

A pulsar-based time-scale from the International Pulsar Timing Array

G. Hobbs¹,¹★ L. Guo,² R. N. Caballero,^{3,4} W. Coles,⁵ K. J. Lee,³ R. N. Manchester,¹★
 D. J. Reardon⁶,⁶ D. Matsakis,⁷ M. L. Tong,⁸ Z. Arzoumanian,⁹ M. Bailes,⁶
 C. G. Bassa,¹⁰ N. D. R. Bhat,¹¹ A. Brazier,^{12,13} S. Burke-Spolaor,^{14,15}
 D. J. Champion⁴,⁴ S. Chatterjee,¹² I. Cognard,^{16,17} S. Dai,¹ G. Desvignes,⁴ T. Dolch,¹⁸
 R. D. Ferdman,¹⁹ E. Graikou,⁴ L. Guillemot,^{16,17} G. H. Janssen,^{10,20} M. J. Keith,²¹
 M. Kerr²²,²² M. Kramer,⁴ M. T. Lam,¹⁴ K. Liu,⁴ A. Lyne,²¹ T. J. W. Lazio,²²
 R. Lynch,^{14,24} J. W. McKee⁴,⁴ M. A. McLaughlin,^{14,15} C. M. F. Mingarelli,²⁵
 D. J. Nice,²⁶ S. Osłowski⁶,⁶ T. T. Pennucci,²³ B. B. P. Perera²¹,²¹ D. Perrodin,²⁷
 A. Possenti,^{27,28} C. J. Russell,²⁹ S. Sanidas,²¹ A. Sesana,³⁰ G. Shaifullah¹⁰,¹⁰
 R. M. Shannon,^{6,31}★ J. Simon,³² R. Spiewak,⁶ I. H. Stairs,³³ B. W. Stappers,²¹
 J. K. Swiggum,³⁴ S. R. Taylor,³⁵ G. Theureau,^{16,17,36} L. Toomey,¹ R. van Haasteren,³²
 J. B. Wang,³⁷ Y. Wang³⁸ and X. J. Zhu³⁹

Affiliations are listed at the end of the paper

Accepted 2019 October 28. Received 2019 October 7; in original form 2019 January 15

ABSTRACT

We have constructed a new time-scale, TT(IPTA16), based on observations of radio pulsars presented in the first data release from the International Pulsar Timing Array (IPTA). We used two analysis techniques with independent estimates of the noise models for the pulsar observations and different algorithms for obtaining the pulsar time-scale. The two analyses agree within the estimated uncertainties and both agree with TT(BIPM17), a post-corrected time-scale produced by the Bureau International des Poids et Mesures (BIPM). We show that both methods could detect significant errors in TT(BIPM17) if they were present. We estimate the stability of the atomic clocks from which TT(BIPM17) is derived using observations of four rubidium fountain clocks at the US Naval Observatory. Comparing the power spectrum of TT(IPTA16) with that of these fountain clocks suggests that pulsar-based time-scales are unlikely to contribute to the stability of the best time-scales over the next decade, but they will remain a valuable independent check on atomic time-scales. We also find that the stability of the pulsar-based time-scale is likely to be limited by our knowledge of solar-system dynamics, and that errors in TT(BIPM17) will not be a limiting factor for the primary goal of the IPTA, which is to search for the signatures of nano-Hertz gravitational waves.

Key words: time – pulsars: general.

1 INTRODUCTION

International Atomic Time (TAI) is the basis for terrestrial time. It is also used to form Coordinated Universal Time (UTC), which approximates the rotational phase of the Earth and provides a realization of the theoretical time-scale Terrestrial Time (TT) through $TT(TAI) = TAI + 32.184$ s. Once TAI is defined, it is not changed,

but it is reviewed annually and departures of TAI from the SI second are incorporated into a post-processed time-scale published by the Bureau International des Poids et Mesures (BIPM; Petit 2003) as TT(BIPMXY), where XY indicates the year of creation. The version we have used for this analysis is TT(BIPM17), which is available from <ftp://ftp2.bipm.org/pub/tai/ttbipm/TTBIPM.17>.

Atomic time-scales continue to improve in accuracy and stability (Petit 2013). Such time-scales are created from an ensemble of atomic clocks (e.g. hydrogen masers) which have good short-term stability but are weaker over longer periods. To provide stability over years and decades TAI is ‘steered’ by comparison with ‘primary and

* E-mail: george.hobbs@csiro.au (GH); dick.manchester@csiro.au (RNM); rshannon@swin.edu.au (RMS)

secondary representations of the second’ (presently caesium and rubidium fountains respectively) to keep the TAI second as close as possible to the SI second (Arias, Panfilo & Petit 2011). This leads to a time-scale which is statistically non-stationary and for which the stability over decades is difficult to determine. It would be valuable to have an independent time-scale with comparable stability over decades.

Guinot & Petit (1991), Matsakis & Foster (1996), Petit & Tavella (1996), Rodin (2011), Hobbs et al. (2012), Manchester et al. (2017), and Yin, Gao & Zhao (2017) have shown that a time-scale based on the spin of radio pulsars with millisecond periods (MSPs) can have a stability comparable to that of atomic time-scales. Whereas normal (longer period) pulsars often show irregular rotation, such as glitch events or low-frequency timing noise, MSPs are significantly more stable. Pulse times of arrival (ToAs) from MSPs can also be determined more precisely than those from normal pulsars because they are shorter and there are more pulses to average.

More than 50 MSPs are currently being observed as part of the International Pulsar Timing Array (IPTA) project, with the primary goal of detecting ultralow frequency gravitational waves (e.g. Hobbs et al. 2010; Verbiest et al. 2016). The IPTA combines the Parkes PTA (PPTA; Manchester et al. 2013) in Australia with the European PTA (EPTA; Desvignes et al. 2016) and with the North American PTA (NANOGrav; Arzoumanian et al. 2018, The NANOGrav Collaboration 2015) and it will continue to expand in the future.

The IPTA team expects to identify the quadrupolar signature of gravitational waves unambiguously by searching for the correlations between the timing residuals of different pulsars (see e.g. Verbiest et al. 2016). As shown by Tiburzi et al. (2016) and references therein, other processes can also produce correlated timing residuals. Errors in the planetary ephemeris will introduce an approximately dipolar correlation between pulsars, and time-scale errors will introduce a monopolar correlation. Since the strongest gravitational waves are expected to have long periods, the stability of the reference time-scale over decades is important for gravitational wave searches.

Two methods have been used for extracting a pulsar-based time standard from PTA data, and for comparing it with a given realization of TT. The first, described in Hobbs et al. (2012), uses a frequentist-based method that includes the clock signal as part of the pulsar timing model and carries out a global, least-squares-fit to estimate that signal. The second, a Bayesian technique described in Caballero et al. (2016), uses a maximum-likelihood estimator and optimal filtering, described in Lee et al. (2014), to estimate the clock signal.

In this paper, we:

- (i) make use of the first IPTA data set to produce a pulsar-based time standard TT(IPTA16),
- (ii) improve the Hobbs et al. (2012) algorithm by accounting for non-stationarity in the noise processes (as described by Reardon et al. 2016),
- (iii) extend the Caballero et al. (2016) algorithm to estimate both the covariance of the noise processes and the clock signal using independent Bayesian algorithms. This improves the accuracy of the uncertainty of clock signal waveform,
- (iv) compare the two different methods for developing the pulsar-based time standard, and
- (v) provide a direct comparisons between TT(IPTA16) with respect to TT(BIPM17), and the best atomic frequency standards at the US Naval Observatory.

2 THE DATA SET AND ANALYSIS

The pulsar timing analysis is a well-developed process of fitting a model, which describes both the pulsar and the propagation of the pulses to Earth, to a set of observed pulse ToAs (see Hobbs, Edwards & Manchester 2006, for details of how this process is implemented within the TEMPO2 software package). For this work, we use ToAs from the first IPTA data release by Verbiest et al. (2016) and Lentati et al. (2016). The data set consists of observations of 49 pulsars from the EPTA telescopes (Lovell, Effelsberg, Westerbork and Nançay), the NANOGrav telescopes (Arecibo and Green Bank), and the Parkes telescope. Of the 49 pulsars,¹ 13 are solitary and 36 are in binary systems.² The pulsars have been observed in numerous observing bands from around 300 to 3000 MHz. The observations are performed with an irregular observing cadence of 2 to 4 weeks (depending on the pulsar and the telescope). For the basic timing analysis we used TEMPO2 (Hobbs et al. 2006) with the DE421 solar-system ephemeris (Folkner, Williams & Boggs 2009).

The ToAs were estimated by fitting a pulse template to the observed average pulse profile. This provides both a ToA and an estimate of its measurement error. The uncertainties of the observed ToAs are more complex than the measurement errors (see Verbiest & Shaifullah 2018 for a review). For instance, the strength of the observed pulse varies rapidly because of interstellar scintillation (see e.g. Rickett 1970), and so the signal-to-noise ratio changes rapidly. The pulse shape itself shows significant jitter from pulse to pulse (see e.g. Liu et al. 2012; Shannon et al. 2014; Lam et al. 2019). Such effects cause independent errors in each measurement, which we refer to as ‘white noise’.

The timing residuals are also affected by low frequency noise processes, which we refer to as ‘red noise’. The pulsar timing method (details are provided in Edwards, Hobbs & Manchester 2006) relies on an initial set of models that includes the pulsar spin frequency and its spin-down rate, its position on the sky and proper motion and parallax, the binary orbital parameters if any, and the electron column density of the interstellar medium (known as the dispersion measure, DM). It also includes instrumental parameters such as phase differences between receivers. However, pulsars often spin irregularly, the interstellar medium changes, and the observing hardware and software change (see details in e.g. Manchester et al. 2013; Lentati et al. 2016; Arzoumanian et al. 2018). Such effects lead to red noise in the timing residuals. The covariance matrices of all sources of noise must be estimated to optimize the least-squares (or maximum-likelihood) fits and to obtain valid uncertainties on the parameters of the timing model. Recent reviews of such issues in pulsar timing residuals in relation to PTA experiments were recently published by Hobbs & Dai (2017) and Tiburzi (2018).

Verbiest et al. (2016) provided the first IPTA data release in three different forms: (i) a ‘raw’ format with only minimal pre-processing, (ii) a format usable for standard TEMPO2 analysis, and (iii) a data set using the Bayesian parametrization for TEMPNESST analysis. These data sets are available from <http://www.ipta4gw.org>. The primary difference between these three data sets is the way in which the noise models are included. For our analysis we started with the IPTA data combination ‘A’, which is the raw form without any noise models.

¹The data set for one of these pulsars, PSR J1939+2134, was removed from our sample. The data are affected by significant low-frequency timing noise and the pulsar does not contribute to determination of the clock signal.

²This includes PSR J1024–0719, which is thought to be in a long-period orbit; see Bassa et al. (2016) and Kaplan et al. (2016).

The derived pulsar time-scale described in this paper is named TT(IPTA16) as our data set is based on Verbiest et al. (2016), which was published in 2016. However, we note that the data set only extends until the end of 2012.

2.1 Data and processing for the frequentist analysis

We have independently processed the data using both frequentist and Bayesian methodologies. Here we describe the frequentist processing. The Bayesian analysis is presented in Section 2.2.

For our data set, any irregularities in TT(BIPM17) are much smaller than the noise in the ToAs for any particular pulsar.³ Consequently, we made noise models from the residuals formed from ToAs with respect to TT(BIPM17) and used those models in all fits. This allowed us to avoid iteration of the global clock fit. Modelling of the noise is perhaps the most difficult part of the timing analysis. The fundamental reason for this is that all pulsars are different and we only have one realization of each pulsar. Thus it is difficult to estimate the covariance matrix without making some constraining assumptions, for example, assuming that the red timing noise is wide-sense stationary and parametrizing its power spectrum with a simple analytical model.

The noise models refer to the residuals after all the deterministic effects have been removed. Variations in the reference clock can be included in the timing model, estimated and removed from the residuals. So in our frequentist approach to fitting the timing model, we estimate the noise models iteratively. Fortunately such iterations converge quite quickly because least-squares solutions are not strongly sensitive to small changes in the noise models.

In Table 1 we summarize the data set, providing, in column order, the pulsar J2000 name, the MJD range, and corresponding span in years, the range of the observing frequencies available, the number of arrival times, and the telescopes used to provide observations. We also provide three rankings to indicate which of these pulsars contributes to the clock signals (these final three columns in the table will be described later). The observational coverage is shown graphically in Fig. 1. Appendix A describes our scripts for processing the data and forming the initial noise model, with Table A1 listing the noise model parameters. We note that, as described in the Appendix, some of the PPTA observations were missing from the IPTA data release. These observations were included in our frequentist-based analysis.

With the final parameter and arrival time files, we applied the Hobbs et al. (2012) technique to derive the common mode, $CM(t)$, clock signal using a simultaneous global fit to all pulsar data sets.⁴ We constrained the fit for the clock signal so that it does not include an offset, linear or quadratic component. We also switch off fitting for the pulsar parameters (apart from the pulse frequency and its first time derivative). The clock, $CM(t)$, is parametrized by a grid of samples (separated by 0.5 yr intervals) and linear interpolation between the samples.⁵

During the work described here, the software TEMPO2 (available from <https://bitbucket.org/psrsoft/tempo2>) was updated and the default least-squares solver was changed to a scheme using the QR

decomposition⁶ from the previous scheme based on the singular value decomposition (SVD). The QR based solver is faster, but we found a small difference in the clock estimate for the full sample of pulsars. The worst-case difference was 0.6 times the error bar. When we used only the eight most significant pulsars the difference between QR and SVD based schemes became insignificant. Further testing revealed that the covariance matrix of the parameters had full rank, but the condition number of that matrix had increased by a factor of 40 when all pulsars were used. This behaviour is not unexpected. The SVD approach is particularly useful when the covariance matrix of the observations is singular, but it is also valuable when the whitened model matrix is ill-conditioned. Users of TEMPO2 who stress the algorithm with global solutions for many pulsars should take care to test both QR and SVD methods on their problem. The results presented in this paper were obtained with the SVD approach.

2.2 Data and processing for the Bayesian analysis

The frequentist method, as described above, relies on the manual modelling of the noise properties, as well as the careful subjective checking of the resulting data sets and fits. The Bayesian method is independent: it fits both the timing model and the noise models for the pulsars simultaneously. This avoids the iteration in the frequentist method, however, it requires that the analyst make a choice of prior distributions for the model parameters.

For the Bayesian analysis, we used the IPTA data set without excluding any data from specific observing systems or including the additional PPTA data used in the frequentist method. We did, however, keep the same time-spans as in the frequentist analysis. Because of the significant computational cost in carrying out the Bayesian analysis, we used a restricted list of eight pulsars. The selection process is discussed in Section 3.3.

Bayesian approaches to analysing pulsar-timing data, whether to search for spatially correlated signals or single-pulsar timing and noise analysis, have appeared in various publications (e.g. van Haasteren et al. 2009; Lee et al. 2014; Lentati et al. 2014) and have been applied to real data sets (e.g. Caballero et al. 2016; Desvignes et al. 2016; Arzoumanian et al. 2018). The results from Bayesian and frequentist methods have been compared and have shown to be generally consistent (for instance, Babak et al. 2016; Li et al. 2016).

With Bayesian parameter estimation, we estimate the probability distribution of the values of a set of parameters, ζ , taking into account the data, X , a model (hypothesis), H , such as a physical model, describing how the parameters are related with each other. Bayesian inference is the process by which we use the information from the observed data to update our knowledge of the probability distribution of the unknown parameters, for which we necessarily have a prior belief, described by the prior probability distribution. The inferred distribution is called the posterior probability distribution. Via Bayes' theorem, we perform the parameter estimation using the relation

$$P_{\text{pos}} \propto P_{\text{pr}} \Lambda. \quad (1)$$

Using P to denote probabilities, the posterior probability distribution, $P_{\text{pos}} = P(\zeta | X, H)$, is therefore the conditional probability of the parameters of interest given the model and the data. The likelihood function, $\Lambda = \Lambda(X | \zeta, H)$, is the conditional probability of the data given the model and its parameters. Finally, the prior

³This is clear from the results of Hobbs et al. (2012), which covered approximately the same data span as the IPTA data processed here.

⁴Note that the term 'common-mode' is being used here in a general manner and refers to the signal common to a group of signals.

⁵We have also re-run the algorithm with a 1 yr and a 100 d sampling. These results are available as part of our data release; see Appendix C.

⁶Note that 'QR' is not an acronym. See e.g. Press et al. (1996) for details.

Table 1. Summary of the data used.

Pulsar	MJD range	Span (yr)	Frequency range (MHz)	Ntoa	Telescopes	Max diff. (ns) (Significance)	Ratio (Rank)	Bayesian rank
J0030+0451	53333–55925	7.1	1346–2628	723	A,E,N	11 (#23)	1.00 (#25)	#9
J0034–0534	53670–55808	5.9	1394–1410	57	N	1 (#48)	1.00 (#26)	#10
J0218+4232	50371–55925	15.2	1357–2048	757	E,J,N,W	8 (#27)	0.98 (#42)	–
J0437–4715	50191–55619	14.9	1295–3117	4320	P	254 (#3)	1.54 (#3)	Top
J0610–2100	54270–55926	4.5	1366–1630	347	J,N	5 (#29)	0.99 (#38)	#20
J0613–0200	50932–55927	13.7	1327–3100	1460	E,G,J,N,P,W	24 (#17)	0.93 (#46)	–
J0621+1002	50693–55922	14.3	1354–2636	568	E,J,N,W	3 (#39)	0.99 (#40)	–
J0711–6830	49589–55619	16.5	1285–3102	427	P	74 (#7)	1.15 (#6)	–
J0751+1807	50363–55922	15.2	1353–2695	1124	E,J,N,W	34 (#10)	1.08 (#10)	–
J0900–3144	54285–55922	4.5	1366–2206	575	J,N	4 (#31)	1.00 (#29)	#15
J1012+5307	50647–55924	14.4	1344–2636	1117	E,G,J,N,W	44 (#8)	0.79 (#48)	–
J1022+1001	50361–55923	15.2	1341–3102	1133	E,J,N,P,W	31 (#11)	1.25 (#4)	–
J1024–0719	50118–55922	15.9	1285–3102	804	E,J,N,P,W	11 (#25)	1.00 (#31)	–
J1045–4509	49406–55620	17.0	1260–3102	510	P	14 (#22)	0.96 (#44)	–
J1455–3330	53375–55926	7.0	1246–1699	395	J,N	3 (#34)	0.99 (#37)	#6
J1600–3053	52302–55919	9.9	1341–3104	999	G,J,N,P	28 (#15)	1.01 (#17)	#8
J1603–7202	50026–55619	15.3	1285–3102	381	P	18 (#19)	0.97 (#43)	–
J1640+2224	50459–55924	15.0	1353–2636	532	A,E,J,N,W	27 (#16)	1.09 (#9)	#13
J1643–1224	49422–55919	17.8	1280–3102	1066	E,G,J,N,P,W	23 (#18)	1.02 (#15)	–
J1713+0747	49422–55926	17.8	1231–3102	1833	A,E,G,J,N,P,W	1319 (#1)	3.12 (#1)	Top
J1721–2457	52076–55854	10.3	1360–1412	152	N,W	2 (#45)	1.00 (#32)	–
J1730–2304	49422–55921	17.8	1285–3102	480	E,J,N,P	37 (#9)	1.10 (#8)	–
J1732–5049	52647–55582	8.0	1341–3102	190	P	10 (#26)	1.04 (#12)	–
J1738+0333	54103–55906	4.9	1366–1628	206	N	3 (#35)	0.99 (#39)	#17
J1744–1134	49921–55925	16.4	1264–3102	823	E,J,N,P	260 (#2)	2.39 (#2)	#4
J1751–2857	53746–55837	5.7	1398–1411	78	N	3 (#40)	1.00 (#28)	#21
J1801–1417	54184–55921	4.8	1396–1698	86	J,N	3 (#38)	1.00 (#19)	–
J1802–2124	54188–55903	4.7	1366–2048	433	J,N	2 (#44)	1.00 (#24)	#14
J1804–2717	53747–55915	5.9	1397–1520	76	J,N	2 (#43)	1.00 (#20)	–
J1824–2452A	53519–55619	5.7	1341–3100	234	P	4 (#33)	1.00 (#23)	–
J1843–1113	53156–55924	7.6	1374–1630	174	J,N,W	5 (#30)	0.99 (#36)	–
J1853+1303	53371–55923	7.0	1370–1520	102	A,J,N	11 (#24)	1.01 (#16)	#11
J1857+0943	50459–55916	14.9	1341–3102	625	A,E,J,N,P,W	29 (#12)	1.10 (#7)	–
J1909–3744	52618–55914	9.0	1341–3256	1398	G,N,P	172 (#5)	1.24 (#5)	Top
J1910+1256	53371–55887	6.9	1366–2378	106	A,J,N	5 (#28)	1.00 (#21)	#12
J1911–1114	53815–55881	5.7	1398–1520	81	J,N	4 (#32)	1.00 (#18)	–
J1911+1347	54093–55869	4.9	1366–1408	45	N	17 (#20)	1.03 (#14)	–
J1918–0642	52095–55915	10.5	1372–1520	265	G,J,N,W	29 (#13)	1.08 (#11)	#5
J1955+2908	53798–55919	5.8	1386–1520	132	A,J,N	2 (#41)	1.00 (#33)	#16
J2010–1323	54087–55918	5.0	1366–2048	296	J,N	17 (#21)	1.03 (#13)	#18
J2019+2425	53446–55921	6.8	1366–1520	80	J,N	2 (#47)	1.00 (#27)	–
J2033+1734	53894–55918	5.5	1368–1520	130	J,N	2 (#46)	1.00 (#30)	–
J2124–3358	49490–55925	17.6	1260–3102	1028	J,N,P	97 (#6)	0.98 (#41)	–
J2129–5721	49987–55618	15.4	1252–3102	343	P	29 (#14)	0.96 (#45)	–
J2145–0750	49518–55923	17.5	1285–3142	1476	E,J,N,P,W	198 (#4)	0.92 (#47)	–
J2229+2643	53790–55921	5.8	1355–2638	234	E,J,N	3 (#36)	1.00 (#22)	#19
J2317+1439	50459–55918	14.9	1353–2638	409	E,J,N,W	3 (#37)	1.00 (#34)	#7
J2322+2057	53916–55921	5.5	1395–1698	199	J,N	2 (#42)	1.00 (#35)	–

Note. Telescope codes: (A) Arecibo, (E) Effelsberg, (G) Green Bank, (J) Jodrell Bank, (N) Nancay, (P) Parkes, and (W) Westerbork

probability distribution, $P_{\text{pr}} = P(\zeta|H)$, is the conditional probability of the parameters given only the model hypothesis. The prior is therefore a mathematical tool to express the degree of knowledge or ignorance on the real probability distribution and is mandatory in Bayesian inference. Bayesian inference is often used when the

model's complexity is high and closed-form solutions are not readily available. In such cases, as is the case with our analysis, the posterior distribution is then inferred by randomly drawing samples from the prior distribution using Monte Carlo methods and testing their likelihood.

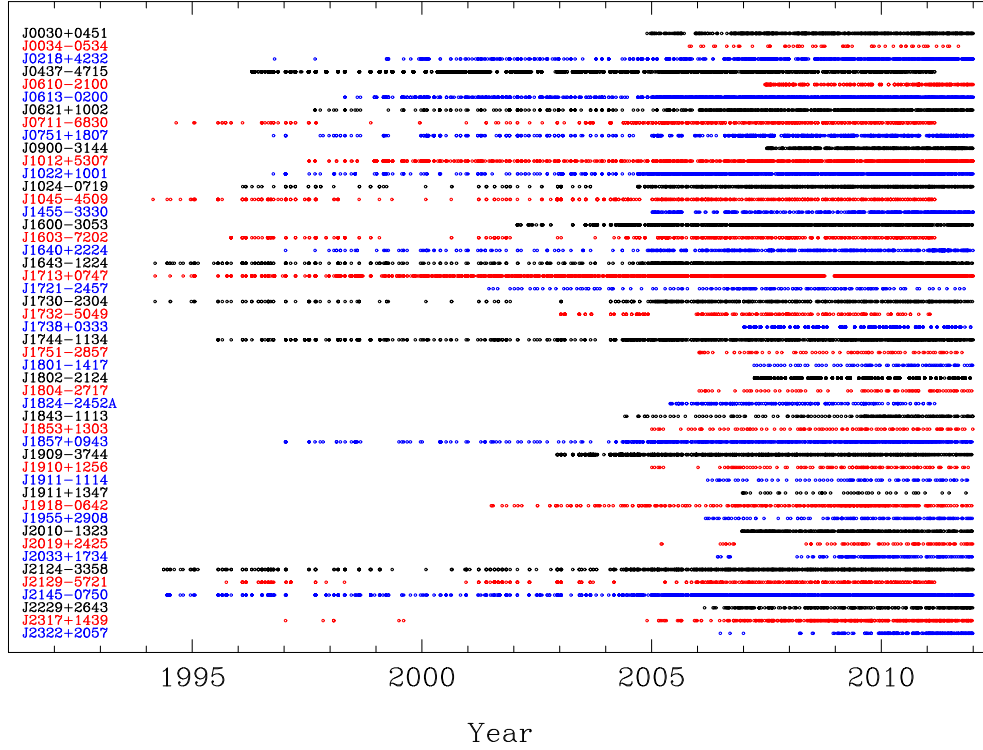


Figure 1. Data spans and sampling for the IPTA pulsars used in our analysis. Each point indicates an observation, but neither the ToA nor its uncertainty are shown. The uncertainties typically are smaller for more recent data.

For the Bayesian analysis in this work, the pulsar timing and noise analysis was performed in the same way as in Caballero et al. (2018). The noise model consists of three basic components: the white-noise parameters which regulate the uncertainties of the ToAs, as well as the red- and DM-noise components. We summarize the noise model and mathematical details in Appendix B (see Caballero et al. 2018 for more details).

As in Lentati et al. (2015) and Caballero et al. (2016), we model the clock signal as a stationary, power-law, red noise process correlated between the pulsars (a monopolar correlation). The power spectrum of the clock signal is modelled as

$$S_{\text{clk}}(f) = \frac{A_{\text{clk}}^2}{f} \left(\frac{f}{f_r} \right)^{2\alpha_{\text{clk}}}, \quad (2)$$

where f is the Fourier frequency and f_r is a reference frequency set to 1 yr^{-1} . In Caballero et al. (2016) the pulsar noise parameters were held fixed at their maximum-likelihood values and the maximum-likelihood values for the clock-noise parameters were calculated in a frequentist way, with a maximum-likelihood estimator. In this study, while we still keep the pulsar noise parameters fixed, we perform Bayesian inference for the clock-error signal parameters while we analytically marginalize over the timing parameters (the analytical marginalization of the timing parameters is also implicitly assumed in Caballero et al. 2016 in the formation of the likelihood function, as in this study). Fixing the noise offered the ability to perform the analysis faster, having first verified that this approach was sufficient to detect and reconstruct the simulated TT(TAIx2) signal, as discussed in Section 3.

Once we obtain the posterior distributions of the clock-signal parameters, we estimate the waveform and uncertainties of the clock signal with the following procedure:

(i) we calculated the 1σ (68 per cent credible interval) boundary of the 2D probability distribution of the two clock parameters using the posterior distribution obtained from the Bayesian inference.

(ii) we constructed the updated total covariance matrix, \mathbf{C}_{upd} , by summing the inferred covariance matrix of the clock signal (\mathbf{C}_{clk}) and pulsar-noise covariance matrices. We used this updated covariance matrix to re-fit the linear timing parameters using generalized least-squares fitting in the presence of correlated noise and calculated the post-fit timing residuals, described by the vector \mathbf{t}_{gls} .

(iii) we used the method described in Lee et al. (2014) to reconstruct clock waveforms for all values of the clock signal's posterior distribution within the 1σ boundary. The clock waveform for each case is then

$$\mathbf{t}_{\text{clk}} = \mathbf{C}_{\text{clk}} \mathbf{C}_{\text{upd}}^{-1} \mathbf{t}_{\text{gls}}. \quad (3)$$

The final clock waveform is the average of these waveforms and the upper and lower envelope of the waveforms provides the reported 1σ boundary. This is different from the original method, which estimated the maximum-likelihood waveform and the 1σ uncertainty levels as the standard deviation of the maximum-likelihood estimator, an approximation that is valid only if the noise is uncorrelated. Since the noise is in general correlated, this approach gives a more reliable estimate of the clock waveform uncertainties.

3 RESULTS AND DISCUSSION

Fig. 2 shows TT(IPTA16) with respect to three different reference time-scales: TT(TAI), TT(BIPM17), and a test case labelled

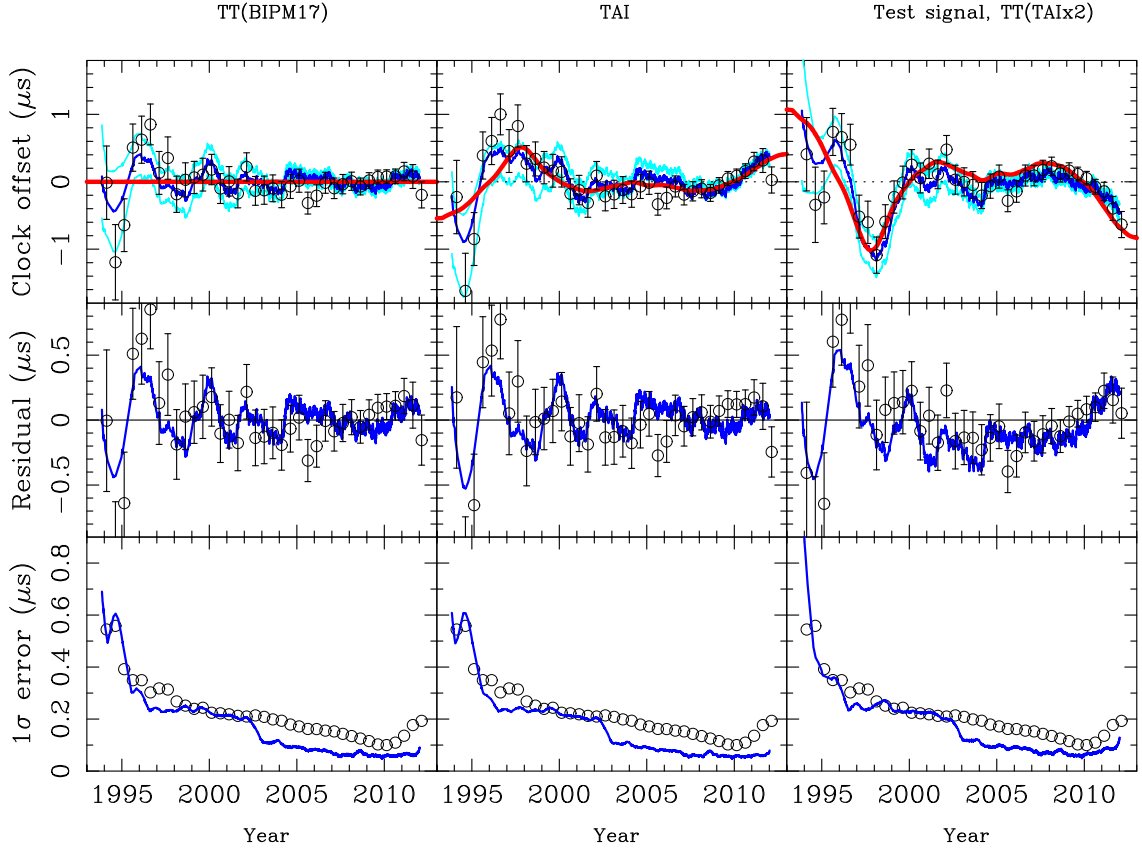


Figure 2. The derived clock signal, TT(IPTA16) (top row) referred to TT(BIPM17) (left-hand column), TT(TAI) (central column), and a test reference which is effectively $\text{TT}(\text{BIPM17}) - 2 \times [\text{TT}(\text{TAI}) - \text{TT}(\text{BIPM17})]$ (right-hand column). The open circles with error bars result from the frequentist analysis and the dark blue lines are the Bayesian determinations of the signal (with 1σ uncertainties indicated with the cyan lines). The red line gives the expected signal in each case. The difference between the expected signal and the derived clock signal is shown in the middle row. The uncertainties on the clock signal as a function of time are given in the bottom row for the two analysis methods.

TT(TAIx2).⁷ The three columns correspond to the different reference time-scales. The top row shows the derived clock signal for each reference time-scale. The black points, with uncertainties, give the frequentist-derived clock signal. The blue, solid curve indicates the Bayesian waveform with its 1σ boundaries in light blue lines. The expected clock waveform is shown as a solid red line. The second row shows the difference from the expected waveform. The frequentist-derived values are points with errors and the Bayesian values are solid lines. The third row provides the one-sided error bar, σ .

We assume that TT(BIPM17) is an accurate time-scale over the sampled data span so the expected clock waveform is zero for time reference TT(BIPM17). For time reference TT(TAI), it is $-\text{TT}(\text{BIPM17}) - \text{TT}(\text{TAI})$ but with a best-fitting quadratic removed. This is necessary because the pulsar timing model must always fit for the spin period and spin-down rate as these are not known a priori. This removes a quadratic polynomial from the residuals of every pulsar, so there can be no quadratic in the resulting clock waveform. This is handled differently in the Bayesian procedure, so we carried out unweighted fits for the quadratic coefficients in both frequentist and Bayesian procedures to

make the results comparable. For the third case, the expected signal is $+2 \times [\text{TT}(\text{BIPM17}) - \text{TT}(\text{TAI})]$ with the quadratic removed.

The error bars on the two analysis schemes are less comparable. The Bayesian lines are 68 per cent confidence limits on the estimated signal at a given time. The frequentist bars are the estimated standard deviation of that particular sample. They depend on the sampling interval. We also note that the clock signal measurements and uncertainties are not independent. As described by Hobbs et al. (2012), the effect of fitting, irregular sampling, differing data spans, and the linear interpolation between adjacent grid points all lead to correlated values. The Bayesian procedure constrains the variation of the clock estimate by its statistical model as a stationary power-law red process, so it is harder to define an interval between independent estimates. The frequentist estimate has a normalized chi-square of 1.08, indicating that the error bars are consistent with the variation in the clock estimate. The Bayesian estimates are also self-consistent.

The following results are easily seen in the figure:

(i) the test signals TAI and TAIx2 are recovered within the confidence limits using both frequentist and Bayesian schemes (more details are given below),

(ii) the residuals (the difference between the expected signal and the derived clock signal) for both methods with TT(BIPM17) and TAI are almost the same, indicating that both algorithms are

⁷This test case was formulated using $\text{TT}(\text{TAIx2}) = \text{TT}(\text{TAI}) - 3 \times [\text{TT}(\text{TAI}) - \text{TT}(\text{BIPM17})]$ as the time reference.

operating linearly at that signal level – a much larger signal than expected in TT(BIPM17),

(iii) the residuals for TAIx2 are slightly different to those for the preceding two cases, indicating that some non-linearity is becoming evident at this level, i.e. the assumption that the clock signal is small compared to the pulsar noise level starts to fail for this case,

(iv) the pulsar time-scale, TT(IPTA16), with respect to TT(BIPM17) is consistent with zero. Note that the variations of ± 500 ns before 1998 are not as significant as they appear because the frequentist error bars are 25 per cent correlated (see Section 3.2).

(v) the IPTA data rate and the receiver performance improved markedly in 2003 and this is reflected in the error bars for both methods, especially in the Bayesian estimates.

Although the known signal in the top central panel of Fig. 2 is clearly detected, an unknown signal would be less obvious. However an unknown signal of the same variance could be detected purely from the χ^2 of the frequentist samples. The reduced χ^2 of the left-hand panel TT(BIPM17) is 1.08, confirming that the error estimates are good. The reduced χ^2 of the central panel is 2.0 and the number of degrees of freedom is reduced from 37 to 33 by the correlation between samples. Thus a χ^2 of 2.0 indicates a 3.6σ detection of a variance increase.

The mean frequentist error bar size in our clock comparison drops from ~ 300 ns in the first half of the data to ~ 150 ns in the second half. This implies that our existing data sets could have made a 3σ detection of a common signal of ~ 900 ns that lasted for 6 months in the early data or ~ 450 ns in the later data. Longer lasting events could have been detected at lower amplitudes. For instance, an event lasting 10 yr with an amplitude ~ 70 ns could be detected in the recent data. We note that large time offsets are not expected from atomic time-scales, however frequency instabilities are possible. At the level that we could detect, such instabilities would likely be caused by planned steering, as in TT(TAI). We have no prior reason to expect any such detectable instabilities in TT(BIPM17).

We have produced a publicly available data collection containing our input data, processing pipelines, and results. A description of this data collection is given in Appendix C.

3.1 Noise power spectra

The power spectral densities of the noise and the clock estimate are shown in Fig. 3. The blue solid line, which is the same in each panel of Fig. 3, represents the power spectrum of the frequentist-derived clock (the spectrum of the Bayesian-derived result is shown in Fig. 4 and discussed later). The spectrum was formed for data with MJD $> 51\,000$ (i.e. from the top left-hand panel in Fig. 2 ignoring the oldest clock samples; this date corresponds to 1998 July 6). The horizontal blue dashed lines represent the mean of the spectrum and 95 per cent confidence levels assuming χ^2 statistics (i.e. exponential statistics, with two degrees of freedom). Although there is a suggestion that the spectrum might be starting to rise at the lowest frequency sample, the spectrum is consistent with white noise and we will assume it is white in further analyses.

The four panels in this Figure provide noise spectra for important pulsars in the sample. The white noise spectrum corresponding to the measured ToA uncertainties that have been corrected for the EFAC and EQUAD parameters (see equation A1) are shown as the horizontal black, dotted line for the four listed pulsars. The red noise model as used in the frequentist analysis is shown as the black, solid curve and that from the Bayesian analysis as the red curve. For comparison we also show the red-noise model for J1713+0747

and J1744–1134 given in the IPTA data release B (blue dot-dashed curve). The red-noise estimates for PSR J0437–4715 in the IPTA B data set is negligible and no red-noise model was provided for PSR J1909–3744.

The frequentist noise modelling includes a corner frequency in the power spectrum; see equation (A2). There was no evidence for a turnover in the spectrum for PSRs J0437–4715, J1713+0747, and J1909–3744, but we did not wish to assume a low-frequency variation that we were unable to verify. We therefore used a corner frequency of $f_0 = 1/[\text{dataspan}]$. We note that the quadratic fitting procedure applied to all pulsars implies that our choice of noise model below a frequency of f_0 has little effect on the final results.

In contrast, the Bayesian analysis does not include a corner frequency and the Bayesian procedures will estimate signal components with $f < f_0$, albeit with reduced confidence. Comparison between our frequentist-derived noise models and the previous IPTA analysis shows basic consistency, but also that the match is not always close. This highlights the subjective component to the noise analysis. The Bayesian procedure provides a reproducible method to produce the noise models, but a subjective bias remains in the choice of Bayesian prior. One key result from our work is that the clock estimates computed with different noise models and different algorithms are consistent within their confidence limits. Clearly the clock fitting process is robust to these changes.

Fig. 3 shows that the clock spectrum is much higher than the white noise spectrum for these pulsars below $f = 1 \text{ yr}^{-1}$. This shows the importance of the red noise models. The clock spectrum for such frequencies is completely dominated by red noise in the pulsar timing residuals.

3.2 Anomalies

In analysing the data we found occasional anomalies that depended on a few observations of a single pulsar. On checking these carefully we found some errors in the data and either corrected or removed them. However there were a few anomalies remaining for which we could find no reason to remove the data. One particular example relates to the apparent offset seen in Fig. 2 around the year 1996. This offset is very similar to that reported by Hobbs et al. (2012), which only used PPTA data sets.

In order to study this anomalously high clock waveform around 1996, we have obtained the frequentist-based clock waveform after removing each pulsar in turn. We find that only PSRs J0437–4715 and J1713+0747 significantly affect the pulsar-derived clock signal around this date. If PSR J1713+0747 is removed then the anomaly in the year 1996 disappears (although the uncertainties on the clock signal significantly increase). Around this time the IPTA data set contains observations from both the Parkes and Effelsberg telescopes and they both indicate a dip of around $2 \mu\text{s}$. The observations of PSR J1713+0747 around this time are complex. The Parkes data were recorded at multiple observing frequencies from 1285 to 1704 MHz and there is very sparse sampling around the centre of the dip. However, it does seem that an event occurred around this time that is not included in our noise modelling procedures. As we only have a few pulsars contributing to the clock signal at this time, this does lead to an apparent significant clock offset. We know that the timing residuals for PSR J1713+0747 occasionally show unusual behaviour (see e.g. Lentati et al. 2016) that cannot be modelled using simple, power-law red noise models.

The ‘bump’ in the clock waveform spectrum seen in Fig. 3 around frequency 0.4 yr^{-1} is entirely caused by PSR J0437–4715 (note that the anomaly is missing from the dashed spectrum in Fig. 3 produced

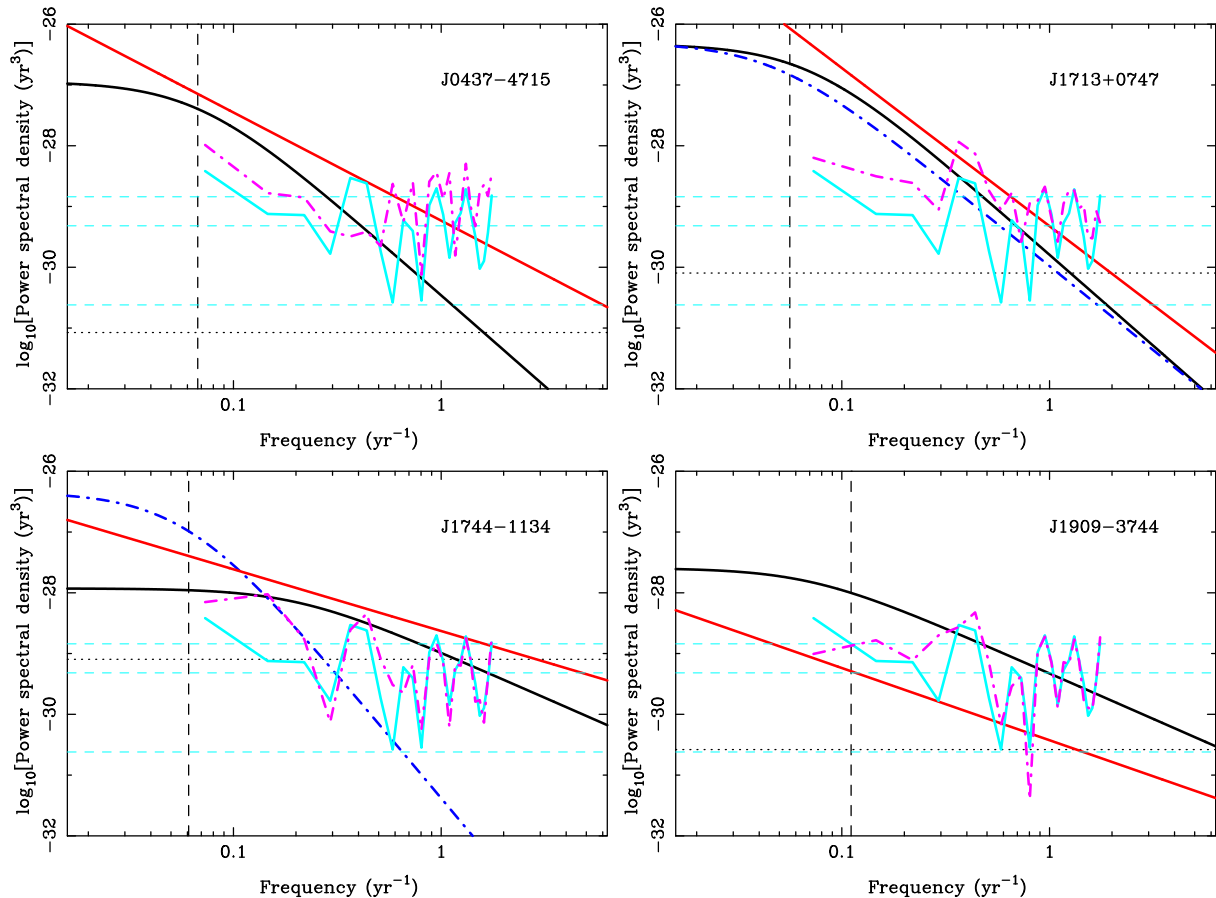


Figure 3. Power spectrum of the IPTA clock signal (light blue) with its mean and 95 per cent confidence levels (horizontal light blue dashed lines) overlaid. Each panel also contains the red noise models for the specified pulsar from this paper (black for the frequentist analysis and red for the Bayesian analysis) and in the IPTA data release (dot-dashed, blue). The spectral density corresponding to the ToA uncertainties for each pulsar scaled by the white noise error factors is shown as the dotted line. The purple dot-dashed line indicates the power spectrum of the clock signal after the specified pulsar has been removed from the sample. The spectral frequency corresponding to $f_0 = 1/[\text{data span}]$ for the specified pulsar is shown as a vertical dashed line.

when PSR J0437–4715 is removed). We have been unsuccessful in determining the cause of this anomaly. We note that this pulsar is only observed by one telescope in the IPTA and so we have no ability, as we do for other pulsars, either to identify and then fix, or to average over, any instrumental effects that may be occurring for a single telescope.

PSR J1909–3744 does not contribute as much as expected in the frequentist analysis, but leads to an anomaly in the clock waveform in the Bayesian analysis around the year 2004. Shannon et al. (2015) indicated that the PSR J1909–3744 timing residuals were statistically white, whereas our work indicates red noise. This is because Shannon et al. (2015) showed that PPTA observations at high frequencies, which were not corrected for DM variations, produced whiter (and lower rms residual) data sets compared with the DM-corrected lower frequency data. Lam et al. (2017) also identified excess, observing-frequency-dependent noise in the timing residuals for this pulsar. We therefore believe that uncorrected interstellar medium effects, or instrumental effects are leading to the observed red noise in the IPTA data set for this pulsar. We repeated the frequentist data processing to re-form the clock waveform after removing the IPTA data set for PSR J1909–3744 and replacing it with the Shannon et al. (2015) data set. Variations of $\lesssim 250$ ns are seen comparing this clock signal with that shown in Fig. 2 between the years 2005 and ~ 2008 , which covers the range where

this pulsar was observed by the Green Bank Telescope. The changes that occur when using the IPTA and the Shannon et al. (2015) data sets and when comparing the Bayesian and frequentist noise estimates highlights both the importance and the challenges of such noise modelling. Such issues have relatively little effect on the pulsar time-scale with existing data, but as data sets become longer and ToA precision continues to improve, such modelling will become more important.

3.3 Which pulsars contribute?

The time taken to produce the pulsar-based time-scale from both the frequentist and Bayesian algorithms significantly increases as more pulsars are included in the analysis, but not all pulsars are equal. The pulsars that most constrain the pulsar time-scale are those with (1) long data spans, (2) minimal red noise, and (3) few data gaps.

Our next analysis is based on the data shown in the top left-hand panel of Fig. 2. The simplest measure of whether a pulsar contributes is therefore to determine whether the data points in the Figure would change if that pulsar were removed from the analysis. The maximum change in the frequentist-derived clock signal with the removal of each pulsar is listed in the seventh column of Table 1 along with the ranking of the significance of the pulsar using this measure (from 1 to 48). This ranking highlights that the

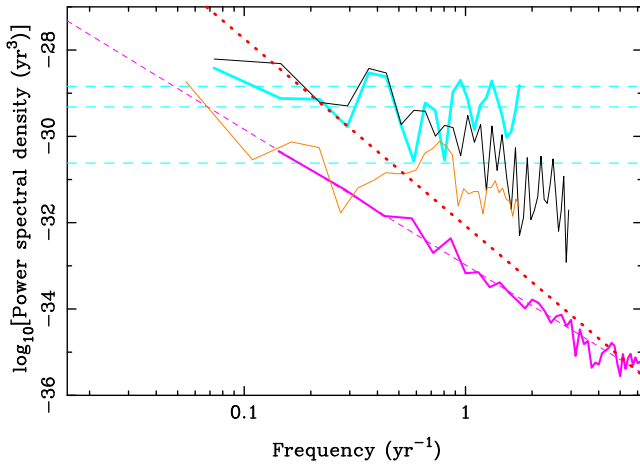


Figure 4. Power spectra of the frequentist-derived (cyan, with mean and 68 percent confidence levels indicated by horizontal dashed lines) and Bayesian-derived (black) pulsar time standard overlaid with the expected power-spectral density of the timing residuals induced by a gravitational wave background with amplitude 10^{-15} (red dotted line). Note that only the monopolar contribution of the gravitational wave background signal would affect the measurement of the clock signal. The purple solid line indicates the average power-spectral density of the four USNO rubidium fountain clocks after subtraction of the common-mode signal. The orange spectrum is derived from the difference in the clock signal when using JPL ephemeris DE436 compared to the default analysis using DE421.

pulsars with the largest influence on the resulting clock signal are PSRs J1713+0747, J1744–1134, J0437–4715, J2145–0750, and J1909–3744 each producing variations to the pulsar time-scale at a level >100 ns for at least one time sample. PSR J1713+0747 has the largest contribution because the IPTA data release includes the early pre-NANOGrav Green Bank and Arecibo data for this pulsar providing a very long span of high-quality data. There are 22 pulsars whose removal only changes the resulting time-scale by <10 ns. Removing all 22 causes no visible change (the maximum change being 13 ns) to the pulsar-based time-scale.

The analysis described above shows that the removal of some pulsars changes the resulting clock signal. It does not indicate whether the removal of a pulsar improved the pulsar time-scale, or not. We have considered various statistics to determine whether the inclusion of a pulsar has a positive or detrimental effect. We have found that comparing the power spectrum of the resulting clock signal (determined on a 100 d grid) with and without the inclusion of each pulsar in turn indicates most clearly which pulsars are contributing.

The ratios of the mean of the power spectral densities for $f \leq 0.5 \text{ yr}^{-1}$ without and with each pulsar in turn are listed in the second last column of Table 1. The pulsars that are most significant by this measure (in order) PSRs J1713+0747, J1744–1134, J0437–4715, J1022+1001, J1909–3744, and J0711–6830. In contrast to the first ranking procedure, PSR J2145–0750 is poorly ranked using this method as its inclusion slightly increases the power spectral density of the pulsar-derived clock signal. We highlight using a bold font (in column 7) those pulsars for which this ratio indicates that the inclusion of the pulsar does not improve the pulsar time-scales.

The Bayesian analysis also provided a ranking of pulsars. This ranking was based on the contribution of each pulsar to the signal-to-noise ratio of the recovered TT(TAIx2) simulated clock signal using the Bayesian procedure. The procedure was iterative. Starting

with only the three pulsars that dominated the initial frequentist analysis (listed as ‘Top’ in Table 1), we determined the simulated clock signal. We then repeated the analysis, but each time adding in a different fourth pulsar. We chose as the fourth pulsar of the subset the one that increased the signal-to-noise ratio of the detection the most. We then continued following this method to add the fifth, sixth, etc. pulsars. The contribution beyond the eighth pulsar was negligible, while adding significant computational time and therefore we used the eight most dominant pulsars to our analysis. These were the top three, PSRs J0437–4715, J1713+0747, and J1909–3744, followed by (in order) J1744–1134, J1918–0642, J1455–3330, J2317+1439, and J1600–3053 (the full ranking is given in the last column of Table 1). The exact choice of pulsars was therefore slightly different in the Bayesian and frequentist methods, but the final results are consistent.

Both the frequentist and Bayesian tests highlight the importance of PSR J1713+0747 as the most significant pulsar and then PSRs J0437–4715 and J1744–1134 as other key pulsars. This is not a surprise. We have long, high-quality data sets on all these pulsars. PSR J1909–3744 is also important, but it has a shorter data span than the preceding three.

3.4 Implications for terrestrial time standards

Until 2012 the primary clocks used to steer TAI were not operated continuously, so evaluating the stability of the periodically steered time-scale required a very sophisticated analysis. However in 2012 the US Naval Observatory (USNO) began operating a set of four rubidium (Rb) fountains continuously (Peil et al. 2014). Their time offsets with respect UTC(USNO), which is very close to TAI, were included in the five-day reports which are published monthly by BIPM. We have retrieved these data sets,⁸ which have a 7-yr data span, and used them to estimate the stability of the atomic clocks used in generating TAI. The USNO Rb fountains are not ‘secondary representations of the second’, they are operated as clocks, but they are among the most stable clocks in the TAI ensemble (Peil et al. 2014). We expect that their stability is at least as good as that of TAI.

The data are uniformly sampled and the noise is stationary so we have used standard spectral analysis to compare the fountain clocks. As the spectral exponents are of the order of -3 we pre-whitened the data with a first difference linear filter to avoid spectral leakage and post-darkened it after the Fourier transformation. We can eliminate the effects of variations in the reference time-scale, any environmental variations common to all of the Rb fountains, and any variations in the time transfer between USNO and BIPM, by comparing the clocks with each other. Rather than do this pairwise, we find a signal common to all four clocks.

We subtract this common-mode signal from each clock independently and, for comparison with the pulsar-derived time-scale, we also fit and remove a quadratic before estimating the power spectra. We find that the four Rb fountains corrected for the common-mode signal have similar power spectra, so we took an average of all four to provide the best spectral estimate. The average spectrum, corrected for the common signal, is shown as a solid purple line

⁸The USNO Rb fountains have been submitted to BIPM with the other USNO clocks since 2012. They can be identified by number: 1930002 through 1930005. The data are filed by year and month on the BIPM ftp site, e.g. ftp://62.161.69.5/pub/tai/data/2018/clocks/usno1801.clk corresponds to 2018 January.

in Fig. 4. The spectrum of the common-mode signal, which is not shown in the figure, is very similar to that of the average but about a factor of ten higher.

We fitted a power-law model to the corrected average spectrum using a weighted least-squares fit, assuming that the spectral estimates were distributed as χ^2_8 . The fitted power-law model, which has a spectral exponent of -3.0 , is shown as a dashed purple line. The first sample was not included in the fit because it is partially suppressed by removing the quadratic. Frequencies above 20 y^{-1} were not included in the fit either, as the spectrum begins to flatten at higher frequencies. A spectral exponent of -3 is characteristic of ‘flicker’ frequency modulation. The spectrum at frequencies above 20 y^{-1} appears to have an exponent ~ -2 , which is characteristic of white frequency modulation, but we do not have enough frequency range to establish this clearly.

The extrapolated mean spectrum of the IPTA pulsar-based time-scale (central dashed cyan line) and the extrapolated average spectrum of the Rb fountains intersect at a period of $\sim 15 \text{ yr}$. However, no individual clock is likely to be operating over this time-scale. Consequently, pulsar timing provides a valuable, independent, and long-term, check on terrestrial time-scales. The question of whether TT(IPTA) can contribute to the stability of TT(BIPM) is unclear, and probably premature, as it will be more than 10 yr before its long-term stability can match that of the best Rb fountains – see e.g. Arias & Petit (2019) for a recent analysis of the frequency stability of TT(BIPM).

3.5 The Solar system ephemeris and gravitational waves

The timing residuals processed in this paper were determined using the DE421 solar-system ephemeris (which was also used in the Verbiest et al. 2016 publication describing the IPTA data set). Recent work within the PTAs have highlighted that the choice of solar-system ephemeris has a large effect on the amount of noise seen in the timing residuals (e.g. Arzoumanian et al. 2018). For currently unknown reasons, some of the more recent ephemerides produce noisier pulsar timing residuals than earlier ones for decade-long data sets. We have reprocessed the frequentist-based analysis using the DE436 solar-system ephemeris⁹ and find deviations in the resulting clock waveform at the $\pm 50 \text{ ns}$ level. The spectrum of the difference between the frequentist-based clock waveforms obtained using DE421 and DE436 is shown as the orange line in Fig. 4. These deviations have no significant effect on the results presented here, but unless the solar-system ephemerides are better understood, they will have a major effect within the next few years. We note that some of the signatures seen in this spectrum may be related to individual Solar system objects and so periodicities may emerge from the spectrum with longer data spans (instead of the low-frequency, red-noise signal, that we currently see).

The primary goal of the IPTA project is to detect ultralow frequency gravitational waves (see e.g. Burke-Spolaor et al. 2018). The power spectral density of the timing residuals induced by an isotropic, stochastic, gravitational wave background (GWB) can be approximated as (e.g. Hobbs et al. 2009)

$$P(f) = \frac{A^2}{12\pi^2} \left(\frac{f}{f_{\text{yr}}} \right)^{-13/3}. \quad (4)$$

Current bounds (e.g. Shannon et al. 2015) suggest that $A < 10^{-15}$ and predictions give $A \sim 5 \times 10^{-16}$ (Sesana et al. 2016). The present

upper bound on the power spectrum of timing residuals that would result from a GWB is shown as a dotted line in Fig. 4. This spectrum rises above the pulsar clock noise spectrum at a frequency of about 0.2 yr^{-1} and remains well above the spectrum of the best atomic clocks. This suggests that, unless the GWB amplitude is orders-of-magnitude lower than the current predictions, errors in terrestrial time standards will not be a limiting factor in the detection of gravitational waves.

An unambiguous detection of a GWB requires that the timing residuals for multiple pulsars are shown to be correlated as described by Hellings & Downs (1983). The expected angular correlation has quadrupolar and higher order terms, but its measurement is also affected by (and a GWB affects the measurement of) other spatially correlated processes, including monopolar and dipolar signals (see e.g. Tiburzi et al. 2016). In particular, the expected angular correlation has a non-zero mean of 0.08 . This term would therefore appear as a clock error with a spectrum significantly below that of the GWB itself. However if the GWB signal is detected then its effects on the monopole are known and can be subtracted, so the pulsar-based clock can be recovered without distortion by a GWB.

3.6 Importance of the international pulsar timing array

The most recent data in the current IPTA data release were obtained in 2012. The upcoming second data release will have a further 5 yr of data, more pulsars, and improved timing precision on the most recent observations. This new data set should significantly improve the pulsar-based time-scale as (1) any extra observations prior to 2005 will help constrain the variations seen in the early data, (2) new, high-quality data on a large number of pulsars since 2012 will enable the extension of the time-scale and an improved understanding of its long-term stability, and (3) up-to-date analysis methods will be applied when producing the data set, which should reduce the number of artefacts and anomalies within the data. Of course, the longer data spans and improved timing precision will also mean that low-frequency noise processes (such as pulsar timing noise) will become more important.

Millisecond pulsars are known to undergo sudden changes, including small glitch events (McKee et al. 2016 described a glitch event in the timing residuals of PSR J0613–0200), magnetospheric changes (Shannon et al. 2016 reported on such changes for PSR J1643–1224 although note that Brook et al. 2018 suggest that profile variations in this pulsar may be caused by the interstellar medium and not magnetospheric variations) and effects relating to sudden changes in the interstellar medium (see e.g. fig. 13 in Lentati et al. 2016; Lam et al. 2018). As our data sets become longer and more extensive, similar events will occur in more pulsars. Many such events will be challenging to model deterministically and the noise modelling will need to be updated to account better for such non-stationary noise processes. Of course, with sufficient numbers of equally contributing pulsars, any individual event will be averaged-down, but unless such events can be better understood and modelled they may present a fundamental limit to the stability of the pulsar-based time-scale.

In Hobbs et al. (2012) only observations from the Parkes telescope were used. This meant that any discrepancies found between the pulsar-based time-scale and a terrestrial time-scale could arise from errors in the observatory time-scale. In contrast the IPTA sample includes data from many observatories, often for a given pulsar. The most dominant pulsar in the IPTA sample, PSR J1713+0747, includes observations from seven telescopes each with their own independent observatory time-scale. We note that the participating

⁹Available from <https://ssd.jpl.nasa.gov/?ephemerides>

observatories in the IPTA do not have a well-defined common reference time. For instance, the data processing to form pulse ToAs was carried out using by cross-matching the observations using reference template profiles. These reference profiles were not identical between the observatories leading to constant phase offsets between results from different observatories. Similarly, Manchester et al. (2013) describes how derived pulse ToAs from the Parkes observatory are related to the intersection of the azimuth and elevation axes (the topocentric reference point) of the telescope; however, a similar process is not carried out at all the IPTA observatories. These offsets are accounted for in the fitting process by allowing for an arbitrary phase offset between data from different observatories for each pulsar. Apart from the very earliest data, the observatory-based time-scales are transferred to UTC using GPS, and from there to TT(BIPM17). Any inaccuracy in that time transfer would be common to all observatories and any variations that do not take the form of a quadratic-polynomial (and hence, not affected by the fitting procedures) would affect our comparisons between TT(IPTA16) and terrestrial time standards. An analysis in Hobbs et al. (2012) suggested that any such errors will be <10 ns and therefore have no effect on the results presented here. More detailed studies of the precision for each step in the process from measuring the ToA to the resulting timing residuals have been presented elsewhere (for instance, see the analysis of instrumental noise in Lentati et al. 2016). We currently see no evidence for a statistically significant common signal in our results (when referred to TT(BIPM17)) and note that it is becoming easier to search for common offsets in PTA data sets as observatory instrumentation becomes more standardized and more telescopes observe the same set of pulsars with similar timing precision.

Pulsar searches are underway at many of the major observatories and we expect the discovery of a large number of new millisecond pulsars. However, as we have shown in this paper, only a few bright nearby pulsars contribute significantly to the pulsar time-scale. We are not confident that a large number of similarly bright pulsars that can be timed with high precision remain to be discovered. Improving the pulsar time scale by adding more bright pulsars will require use of the next generation of telescopes such as FAST and MeerKAT.

4 CONCLUSIONS

We have used the IPTA DR1 data set with two different analysis algorithms, one frequentist and one Bayesian, to establish a pulsar-based time-scale which we call TT(IPTA16). Consistent results are obtained from the two analysis methods. We show that TT(IPTA16) is consistent with the time-scale TT(BIPM17). We confirm that TT(BIPM17) is currently the most stable time-scale in existence and that current efforts to detect nano-Hertz gravitational waves through pulsar timing using a post-corrected BIPM time-scale as a reference are not limited by instabilities in the reference time-scale.

The pulsar-based time standard will improve with continued observations using the existing IPTA programs. The IPTA is currently preparing a second data release that includes significantly more pulsars, uses more telescopes and upgraded instruments at the observatories. In the longer term, even more telescopes will start to contribute to the IPTA data set. Within a few years, we expect high precision observations from the FAST telescope currently being commissioned in China, and the MeerKAT telescope in South Africa.

Of course, atomic clocks will also continue to improve, as fast, or faster, than pulsar observations. However, since pulsar time-scales are based on entirely different physics compared to atomic time-

scales, and are essentially unaffected by any terrestrial phenomena, they form a valuable independent check on atomic time-scales. In addition, they provide a time-scale that is, in principle, continuous over billions of years.

ACKNOWLEDGEMENTS

The National Radio Astronomy Observatory and Green Bank Observatories are facilities of the NSF operated under cooperative agreement by Associated Universities, Inc. The Arecibo Observatory is operated by SRI International under a cooperative agreement with the NSF (AST-1100968), and in alliance with Ana G. Méndez-Universidad Metropolitana, and the Universities Space Research Association. The Parkes telescope is part of the Australia Telescope which is funded by the Commonwealth Government for operation as a National Facility managed by CSIRO. 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. The 100-m Effelsberg Radio Telescope is operated by the Max-Planck-Institut für Radioastronomie at Effelsberg. Some of the work reported in this paper was supported by the ERC Advanced Grant ‘LEAP’, Grant Agreement Number 227947 (PI Kramer). Pulsar research at the Jodrell Bank Centre for Astrophysics is supported by a consolidated grant from STFC. The Nancay radio telescope is operated by the Paris Observatory, associated with the Centre National de la Recherche Scientifique (CNRS) and acknowledges financial support from the ‘Programme National de Cosmologie et Galaxies (PNCG)’ and ‘Gravitation, Références, Astronomie, Métrologie (GRAM)’ programmes of CNRS/INSU, France. LG is supported by National Natural Science Foundation of China No.11873076. The Flatiron Institute is supported by the Simons Foundation. The NANOGrav work in this paper was supported by National Science Foundation (NSF) PIRE program award number 0968296 and NSF Physics Frontiers Center award number 1430824. KJL is supported by XDB23010200, National Basic Research Program of China, 973 Program, 2015CB857101, and NSFC U15311243, 11690024, and the MPG funding for the Max-Planck Partner Group. LG was supported by National Natural Science Foundation of China (NSFC) (nos. U1431117) and #11010250. EG acknowledges the support from IMPRS Bonn/Cologne and the Bonn–Cologne Graduate School (BCGS). YW acknowledges the support from NSFC under grants 11503007, 91636111, and 11690021. SBS acknowledges the support of NSF EPSCoR award number 1458952. KL and GD acknowledge the financial support by the European Research Council for the ERC Synergy Grant BlackHoleCam under contract no. 610058. Part of this research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Work at NRL is supported by NASA. We thank Patrizia Tavella from the BIPM for helping us understand the nomenclature used when describing time standards.

REFERENCES

- Arias E. F., Petit G., 2019, *Ann. Phys.*, 531, 1900068
 Arias E. F., Panfilo G., Petit G., 2011, *Metrologia*, 48, S145
 Arzoumanian Z. et al., 2018, *ApJ*, 859, 47
 Babak S. et al., 2016, *MNRAS*, 455, 1665
 Bassa C. G. et al., 2016, *MNRAS*, 460, 2207
 Brook P. R. et al., 2018, *ApJ*, 868, 122
 Burke-Spolaor S. et al., 2018, *A&A*, 27, 5

- Caballero R. N. et al., 2016, *MNRAS*, 457, 4421
 Caballero R. N. et al., 2018, *MNRAS*, 481, 5501
 Desvignes G. et al., 2016, *MNRAS*, 458, 3341
 Edwards R. T., Hobbs G. B., Manchester R. N., 2006, *MNRAS*, 372, 1549
 Feroz F., Hobson M. P., Bridges M., 2009, *MNRAS*, 398, 1601
 Folkner W. M., Williams J. G., Boggs D. H., 2009, IPN Progress Report 42-178, ‘The Planetary and Lunar Ephemeris DE 412’. NASA Jet Propulsion Laboratory, Pasadena, CA
 Guinot B., Petit G., 1991, *A&A*, 248, 292
 Hellings R. W., Downs G. S., 1983, *ApJ*, 265, L39
 Hobbs G., Dai S., 2017, preprint ([arXiv:1707.01615](https://arxiv.org/abs/1707.01615))
 Hobbs G. et al., 2009, *MNRAS*, 394, 1945
 Hobbs G. et al., 2010, *Class. Quantum Gravity*, 27, 084013
 Hobbs G. et al., 2012, *MNRAS*, 427, 2780
 Hobbs G. B., Edwards R. T., Manchester R. N., 2006, *MNRAS*, 369, 655
 Kaplan D. L. et al., 2016, *ApJ*, 826, 86
 Keith M. J. et al., 2013, *MNRAS*, 429, 2161
 Lam M. T. et al., 2017, *ApJ*, 834, 35
 Lam M. T. et al., 2018, *ApJ*, 861, 132
 Lam M. T. et al., 2019, *ApJ*, 872, 193L
 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., Feroz F., van Haasteren R., Lee K. J., Shannon R. M., 2014, *MNRAS*, 437, 3004
 Lentati L. et al., 2015, *MNRAS*, 453, 2576
 Lentati L. et al., 2016, *MNRAS*, 458, 2161
 Li L., Wang N., Yuan J. P., Wang J. B., Hobbs G., Lentati L., Manchester R. N., 2016, *MNRAS*, 460, 4011
 Liu K., Keane E. F., Lee K. J., Kramer M., Cordes J. M., Purver M. B., 2012, *MNRAS*, 420, 361
 Manchester R. N. et al., 2013, *Publ. Astron. Soc. Aust.*, 30, e017
 Manchester R. N., Guo L., Hobbs G., Coles W. A., 2017, *Science of Time*. Springer Publishing, Astrophysics and Space Science Proceedings, New York
 Matsakis D. N., Foster R. S., 1996, in Chiao R. Y., ed., *Amazing Light*. Springer, New York, NY, p. 445
 McKee J. W. et al., 2016, *MNRAS*, 461, 2809
 Peil S., Hanssen J. L., Swanson T. B., Taylor J., Ekstrom C. R., 2014, *Metrologia*, 51, 263
 Petit G., 2003, in 35th Annual Precise Time and Time Interval (PTTI) Meeting, San Diego, December 2003. Institute of Navigation, Manassas, VA, p. 307
 Petit G., 2013, in Capitaine N., ed., *Journées 2013 Systèmes de Référence Spatio-Temporels*, Observatoire de Paris, p. 120
 Petit G., Tavella P., 1996, *A&A*, 308, 290
 Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 1996, *Numerical Recipes in Fortran 77: The Art of Scientific Computing*. 2nd edn., Cambridge Univ. Press, Cambridge
 Reardon D. J. et al., 2016, *MNRAS*, 455, 1751
 Rickett B. J., 1970, *MNRAS*, 150, 67
 Rodin A. E., 2011, in Capitaine N., ed., *Proceedings of the Journées 2010 Systèmes de Référence Spatio-Temporels*. Observatoire de Paris, p. 243
 Sesana A., Shankar F., Bernardi M., Sheth R. K., 2016, *MNRAS*, 463, L6
 Shannon R. M. et al., 2014, *MNRAS*, 443, 1463
 Shannon R. M. et al., 2015, *Science*, 349, 1522
 Shannon R. M. et al., 2016, *ApJ*, 828, L1
 The NANOGrav Collaboration, 2015, *ApJ*, 813, 65
 Tiburzi C., 2018, *Publ. Astron. Soc. Aust.*, 35, e013
 Tiburzi C. et al., 2016, *MNRAS*, 455, 4339
 van Haasteren R., Levin Y., McDonald P., Lu T., 2009, *MNRAS*, 395, 1005
 Verbiest J. P. W., Shaifullah G. M., 2018, *Class. Quantum Gravity*, 35, 133001
 Verbiest J. P. W. et al., 2016, *MNRAS*, 458, 1267
 Yin D.-S., Gao Y.-P., Zhao S.-H., 2017, *Chin. Astron. Astrophys.*, 41, 430

APPENDIX A: FREQUENTIST NOISE MODELLING

Table A1 contains, in column order, each pulsar name, the DM, its first time derivative, annual sinusoidal, and cosinusoidal components to DM(t), the grid spacing for measuring DM(t), extra DM-covariance parameters (a and b ; see below), and the band-independent low-frequency noise parameters (α , P_0 and f_c ; see below).

Scripts were used to produce the data products based on the original IPTA data sets. The scripts are available from our data release (see Section C) and contains RUNPROCESSSCRIPT.TCSH, which provides the basic processing steps as follows:

(i) A check is carried out to ensure that the required software packages are installed. The frequentist-style processing described here was carried out within a virtual machine on a single laptop.

(ii) The arrival time and parameter files are corrected for known issues. Such issues include correcting telescope observing codes, removing any very early, or very recent observations (to ensure a well-defined time range for obtaining the clock signal). IPTA-defined parameters such as phase jumps, white noise models etc. are removed. The arrival times are subsequently sorted into time order. PSRs J0437–4715, J0711–6830, J1045–4509, J1603–7202, J1732–5049, J1824–2452A, and J2129–5721 are only observed by the PPTA. For these pulsars we use the more recent Reardon et al. (2016) arrival time and parameter files instead of the IPTA data release. The Parkes observations for PSR J1909–3744 were also taken directly from Reardon et al. (2016).

(iii) The parameter files were updated to ensure that the JPL DE421 solar ephemeris was used in the timing procedure and the ToAs referred to TT(BIPM17). Default TEMPO2 parameters (such as models for the dispersion measure variations caused by the solar wind) were used.

(iv) The NANOGrav data sets are provided in multiple sub-bands. We have used the AVERAGEDATA plugin in TEMPO2 to produce a single ToA and corresponding uncertainty for each independent observation. The AVERAGEDATA plugin (1) identifies observations close in time (for us we select all simultaneous observations over different bands observed by NANOGrav), (2) for each block of data the non-weighted mean of the observation times are determined and all the time-dependent parameters in the timing model updates to this epoch, (3) a weighted fit is carried out for a phase offset and its uncertainty using only the specific block of data, (4) a pseudo-observation is added with the arrival time being the new epoch and the frequency and observatory site-code being determined from the closest, in time, actual observation. The uncertainty of this point is set to the uncertainty on the fitted phase offset, (5) timing residuals are re-formed using this pseudo-data point and the residual subtracted from its arrival time and this process repeated until convergence. A new arrival time file is then produced with only the pseudo-observations. This process has been carefully tested and the long-term timing results (as used when determining the clock signal) are unchanged when the averaged data points are used.

(v) The EFACEQUAD plugin to TEMPO2 is run on each observing system for each pulsar. This plugin determines the white noise corrections ‘EFAC’ and ‘EQUAD’ to allow for a scaling error in each ToA uncertainty along with an independent source of white noise. They have the form

$$\sigma'_i = \text{EFAC} (\sigma_i^2 + \text{EQUAD}^2)^{0.5}, \quad (\text{A1})$$

Table A1. Parameters describing the DM model (columns 2–8) and the red-noise model (last 3 columns) used for each pulsar. Note that the DM modelling was carried out using all the available observations (including the lowest frequency data, which was removed for subsequent processing).

Pulsar	DM ₀ (cm ⁻³ pc)	dDM/dt (cm ⁻³ pc yr ⁻¹)	Sine (cm–3pc)	Cosine (cm ⁻³ pc)	Grid (Δt _{DM} yr)	a (s ²)	b (d)	α	P ₀ (yr ³)	f _c (yr ⁻¹)
J0030+0451	4.33	3.5 × 10 ⁻⁵	–	–	–	–	–	2	6.79851 × 10 ⁻²⁸	0.3
J0034–0534	13.77	–3.5 × 10 ⁻⁵	–	–	–	–	–	4	9.5249 × 10 ⁻²⁶	0.5
J0218+4232	61.25	–8.9 × 10 ⁻⁴	–	–	–	–	–	4	2.41121 × 10 ⁻²⁶	0.6
J0437–4715	2.64	–	–	–	0.2	–	–	3	1.14 × 10 ⁻²⁷	0.0673
J0610–2100	60.64	–	–	–	–	–	–	3	3.04161 × 10 ⁻²⁷	0.4
J0613–0200	38.78	–1.7 × 10 ⁻⁴	–1.0 × 10 ⁻⁴	–1.0 × 10 ⁻⁴	0.3	–	–	3	9.89339 × 10 ⁻²⁷	0.2
J0621+1002	36.47	–1.1 × 10 ⁻²	–	–	–	–	–	4	3.4345 × 10 ⁻²⁵	0.8
J0711–6830	18.41	1.0 × 10 ⁻⁴	–	–	0.5	–	–	2.5	1.03113 × 10 ⁻²⁷	0.6
J0751+1807	30.25	–2.2 × 10 ⁻⁴	–	–	–	–	–	6	4.84735 × 10 ⁻²⁷	0.6
J0900–3144	75.70	–6.2 × 10 ⁻⁵	–	–	–	–	–	2	5.0048 × 10 ⁻²⁷	0.2
J1012+5307	9.02	3.4 × 10 ⁻⁵	–	–	0.3	–	–	4	1.02144 × 10 ⁻²⁷	1
J1022+1001	10.25	–4.3 × 10 ⁻⁶	–	–	0.5	–	–	6	2.295 × 10 ⁻²⁷	1
J1024–0719	6.49	1.3 × 10 ⁻⁴	–	–	–	–	–	4	6.70908 × 10 ⁻²⁴	0.07
J1045–4509	58.14	–3.7 × 10 ⁻³	–7.3 × 10 ⁻⁴	1.9 × 10 ⁻⁴	0.5	1.668 × 10 ⁻¹¹	179	3	2 × 10 ⁻²⁴	0.0587
J1455–3330	13.56	7.9 × 10 ⁻⁴	–	–	–	–	–	2	1.76702 × 10 ⁻²⁶	0.2
J1600–3053	52.33	–5.9 × 10 ⁻⁴	–	–	0.3	–	–	4	1.22878 × 10 ⁻²⁷	0.4
J1603–7202	38.05	–	–	–	0.3	6.276 × 10 ⁻¹³	64	2.5	1.2 × 10 ⁻²⁵	0.065
J1640+2224	18.42	1.5 × 10 ⁻⁴	–	–	–	–	–	4	1.42107 × 10 ⁻²⁷	1
J1643–1224	62.41	–1.2 × 10 ⁻³	–2.9 × 10 ⁻⁴	–5.9 × 10 ⁻⁴	1.0	–	–	3	1.59602 × 10 ⁻²⁵	0.1
J1713+0747	15.99	2.7 × 10 ⁻⁵	–	–	0.3	–	–	3	4.71009 × 10 ⁻²⁷	0.07
J1721–2457	48.62	–	–	–	–	–	–	4	2.4091 × 10 ⁻²⁵	1
J1730–2304	9.61	3.9 × 10 ⁻⁵	–	–	0.3	–	–	4	2.01716 × 10 ⁻²⁶	0.3
J1732–5049	56.84	8.8 × 10 ⁻⁴	–	–	–	–	–	4	1.82694 × 10 ⁻²⁶	0.2
J1738+0333	33.79	–	–	–	–	–	–	3	1.8891 × 10 ⁻²⁷	0.5
J1744–1134	3.14	–1.7 × 10 ⁻⁴	–	–	0.3	–	–	1.5	1.17822 × 10 ⁻²⁸	0.2
J1751–2857	42.89	–	–	–	–	–	–	3	1.05861 × 10 ⁻²⁶	0.4
J1801–1417	57.21	–	–	–	–	–	–	4	1.33307 × 10 ⁻²⁵	0.3
J1802–2124	149.62	3.3 × 10 ⁻⁴	–	–	–	–	–	4	4.99447 × 10 ⁻²⁶	0.4
J1804–2717	24.63	–	–	–	–	–	–	3	1.95346 × 10 ⁻²⁶	0.5
J1824–2452A	119.89	1.1 × 10 ⁻³	–	–	0.5	–	–	3.5	6.26696 × 10 ⁻²⁵	0.17
J1843–1113	59.97	–	–	–	–	–	–	4	2.69402 × 10 ⁻²⁶	0.2
J1853+1303	30.55	–	–	–	–	–	–	4	4.37242 × 10 ⁻²⁸	2
J1857+0943	13.30	3.6 × 10 ⁻⁴	–	–	–	–	–	1.5	4.26553 × 10 ⁻²⁷	0.07
J1909–3744	10.39	–3.1 × 10 ⁻⁴	–	–	0.3	–	–	1.5	2.55001 × 10 ⁻²⁸	0.07
J1910+1256	38.09	3.8 × 10 ⁻³	–	–	–	–	–	3	5.86889 × 10 ⁻²⁷	0.3
J1911–1114	31.07	–	–	–	–	–	–	3	9.95153 × 10 ⁻²⁸	0.8
J1911+1347	30.98	–	–	–	–	–	–	3	3.77936 × 10 ⁻²⁸	0.4
J1918–0642	26.61	–	–	–	–	–	–	4	1.71992 × 10 ⁻²⁷	0.5
J1955+2908	104.47	–	–	–	–	–	–	5	9.46096 × 10 ⁻²⁷	0.8
J2010–1323	22.18	5.5 × 10 ⁻⁴	–	–	–	–	–	4	8.09517 × 10 ⁻²⁸	0.3
J2019+2425	17.10	–	–	–	–	–	–	4	3.40917 × 10 ⁻²⁶	0.4
J2033+1734	25.00	–	–	–	–	–	–	3	8.65127 × 10 ⁻²⁶	0.4
J2124–3358	4.60	9.9 × 10 ⁻⁵	–	–	0.3	–	–	3	2.61433 × 10 ⁻²⁵	0.07
J2129–5721	31.85	–1.6 × 10 ⁻⁴	–	–	–	–	–	3	1.3839 × 10 ⁻²⁷	0.3
J2145–0750	9.00	2.1 × 10 ⁻⁴	–	–	0.3	–	–	3.5	2.10773 × 10 ⁻²⁵	0.07
J2229+2643	22.68	4.5 × 10 ⁻⁴	–	–	–	–	–	4	4.3802 × 10 ⁻²⁷	0.7
J2317+1439	21.90	–5.5 × 10 ⁻⁴	–	–	0.3	–	–	3	2.04548 × 10 ⁻²⁴	0.07
J2322+2057	13.56	5.8 × 10 ⁻²	–	–	–	–	–	2	1.55902 × 10 ⁻²⁵	0.2

where σ_i is the initial uncertainty for the specified ToA. The residuals for a specific observing instrument and telescope site are ‘whitened’ using an interpolation scheme (known as ‘IFUNCS’). The plugin then trials a well-defined number of EFAC and EQUAD parameters. For each parameter pair a Kolmogorov–Smirnov (KS) test is carried out to compare the normalized, whitened residuals

with Gaussian, white noise. The EFAC/EQUAD parameter pair that yields residuals closest to white, Gaussian noise is subsequently recorded and used for later processing.

(vi) A new set of arrival time flags are introduced including specific observing bands with bandwidths of 500 MHz (i.e. 0–500 MHz, 500–1000 MHz etc.).

(vii) We add the data from each observing system into a final arrival-time file in order of data span. We start with the longest data span and then identify an overlapping data set in the same observing band. We use the overlapping region to obtain an initial measurement of the offset between the observing systems and then repeat.

(viii) Some pulsars have only been observed in a single band. However, other pulsars have been observed in multiple bands. For such pulsars we fit for the dispersion measure and its first time derivative as part of the timing model. Where necessary we also fit for dispersion measure time variations following the Keith et al. (2013) procedure.

(ix) Any observations relating to the lowest two bands (0–500 MHz and 500–1000 MHz) are then removed as the high-precision timing observations needed for determining the clock signal are dominated by the higher frequency observations.

(x) For the majority of the pulsars we found that a single red noise model (obtained with the SPECTRALMODEL plugin) was sufficient for modelling the band-independent, low-frequency noise. The power spectral density, $P(f)$, of the red noise was assumed to have the form

$$P(f) = P_0 (1 + (f/f_c)^2)^{-\alpha/2}. \quad (\text{A2})$$

An iterative process was subsequently carried out in which the low-frequency noise modelling, jump fitting, and parameter estimation were repeatedly checked. In a few cases, in which the noise was significant, and there was only single-band data in the earlier observations we used the Reardon et al. (2016) split-modelling method in which the covariance function of the DM variations were modelled by:

$$\text{Cov}[\text{DM}(\tau)] = a \exp(-(\tau/b)^2). \quad (\text{A3})$$

For PSRs J0613–0200, J1045–4509, and J1643–1224 we also modelled annual sinusoidal dispersion measure variations (see Reardon et al. 2016 for details).

APPENDIX B: BAYESIAN MODELLING

Following the same reasoning as Caballero et al. (2018), we used a simpler noise model for the individual pulsars than the ones published in Lentati et al. (2016). Initial timing and noise models are produced with a joint timing and noise analysis using MULTINEST (Lentati et al. 2014), which utilizes the TEMPO2 routines to evaluate the timing model and Feroz, Hobson & Bridges (2009) to perform Bayesian inference of the timing and noise parameters via a nested-sampling Monte Carlo sampling. The analysis and noise model is the same as in Caballero et al. (2016). The uncertainties of the observing systems were weighted using, as in the frequentists analysis, a combination of EFAC and EQUAD per observing system (grouped as in Lentati et al. 2016). Following the same notation as in equation (A1), the rescaled uncertainties for each observing system is

$$\sigma'_i = [(\sigma \cdot \text{EFAC})^2 + \text{EQUAD}^2]^{0.5}. \quad (\text{B1})$$

The final noise models were produced using the FORTYTWO package. The white-noise levels in this stage are then regulated by a ‘global EFAC’ per pulsar, a method that is shown to perform adequately well (Lentati et al. 2015). The pulsar red noise and stochastic DM variations were modelled as stochastic, wide-sense stationary

processes, with power-spectra of the form

$$S(f) = \frac{A^2}{f} \left(\frac{f}{f_c} \right)^{2\alpha}, \quad (\text{B2})$$

where the spectrum is fully described by two parameters, namely the amplitude, A , and spectral index, α . These spectra have a sharp cut-off at $1/[\text{dataspan}]$. For values of spectral indices that are typical for MSPs, this is sufficient as power at frequencies lower than $1/[\text{dataspan}]$ is fitted out by timing parameters. In the red noise case, this is always true due to the presence of the spin and spin-down parameters (van Haasteren et al. 2009; Lee et al. 2012). To make the introduction of such a power-law spectrum in the model for the stochastic DM-variations, we introduce in our timing models for every pulsar a linear and a quadratic temporal variation term for the DM, that act as analogues to the pulsar spin and spin-down terms (Lee et al. 2014). The covariance matrices for the stochastic red noise and DM-variation noise have elements calculated as (Lee et al. 2014)

$$C_{r,ij} = \int_{1/T}^{\infty} S_r(f) \cos(2\pi f t_{ij}) df, \quad \text{and} \quad (\text{B3})$$

$$C_{\text{dm},ij} = \frac{\kappa^2 \int_{1/T}^{\infty} S_d(f) \cos(2\pi f t_{ij}) df}{v_i^2 v_j^2}. \quad (\text{B4})$$

In the above equations, the r and dm subscripts denote the case for red noise or DM variations, respectively. The i, j indices denote the time epochs, and t_{ij} is the time lag between the two respective time epochs, ν denotes the observing frequency, f is as usual the Fourier frequency, and $\kappa = 4.15 \times 10^{-3}$ s. The additional terms in the case of DM variations are due to the fact that the ToA delays from the dispersive effects of the interstellar medium are modelled to follow the dispersive law of cold homogeneous plasma, i.e. the time delay of a signal at two observing frequencies is proportional to the difference of the inverse squares of those frequencies.

In the Bayesian determination of the clock signal, we reduced the computational cost by keeping the pulsar noise properties fixed to their maximum-likelihood values, while performing Monte Carlo sampling of the clock parameters using MULTINEST. As detailed in Caballero et al. (2016), the likelihood function of the problem can be written separating the parameters of stochastic signals of interest, i.e. the clock and pulsar noise parameters, and the timing, deterministic terms for which we do not need to estimate the posterior distributions (nuisance terms). This approach is identical to the approach of van Haasteren et al. (2009) estimating the parameters of stochastic gravitational-wave background. Because the timing parameters are linear, we can marginalize over them analytically and use the reduced likelihood function (van Haasteren et al. 2009) as

$$\Lambda \propto \frac{1}{\sqrt{|\mathbf{C}\mathbf{C}'|}} \times \exp \left(-\frac{1}{2} \sum_{i,j,I,J} (\delta t_{I,i})^T \mathbf{C}'_{I,J,i,j} (\delta t_{J,j}) \right). \quad (\text{B5})$$

The I, J indices denote the different pulsars while the indices i, j denote the different time epochs. The timing model is now expressed via the vector $\delta \mathbf{t} = \delta \mathbf{t}_{\text{post}} - \mathbf{D}\delta(\epsilon)$, where \mathbf{D} is the timing model’s design matrix, $\delta \mathbf{t}_{\text{post}}$ is the post-fit residual vector, and $\delta(\epsilon)$ is a linear perturbation. The covariance matrix, \mathbf{C} , is the sum of the clock covariance matrix and the covariance matrices of the pulsar noise components (see Appendix B for relevant equations) and $\mathbf{ss}\mathbf{C}' = \mathbf{ss}\mathbf{C}^{-1} - \mathbf{ss}\mathbf{C}^{-1} \mathbf{ss}\mathbf{D} (\mathbf{ss}\mathbf{D}^{\text{ssT}} \mathbf{ss}\mathbf{C}^{-1} \mathbf{ss}\mathbf{D})^{-1} \mathbf{ss}\mathbf{D}^{\text{ssT}} \mathbf{ss}\mathbf{C}^{-1}$. As a red-noise process, the clock signal will induce on each pulsar an autocorrelation effect described by $\mathbf{C}_{\text{clk},ij}$, calculated with an

equation analogous to equation (B3) for the pulsar red noise. The clock signal is identical for all pulsars so we can easily add the cross-correlation effect and denote the elements of the clock covariance matrix, \mathbf{C}_{clk} as

$$C_{\text{clk},I,J,i,j} = C_{\text{clk},ij}C_{\text{clk},I,J}, \quad C_{\text{clk},I,I} = 1, \forall I \neq J. \quad (\text{B6})$$

APPENDIX C: RELEASING OUR DATA, PROCESSING SCRIPTS, AND RESULTS

The first IPTA data release is available from <http://www.ipta4gw.org>. We have also made available the exact input data used for the frequentist and Bayesian data processing, along with the script used to carry out the frequentist data processing. Our data release also provides the pulsar-derived clock waveforms and spectra (such as those shown in Fig. 3) for each pulsar. This data release can be obtained from the IPTA website (<http://www.ipta4gw.org>) and also from CSIRO's data archive (<https://doi.org/10.25919/5c354f2623ac5>).

¹CSIRO Astronomy and Space Science, Australia Telescope National Facility, Box 76, Epping, NSW 1710, Australia

²Shanghai Astronomical Observatory, CAS, Shanghai 200030, China

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

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

⁵Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093, USA

⁶Centre for Astrophysics and Supercomputing, Swinburne University of Technology, PO Box 218, Hawthorn, VIC 3122, Australia

⁷U.S. Naval Observatory, Naval Research Laboratory, 4555 Overlook Ave., SW, Washington D.C., 20375, USA

⁸National Time Service Center, CAS, Xi'an, Shaanxi 710600, China

⁹Astrophysics Science Division, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA

¹⁰ASTRON, the Netherlands Institute for Radio Astronomy, Oude Hoogeveensedijk 4, NL-7991 PD Dwingeloo, the Netherlands

¹¹International Centre for Radio Astronomy Research, Curtin University, Bentley, WA 6102, Australia

¹²Cornell Center for Astrophysics and Planetary Science, Cornell University, Ithaca, NY 14853, USA

¹³Cornell Center for Advanced Computing, Cornell University, Ithaca, NY 14853, USA

¹⁴Department of Physics and Astronomy, West Virginia University, Morgantown, WV 26506, USA

¹⁵Center for Gravitational Waves and Cosmology, West Virginia University, Chestnut Ridge Research Building, Morgantown, WV 26505, USA

¹⁶Laboratoire de Physique et Chimie de l'Environnement et de l'Espace LPC2E CNRS-Université d'Orléans, F-45071 Orléans, France

¹⁷Station de radioastronomie de Nançay, Observatoire de Paris, PSL University, CNRS/INSU F-18330 Nançay, France

¹⁸Department of Physics, Hillsdale College, 33 E. College Street, Hillsdale, MI 49242, USA

¹⁹University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK

²⁰Department of Astrophysics/IMAPP, Radboud University, PO Box 9010, NL-6500 GL Nijmegen, the Netherlands

²¹Jodrell Bank Centre for Astrophysics, University of Manchester, Manchester M13 9PL, UK

²²Space Science Division, Naval Research Laboratory, Washington, DC 20375-5352, USA

²³Hungarian Academy of Sciences MTA-ELTE 'Extragalactic Astrophysics' Research Group, Institute of Physics, Eötvös Loránd University, Pázmány P. s. 1/A, Budapest 1117, Hungary

²⁴Green Bank Observatory, PO Box 2, Green Bank, WV 24944, USA

²⁵Center for Computational Astrophysics, Flatiron Institute, 162 Fifth Ave, New York, NY 10010, USA

²⁶Physics Department, Lafayette College, Easton, PA 18901, USA

²⁷INAF-Osservatorio Astronomico di Cagliari, via della Scienza 5, I-09047 Selargius (CA), Italy

²⁸Università degli Studi di Cagliari, Dip. Fisica, S.P. Monserrato-Sestu Km 0.700, I-09042 Monserrato, Italy

²⁹CSIRO Scientific Computing, Australian Technology Park, Locked Bag 9013, Alexandria, NSW 1435, Australia

³⁰School of Physics and Astronomy, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK

³¹Australian Research Council Centre for Excellence for Gravitational-Wave Discovery

³²Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA

³³Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, BC V6T 1Z1, Canada

³⁴Center for Gravitation, Cosmology and Astrophysics, Department of Physics, University of Wisconsin-Milwaukee, PO Box 413, Milwaukee, WI 53201, USA

³⁵TAPIR, California Institute of Technology, MC 350-17, Pasadena, CA 91125, USA

³⁶Laboratoire Univers et Théories LUTH, Observatoire de Paris, PSL University, CNRS/INSU, Université Paris Diderot, 5 Place Jules Janssen, F-92190 Meudon, France

³⁷Xinjiang Astronomical Observatory, Chinese Academy of Sciences, 150 Science 1-Street, Urumqi, Xinjiang 830011, China

³⁸School of Physics, Huazhong University of Science and Technology, Wuhan, Hubei Province 430074, China

³⁹Monash Centre for Astrophysics (MoCA), School of Physics and Astronomy, Monash University, VIC 3800, Australia

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.