 to −3 for the logarithm of the EQUADs and 0.3 to 30 for the EFACs. For EQUADs, TN and DM variations amplitudes we used two different types of pPDs. The first is a uniform distribution in log space (log-uniform) and the second is a uniform distribution in linear space (uniform). Log-uniform pPDs assume that all orders of magnitude are equally likely for the parameter value while for uniform pPDs, we assign the same probability for all values. The uninformative log-uniform pPDs will result in PPDs for the parameters that are the least affected by the pPD and therefore are what we consider as the parameter measurement. If no significant noise can be detected in the data, the PPDs are unconstrained and the distribution's upper limit is dependent on the lower limit of the pPD. Therefore, a separate analysis is required using uniform pPDs in order to obtain robust upper limits. If the signal is strong and the result from a log-uniform-pPD analysis is a well-constrained PPD, then the change of the pPD should not affect the result significantly and the PPDs should be almost identical. As a result, we performed the analysis with the following combination of pPDs:

a) Uniform EQUAD pPDs and log-uniform pPDs for TN and DM variation amplitudes. This set of pPDs results in upper limits for EQUADs. As such, the solutions have the highest possible timing residuals uncertainties, resulting in weaker TN and DM variations detections. The TN and the DM variations are treated in the same way, giving no prior information that can favour the one over the other when multi-frequency data are not sufficient to de-couple them. In the absence of multi-frequency data one can therefore expect that their PPDs will not be well-constrained.

b) Uniform TN amplitude and log-uniform pPDs for EQUADs and DM variation amplitudes: The total white noise levels of these solutions are lower, since EQUAD PPDs can be flat if the data do not support them to be measurable. The use of uniform pPDs for the TN amplitude and log-uniform for the DM variations results in solutions in favour of the TN against the DM variations in the absence of multi-frequency data. This set of pPDs will provide the strictest upper limits on the TN amplitudes. We used the PPDs from this analysis to calculate the amplitude upper limits at the 95% confidence level (C.L.).

Table 2. Timing-noise characteristics of EPTA MSPs based on Bayesian inference for a single power-law model as described by Equation (3). The results are divided based on the quality of the posterior probability distributions (PPDs) as described in Section 4.1. We tabulate the maximum likelihood (ML) and median (med) values of the dimensionless amplitude, A, at reference frequency of 1 yr^{-1} and the the spectral index, γ. For A, we also tabulate the 95% confidence upper limits. The 1-σ uncertainties are calculated such that the 68% of the area under the 1-dimensional marginalised PPS of the parameter is symmetrically distributed around the median. As described in Section 4.1, for unconstrained PPDs we only consider the upper-limits analysis results.

PSR J-Name	$\log(\mathrm{A_{ML}})$	$\log(\mathrm{A_{med}})$	$\log(\mathrm{A_{UL}^{95\%}})$	γ_{ML}	γ_{med}
		Well-constrained PPDs			
J0621+1002	−12.029	$-12.07^{+0.06}_{-0.06}$	−11.9	2.5	$2.4^{+0.3}_{-0.2}$
J1012+5307	−13.20	$-13.09^{+0.07}_{-0.07}$	−12.94	1.7	$1.7^{+0.3}_{-0.2}$
J1022+1001	−13.2	$-13.0^{+0.1}_{-0.2}$	−12.8	2.2	$1.6^{+0.4}_{-0.4}$
J1600−3053	−13.35	$-13.28^{+0.06}_{-0.06}$	−13.11	1.2	$1.7^{+0.3}_{-0.2}$
J1713+0747	−14.7	$-15.2^{+0.5}_{-0.5}$	−13.8	4.8	$5.4^{+0.9}_{-1.0}$
J1744−1134	−13.7	$-13.8^{+0.2}_{-0.3}$	−13.3	2.2	$2.7^{+0.7}_{-0.6}$
J1857+0943	−13.3	$-13.3^{+0.2}_{-0.3}$	−12.9	2.6	$2.4^{+0.7}_{-0.6}$
J1939+2134	−14.2	$-14.5^{+0.3}_{-0.3}$	−13.7	5.9	$6.2^{+0.5}_{-0.6}$
		Semi-constrained PPDs			
J0030+0451	−14.9	$-14.9^{+0.8}_{-2.1}$	−13.0	6.3	$5.2^{+1.2}_{-2.1}$
J0218+4232	−13.1	$-14.1^{+1.0}_{-1.7}$	−12.4	2.7	$3.9^{+1.7}_{-1.6}$
J0610−2100	−18.7	$-16.0^{+2.9}_{-2.7}$	−12.4	1.4	$2.7^{+2.8}_{-2.1}$
J0613−0200	−13.7	$-14.4^{+0.7}_{-0.9}$	−13.0	2.8	$4.1^{+1.6}_{-1.5}$
J0751+1807	−18.8	$-15.9^{+2.6}_{-2.7}$	−12.9	6.5	$3.0^{+2.0}_{-1.4}$
J1024−0719	−14.0	$-16.3^{+2.1}_{-2.4}$	−13.1	5.3	$3.9^{+2.0}_{-2.5}$
J1455−3330	−19.8	$-14.2^{+1.0}_{-3.7}$	−12.7	0.8	$3.6^{+1.9}_{-1.6}$
J1640+2224	−13.2	$-13.1^{+0.2}_{-3.4}$	−12.8	0.01	$0.4^{+1.7}_{-0.3}$
J1643−1224	−17.7	$-13.3^{+0.6}_{-2.4}$	−12.5	1.8	$1.7^{+0.9}_{-0.6}$
J1721−2457	−11.7	$-13.5^{+1.7}_{-4.5}$	−11.5	1.1	$1.9^{+2.7}_{-1.0}$
J1730−2304	−12.8	$-14.7^{+1.7}_{-3.6}$	−12.6	1.8	$2.9^{+1.9}_{-1.3}$
J1801−1417	−14.4	$-15.1^{+2.5}_{-3.4}$	−12.2	6.3	$3.3^{+2.2}_{-1.8}$
J1802−2124	−17.0	$-15.6^{+3.2}_{-3.0}$	−12.2	4.5	$2.3^{+2.9}_{-0.8}$
J1843−1113	−13.0	$-12.9^{+0.2}_{-3.3}$	−12.5	0.6	$1.5^{+3.1}_{-0.5}$
J1909−3744	−14.1	$-14.1^{+0.2}_{-1.9}$	−13.8	2.4	$2.3^{+1.0}_{-0.6}$
J1918−0642	−16.9	$-14.5^{+0.7}_{-0.5}$	−12.6	1.7	$5.4^{+1.1}_{-1.6}$
J2145−0750	−14.4	$-14.0^{+0.6}_{-0.8}$	−12.9	5.2	$4.1^{+1.6}_{-1.3}$
		Unconstrained PPDs			
J0034−0534	-	-	−12.3	-	-
J0900−3144	-	-	−12.7	-	-
J1738+0333	-	-	−12.7	-	-
J1751−2857	-	-	−12.4	-	-
J1804−2717	-	-	−12.3	-	-
J1853+1303	-	-	−12.4	-	-
J1910+1256	-	-	−12.1	-	-
J1911−1114	-	-	−12.1	-	-
J1911+1347	-	-	−12.9	-	-
J1955+2908	-	-	−12.1	-	-
J2010−1323	-	-	−12.8	-	-
J2019+2425	-	-	−11.9	-	-
J2033+1734	-	-	−12.0	-	-
J2124−3358	-	-	−12.8	-	-
J2229+2643	-	-	−12.7	-	-
J2317+1439	-	-	−13.1	-	-
J2322+2057	-	-	−12.3	-	-

Table 3. Timing-noise characteristics of EPTA MSPs based on power-spectral analysis for a single power-law model as described by Equation (3). We tabulate the dimensionless amplitude, A, at reference frequency of $1\mathrm{yr}^{-1}$, the spectral index, γ, and the white-noise power level, S_{W}, and their respective 1-σ uncertainties. We also tabulate the pre-whitening level used ($\mathrm{level_{pw}}$). For the pulsars where the measurement of timing noise was not possible, we quote the 95% confidence upper limits for the amplitude. The table is divided as Table 2 for easier comparison.

Measured				
PSR J-Name	log(A)	γ	$\log(\mathrm{S_W(yr^3)})$	$\mathrm{level_{pw}}$
J0621+1002	−12.3±0.1	2.8±0.6	−26.94±0.04	1
J1012+5307	−13.01±0.07	1.7±0.3	−28.60±0.02	1
J1022+1001	−13.2±0.2	2.0±0.6	−27.97±0.03	0
J1600−3053	−13.6±0.1	1.3±0.5	−29.36±0.05	0
J1713+0747	−14.2±0.2	4.9±0.6	−30.146±0.02	2
J1744−1134	−13.6±0.2	3.0±0.6	−28.90±0.03	1
J1857+0943	−13.2±0.2	2.3±0.7	−27.97±0.04	1
J1939+2134	−14.3±0.1	6.7±0.5	−30.27 ± 0.02	2
J0030+0451	−13.2 ± 0.4	4.5 ± 1.0	−27.78±0.03	2
J0218+4232	−12.6±0.2	2.3±0.6	−26.69±0.03	0
J0610−2100	-13.6±0.1	2.1±0.6	−29.62±0.03	0
J0613−0200	-14.9±0.9	5.2±1.8	−28.45±0.03	0
J0751+1807	-14.3±0.7	5.2±1.6	−27.86±0.03	1
J1024-0719	−13.0±0.1	4.1±0.5	−28.15±0.03	2
J1455−3330	−13.4±0.4	3.5±1.2	−27.59±0.03	0
J1640+2224	−13.0±0.1	1.4±0.4	−27.96±0.05	0
J1643−1224	−13.2±0.1	3.5±0.4	−28.25±0.03	0
J1721−2457	−12.3±0.3	2.7±0.8	−26.01±0.09	0
J1730−2304	−12.8±0.2	1.7±0.5	−27.31±0.06	0
J1801−1417	−13.3±0.3	2.4±1.1	−28.41±0.10	0
J1802−2124	−12.8±0.2	2.9±0.7	−27.93±0.04	0
J1843−1113	−12.8±0.1	3.0±0.6	−27.93±0.05	1
J1909−3744	−14.5±0.7	1.6±1.7	−30.05±0.04	0
J1918−0642	−13.0±0.2	2.8±0.8	−27.72±0.05	1
J2145-0750	−13.7±0.3	3.5±0.7	−28.36 ± 0.03	0
Upper Limits				
PSR J-Name	$\log(\mathrm{A_{UL}^{95\%}})$		$\log(\mathrm{S_W(yr^3)})$	$\mathrm{level_{pw}}$
J0034−0534	−12.4	-	−27.02 ± 0.05	0
J0900−3144	−12.8	-	−28.0 ± 0.1	0
J1738+0333	−12.6	-	−27.36 ± 0.04	0
J1751−2857	−12.1	-	−27.3 ± 0.6	0
J1804−2717	−12.2	-	−26.57 ± 0.09	0
J1853+1303	−12.7	-	−27.7 ± 0.1	0
J1910+1256	−12.6	-	−27.38 ± 0.06	0
J1911−1114	−12.2	-	−26.7 ± 0.1	0
J1911+1347	−12.8	-	−27.88±0.1	0
J1955+2908	−12.1	-	−26.46 ± 0.06	0
J2010−1323	−12.9	-	−27.95 ± 0.04	0
J2019+2425	−12.0	-	−26.14 ± 0.08	0
J2033+1734	−12.0	-	−26.15 ± 0.06	0
J2124−3358	−12.8	-	−27.69 ± 0.04	0
J2229+2643	−12.7	-	−27.66 ± 0.05	1
J2317+1439	−12.8	-	−27.678 ± 0.03	0
J2322+2057	−12.3	-	−26.78 ± 0.05	0

3.3 Noise Parameter Estimation Using Power-Spectral Analysis

Power-spectral analysis of pulsar timing data using standard discrete Fourier transforms is complicated by highly variable error bars, irregular sampling, data gaps (due to difficulties in being granted telescope time at exact regular intervals but also due to loss of data from technical difficulties, weather conditions, telescope maintenance or from weak pulses on particular days due to unfavourable interstellar scintillation) and the presence of TN which has a steep red spectrum. Fourier transforms require equispaced data points. Interpolation of data points on regular grids intro-

duces time-correlations in data points and the presence of strong TN introduces spectral leakage. In order to bypass such problems, Coles et al. (2011) introduced an algorithm for pulsar-timing analysis in the presence of correlated noise which employs the use of generalised least-squares (GLS) analysis of the timing data using the covariance matrix of the residuals (as described in Section 3.2). In brief, the covariance matrix of the residuals is used to perform a linear transformation that whitens both the residuals and the timing model. The transformation is based on the Cholesky decomposition of the covariance matrix.

For this algorithm, initial estimates of the residuals covariance matrix are necessary, and are obtained using the Lomb-Scargle periodogram (LSP), which can calculate the power spectrum of irregularly sampled data. Spectral leakage in the presence of strong TN with steep power-law spectra is mitigated with pre-whitening using the difference filter. The difference pre-whitening filter of any order, k, can be described by $y_{w,k} = y_{w,k-1}(t_i) - y_{w,k-1}(t_{i-1})$, where t_i is the i-th sampling time and $y_{w,k}$ is the whitened residual of difference order k ($k = 0$ corresponds to the original residuals). It was suggested to use the lowest order necessary to whiten the data enough to mitigate spectral leakage. Effectively, this filter is equivalent to multiplying the power spectrum by a filter (e.g. for first order difference, the filter is the square of the transfer function). After the spectrum is estimated using the pre-whitened data, one corrects the power spectrum by dividing it with the same filter, a process known as post-darkening. The low-frequency spectrum can be fitted with a power-law model leading to the first estimation of the covariance matrix. Through an iterative process, new estimates of the spectrum can be achieved by using LSP after whitening the data using the Cholesky decomposition of the covariance matrix.

Coles et al. (2011) have demonstrated that the implementation of this method allows better timing solutions with more robust timing parameters and uncertainty calculations. In particular the measured spin and spin-down of the pulsar show the largest improvements, since they have low-frequency signatures in the Fourier domain and correlate with TN. However, this method is not optimised to accurately estimate the TN properties through detailed fitting of a noise model to the power-spectrum. The algorithm described in Coles et al. (2011) focuses on obtaining a linear, unbiased estimator of the timing parameters. For this purpose, they demonstrate that using the GLS timing solutions using the covariance matrices of any TN models which whiten the data sufficiently to remove spectral leakage, are statistically consistent. In this work, we extend the algorithms of Coles et al. (2011), focusing on the precise evaluation of the power spectra and the power-law model parameters. To this end, we have developed an independent power spectral analysis and model fitting code.

A fully frequentist analysis should include a white-noise and DM-correction analysis. However, in order to focus on comparing the methods with regards to the estimation of the TN properties, we use the ML EFAC and EQUAD values and subtract the ML DM variations waveforms derived from the Bayesian analysis.

Our spectral analysis code calculates a generalised LSP, i.e. it performs a wLS fit of sine and cosine pairs at each frequency. We follow an iterative procedure as follows: (1) We first use `Tempo2` to obtain the wLS post-fit residuals, while subtracting the ML DM variations signal estimated with the Bayesian methods described in Section 3.2. (2) We calculate the spectrum of these residuals using a chi-squared minimisation fit on all frequency points. (3) `Tempo2` is re-run using the covariance matrix of the initial noise model to perform a GLS fit. (4) Finally, we re-run the spectral analysis code on the residuals from the GLS timing solution to update the TN model and repeat steps 3 and 4 until the solution converges. Typically, this required no more than one iteration.

Our code implements a generalised LSP to account for the timing residual uncertainties. Denoting each pair of time and residual as (t_i, y_i), the LSP is formed by fitting sine-cosine pairs of the form $\hat{y}(\omega_k, t_i) = a_k \cos(\omega_k t_i) + b_k \sin(\omega_k t_j)$ at all angular frequencies, $\omega_k = 2\pi f_k$, with f_k the frequency. The solution is obtained by minimising the chi-squared for each ω_k, weighted by the summed uncertainties of the timing residuals as:

$$\chi^2_k = \sum_i \left(\frac{y_i - a\sin(\omega t_i) - b\cos(\omega t_i)}{\hat{\sigma}_i} \right)^2 . \tag{11}$$

Once the LSP is calculated, noting the number of timing residuals as N, the spectral density is finally computed as:

$$S(f) = \frac{2|\hat{y}|^2 T}{N^2}. \tag{12}$$

We examine whether spectral leakage is present following the same routine as in Coles et al. (2011). Visual inspection of the original spectrum allows to approximately define the frequency where the red component of the spectrum intersects the flat, white component. We apply a low-pass filter in time-domain to separate the high-frequency from the low-frequency residuals and calculate their individual spectra. The high-frequency spectrum should be consistent with the high-frequency part of the spectrum of the original data. If that is not the case, and instead the high-frequency spectrum is significantly weaker, then leakage is important and we need to apply the pre-whitening filter. The code allows for any order of difference whitening. For this data set, we required only up to second order. We then proceed with calculating the LSP as before and finally post-darken the spectrum before calculating the final spectral density.

We fit the power spectrum with the following function:

$$S(f) = S_0 \Big(\frac{f}{f_r} \Big)^{-\gamma} + S_{\rm W}, \quad S_0 = \frac{A^2}{12\pi^2}. \tag{13}$$

Here, $S_{\rm W}$ is the spectral density of the high-frequency (white) component. The power-law description of the low-frequency component is equivalent to Eq. (3), with S_0 the spectral density at reference frequency, f_r, which is set to $1\mathrm{yr}^{-1}$. A fit of only the low-frequency component is proven difficult; due to the steepness of the spectrum at low frequencies and moderate power of the TN in many MSPs, only about five frequency points would be included in a pure power-law fit of only the red part of the spectrum. This leads to unstable fits without meaningful error estimations.

The fit minimises the chi-squared, χ^2_S. Chi-squared minimisation assumes that the spectrum is normally distributed. In principle, the power spectrum is a chi-squared distribution. However, in logarithmic space, the distribution is approximately Gaussian with variance of order unity. Therefore this is a good approximation if we fit the power-law model to the spectrum in logarithmic space. By doing so, we minimise the chi-squared defined as:

$$\chi^2_S = \sum_{i=1}^{N} \left\{ \log S_i - \log \left(S_0 \left(\frac{f_i}{f_r} \right)^{-\gamma} + S_{\rm W} \right) \right\}^2 , \qquad (14)$$

where S_i and f_i define the points of the spectrum for each frequency bin, i, while simultaneously fitting for S_0, γ, and $S_{\rm W}$. We first fit the spectrum while setting the uncertainties of the LSP points to one and then scale the uncertainties to achieve a reduced chi-squared of unity.

Once we obtain the values for the noise parameters, we construct the covariance matrix of the TN, $\mathbf{C}_{\rm TN}$. The Fourier transform of the TN power-law model gives the covariance function, $c_{\rm TN}(\tau) = \langle {\rm t}_{TN,i} {\rm t}_{TN,j} \rangle$. The i and j indices refer to the time epoch of the observation and $\tau = i - j$. The TN covariance matrix is then formed by the elements $C_{{\rm TN},ij} = c(\tau_{{\rm TN},ij})$, where $\tau_{ij} = |i-j|$. Using the total covariance matrix (Eq. 7), we then perform a `Tempo2` GLS fit on the TOAs, repeat the power-spectrum analysis and power-law fit to update the model parameter values and iterate these steps until we converge to a stable solution.

For the cases where the spectra are white-noise dominated and no measurement of the TN parameters can be achieved on a 3-σ level, we derived upper limits for the TN amplitude. The limits are at the 95% C.L. and are calculated as the 2-σ upper limit of the white-noise level ($S_{\rm W}$ in Eq. (13) and Table 3).

4 RESULTS

Table 2 summarises the results of the noise properties determined using `TempoNest`, while Table 3, summarises the results from the power spectral analysis. The reader can find online[4] the PPDs of the TN properties from the Bayesian analysis, the power spectra and the TN waveforms from both methods. In the rest of this Section, we first discuss the framework under which we compare the results from the two methods and then proceed with the comparison of the results in more detail. We conclude this Section by presenting and discussing the results on the white noise parameters from the Bayesian analysis.

4.1 Comparing Bayesian and Frequentists Results

Bayesian analysis is based on the principle that we test a hypothesis (model) given the data and a pPD. The latter is essential in Bayesian inference and states our prior degree of confidence on what the PD of the parameter is. The inference results in the PPD, which is the updated probability distribution for the unknown parameter, based on the information provided by the data. Bayesian inference also assigns the likelihood value for each model (i.e. for each set of values for all unknown parameters), providing a measure of how well the model describes the data. To evaluate the `TempoNest` results, we report in Table 2 the ML values of the TN parameters and the median value and 1-σ uncertainties of the one dimensional marginalised PPDs. The uncertainties are calculated such that 68% of the area under the distribution is symmetrically distributed around the median. The asymmetry of many PPDs will result in asymmetric error bars.

We sort the PPDs in three categories, and show representative examples in Figure 1. We name the first category of distributions "well-constrained"; this represents cases where the data were sufficient to obtain good measurements of the noise parameters. As seen in Figure 1 for the case of PSR J1012+5307, the PPDs are well defined and very close to symmetric. As a result, the median values of the 1-dimensional PPDs coincide well with the ML solution. There are cases where the PPD of at least one of the TN parameters suffers from long tails due to strong covariances between unknown parameters (e.g. amplitude of TN and amplitude of DM variation noise in the absence of sufficient multi-frequency data). We refer to these distributions as "semi-constrained". As seen for the case of PSR J0751+1807 in Figure 1, the two-dimensional distribution shows a main area of high probability as well as many smaller regions of local maxima. The tails in the one dimensional distribution of amplitude (which in general extend to $\pm\infty$), causes the median value to vary significantly from the ML model. Moreover, the large amount of area under the curve, along the tail, causes the uncertainties around the median to have large and very asymmetric values. Finally, when the data do not support any evidence of TN, the PPDs are flat. We refer to these as "unconstrained". As seen for the case of PSR J2229+2643 in Figure 1, the reported median and ML values do not hold a strong significance. The only meaningful result to report in such cases is the upper limits for the amplitude, as seen in the bottom right panel of Figure 1.

Power-spectral analysis provides single-value results-from the power-law model fit to the power spectrum. This fit is performed under the assumption of Gaussian statistics. As discussed above, in the case of power spectra, this is only an approximation. Finally, the fit is dependent on the estimation of the uncertainties of the power spectrum points, which was ensured to be properly calculated by pre-whitening the data when TN caused spectral leakage. The comparison of the results derived with these two methods should also consider the effects of the Bayesian ML DM variations waveform subtraction from the residuals before performing the power-spectral analysis. In the case of semi-constrained PPDs, the amplitude parameters for the two TN and DM variations are naturally highly correlated. When this is the case, the ML parameter estimates are not as reliable, as the particular ML solution might correspond to either significant DM variations and no TN, or significant TN and no DM variations. This can lead to over- or under-estimations of the DM variations which will lead to either part of the TN being subtracted as well or part of the DM signal leaking into the TN.

As an example, we show in the left panel of Figure 2 the two- and one-dimensional marginalised PPDs for the amplitudes of the timing noise and DM variations for PSR

[4] http://www.epta.eu.org/aom/DR1noise.html

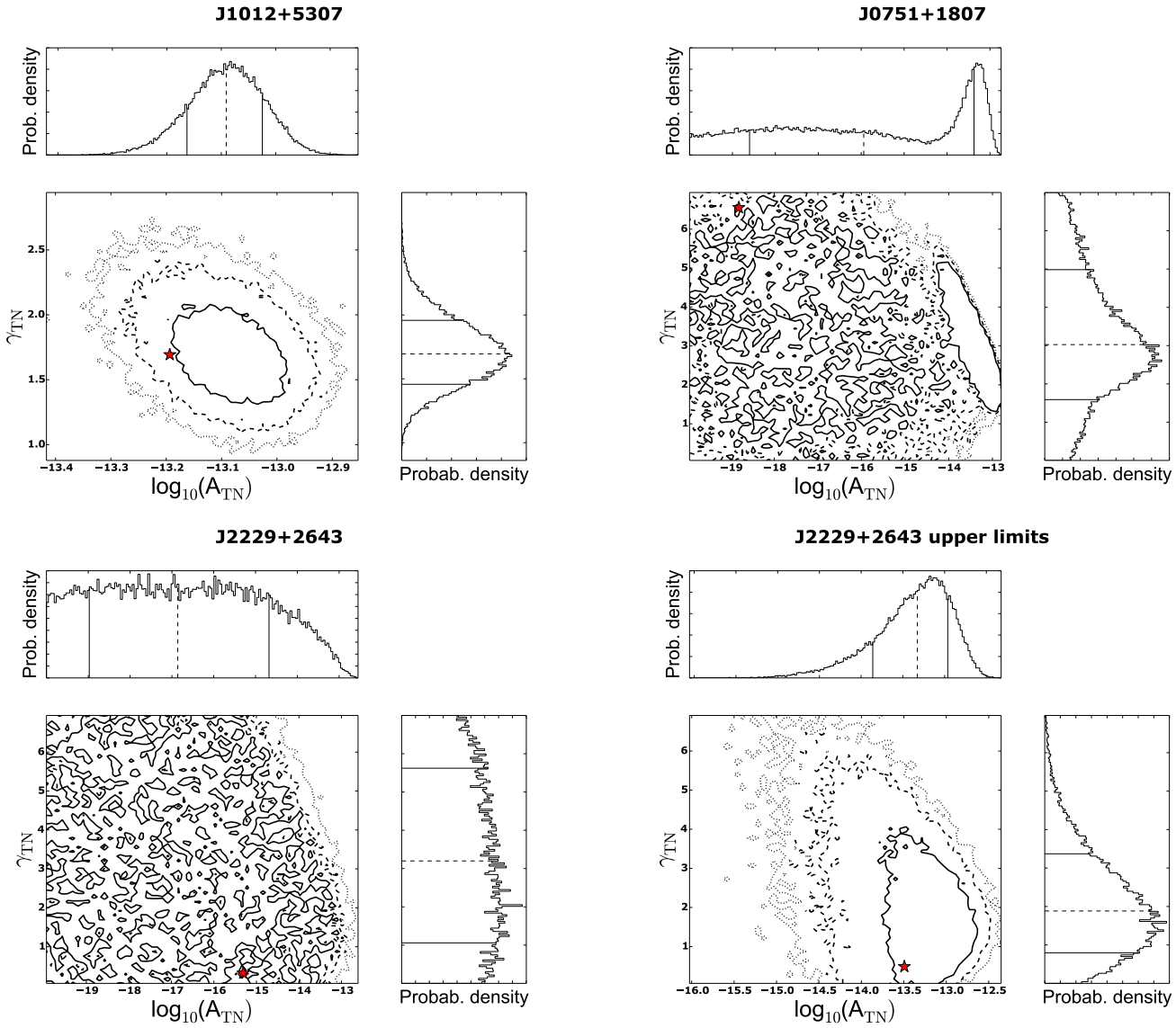

Figure 1. Two- and one-dimensional marginalised PPDs for the timing noise parameters of three pulsars: J1012+5307, J0751+1807, and J2229+2643. In the two-dimensional distributions, the solid, dashed and dotted contours represent the 68%, 95% and 99.7% (1-, 2- and 3-σ) confidence intervals and the red star marks the maximum likelihood solution. The 1-dimensional distributions have the median and 1-σ uncertainties marked as dashed and solid lines respectively. For J2229+2643, the right figure shows the distribution of the noise parameters from the upper limits analysis. Note the different ranges on the amplitude axes. See text in Section 4.1 for discussion.

J0751+1807 (semi-constrained PPDs case). One can see the strong covariance between the two parameters. The data support that the TN amplitude is more likely to be very low (the TN tail has more probability than the DM tail), however, there is still a non-zero probability that the DM variations signal is weaker than the ML model suggests. For well-constrained PPDs, DM variations and TN are decoupled, as seen in the right panel of Figure 2 for the case of J1012+5307, and the DM ML waveform subtraction is more reliable. If the statistical assumptions of the Bayesian and frequentist analysis are valid, the results for the TN of pulsars with well-constrained PPDs should be consistent between the two methods.

4.2 Timing-Noise Parameters

Out of the 42 sources, the Bayesian analysis resulted in well-constrained PPDs for both the amplitude and the spectral index of the TN power-law model for eight sources. For these, the Bayesian ML and median values are always consistent at the 1-σ level. The two methods are always consistent at the 1-σ level for the spectral index, while for the amplitude, three sources show deviations, though consistency remains at the 2-σ level. (Figure 3, top row).

For 17 MSPs, the PPDs of at least one of the timing noise parameters is semi-constrained. The Bayesian ML and median values show inconsistencies at the 1-σ level in four pulsars (Figure 3, middle row). The power-spectral analysis

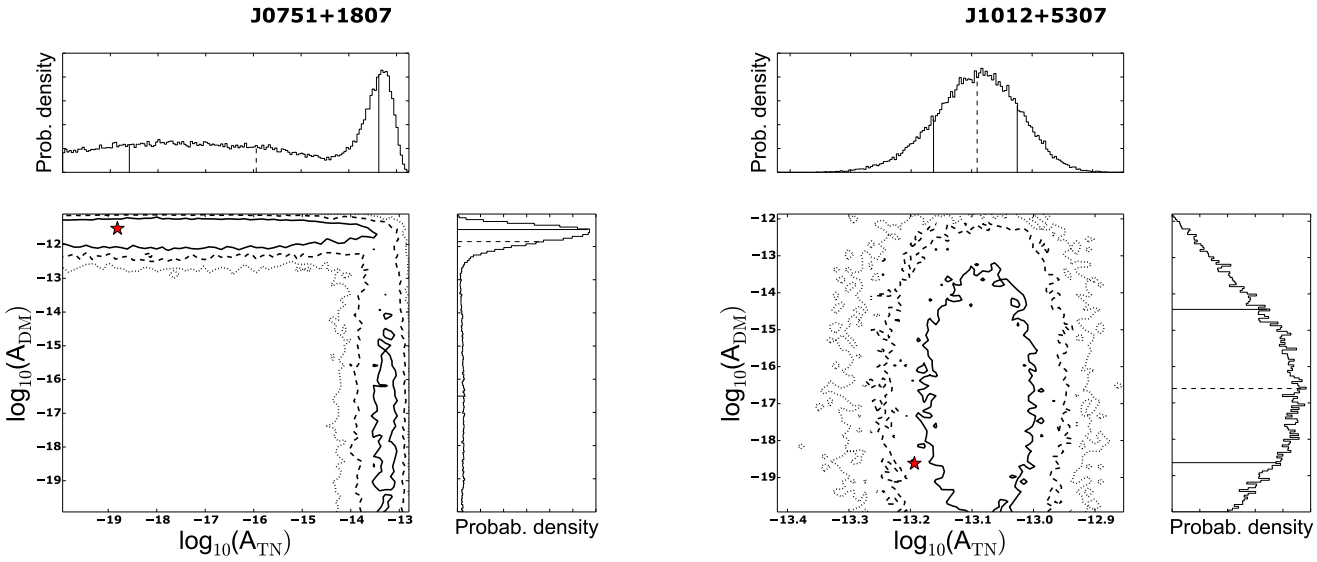

Figure 2. Two- and one-dimensional marginalised PPDs for the timing noise and DM variations amplitudes for J0751+1807 and J1012+5307. In the two-dimensional distributions, the solid, dashed and dotted contours represent the 68%, 95% and 99.7% (1-, 2- and 3-σ) confidence intervals and the red star marks the maximum likelihood solution. The 1-dimensional distributions have the median and 1-σ uncertainties marked as dashed and solid lines respectively. Note the different ranges on the amplitude axes. See text in Section 4.1 for discussion.

results are in agreement with the Bayesian median values. All Bayesian upper limits are in agreement with the rest of the results. We note that for PSR J1909−3744, we did not achieve a 3-σ measurement for spectral index with the power-spectral analysis.

The rest of the sources, 17 in total, show flat, unconstrained PPDs. The bottom row of Figure 3 shows the 95% C.L. upper limits from the two methods. Given the low significance of the TN measurement in these cases, inconsistencies in the amplitudes do not have statistically significant effects on the timing solutions when using the total covariance matrix to perform GLS timing solutions.

The agreement between the two methods for the sources with statistically significant TN measurements, supports the confidence in the methods and the results. When covariances between noise properties cannot be decoupled by the data, the interpretation of the results requires more attention. For this reason, we propose that cross-checks of the results with various methods should become common practice.

4.3 White-Noise Parameters

Radiometer noise estimation is typically robust when the pulse has a medium to high signal-to-noise ratio (S/N) (Taylor 1992), so EFACs are expected to be close to unity for most observing systems. The EQUADs results indicate for which observing systems there may be additional scatter in the residuals from physical processes related to the pulsars (e.g. pulse phase jitter) or RFI.

Figure 4 shows the distribution of the ML EFAC values. As expected, the distribution strongly peaks around unity. A few systems show EFAC values up to ∼5. These are typically high-frequency observations with very weakly detected pulses. The cases where EFACs take values significantly lower than one are either due to strong overestimation of the uncertainties or when a system's EFAC and EQUAD are highly correlated.

We examine in a similar way the distribution of EQUAD values. Figure 5 shows the distribution of the measured ML EQUAD values from the analysis using log-uniform EQUAD pPDs, and the distribution of their upper limits. As expected, in the vast majority of cases, the EQUADs are much below the TOA precision, which typically ranges from 0.5-10μs (D15).

We have examined the EQUAD PPDs from the analysis with log-uniform pPDs to determine the cases where EQUADs have well-constrained PPDs and therefore show measurable EQUADs. For some of these cases, this could reflect signs of jitter noise present in the data. We list these pulsars and observing systems in Table 4. We note that there are 49 cases where the EQUAD PPDs are semi-constrained and significantly covariant with EFACs, and therefore cannot be considered as significant EQUAD measurements. From Table 4 we can see that the vast majority of EQUADs come from L-band systems, which typically have the most sensitive data. For each pulsar there are usually only one or two systems with clear EQUAD measurement with the exception of PSR J1022+1001. This source is known to require a high level of polarimetric calibration (van Straten 2013) and to show phase jitter noise (Liu et al. 2015). Only part of the NRT data were calibrated and this may explain the high levels of EQUADs in this source. We stress once again, that more detail investigation is required to comment on the origin of the EQUAD measurements. It is likely that EQUADs could reflect additional scatter in the residuals from instrumental instabilities or analysis systematics, which could explain the EQUAD measurements in systems where the TOA precision is too low to expect any measurements of pulse

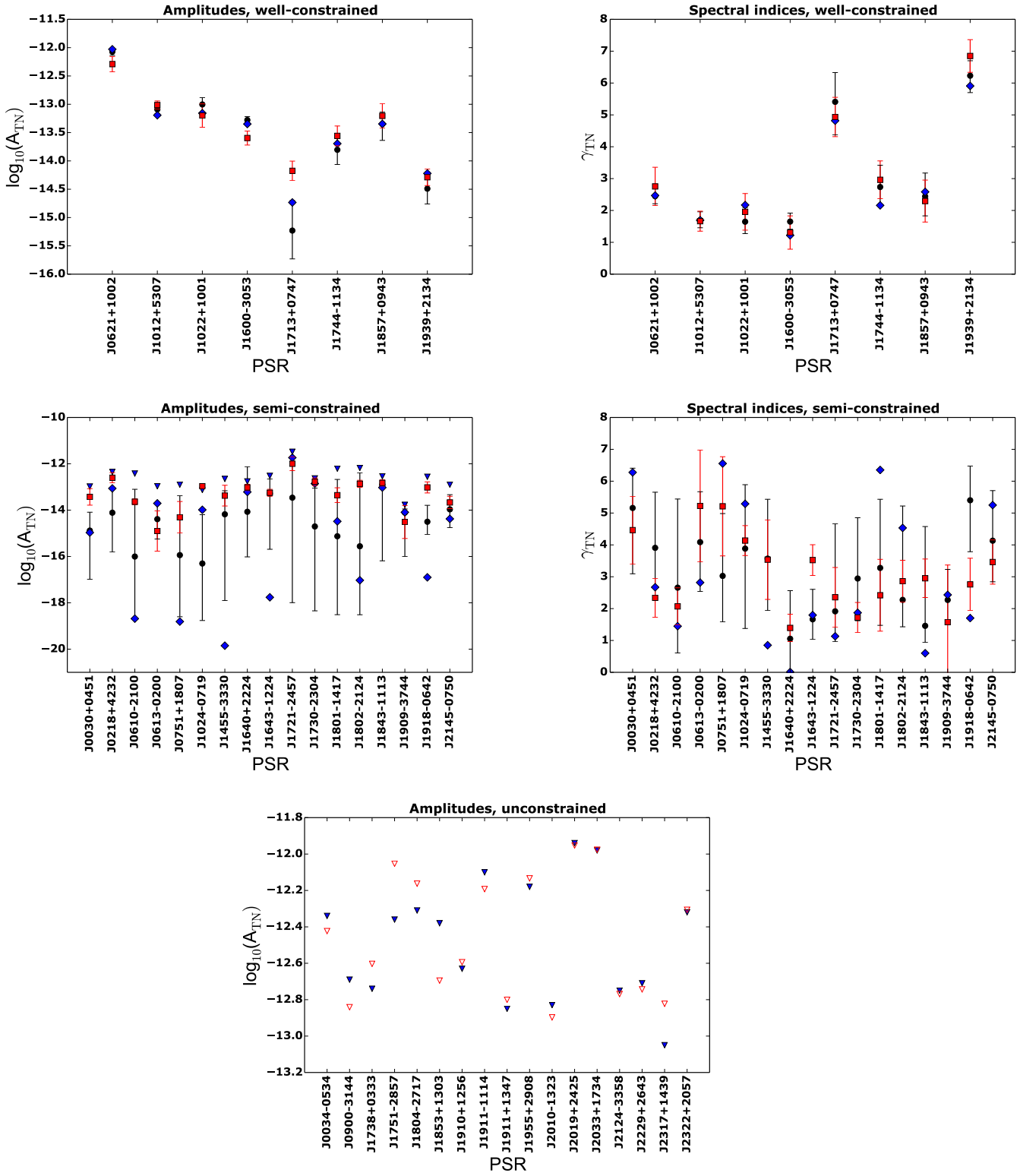

Figure 3. Comparison of the timing noise parameters estimated with the Bayesian (blue, filled diamonds for the maximum likelihood values, black, filled circles for the median values with 1-σ error bars and blue, filled triangle for upper limits) and frequentist method (red, filled squares and red, open triangles for upper limits). *Top Row*: Results for the cases where the Bayesian analysis resulted in well-constrained posterior probability distributions for both parameters. *Middle Row*: Results for the cases where the Bayesian analysis resulted in semi-constrained posterior probability distributions for at least one of the parameters. *Bottom Row*: Results for the cases where the Bayesian code resulted in unconstrained posterior probability distributions for at least one of the parameters.

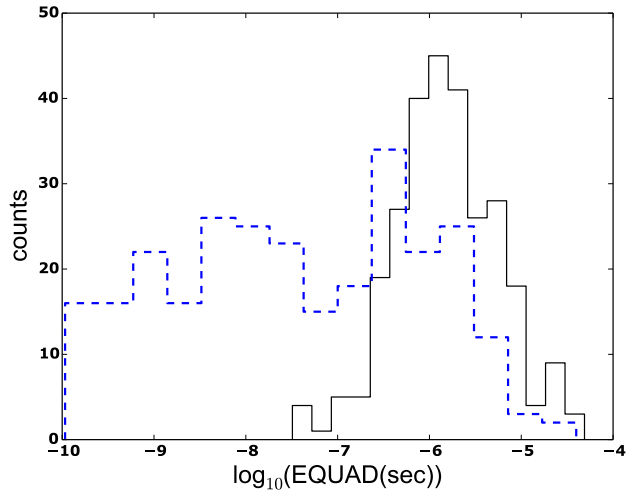

Figure 4. Distribution of EFAC values for all MSPs and observing systems. The black, solid line refers to the results of the Bayesian analysis for which the EQUAD priors are set to uniform to get their upper limit values, while the blue, dashed line is for the analysis were EQUAD priors are uninformative log-uniform.

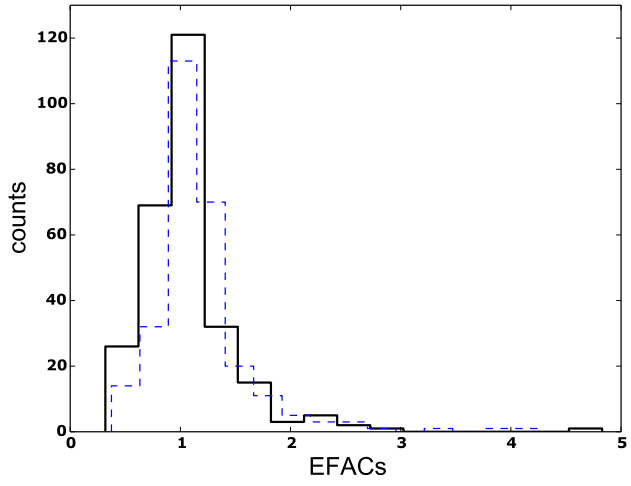

Figure 5. Distribution of EQUAD values for all MSPs and observing systems. The solid line refers to the results of the Bayesian analysis for which the EQUAD priors are set to uniform to get their upper limit values, while the dashed line is for the analysis were EQUAD priors are uninformative log-uniform.

jitter noise (as in the case e.g. of PSR J2033+1734, see Table 4.)

5 TIMING NOISE FROM INDIVIDUAL OBSERVING SYSTEMS

For MSPs which have large enough data span with overlapping data from various observing systems we examine whether part of the measured TN is present only in specific observing systems. We perform the noise analysis on selected pulsars with data from one telescope removed at a time. For the Effelsberg data, this is more complicated for many MSPs where it is the only telescope with data in the first half of the data set, so removing its data automatically means a loss of about half the data span. We note that this test may not be feasible in some cases with this data set, e.g. when a significant fraction of the residuals sensitivity to the TN is lost when removing a set of dominant, very precise data points. When the TN was absent after removing data from one telescope, we confirmed that the rest of the data would be sufficient to detect the noise by simulating realisations of the new data and performing the noise analysis after injecting TN with the measured properties.

Our analysis shows evidence for TN specific to the NRT data. Figure 6 shows the PPDs for the TN parameters when using the full data set and when excluding the NRT data, and the respective ML TN waveforms. For PSR J1022+1001 the PPDs become significantly broader when excluding the NRT data. The mean value of the amplitude reduces by two orders of magnitude and the TN waveform becomes smoother, although the waveform has almost unchanged peak-to-peak variations. The TN parameters PPDs of PSR J2145−0750 show a bimodality, which is not present when removing the NRT data. The two TN waveforms are almost identical, apart from the fact that the waveform of the full data set shows a bump around MJD 56000, which is not present when removing the NRT data. These effects are most likely caused either by additional noise in the NRT data from instrumental instabilities or by some additional non-instrumental noise component that only the NRT data are sensitive to, having indeed the highest precision TOAs. We

Table 4. List of the pulsars and observing systems for which we have well-constrained posterior probability distributions for the EQUADs. The last column shows the EQUADs maximum likelihood values from a Bayesian analysis with log-uniform EQUAD prior distribution. The telescope and backend acronyms are as introduced in Section 2.

PSR J-Name	Telesc.	Backend	Freq. (MHz)	$\mathrm{EQUAD_{ML}}$ (μs)
J0751+1807	EFF	EBPP	1360	5.0
J1012+5307	EFF	EBPP	1360	3.4
J1022+1001	JBO	DFB	1520	1.4
	NRT	BON	1400	1.3
	EFF	EBPP	1410	3.9
J1643−1224	JBO	DFB	1520	2.5
J1744−1134	JBO	DFB	1520	1.0
J1857+0943	NRT	BON	1400	0.9
J1939+2134	NRT	DDS	1400	0.3
	EFF	EBPP	1410	0.3
J2033+1734	NRT	BON	1400	25
J2145−0750	NRT	BON	2000	0.3
	JBO	DFB	1520	0.9

stress that since we have assumed the TN to be stationary, the properties of instrumental noise during a specific time-interval can leak into the estimated TN waveform throughout the pulsar data set. We note that there were known instrumental instabilities at the NRT during the period between MJD 54300-54500 (July 2007 to February 2008).

This analysis can be better performed using the International Pulsar Timing Array (IPTA) data set (Verbiest et al., submitted) where data from another 3 telescopes are included, offering a larger amount of multi-telescope overlapping data. The presence of observing system-dependent noise is more extensively investigated in the paper examining the noise properties of the IPTA data set (Lentati et al., submitted.).

6 TIMING NOISE FROM ERRORS IN TERRESTRIAL TIME STANDARDS

During pulsar timing observations, the TOAs are referenced against the local atomic clock (e.g. Hydrogen maser clock) or a Global Positioning System (GPS) clock[5]. These clocks are stable on timescales of weeks, allowing good phase keeping (1-pulse-per-second signal) during observations. These clocks, however, show instabilities on timescales of months to years and the TOAs recorded using them, are therefore not suitable for high-precision pulsar timing projects. This problem can be mitigated through the application of a series of corrections based on monitoring the offsets between pairs of clocks (see e.g. Hobbs et al. 2006, 2012).

Cross-correlating the pulse profiles with the template profile references all arrival times to the same (arbitrary) phase, forming the topocentric TOAs. Unless the time-stamping was performed using a GPS clock, the TOAs are then converted to GPS-based Universal Coordinated Time (UTC) time, using clock correction files, created by calculating the difference between the local atomic clock and the GPS times. This is then converted to UTC and subsequently to the International Atomic Time (TAI) standard. TAI is formed by the weighted average of the time-scales of several hundred atomic clocks around the world and subsequent frequency adjustments using primary frequency standards. These adjustments are made over timescales of years, a process known as "steering". As a result, TAI can have errors during the steering periods which are never retroactively corrected. For these reasons, for pulsar timing we use the corrections on TAI provided by the Bureau International des Poids et Mesures (BIPM)[6]. These corrections are made through measuring offsets between various clock pairs to achieve the best possible precision and are regularly updated.

Any possible remaining errors in the BIMP terrestrial time standard or error propagated to the TOAs by systematics when referencing the TOAs to the various time standards, will lead to a "clock error" signal, a monopolar correlated signal in the PTA sources, i.e. a signal with the same waveform in all pulsars and observing systems. As discussed in Tiburzi et al. (2016), the mitigation of the clock error signal is of central importance in PTA efforts for GW detection.

[5] This is the case for the NRT data
[6] http://www.bipm.org/

In this section, we search for a terrestrial clock error in the data set to determine how much of the measured noise can be attributed to clock error noise. Previously, Hobbs et al. (2012) presented their measurement of the clock error using data from the PPTA and discussed how pulsars can serve as an independent, non-terrestrial time standard.

6.1 Methodology and results

We use a maximum likelihood estimator to infer the clock error signal. The clock-error noise is modelled as a red-noise process power-law with power spectral density described by Eq. (3), with amplitude $A_{\rm clk}$ and spectral index $\gamma_{\rm clk}$. Using the results on these parameters, we subsequently construct the ML signal waveform.

For this analysis, we set the TN parameters of the MSPs to the ML values from the Bayesian analysis with uniform pPDs on the TN amplitude (as described in Section 3.2). In this way, we derive the ML solution for the clock-error noise with the higher possible amplitude, given our TN results. We use the residuals after subtracting the ML DM variations signal as in Section 3.3, to focus on the TN only. The likelihood function is similar to Eq. (8) but with the extension to multiple pulsars to investigate the clock signal, which is identical among all pulsars, as:

$$L \propto \frac{1}{\sqrt{|\mathbf{C}|}} e^{-\frac{1}{2}\sum_{i,j,I,J}(t_{I,i}-\tau_{I,i})C^{-1}_{I,J,i,j}(t_{J,j}-\tau_{J,j})}, \quad (15)$$

where the index I, J are for pulsars, and index i, j are for the time epoch. The total covariance matrix now includes the covariance matrix of the clock error signal, $\mathbf{C}_{\rm clk}$, while not including the matrix of the DM variations such that, $\mathbf{C} = \mathbf{C}_{\rm w} + \mathbf{C}_{\rm TN} + \mathbf{C}_{\rm clk}$. The intrinsic noise of pulsars is not correlated between pulsar pairs, so $\mathrm{C_{w}}_{I,J} = 0$ and $\mathrm{C_{TN}}_{I,J} = 0$ for $I \neq J$. The clock error waveform is identical in all pulsars, therefore its covariance matrix components can be expressed as $C_{\rm clk\,I,J,i,j} = C_{\rm clk}(t_i - t_j)\mathrm{C_{clk}}_{I,J}$, with $\mathrm{C_{clk}}_{I,J} = 1$ for all I,J pairs. The likelihood function shows that for the estimation of the clock noise parameters we consider both the the clock error signal on the residuals of each pulsar (autocorrelation effect) and the cross-correlation of the residuals between pulsar pairs.

We make the linear approximation of the timing model as described in van Haasteren et al. (2009), i.e. considering linear deviations of the true timing parameter values, $\boldsymbol{\epsilon}$, from the least-square-fit timing model values, $\boldsymbol{\epsilon_0}$, via the linear relation $\delta(\boldsymbol{\epsilon}) = \boldsymbol{\epsilon}$ - $\boldsymbol{\epsilon_0}$. We therefore substitute the expression for the residuals in Eq. (15), $\mathbf{t} - \boldsymbol{\tau}(\boldsymbol{\epsilon})$, with $\delta\mathbf{t} = \delta\mathbf{t}_{post} - \mathbf{M}\delta(\epsilon)$; $\delta\mathbf{t}_{post}$ are the post-fit timing residuals and $\mathbf{M}$ is the design matrix of the timing parameters. We marginalise analytically over all timing parameters and get the reduced likelihood function:

$$L \propto \frac{1}{\sqrt{|\mathbf{C}|}} e^{-\frac{1}{2}\sum_{i,j,I,J}(\delta t_{I,i})C'^{-1}_{I,J,i,j}(\delta t_{J,j})}, \quad (16)$$

with $\mathbf{C}' = \mathbf{C}^{-1} - \mathbf{C}^{-1}\mathbf{M}(\mathbf{M^T}\mathbf{C}^{-1}\mathbf{M})^{-1}\mathbf{M^T}\mathbf{C}^{-1}$. Going one step further, we split the deterministic signal between that of parameters for which we want to marginalise over (usually the timing model parameters), $\delta\mathbf{t}'$ and the signal of parameters we assume unknowns of the likelihood function (see e.g. Section 7.2). We note the latter parameters with the vector

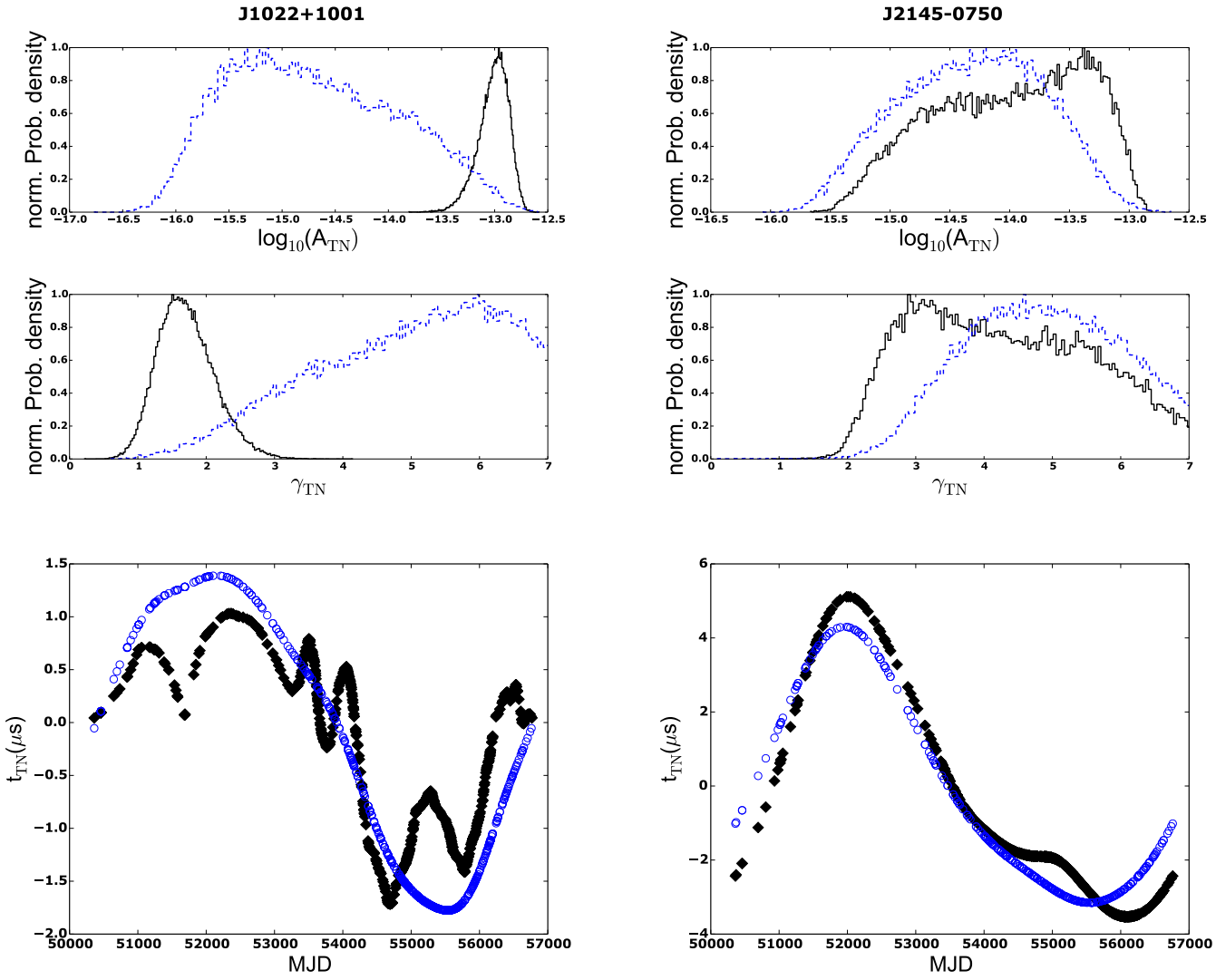

Figure 6. *Top panel:* Comparison of the 1-dimensional marginalised posterior probability distributions of the timing noise parameters when using the full EPTA data set (solid, black lines) and the data subset which does not include any NRT-BON data (blue, dashed lines). *Bottom panel:* Comparison of the timing noise waveforms ($\mathbf{t}_{\rm TN}$) when performing the noise analysis on the full EPTA data set (filled black diamonds) and the data subset which excludes the NRT-BON data (open blue circles). See Section 5 for a discussion.

lambda, and assume their waveforms to be described by the $\boldsymbol{S(\lambda)}$. The likelihood function is then re-written as:

$$L \propto \frac{1}{\sqrt{|\mathbf{C}|}} e^{-\frac{1}{2}\sum_{i,j,I,J}(\delta t'_{I,i}-S(\lambda)_{I,i})C'^{-1}_{I,J,i,j}(\delta t'_{J,j}-S(\lambda)_{J,j})}, \tag{17}$$

We sample A_{clk} and γ_{clk} over a uniform grid of values and search for the model that maximises the likelihood. The amplitude is sampled with values of $\log(\mathrm{A}_{clk})$ ranging from -17.0 to -14.0 with a step of 0.1, while the spectral index values range from 0.5 to 5 with a step of 0.1. Due to the large condition number of the clock error's covariance matrix, the individual likelihood computations are unstable. As such, the direct search for the ML solution with uniform grids produces non-desirable artefact (non-physical likelihood maxima). To avoid these effects, we performed a large number of trials by dithering noise parameters with randomised offset values within each search grid. The likelihood value of the grid is taken to be the maximum of all trials.

To reduce the computational cost of the analysis we use the "restricted data set" proposed in Babak et al. (2016). This consists of six MSPs from the full data set, which give 90 % of the sensitivity to CGWs. This "restricted data set" has also been used in the derivation of upper limits to the amplitude of GWs with the EPTA Legacy data set (Lentati et al. 2015; Taylor et al. 2015; Babak et al. 2016). The "restricted data set" contains the pulsars PSRs J0613−0200, J1012+5307, J1600−3053, J1713+0747, J1744−1134, and J1909−3744.

We find a ML solution at $\mathrm{A}_{clk} = -15.2$ and $\gamma_{clk} = 4.8$. We use these values to calculate the ML waveform of the signal, shown in Fig. 7. The clock signal waveform can be

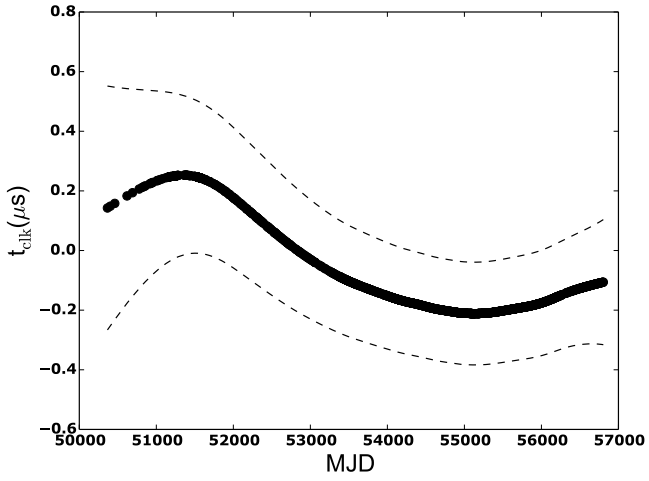

Figure 7. The estimated waveform of the clock-error noise. The filled circles is the maximum likelihood waveform ($\mathbf{t}_{\rm clk}$). The dashed lines indicate the 68% confidence intervals. For the estimation of the waveform, we used the upper limits for the values of the individual pulsar timing noise parameters providing upper limits for the clock error noise parameters.

estimated in the same way we did for TN:

$$\mathbf{t}_{\rm clk} = \mathbf{C}_{\rm clk}\mathbf{C}^{-1}\mathbf{t}\,. \tag{18}$$

with uncertainties estimated as:

$$\sigma_{\rm clk} = \mathbf{C}_{\rm clk} - \mathbf{C}_{\rm clk}\mathbf{C}^{-1}\mathbf{C}_{\rm clk} \tag{19}$$

The upper limit clock error waveform has an rms value of $0.17\mu s$. By integrating Eq. (3) from the lowest to the highest spectral frequency for the clock error noise we derive the average power of the signal. We can compare this to the average power of the noise for each MSP, which is calculated by adding the TN average power and the white noise average power ($S_{\rm W}$, as in Eq. 13). We find that the contribution of the clock error noise to the total noise levels of the individual pulsars is less than 1%.

7 EFFECTS OF TIMING NOISE ON PROSPECTS FOR GW DETECTION

Various studies have examined the sensitivity of PTAs to GWB signals (e.g. Jenet et al. 2005; Lee et al. 2012; Siemens et al. 2013). These studies focus on making detection significance estimations and projections based on analytic formulae or scaling laws, which are derived assuming a given detection technique. These estimates are usually made based on specific assumptions, such as: the TOAs are regularly sampled and simultaneous across pulsars, that the measurement precision is constant and identical for all pulsars and the absence of low frequency noise. The detection significance is usually expressed as the precision by which the dimensionless amplitude can be measured for a given spectral index.

In this paper we make use of the Crámer-Rao lower bound (CRLB) to investigate the limitations of using the present data set in detecting GWs, both for stochastic isotropic GWBs and CGWs from SMBHBs. The advantage of this method is that it takes into account all the observational properties of the data, such as cadence, white and TN levels, while still using analytic calculations that demand very few computational resources and does not require data simulations. The impact of the TN present in the data on the PTA's sensitivity to GWs can then be estimated by comparing the CRLB when using the full covariance matrix and when omitting the TN component.

The CRLB states that, for any unbiased estimator, the variance is equal to or higher than the inverse of the Fisher information matrix, $\boldsymbol{\mathcal{I}}$. When the equality is valid the estimator is also "fully efficient" (Fisz 1963). As discussed in Vallisneri (2008), the ML estimator (which we use in this analysis for the GW amplitude as described below) can achieve the bound in the high S/N regime. For the amplitude of GW signals, the CRLB represents the lowest uncertainties (in case of detection) or upper limits (in case on non-detection) any unbiased estimator can achieve. We note, that although the CRLB is underperformed by all unbiased estimator, in principle it can be outperformed by a biased estimator (Vallisneri 2008). The interpretation of the bound as the amplitudes upper limit in the non-detection case warrants more caution, since by default it assumes we are outside the high S/N regime. Nevertheless, it is unlikely that other estimators can provide lower upper limits than the CRLB under the same assumptions. For the purpose of evaluating the role of TN on the data's sensitivity to GWs, we are primarily interested in the ratio of the CRLB when assuming only white noise in the data and when the TN is taken into account. Therefore, even if the individual CRLB results are not optimal, their ratio should be representative of the effects of TN. The CRLB calculated in the presence of TN are in fact comparable[7] to the amplitude limits derived in Lentati et al. (2015) and Babak et al. (2016) using more rigorous algorithms.

In its general form, the CRLB is formulated as follows. Given a likelihood function, $f(\boldsymbol{\lambda}, \mathbf{x})$, where $\mathbf{x}$ is the data and $\boldsymbol{\lambda}$ are the model parameters, the CRLB is:

$$\mathrm{Cov}(\boldsymbol{\lambda}) = \langle \sigma_{\lambda_i}\sigma_{\lambda_j} \rangle \geq \boldsymbol{\mathcal{I}}^{-1}_{ij}, \tag{20}$$

where the indices i and j denote the different parameters and $\mathcal{I}_{ij}$ is:

$$\mathcal{I}_{ij} = \left\langle \frac{\partial \ln f(\boldsymbol{x},\boldsymbol{\lambda})}{\partial\lambda_i}\frac{\partial \ln f(\boldsymbol{x},\boldsymbol{\lambda})}{\partial\lambda_j} \right\rangle \equiv -\left\langle \frac{\partial^2 \ln f(\boldsymbol{x},\boldsymbol{\lambda})}{\partial\lambda_i\partial\lambda_j} \right\rangle \tag{21}$$

It is well-known that $\mathcal{I}$ can be analytically calculated for Gaussian likelihood functions (as is Eq. 8), and results in the so-called Slepian-Bangs formula (Slepian 1954; Bangs 1971):

$$\mathcal{I}_{ij} = \frac{1}{2}\left\{ \mathbf{tr}\left[\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial\beta_i}\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial\beta_j} \right] + \frac{\partial \mathbf{S}(\lambda)}{\partial\lambda_i}\mathbf{C}^{-1}\frac{\partial \mathbf{S}(\lambda)}{\partial\lambda_j} \right\}. \tag{22}$$

[7] Note that the CRLB refers to the equivalent of a 68% C.L. Typically, the 95% C.L. is reported in the PTA literature for the amplitude of GWs.

Here, β_i are the model parameters describing the covariance matrix, λ_i, are the parameters describing the unknown waveform $\mathbf{S}$ and $\mathbf{tr}$ is the matrix trace.

We make use of the same maximum likelihood estimator as in Section 6 (Eq. 17), but we replace the stochastic clock error signal with that of a stochastic and isotropic GWB and we set $\boldsymbol{S}(\boldsymbol{\lambda})$ to be the CGW signal from a single SMBHB, as detailed in Section 7.2. The likelihood function (Eq. 17) uses a total covariance matrix which includes the covariance matrix of the GWB, such that $\mathbf{C} = \mathbf{C}_{\rm w} + \mathbf{C}_{\rm TN} + \mathbf{C}_{\mathbf{gwb}}$. The GWB's covariance matrix, is dictated by the expected correlation coefficient in the residuals of every pulsar pair, described by the overlap reduction function (Finn et al. 2009), $\Gamma(\zeta)$, defined as:

$$\Gamma(\zeta) = \frac{3}{8}\Big[1 + \frac{\cos\zeta_{IJ}}{3} + 4(1-\cos\zeta_{IJ})\ln\Big(\sin\frac{\zeta_{IJ}}{2}\Big)\Big](1+\delta_{IJ}). \tag{23}$$

Here, ζ_{IJ} is the angular separation between the I-th and the J-th pulsar, and δ_{IJ} is the Kronecker delta. In principle, both an Earth and a pulsar term contribute to the correlation and δ_{IJ} accounts for the latter. In the short-wavelength approximation, i.e. when the pulsars are separated from the Earth and from each other by many GW wavelengths, the overlap reduction function is also known as the Hellings-Downs curve (Hellings & Downs 1983). The components of the covariance matrix of the GWB are then expressed as $\mathrm{C}_{\mathrm{gwb}I,J,i,j} = \mathrm{C}_{\mathrm{gwb}}(t_i - t_j)\Gamma(\zeta_{IJ})$. As in the case of the clock error covariance matrix (Section 6.1), the form of the covariance matrix allows the calculation of the CRLB to include both the autocorrelation and cross-correlation effects of the GW.

For this analysis, we use the same six MSPs that we used to estimate the clock error noise parameters in Section 6.1 and we set the TN properties to their ML values as estimated with the Bayesian pulsar noise analysis described in Section 3.2 and presented in Table 2. As discussed in Section 6.1, the estimation that the sensitivity loss to GWs when using this data subset is below 10 % was made for the case of CGWs. For low-frequency stochastic signals such as the GWB or the clock error signal, the sensitivity loss should be less. For CGWs, adding a pulsar with precise data only in part of its data span can increase the S/N of a detection significantly if the SMBHB orbit is fully sampled. In the case of the GWB, however, the targeted correlated signal must be found in cross-correlations of TOAs across a long timespan of order equal to the inverse of the GW frequency, with sufficient precision. We have verified this by calculating the CRLB for the GWB using 40 MSPs and noting an improvement in the amplitude limit of order 2 %. The scaling of the sensitivity to GWs with the number of MSPs, the S/N regime of the targeted signal and other factors have been studied elsewhere (e.g. Babak & Sesana 2012; Siemens et al. 2013) and is outside the scope of this work.

In order to focus on the impact of TN only, we mitigate the DM variations beforehand by subtracting the ML DM variations waveforms from the residuals. For detailed derivations and astrophysical interpretations on GW limits using the EPTA Legacy data set, we refer the reader to Lentati et al. (2015), Taylor et al. (2015) and Babak et al. (2016) for the cases of a stochastic and isotropic GWB, the anisotropy in the GWB and the CGW from individual SMBHBs respectively.

7.1 Stochastic Gravitational-Wave Background

When estimating the CRLB for the GWB amplitude, the terms with partial derivatives of $\mathbf{S}$ are zero and Eq. (22) reduces to

$$\mathcal{I}_{ij} = \frac{1}{2}\mathbf{tr}\Big[\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial \beta_i}\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial \beta_j}\Big]. \tag{24}$$

We calculate the CRLB for the GWB amplitude, keeping each time the GWB spectral index fixed. We do so for a range of spectral indices, from -2 to 1, which covers GWB signals often discussed in PTAs literature, e.g. from SMBHBs, cosmic strings and the relic GWB from the inflationary era.

This simplified approach intends to provide an understanding of the difficulties the TN imposes on the detection of the various GWBs probed by PTAs. It is not exhaustive, since each of these GWBs can in general have a range of possible spectral index values. In the case of SMBHBs, this depends on the orbital eccentricities and whether the SMBHBs are coupled to their stellar and gaseous environment or they are driven by GW emission only (Sesana 2013). The often used power-law index of $-2/3$ refers to circular, GW-driven SMBHBs (Rajagopal & Romani 1995; Jaffe & Backer 2003). Strong environment coupling and high orbital eccentricities can cause a turnover of the spectrum at low-frequencies (e.g. Fig. 2 in Sesana 2013). The value -7/6 we have used for the spectral index of the cosmic string GWB has been analytically derived using a simplified approximation of the loop number density and assuming cusp emission (e.g. Damour & Vilenkin 2005). However, especially in the frequencies probed by PTAs, a wide range of spectral indices is possible, depending on some characteristic parameters used to describe the evolution of the network and the details of the dominant GW emission mechanism, and one typically sets limits on the amplitude for a range of these parameters (Sanidas et al. 2012) For the cosmological relic GWB, a spectral index of -1 is often cited (Grishchuk 2005). For more details on the sources of the various GWBs and details on the derivation of amplitude limits as function of the spectral index and other physical parameters, we refer the reader to Lentati et al. (2015); Arzoumanian et al. (2015)

The CRLBs are calculated using the TN parameters from the two Bayesian analyses, using different types of pPDs on the TN noise amplitude. For each set of TN results, we calculate the CRLB for two cases, namely assuming the presence of the measured white and TN, or assuming only the measured white noise levels, and finally, calculate their ratios. Figure 8 shows the results for both cases. The results for the spectral indices representative of GWBs from SMBHBs, cosmic strings and relic GWs are presented in Table 5. The improvement factor on the lower bound when assuming no TN in the data is always more than an order of magnitude, ranging from 9.1 to 11.4. These results demonstrate how strongly TN can reduce the data's sensitivity to GWs. To stress this even further, we note that the upper limits on the GWB amplitude by SMBHBs (spectral index $-2/3$)

Table 5. Results for the Crámer-Rao bound (CRLB) on a GWB for the expected signals from SMBHBs ($\alpha = -2/3$), cosmic strings ($\alpha = -7/6$) and cosmological relic GWs ($\alpha = -1$) (see Section 7.1 for a discussion on the noted spectral indices). We tabulate the CRLB assuming both the measured white- and timing-noise levels, ($A_{GWB_{\rm wr}}$) for measured and upper limit values and white-noise levels only ($A_{GWB_{\rm w}}$).

α_{GWB}	$A_{GWB_{\rm wr}}$	$A_{GWB_{\rm w}}$	$\frac{A_{GWB_{\rm wr}}}{A_{GWB_{\rm w}}}$
Max. likelihood TN			
-2/3	8.3×10^{-16}	9.1×10^{-17}	9.1
-7/6	4.6×10^{-17}	4.4×10^{-18}	10.3
-1	3.5×10^{-16}	3.5×10^{-18}	10.0
Red noise upper limits			
-2/3	9.2×10^{-16}	8.5×10^{-17}	10.7
-7/6	4.7×10^{-17}	4.1×10^{-18}	11.4
-1	3.7×10^{-16}	3.3×10^{-17}	11.1

by PTAs have improved by a factor of ten over the past ten years.

7.2 Gravitational Waves from single SMBHBs

Here we focus on CGWs from resolvable, GW-driven SMBHBs with circular orbits and without measurable frequency evolution of the signal over the observing interval due to energy loss from the binary by GW emission (an effect known as frequency chirping, see e.g. Hughes (2009)). The waveform ($\mathbf{S}$) of CGWs has been calculated by many independent studies (e.g. Wahlquist 1987; Blanchet 2006; Hughes 2009). For each SMBHB, the waveform is characterised by seven parameters, namely the GW amplitude, frequency and phase, the SMBHB's sky co-ordinates (right ascension and declination), orbital inclination, and direction of the binary's ascending node on the sky. Clearly, the terms with partial derivatives of $\mathbf{C}$ are zero for the single SMBHB signal and Eq. (22) reduces to

$$\mathcal{I}_{ij} = \frac{1}{2}\frac{\partial \mathbf{S}}{\partial \lambda_i}\mathbf{C}^{-1}\frac{\partial \mathbf{S}}{\partial \lambda_j}\,. \tag{25}$$

Due to the seven parameters, the covariance matrix for the single GW source is a 7×7 matrix. The CRLB of the single source amplitude depends on the GW frequency, source position, orbital inclination and orientation. It has been shown (Lee 2013) that the precision estimation of the GW source position using CRLB would be poor, due to the lack of a unique un-biased estimator for the single source problem. The statistics of the amplitude estimator, on the other hand, can be well described by the CRLB, which determines the sensitivity of a PTA as function of frequency. The sensitivity depends on the GW source position. We estimate the CRLB for three scenarios: placing the SMBHB at the sky position where the PTA has the minimum and maximum sensitivity as well as the average of all positions on the sky. Our results are given in Fig. 9. The low-frequency sensitivity extends to values lower than the frequency resolution (1/T) because the GW low frequency signal still leaks power into the observing window after the pulsars' spin and spin-down fitting.

Table 6. Results for the Crámer-Rao lower bound (CRLB) on the strain amplitude of continuous GWs from resolvable SMBHBs with circular orbits and without measurable frequency chirping. We quote the limits for the cases when the SMBHB is at the sky location where the PTA has the maximum (max) and minimum (min) sensitivity, and the average of all sky positions (avg) at GW frequencies of 5 and 7 nHz. For each case we quote the limits when accounting for the white and the TN of the data, $A_{\rm CGW_{wr}}$ and for the white noise only, $A_{\rm CGW_{w}}$. The last column shows the ratio of the limits for these two cases.

GW freq. (nHz)	$A_{\rm CGW_{wr}}$	$A_{\rm CGW_{w}}$	$\frac{A_{\rm CGW_{wr}}}{A_{\rm CGW_{w}}}$
	Max PTA	sensitivity	
5	1.2×10^{-14}	2.1×10^{-15}	5.6
7	9.1×10^{-15}	3.8×10^{-15}	2.4
	Avg PTA	sensitivity	
5	4.0×10^{-15}	8.1×10^{-16}	5.0
7	2.7×10^{-15}	1.1×10^{-15}	2.4
	Min PTA	sensitivity	
5	1.3×10^{-15}	2.4×10^{-16}	5.3
7	1.0×10^{-15}	4.4×10^{-16}	2.3

This causes the curve to rise below the frequency resolution. The rise of the curve at high GW frequencies is due to the PTA frequency response, as the GW induced timing residuals are the time integral of the GW strain. The peak at 1 $\rm yr^{-1}$ (3.17×10^{-8} Hz) is caused by the pulsar sky position fitting.

The improvement in the PTA sensitivity at low frequencies is obvious from Fig. 9. One can clearly notice how the presence of TN flattens the sensitivity below ∼10 nHz, which, in contrast, keeps improving in the case of timing data free of TN. In the absence of TN, the sensitivity at low GW frequencies is only limited by the PTA's frequency resolution. Table 6 summarises the CRLBs for the CGWs amplitude at frequencies of 5 and 7 nHz and the improvement factors to the sensitivity when the data do not have TN, which range from 2.3 to 5.6.

8 CONCLUSIONS

In this paper, we have characterised the noise properties for 42 MSPs, using the EPTA Legacy data set. While the central focus is on the timing noise properties, we have also characterised the white noise properties of the data. The long time-spans of the pulsar data sets (the shortest being 6.9 years and the longest 24.1 years long) of high-quality timing data, are especially valuable for determining the timing noise. In order to increase our confidence in the results, we have employed two established methods, one based on Bayesian and the other one on power-spectral analysis. We used the Bayesian pulsar timing analysis package `TempoNest` to simultaneously determine the time-correlated timing noise, DM variations and uncorrelated noise (white-noise) properties. In order to focus the comparison between the methods on the timing noise characterisation, we used the maximum likelihood `TempoNest` results on DM variations

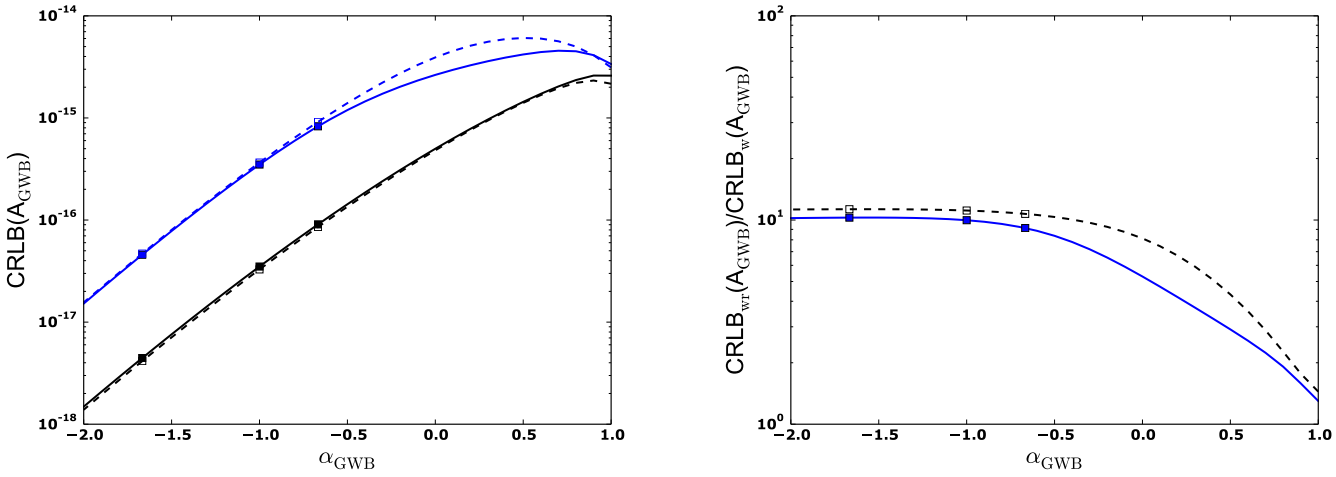

Figure 8. Crámer-Rao lower bounds (CRLBs) for the strain amplitude of GWBs, A_{GWB} for a range of spectral indices, α_{GWB}. Squares denote the values shown in Table 5. *Left panel:* CRLBs calculated bounds using the TN maximum likelihood parameters from the Bayesian analysis using log-uniform priors on the TN amplitude (solid lines and filled squares) and using uniform priors on the TN amplitude (upper limits on TN; dashed lined with open squares). Blue symbols are for limits calculated assuming timing and white noise, while black symbols when only the white noise levels are taken into account. *Right panel:* The ratio of the CRLBs for when assuming timing and white-noise and white noise only in the data. The blue solid line is when using the timing noise properties the analysis with log-uniform priors and the while the black dashed line is for the analysis with uniform priors.

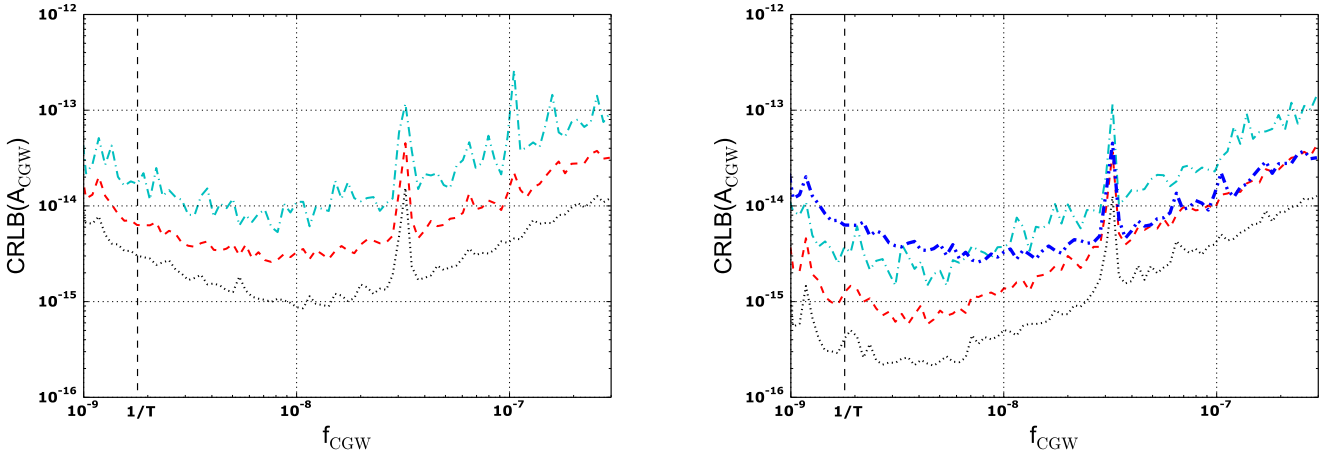

Figure 9. Results for the Crámer-Rao lower bound (CRLB) on the strain amplitude of continuous GWs, A_{CGW}, against the CGW frequency, f_{CGW}, from resolvable SMBHBs with circular orbits and without measurable frequency chirping. The different curves are for the cases where the SMBHB is at the sky location where the PTA has the maximum (cyan, dot-dashed lines) and minimum (black, dotted lines) sensitivity, and the average of all positions on the sky (red, solid lines). The vertical line show the frequency resolution of the PTA, 1/T, where T is time-span of the pulsar with the longest data set. *Left panel:* Sensitivity curves when accounting for the white and the timing noise of the data. *Right panel:* Sensitivity curves when only accounting the white noise of the data. The additional blue, thick double dot-dashed line is the case for mean PTA sensitivity when including the timing noise as in the left panel (red, solid line) for better comparison.

and white-noise parameters as a priori known information when performing the frequentist analysis, based on a developed power-spectral analysis code described in this paper. For pulsars with statistically significant timing noise measurements, the two methods give statistically consistent results.

The lack of sufficient multi-frequency data in 17 pulsars where timing noise is detected leads to strong covariances between the timing noise and DM variations, causing the posterior distributions of the noise parameters derived from the Bayesian analysis to have probability tails extending to $\pm\infty$. These reflect the small probabilities of the noise ampli-

tude to be zero, causing some deviations between the maximum likelihood and mean values of the parameters. The values of the ML and mean parameters as well as the parameter values estimated with the power-spectral analysis, are still however statistically consistent. Upper limit analysis is performed in these cases to set robust upper limits on the timing noise amplitude.

Our analysis shows evidence of timing noise specific to the NRT data, which are likely linked to improper polarisation calibration in a roughly six-month-long epoch. We have also placed an upper limit on clock-error timing noise and find that it contributes at most 1% to the total noise in the MSPs under examination. Finally, we assessed the role of timing noise in the efforts for GW detection using PTAs. We did so by estimating the Crámer-Rao lower bound on the strain amplitude of a stochastic GWB and CGWs from resolvable SMBHBs, accounting only for the measured white noise first and then adding the measured timing noise properties. We find that, for GWBs, the timing noise in this data set reduces the sensitivity of this data set by a factor of 9.1 to 11.4, depending on the GWB spectral index. For CGWs, the sensitivity reduces by a factor of 2.3 to 5.6, depending on the GW frequency and the sky position of the SMBHB with respect to the sky position where the PTA is most sensitive.

The results of this paper stress in a clear way the imperative need of PTAs to improve the noise characterisation and mitigation techniques and the development of good observing and data reduction practices to avoid introducing timing noise due to systematics. It also demonstrates the demand for new discoveries of MSPs that are not only bright, but also exhibit stable rotation. The rotational stability of pulsars can only be evaluated via timing-noise characterisation on data sets that are at least five years long, making the long-term follow-up timing observations of newly discovered MSPs essential for PTA observing campaigns.

ACKNOWLEDGMENTS

We are grateful to Mike Keith, Bill Coles and George Hobbs for useful discussions.

Part of this work is based on observations with the 100-m telescope of the Max-Planck-Institut für Radioastronomie (MPIfR) at Effelsberg. The Nançay Radio Observatory is operated by the Paris Observatory, associated to the French Centre National de la Recherche Scientifique (CNRS). We acknowledge financial support from "Programme National de Cosmologie and Galaxies" (PNCG) of CNRS/INSU, France. Pulsar research at the Jodrell Bank Centre for Astrophysics and the observations using the Lovell Telescope is supported by a consolidated grant from the STFC in the UK. The Westerbork Synthesis Radio Telescope is operated by the Netherlands Institute for Radio Astronomy (ASTRON) with support from The Netherlands Foundation for Scientific Research NWO.

RNC acknowledges the support of the International Max Planck Research School Bonn/Cologne and the Bonn-Cologne Graduate School. KJL gratefully acknowledge support from National Basic Research Program of China, 973 Program, 2015CB857101 and NSFC 11373011. PL acknowledges the support of the International Max Planck Research School Bonn/Cologne. SO is supported by the Alexander von Humboldt Foundation. JWTH acknowledges funding from an NWO Vidi fellowship and from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Starting Grant agreement nr. 337062 ("DRAGNET"). CMFM was supported by a Marie Curie International Outgoing Fellowship within the 7th European Community Framework Programme. AS is supported by the Royal Society. This research was in part supported by ST's appointment to the NASA Postdoctoral Program at the Jet Propulsion Laboratory, administered by Oak Ridge Associated Universities through a contract with NASA. RvH is supported by NASA Einstein Fellowship grant PF3-140116.

REFERENCES

Alpar M. A., Cheng A. F., Ruderman M. A., Shaham J., 1982, Nature, 300, 728
Armstrong J. W., Rickett B. J., Spangler S. R., 1995, ApJ, 443, 209
Arzoumanian Z., et al., 2015, preprint, (arXiv:1508.03024)
Babak S., Sesana A., 2012, Phys. Rev. D, 85, 044034
Babak S., et al., 2016, MNRAS, 455, 1665
Bangs G. W., 1971, PhD thesis, Yale University
Blanchet L., 2006, Living Reviews in Relativity, 9, 4
Britton M. C., 2000, ApJ, 532, 1240
Chamberlin S. J., Siemens X., 2012, Phys. Rev. D, 85, 082001
Champion D. J., et al., 2010, ApJ, 720, L201
Coles W., Hobbs G., Champion D. J., Manchester R. N., Verbiest J. P. W., 2011, MNRAS, 418, 561
Cordes J. M., Downs G. S., 1985, ApJS, 59, 343
D'Alessandro F., McCulloch P. M., Hamilton P. A., Deshpande A. A., 1995, MNRAS, 277, 1033
Damour T., Vilenkin A., 2005, Phys. Rev. D, 71, 063510
Estabrook F. B., Wahlquist H. D., 1975, General Relativity and Gravitation, 6, 439
Favata M., 2009, Phys. Rev. D, 80, 024002
Feroz F., Hobson M. P., 2008, MNRAS, 384, 449
Finn L. S., Larson S. L., Romano J. D., 2009, Phys. Rev. D, 79, 062003
Fisz M., 1963, Probability Theory and Mathematical Statistics. John Wiley and Sons, doi:10.1080/00401706.1964.10490212
Foster R. S., Backer D. C., 1990, ApJ, 361, 300
Freire P. C. C., et al., 2012, MNRAS, 423, 3328
Grishchuk L. P., 2005, Physics Uspekhi, 48, 1235
Hellings R. W., Downs G. S., 1983, ApJ, 265, L39
Hobbs G., 2013, Classical and Quantum Gravity, 30, 224007
Hobbs G. B., Edwards R. T., Manchester R. N., 2006, MNRAS, 369, 655
Hobbs G., et al., 2012, MNRAS, 427, 2780
Hughes S. A., 2009, ARA&A, 47, 107
Jaffe A. H., Backer D. C., 2003, ApJ, 583, 616
Jenet F. A., Lommen A., Larson S. L., Wen L., 2004, ApJ, 606, 799
Jenet F. A., Hobbs G. B., Lee K. J., Manchester R. N., 2005, ApJ, 625, L123
Kaspi V. M., Taylor J. H., Ryba M. F., 1994, ApJ, 428, 713
Keith M. J., et al., 2013, MNRAS, 429, 2161
Kibble T. W. B., 1976, Journal of Physics A Mathematical General, 9, 1387
Kramer M., Champion D. J., 2013, Classical and Quantum Gravity, 30, 224009
Kramer M., Lyne A. G., O'Brien J. T., Jordan C. A., Lorimer D. R., 2006a, Science, 312, 549
Kramer M., et al., 2006b, Science, 314, 97
Lee K. J., 2013, Classical and Quantum Gravity, 30, 224016

Lee K. J., Bassa C. G., Janssen G. H., Karuppusamy R., Kramer M., Smits R., Stappers B. W., 2012, MNRAS, 423, 2642
Lee K. J., et al., 2014, MNRAS, 441, 2831
Lentati L., Alexander P., Hobson M. P., Taylor S., Gair J., Balan S. T., van Haasteren R., 2013, Phys. Rev. D, 87, 104021
Lentati L., Alexander P., Hobson M. P., Feroz F., van Haasteren R., Lee K. J., Shannon R. M., 2014, MNRAS, 437, 3004
Lentati L., et al., 2015, MNRAS, 453, 2576
Liu K., Verbiest J. P. W., Kramer M., Stappers B. W., van Straten W., Cordes J. M., 2011, MNRAS, 417, 2916
Liu K., et al., 2015, MNRAS, 449, 1158
Lorimer D. R., Kramer M., 2005, Handbook of Pulsar Astronomy
Lyne A., Hobbs G., Kramer M., Stairs I., Stappers B., 2010, Science, 329, 408
Matsakis D. N., Taylor J. H., Eubanks T. M., 1997, A&A, 326, 924
McLaughlin M. A., 2013, Classical and Quantum Gravity, 30, 224008
Rajagopal M., Romani R. W., 1995, ApJ, 446, 543
Sanidas S. A., Battye R. A., Stappers B. W., 2012, Phys. Rev. D, 85, 122003
Sazhin M. V., 1978, Soviet Astronomy, 22, 36
Sesana A., 2013, Classical and Quantum Gravity, 30, 224014
Shannon R. M., Cordes J. M., 2010, ApJ, 725, 1607
Shannon R. M., et al., 2013, ApJ, 766, 5
Shannon R. M., et al., 2014, MNRAS, 443, 1463
Shannon R. M., et al., 2015, Science, 349, 1522
Shao L., Caballero R. N., Kramer M., Wex N., Champion D. J., Jessner A., 2013, Classical and Quantum Gravity, 30, 165019
Siemens X., Ellis J., Jenet F., Romano J. D., 2013, Classical and Quantum Gravity, 30, 224015
Skilling J., 2004, in Fischer R., Preuss R., von Toussaint U., eds, American Institute of Physics Conference Series Vol. 735, Bayesian inference and maximum entropy methods in science and engineering. pp 395–405, doi:10.1063/1.1835238
Slepian D., 1954, in Transactions of the IRE Professional Group on Information Theory. pp 68–89, doi:10.1109/IREPGIT.1954.6373401
Taylor J. H., 1992, Philosophical Transactions of the Royal Society of London Series A, 341, 117
Taylor J. H., Weisberg J. M., 1989, ApJ, 345, 434
Taylor S. R., et al., 2015, Physical Review Letters, 115, 041101
Thorsett S. E., Arzoumanian Z., Camilo F., Lyne A. G., 1999, ApJ, 523, 763
Tiburzi C., et al., 2016, MNRAS, 455, 4339
Vallisneri M., 2008, Phys. Rev. D, 77, 042001
Verbiest J. P. W., et al., 2009, MNRAS, 400, 951
Wahlquist H., 1987, General Relativity and Gravitation, 19, 1101
Zhu W. W., et al., 2015, ApJ, 809, 41
van Haasteren R., Levin Y., 2013, MNRAS, 428, 1147
van Haasteren R., Levin Y., McDonald P., Lu T., 2009, MNRAS, 395, 1005
van Haasteren R., et al., 2011, MNRAS, 414, 3117
van Straten W., 2006, ApJ, 642, 1004
van Straten W., 2013, ApJS, 204, 13
van Straten W., Bailes M., 2003, in Bailes M., Nice D. J., Thorsett S. E., eds, Astronomical Society of the Pacific Conference Series Vol. 302, Radio Pulsars. p. 65