Kalman tracking and parameter estimation of continuous gravitational waves with a pulsar timing array: inclusion of pulsar terms

Tom Kimpson[1,2]⋆, Andrew Melatos[1,2], Joseph O'Leary[1,2], Julian B. Carlin[1,2], Robin J. Evans[3], William Moran[3], Tong Cheunchitra[1,2], Wenhao Dong[1,2], Liam Dunn[1,2], Julian Greentree[3], Nicholas J. O'Neill[1,2], Sofia Suvorova[3], Kok Hong Thong[1,2], Andrés F. Vargas[1,2]

[1]*School of Physics, University of Melbourne, Parkville, VIC 3010, Australia*
[2]*OzGrav, University of Melbourne, Parkville, VIC 3010, Australia*
[3]*Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, Victoria 3010, Australia*

Last updated 22 November 2023

ABSTRACT

Pulsar timing arrays may be able to detect continuous nanohertz gravitational waves emitted by individual supermassive black hole binaries. The gravitational wave (GW) signal is composed of two components, the "Earth term" and "pulsar term" which correspond to the GW incident on the Earth and the pulsar respectively. The data analysis procedure can be formulated within a state-space framework, wherein intrinsic achromatic spin wandering is tracked simultaneously with the modulation induced by a single-source gravitational wave in the pulse times of arrival. We show how to extend previous state-space analyses to include the pulsar terms. It is shown via astrophysically representative software injections in Gaussian measurement noise that the inclusion of the pulsar terms eliminates the biases in the parameter estimates, increasing the parameter accuracy by X percent and the minimal detectable strain by Y percent relative to the Earth-terms-only approach.

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

1 INTRODUCTION

Nanohertz gravitational waves (GWs) are produced by the inspiral of supermassive black hole binaries (SMBHBs; Rajagopal & Romani 1995; Jaffe & Backer 2003; Wyithe & Loeb 2003; Sesana 2013; McWilliams et al. 2014; Ravi et al. 2015; Burke-Spolaor et al. 2019; Sykes et al. 2022). A nHz GW influences the time of arrival (TOA) at the Earth of radio pulses from pulsars. A pulsar timing array (PTA; Tiburzi 2018; Verbiest et al. 2021) simultaneously measures the TOAs from multiple pulsars. In this way it is possible to detect GWs at frequencies that are prohibitively low for terrestrial interferometers. Consilient evidence from PTA observations has been presented (Agazie et al. 2023a; Antoniadis et al. 2023a; Reardon et al. 2023; Xu et al. 2023) for a stochastic GW background, which arises from the incoherent superposition of multiple SMBHB sources (Allen 1997; Sesana et al. 2008; Christensen 2019; Renzini et al. 2022). SMBHBs which are sufficiently close and massive may be individually resolvable, though no definitive detection has yet been made (Jenet et al. 2004; Zhu et al. 2014; Babak et al. 2016; Arzoumanian et al. 2023; Antoniadis et al. 2023c).

The modulation in the pulsar TOAs due to the presence of a GW has two contributions: an "Earth term" and a "pulsar term" which correspond to the GW incident on the Earth and the pulsar respectively. Whilst the Earth term is phase coherent between all pulsars, the pulsar terms have uncorrelated phases. Consequently

the pulsar terms are typically considered as a source of self-noise and dropped from many standard PTA analyses (e.g. Sesana & Vecchio 2010; Babak & Sesana 2012a; Petiteau et al. 2013; Zhu et al. 2015; Taylor et al. 2016; Goldstein et al. 2018; Charisi et al. 2023). However, dropping the pulsar terms leads to biases in the inferred parameters, and reduces the detection probability (Zhu et al. 2016; Chen & Wang 2022; Kimpson, Melatos, O'Leary, Carlin et al. Kimpson et al.).

Kimpson, Melatos, O'Leary, Carlin et al. (Kimpson et al.) (K24 hereafter) introduced a novel, state-space method for the detection and parameter estimation of continuous GWs from individual SMBHBs, that self-consistently tracks the intrinsic timing noise in PTA pulsars (e.g. Shannon & Cordes 2010a; Lasky et al. 2015; Caballero et al. 2016; Goncharov et al. 2021) and disentangles it from GW-induced TOA modulations. The method described in K24 follows the example of standard PTA analyses and also drops the pulsar term. In this paper we undertake two key tasks. Firstly we extend the methods presented in K24 to include the pulsar term. We show how to reparametrise the state-space model in terms of an additional pulsars-specific phase term. We demonstrate how the static GW parameters and the additional pulsar phase term can be estimated by combining a Kalman filter (Kalman 1960) with a Bayesian nested sampler (Skilling 2006; Ashton et al. 2022). Secondly we quantify how the inclusion of the pulsar term affects (i) the accuracy with which the static GW parameters can be estimated, and (ii) the minimal detectable GW strain, compared to the case

---

⋆ Contact e-mail: tom.kimpson@unimelb.edu.au

where the pulsar terms are omitted.

The paper is organised as follows. In Section 2 we briefly review the state-space model of PTA data analysis introduced in K24. In Section 3 we show how to include the pulsar term in the model via a convenient reparametrisation. In Section 4 we test the updated model, inclusive of the pulsar term, on synthetic data for both a single representative GW source and across an astrophysically relevant domain of SMBHB source parameters. In Section 5 we compare the parameter estimation accuracy and the minimal detectable GW strain when including the pulsar terms to the case where the pulsar terms are omitted. Conclusions are drawn in Section 6. Throughout the paper we adopt the natural units, with $c = G = \hbar = 1$, and metric signature $(-, +, +, +)$.

2 STATE-SPACE FORMULATION

In this section we briefly review the state-space formulation of PTA analysis presented in K24.

There are N pulsars in the array. The index n labels the n-th pulsar, with $1 \leqslant n \leqslant N$. The intrinsic spin frequency of the n-th pulsar, $f_p^{(n)}(t)$, as measured in the local, freely-falling rest frame of the pulsar's centre of mass, evolves according to a stochastic differential equation. The radio pulse frequency measured by an observer at Earth, $f_m^{(n)}(t)$, is related to $f_p^{(n)}(t)$ via a measurement equation which quantifies the influence of the GW. In Section 2.1 we outline the phenomenological model for the evolution of $f_p^{(n)}(t)$. In Section 2.2 we outline the measurement equation which relates $f_m^{(n)}(t)$ to $f_p^{(n)}(t)$. In Section 2.3 we summarise the parameters of the model. The model is formulated in terms of pulse frequency time series as a proof-of-principle and to maintain consistency with previous work (Meyers et al. 2021a,b; Kimpson, Melatos, O'Leary, Carlin et al. Kimpson et al.). It will be necessary to modify the method to accept pulse TOAs instead of a pulse frequency time series when implementing it to analyze real data, a subtle generalization we defer to future work.

2.1 Spin evolution

The intrinsic pulsar spin frequency evolves according to an Ornstein-Uhlenbeck process, described by a Langevin equation with a time-dependent drift term (Vargas & Melatos 2023),

$$\frac{df_p^{(n)}}{dt} = -\gamma^{(n)} [f_p^{(n)} - f_{em}^{(n)}(t)] + \dot{f}_{em}^{(n)}(t) + \xi^{(n)}(t) , \tag{1}$$

where $f_{em}^{(n)}$ is the deterministic evolution, $\dot{f}_{em}^{(n)}$ is the spin derivative, $\gamma^{(n)}$ is a damping constant whose reciprocal specifies the mean-reversion timescale, and $\xi^{(n)}(t)$ is a white noise stochastic process which satisfies

$$\langle \xi^{(n)}(t) \rangle = 0 , \tag{2}$$

$$\langle \xi^{(n)}(t) \xi^{(n')}(t') \rangle = [\sigma^{(n)}]^2 \delta_{n,n'}(t - t') . \tag{3}$$

In Equation (3) $[\sigma^{(n)}]^2$ is the variance of $\xi^{(n)}$ which parametrizes the noise amplitude and the gives characteristic root mean square fluctuations $\approx \sigma^{(n)}/[\gamma^{(n)}]^{1/2}$ in $f_p^{(n)}(t)$ (Gardiner 2009). The deterministic evolution $f_{em}^{(n)}$ is attributed to magnetic dipole braking,

with braking index $n_{em} = 3$ (Goldreich & Julian 1969). PTAs are typically composed of millisecond pulsars, for which the quadratic correction due to n_{em} in $f_p^{(n)}(t)$ is negligible over the observation time $T_{obs} \sim 10$ yr. Consequently, $f_{em}^{(n)}(t)$ can be approximated accurately by

$$f_{em}^{(n)}(t) = f_{em}^{(n)}(t_1) + \dot{f}_{em}^{(n)}(t_1)t , \tag{4}$$

where an overdot denotes a derivative with respect to t and t_1 labels the time of the first TOA.

Equations (1) – (4) are inherently an idealized phenomenological model to capture the main qualitative features of a typical PTA pulsar's observed spin evolution, i.e. random, mean-reverting, small-amplitude excursions around a smooth, secular trend (Agazie et al. 2023b; Antoniadis et al. 2023b; Zic et al. 2023). Such a phenomenological model is necessary as a predictive, first-principles theory of timing noise does not currently exist c.f. the multiple theorized mechanisms of timing noise (see discussion in Section 1 of Kimpson, Melatos, O'Leary, Carlin et al. (Kimpson et al.)). Similar approaches, analogous to Equations (1) – (4) have been followed successfully in other timing analyses in the context of anomalous braking indices (Vargas & Melatos 2023) and hidden Markov model glitch searches (Melatos et al. 2020; Lower et al. 2021; Dunn et al. 2022, 2023). However, Equations (1) – (4) do contain certain modelling omissions and idealisations which much be recognized (Meyers et al. 2021b,a; Vargas & Melatos 2023). These include the exclusion of phenomenological elements, which are likely to be present in reality, e.g. the classic, two-component, crust-superfluid structure inferred from post-glitch recoveries (Baym et al. 1969; van Eysden & Melatos 2010; Gügercinoğlu & Alpar 2017; Meyers et al. 2021a,b); the exclusion of non-Gaussian excursions such as Lévy flights (Sornette 2004); the suitability of the white noise driver $\xi(t)$ for millisecond pulsars in PTAs; and the formal interpretation of $d^2 f_p/dt^2$ due to $\xi(t)$ in Equation (1) not being differentiable.

2.2 Modulation of pulsar frequency by a GW

In the presence of a GW, the intrinsic pulsar spin frequency for the n-th pulsar is related to the radio pulse frequency measured by an observer on Earth via a a non-linear measurement equation:

$$f_m^{(n)}(t) = f_p^{(n)} \left[t - d^{(n)} \right] g^{(n)}(t) + \varepsilon^{(n)}(t) , \tag{5}$$

where $d^{(n)}$ labels the distance to the n-th pulsar, and $\varepsilon^{(n)}$ is a Gaussian measurement noise which satisfies

$$\langle \varepsilon^{(n)}(t) \rangle = 0 , \tag{6}$$

$$\langle \varepsilon^{(n)}(t) \varepsilon^{(n')}(t') \rangle = \sigma_m^2 \delta_{n,n'} \delta(t - t') , \tag{7}$$

where σ_m is the covariance of the measurement noise at the telescope and is shared between all pulsars. The measurement function $g^{(n)}(t)$ is

$$\begin{aligned} g^{(n)}(t) = 1 &- \frac{H_{ij} [q^{(n)}]^i [q^{(n)}]^j}{2[1 + \boldsymbol{n} \cdot \boldsymbol{q}^{(n)}]} \\ &\times \left[\cos\left(-\Omega t + \Phi_0 \right) \right. \\ &\left. - \cos\left\{ -\Omega t + \Phi_0 + \Omega \left[1 + \boldsymbol{n} \cdot \boldsymbol{q}^{(n)} \right] d^{(n)} \right\} \right] . \end{aligned} \tag{8}$$

where $[q^{(n)}]^i$ labels the i-th coordinate component of the n-th pulsar's position vector $\boldsymbol{q}^{(n)}$, Ω is the constant angular frequency of the GW, $\boldsymbol{n}$ is a unit vector specifying the direction of propagation

of the GW, H_{ij} is the spatial part of the gravitational plane wave amplitude tensor, and Φ_0 the phase offset of the GW with respect to some reference time.

Equation (8) contains two simplifying assumptions which must be recognised:

  (i) The GW is a monochromatic plane wave.
 (ii) The pulsar locations are constant with respect to the observer

Regarding point (i), studies of SMBHB inspirals in the PTA context show that the gravitational wave frequency $f_{\rm gw}$ (= $\Omega/2\pi$) evolves according to (e.g. Sesana & Vecchio 2010),

$$\Delta f_{\rm gw} = 0.05\,{\rm nHz} \left(\frac{M_{\rm c}}{10^{8.5}M_\odot}\right)^{5/3} \left[\frac{f_{\rm gw}(t=t_1)}{50\,{\rm nHz}}\right]^{11/3} \left(\frac{T_{\rm obs}}{10{\rm yr}}\right)\,, \quad (9)$$

where $M_{\rm c}$ is the chirp mass of the SMBHB, $f_{\rm gw}(t=t_1)$ is the GW frequency at the time of the first observation, and $T_{\rm obs}$ is the length of time over which $\Delta f_{\rm gw}$ is measured, which for PTAs is ~ 10 years. A source can be considered monochromatic, if $\Delta f_{\rm gw}$ is less than the PTA frequency resolution $1/T_{\rm obs}$. Equation (9) implies $(M_{\rm c}/10^{8.5}M_\odot)^{5/3}[f_{\rm gw}(t=t_1)/50\,{\rm nHz}]^{11/3}(T_{\rm obs}/10{\rm yr})^2 \gtrsim 20$ for $\Delta f_{\rm gw} \gtrsim 1/T_{\rm obs}$. The majority of SMBHBs detectable with PTAs are expected to satisfy $\Delta f_{\rm gw} < 1/T_{\rm obs}$; for a PTA composed of pulsars with a mean distance of 1.5 kpc, 78% of simulated SMBHBs satisfy this condition for the current IPTA, whilst for the second phase of the Square Kilometer Array this fraction drops to 52%; see Figure 7 in Rosado et al. (2015). We are therefore justified in treating the GW source as monochromatic as a first pass in this introductory paper (Sesana et al. 2008; Sesana & Vecchio 2010; Ellis et al. 2012).

Regarding point (ii) whilst in practice the pulsar locations vary with respect to the Earth, pulsar TOAs are defined relative to the Solar System barycentre (SSB).The barycentering correction is typically applied when generating TOAs, e.g. with TEMPO2 (Hobbs et al. 2006) and related timing software, and is inherited by the frequency time series. Some pulsars, including some PTA pulsars, do have non-negligible proper motions of order 10^2 km s^{-1} after the barycentering corrections have been applied (e.g. Jankowski et al. 2018), but we do not consider this effect in this paper.

2.3 Static parameters of the model

For the model described in Sections 2.1 and 2.2, the set of static parameters which are specific to the pulsars in the array are

$$\boldsymbol{\theta}_{\rm psr} = \left\{\gamma^{(n)}, \sigma^{(n)}, f_{\rm em}^{(n)}(t_1), \dot{f}_{\rm em}^{(n)}(t_1), d^{(n)}\right\}_{1\leqslant n\leqslant N}\,. \quad (10)$$

The parameters which are specific to the GW source are

$$\boldsymbol{\theta}_{\rm gw} = \{h_0, \iota, \psi, \delta, \alpha, \Omega, \Phi_0\}\,, \quad (11)$$

where h_0 is the characteristic wave strain, ι the orbital inclination, ψ the polarisation angle of the wave, δ the declination of the source and α its right ascension. These parameters enter the model through Equation (8) where $H_{ij} = H_{ij}(h_0, \iota, \psi, \delta, \alpha)$ and $\boldsymbol{n} = \boldsymbol{n}(\delta, \alpha)$. The complete set of $7 + 5N$ static parameters is denoted $\boldsymbol{\theta} = \boldsymbol{\theta}_{\rm gw} \cup \boldsymbol{\theta}_{\rm psr}$. While we have no prior information about the values of the members of $\boldsymbol{\theta}_{\rm gw}$, the members of $\boldsymbol{\theta}_{\rm psr}$ do have constraints from electromagnetic observations; for example estimates of $d^{(n)}$ are accurate to $\sim 10\%$ (Cordes & Lazio 2002; Verbiest et al. 2012; Desvignes et al. 2016; Yao et al. 2017).

3 INCLUSION OF THE PULSAR TERM

K24 uses a likelihood-based Bayesian framework for inference of the model parameters and model selection. That is, given a temporal sequence of noisy measurements $\boldsymbol{Y}(t)$, the posterior distribution of $\boldsymbol{\theta}$ is calculated by Bayes' Rule,

$$p(\boldsymbol{\theta}|\boldsymbol{Y}) = \frac{\mathcal{L}(\boldsymbol{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\mathcal{Z}}\,, \quad (12)$$

where $\pi(\boldsymbol{\theta})$ is the prior distribution on $\boldsymbol{\theta}$ and $\mathcal{Z}$ is the marginalised likelihood, or evidence

$$\mathcal{Z} = \int d\boldsymbol{\theta}\, \mathcal{L}(\boldsymbol{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\,, \quad (13)$$

and $\mathcal{L}(\boldsymbol{Y}|\boldsymbol{\theta})$ is the likelihood calculated via a Kalman filter (Kalman 1960). For our purposes $\boldsymbol{Y}(t) = \{f_{\rm m}^{(n)}(t)\}$. Given a specific realisation of $\boldsymbol{Y}$, $\mathcal{L}(\boldsymbol{Y}|\boldsymbol{\theta})$ is solely a function of $\boldsymbol{\theta}$, i.e. $\mathcal{L}(\boldsymbol{Y}|\boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\theta})$. Nested sampling (Skilling 2006) is used to estimate $p(\boldsymbol{\theta}|\boldsymbol{Y})$ and $\mathcal{Z}$. We refer the reader to Appendix A for a review of Kalman filtering, nested sampling and the combined workflow in the context of PTA searches for GWs.

The inference model used by the Kalman filter in K24 drops the pulsar term, consistent with other PTA analyses. In this section we show how to remove this limitation and include the pulsar term in the state-space inference model passed to the Kalman filter. In Section 3.1 we review how the pulsar term is dropped from typical PTA analyses and the modification to Equation (8) used in K24. In Section 3.2 we introduce a convenient reparametrisation of the pulsar term for likelihood-based inference methods like nested sampling. This new parametrisation is validated with synthetic data in Section 4.

3.1 Earth term

The influence of the GW on the pulsar radio pulses is completely described by Equation (8). The GW signal is composed of an "Earth term", corresponding to the $\cos(-\Omega t + \Phi_0)$ component, and a "pulsar term", corresponding to the $\cos\left\{-\Omega t + \Phi_0 + \Omega\left[1 + \boldsymbol{n}\cdot\boldsymbol{q}^{(n)}\right]d^{(n)}\right\}$ component. The Earth term describes the GW phase at the observer on Earth, while the pulsar term describes the GW phase at the individual pulsar. The Earth term depends only on the GW source parameters and so is shared across all pulsars. In contrast the pulsar term is a function of $d^{(n)}$ and $\boldsymbol{q}^{(n)}$ and so varies between pulsars.

Whilst the Earth term is phase coherent between all pulsars and so can be summed across the array to increase the total signal-to-noise, the pulsar terms each have individual phases which cannot be coherently summed and are instead generally considered as a source of self-noise. The pulsar term depends on $d^{(n)}$ which is generally poorly constrained by electromagnetic observations, with uncertainties much greater than the typical target GW wavelength. It is therefore not possible to set the phase of the pulsar term. Consequently the pulsar term is typically dropped from in standard PTA analysis (e.g. Sesana & Vecchio 2010; Babak & Sesana 2012a; Petiteau et al. 2013; Zhu et al. 2015; Taylor et al. 2016; Goldstein et al. 2018; Charisi et al. 2023). Dropping the pulsar term is also convenient computationally because it reduces the dimensionality of the parameter space, since the N values of $d^{(n)}$ do not have to be inferred, making the parameter estimation computationally cheaper. Within the state-space model described in Section 2, dropping the pulsar terms equates to using a modified measurement equation, c.f. Equations (5) and (8),

$$f_{\rm m}^{(n)}(t) = f_{\rm p}^{(n)}(t-d)g_{\rm Earth}^{(n)}(t) + \varepsilon^{(n)}(t)\,, \quad (14)$$

with

$$g_{\text{Earth}}^{(n)}(t) = 1 - \frac{H_{ij}[q^{(n)}]^i[q^{(n)}]^j}{2[1 + \boldsymbol{n} \cdot \boldsymbol{q}^{(n)}]} \cos(-\Omega t + \Phi_0) \, . \tag{15}$$

However, dropping the pulsar term leads to biases in the parameter estimation, in particular the source localisation, and a reduced detection probability (Zhu et al. 2016; Chen & Wang 2022; Kimpson, Melatos, O'Leary, Carlin et al. Kimpson et al.). It is therefore important to remove this limitation and include the pulsar term in PTA data analysis pipelines.

3.2 Reparametrisation of the pulsar term

We define a new parameter

$$\chi^{(n)} \equiv \Omega \left(1 + \boldsymbol{n} \cdot \boldsymbol{q}^{(n)}\right) d^{(n)} \quad \text{mod } 2\pi \, , \tag{16}$$

such that Equation (8) becomes

$$\begin{aligned} g^{(n)}(t) = 1 &- \frac{H_{ij}[q^{(n)}]^i[q^{(n)}]^j}{2[1 + \boldsymbol{n} \cdot \boldsymbol{q}^{(n)}]} \\ &\times \Big[\cos\left(-\Omega t + \Phi_0\right) \\ &- \cos\left\{-\Omega t + \Phi_0 + \chi^{(n)}\right\} \Big] \, . \end{aligned} \tag{17}$$

The reparametrisation in terms of $\chi^{(n)}$ treats the pulsar-dependent phase correction $\Omega \left(1 + \boldsymbol{n} \cdot \boldsymbol{q}^{(n)}\right) d^{(n)}$ in the argument of the cosine of the pulsar term as an explicit parameter to be inferred for each pulsar. The advantage of this approach, compared to trying to infer $d^{(n)}$ directly, is that the resulting likelihood surface across the parameter space (i.e. the shape of the function $\mathcal{L}(\boldsymbol{\theta})$) is smoother, with smaller derivatives. This enables sampling algorithms to efficiently find the likelihood maxima and estimate $p(\boldsymbol{\theta}|\boldsymbol{Y})$. This point is discussed in more detail in Appendix F. The new parametrisation also does not increase the dimension of the parameter space, because we have traded $d^{(n)}$ for $\chi^{(n)}$. The static parameters specific to the pulsars in the array (c.f. Equation (10)) are now

$$\boldsymbol{\theta}_{\text{psr}}' = \left\{\gamma^{(n)}, \sigma^{(n)}, f_{\text{em}}^{(n)}(t_1), \dot{f}_{\text{em}}^{(n)}(t_1)\right\}_{1 \leqslant n \leqslant N} \, , \tag{18}$$

whilst $\boldsymbol{\theta}_{\text{gw}}$ remains unchanged. We exclude $\chi^{(n)}$ from membership of $\boldsymbol{\theta}_{\text{psr}}'$ as the other elements of $\boldsymbol{\theta}_{\text{psr}}'$ depend solely on the properties of the pulsar, with no dependence on the properties of the GW source. In contrast $\chi^{(n)}$ is a cross term that depends on both the individual pulsar and the GW. We define the complete set of $7 + 4N + N$ static parameters to be inferred as $\boldsymbol{\theta}' = \boldsymbol{\theta}_{\text{gw}}' \cup \boldsymbol{\theta}_{\text{psr}}' \cup \chi^{(n)}$.

4 VALIDATION OF PULSAR-TERMS INCLUSION WITH SYNTHETIC DATA

In this section we apply the analysis scheme of K24 while including the pulsar terms using the parametrisation described in Section 3.2. As outlined in Section 3, the analysis scheme consists of a Kalman filter to track the state evolution of the intrinsic pulsar spin frequency $f_{\text{p}}^{(n)}(t)$ which provides a likelihood for nested sampling. Nested sampling then returns estimates of $\boldsymbol{\theta}'$ and $\mathcal{Z}$. We refer the reader to Appendix A and K24 for details of the Kalman filter - nested sampling scheme. We apply the analysis scheme to a PTA perturbed by a GW from an individual quasi-monochromatic SMBHB source.

The analysis of a stochastic background composed of the superposition of multiple sources is more challenging and is postponed to a forthcoming paper. The goal of this section is to validate the analysis scheme when the pulsar terms are included in the inference model.

In Section 4.1 we demonstrate how to generate synthetic data, namely noisy frequency time series $f_{\text{m}}^{(1)}(t), \dots, f_{\text{m}}^{(N)}(t)$, with which to validate the scheme. In Section 4.2 we apply the analysis scheme to the synthetic data to estimate $\boldsymbol{\theta}'$ for a particular, arbitrary, representative choice of $\boldsymbol{\theta}_{\text{gw}}$. The scheme is validated on multiple noise realisations of the pulsar process noise and the measurement noise in order to test the scheme more strongly and quantify the irreducible "cosmic" variance in the parameter estimates. In Section 4.3 we extend the tests of Section 4.2 to cover a broader parameter space and consider a range of astrophysically relevant SMBHB source parameters $\boldsymbol{\theta}_{\text{gw}}$.

4.1 Generating synthetic data

In order to generate synthetic frequency time series $\boldsymbol{Y}(t) = \{f_{\text{m}}^{(1)}(t), \dots, f_{\text{m}}^{(N)}(t)\}$, Equations (1)–(4) are integrated numerically using a Runge-Kutta Itô integrator implemented in the `sdeint` python package [1]. This produces random realizations of $f_{\text{p}}^{(n)}(t)$ for $1 \leqslant n \leqslant N$. The measured $f_{\text{m}}^{(n)}(t)$ are then produced via Equations (5)–(8). Equations (1)–(4) and (5)–(8) depend on the choice of static parameters $\boldsymbol{\theta}'$, as well as the choice of pulsar directions $\boldsymbol{q}^{(n)}$ and detector measurement noise covariance σ_{m} (Equation (20)). We now describe how $\boldsymbol{\theta}_{\text{psr}}$, $\boldsymbol{q}^{(n)}$ and σ_{m} are chosen. This is equivalent to specifying the configuration of a synthetic PTA. In Section 4.2 and Section 4.3 we describe how $\boldsymbol{\theta}_{\text{gw}}$ chosen. This is equivalent to specifying the synthetic SMBHB source.

In this paper we use the same values for $\boldsymbol{\theta}_{\text{psr}}$ as in K24. That is, we adopt the 47 millisecond pulsars (MSPs) in the 12.5 year NANOGrav dataset (Arzoumanian et al. 2020), i.e. $N = 47$. We consider all pulsars to be observed with a weekly cadence, $T_{\text{cad}} = 1$ week over a 10 year period. For each pulsar fiducial values for $\boldsymbol{q}^{(n)}$, $d^{(n)}$, $f_{\text{em}}^{(n)}(t_1)$, and $\dot{f}_{\text{em}}^{(n)}(t_1)$ are obtained via the Australia Telescope National Facility (ATNF) pulsar catalogue (Manchester et al. 2005) using the `psrqpy` package (Pitkin 2018). We identify $f_{\text{em}}^{(n)}(t_1)$ and $\dot{f}_{\text{em}}^{(n)}(t_1)$ with the pulsar barycentric rotation frequency and its time derivative respectively, as quoted in catalogues. No direct measurements exist for $\gamma^{(n)}$, though the mean reversion timescale typically satisfies $[\gamma^{(n)}]^{-1} \gg T_{\text{obs}}$ (Price et al. 2012; Meyers et al. 2021a,b; Vargas & Melatos 2023). In this paper, for the sake of simplicity in the absence of independent measurements, we fix $\gamma^{(n)} = 10^{-13} \text{ s}^{-1}$ for all n. Similarly, no direct measurements exist for $\sigma^{(n)}$. We relate the process noise in the frequency $\sigma^{(n)}$ to the root mean square TOA noise accumulated over one week such that

$$\sigma^{(n)} \approx \sigma_{\text{TOA}}^{(n)} f_{\text{p}}^{(n)} T_{\text{cad}}^{-3/2} \, . \tag{19}$$

where, following K24, we use the empirical timing noise model for MSPs from Shannon & Cordes (2010b) in order to set physically reasonable values on $\sigma_{\text{TOA}}^{(n)}$. The median $\sigma^{(n)}$ calculated in this way is $\sigma^{(n)} = 5.51 \times 10^{-24} \text{ s}^{-3/2}$, $\min[\sigma^{(n)}] = 1.67 \times 10^{-26} \text{s}^{-3/2}$ for PSR J0645+5158 and $\max[\sigma^{(n)}] = 2.56 \times 10^{-19} \text{ s}^{-3/2}$ for PSR J1939+2134.

The measurement noise covariance as defined in Equation (7) can

---

[1] https://github.com/mattja/sdeint

	Parameter	Injected value	Units	Prior
$\boldsymbol{\theta}_{\mathrm{gw}}$	Ω	5×10^{-7}	Hz	$\mathrm{LogUniform}(10^{-9}, 10^{-5})$
	α	1.0	rad	$\mathrm{Uniform}(0, 2\pi)$
	δ	1.0	rad	$\mathrm{Uniform}(-\pi/2, \pi/2)$
	ψ	2.50	rad	$\mathrm{Uniform}(0, 2\pi)$
	Φ_0	0.20	rad	$\mathrm{Uniform}(0, 2\pi)$
	h_0	5×10^{-15}	—	$\mathrm{LogUniform}(10^{-15}, 10^{-9})$
	ι	1.0	rad	$\mathrm{Uniform}(0, \pi)$
$\boldsymbol{\theta}'_{\mathrm{psr}}$	$f_{\mathrm{em}}^{(n)}(t_1)$	$f_{\mathrm{ATNF}}^{(n)}$	Hz	$\mathrm{Uniform}\left[f_{\mathrm{ATNF}}^{(n)} - 10^3 \eta_f^{(n)}, f_{\mathrm{ATNF}}^{(n)} + 10^3 \eta_f^{(n)}\right]$
	$\dot{f}_{\mathrm{em}}^{(n)}(t_1)$	$\dot{f}_{\mathrm{ATNF}}^{(n)}$	s^{-2}	$\mathrm{Uniform}\left[\dot{f}_{\mathrm{ATNF}}^{(n)} - 10^3 \eta_{\dot{f}}^{(n)}, \dot{f}_{\mathrm{ATNF}}^{(n)} + 10^3 \eta_{\dot{f}}^{(n)}\right]$
	$\sigma^{(n)}$	$\sigma_{\mathrm{SC}}^{(n)}$	$\mathrm{s}^{-3/2}$	$\mathrm{LogUniform}\left[10^{-2}\sigma_{\mathrm{SC}}^{(n)}, 10^2 \sigma_{\mathrm{SC}}^{(n)}\right]$
	$\gamma^{(n)}$	10^{-13}	s^{-1}	—
	$\chi^{(n)}$	$\Omega\left(1 + \boldsymbol{n} \cdot \boldsymbol{q}^{(n)}\right) d_{\mathrm{ATNF}}^{(n)}$	rad	$\mathrm{Uniform}(0, 2\pi)$

Table 1. Summary of injected static parameters used for generating synthetic data for validation of the analysis scheme when including the pulsar-term in Section **??**. The prior used for Bayesian inference (Section 4.2) on each parameter is also displayed (rightmost column). The top, middle and bottom sections of the table contain the values for $\boldsymbol{\theta}_{\mathrm{gw}}$, $\boldsymbol{\theta}_{\mathrm{psr}}$ and $\chi^{(n)}$ respectively. The subscript "ATNF" denotes values which have been obtained from the ATNF pulsar catalogue as described in Section 4.1. The subscript "SC" on the pulsar noise amplitude $\sigma^{(n)}$ indicates that the injected value has been calculated using Equation and (19) the empirical timing noise model for MSPs from Shannon & Cordes (2010a). The quantities $\eta_f^{(n)}$ and $\eta_{\dot{f}}^{(n)}$ are the errors in $f_{\mathrm{em}}^{(n)}(t_1)$ and $\dot{f}_{\mathrm{em}}^{(n)}(t_1)$ respectively, as quoted in the ATNF catalogue. We do not try to infer $\gamma^{(n)}$ as $\gamma^{(n)} T_{\mathrm{obs}} \sim 10^{-5}$ and so no prior is set. The choice of priors on $\boldsymbol{\theta}_{\mathrm{psr}}$ is discussed in Appendix C.

be approximately related to the uncertainty in the pulse TOA, σ_{TOA}, as

$$\sigma_{\mathrm{m}} \approx f_{\mathrm{p}}^{(n)} \sigma_{\mathrm{TOA}} T_{\mathrm{cad}}^{-1} . \tag{20}$$

For a millisecond pulsar with $f_{\mathrm{p}}^{(n)} \sim 100$ Hz observed with a weekly cadence and $\sigma_{\mathrm{TOA}} \sim 1\mu$s Equation (20) implies $\sigma_{\mathrm{m}} \sim 10^{-10}$ Hz. The most accurately timed pulsars might have $\sigma_{\mathrm{TOA}} \sim 10$ ns or $\sigma_{\mathrm{m}} \sim 10^{-12}$ Hz. Throughout this paper we fix $\sigma_{\mathrm{m}} = 10^{-11}$ Hz and take it as known *a priori* rather than a parameter to be inferred. When analysing real data this assumption can easily be relaxed. Whilst σ_{m} is the same for every pulsar, $f_{\mathrm{m}}^{(n)}$ is constructed by a different random realisation of $\varepsilon^{(n)}(t)$ for each pulsar.

4.2 Parameter estimation

In this section, we apply the Kalman filter in conjunction with nested sampling to calculate the joint posterior probability distribution $p(\boldsymbol{\theta}'|\boldsymbol{Y})$ for synthetic data $\boldsymbol{Y}$, and compare it to the known, injected values. The procedure is undertaken for multiple realisations of the noise processes $\xi^{(n)}(t)$ and $\varepsilon^{(n)}(t)$. That is, given the injected static parameters $\boldsymbol{\theta}'$, multiple realisations of $\boldsymbol{Y}$ are generated using the methods described in Section 4.1, and for each realisation we independently estimate $\boldsymbol{\theta}'$. The aims are (i) to demonstrate that the analysis scheme works (i.e. that it converges to a well-behaved, unimodal posterior for multiple noise realisations), (ii) to give a preliminary sense of its accuracy, and (iii) demonstrate that quantify the natural random dispersion in the one-dimensional posterior medians between noise realisations. The injected static GW source parameters $\boldsymbol{\theta}_{\mathrm{gw}}$ used for this mock analysis to generate $\boldsymbol{Y}$ are selected to be astrophysically reasonable and representative. Different values of the static GW source parameters are explored in Section 4.3. The static pulsar parameters $\boldsymbol{\theta}'_{\mathrm{psr}}$ are as specified in Section 4.1. All injected static parameters are summarised in the second column of Table 1. The specification of the priors on $\boldsymbol{\theta}'$ is described in Appendix C and summarised in the rightmost column

of Table 1.

Figure 1 displays results for the seven parameters in $\boldsymbol{\theta}_{\mathrm{gw}}$ for YY arbitrary noise realisations in the form of a traditional corner plot. The histograms are the one-dimensional posteriors for each parameter, marginalized over the six other parameters. The solid orange line marks the known injected value. The two-dimensional contours mark the (0.5, 1, 1.5, 2)-sigma level surfaces. All histograms and contours are consistent with a unimodal joint posterior, which peaks near the known, injected values. There is scant evidence of railing against the prior bounds. There is no strong evidence for correlations between parameter pairs, e.g. banana-shaped contours. The injected value is contained within the 90% credible interval of the one-dimensional marginalized posteriors in YY (i.e. YY %) out of the YY possible combinations of the seven parameters and YY realizations. For this representative example, with characteristic strain XX the analysis scheme estimates $\boldsymbol{\theta}_{\mathrm{gw}}$ accurately.

In addition to estimates of $\boldsymbol{\theta}_{\mathrm{gw}}$, it is of interest how well the pulsar-term phase correction parameters $\chi^{(n)}$ can be estimated, since this parameter is a key feature of our approach. Figure 2 displays the results for a subset of five parameters of $\chi^{(n)}$, $\{\chi^{(1)} \ldots \chi^{(5)}\}$ over YY noise realisations, analogously to Figure 1. We do not display the resulting corner plot for all elements of $\chi^{(n)}$ because it is too big (total number of panels is $N(N+1)/2 = 1128$). All elements of $\chi^{(n)}$ are recovered unambiguously across all noise realisations, with the injected value contained within the 90% credible interval of the one-dimensional marginalized posteriors in YY (i.e. YY %) out of the YY possible combinations of the seven parameters and YY realizations. Similar results are obtained for the remaining $4N$ parameters of $\boldsymbol{\theta}'_{\mathrm{psr}}$; the parameters are recovered unambiguously across the XX noise realisations. For the sake of brevity we do not display the calculated posterior distributions here because a) the inference of $\boldsymbol{\theta}_{\mathrm{gw}}$ is the main focus of this paper and PTA analyses, and b) the resulting corner plot is too big.

Figure 1 confirms two points: (i) the analysis scheme returns

accurate answers for arbitrary noise realisations and (ii) there is appreciable natural random dispersion in the one-dimensional posterior medians between noise realisations. Regarding the second point, quantifying the dispersion is important since it is a practical measure of the accuracy of the parameter estimation scheme when it is applied to real astronomical data, where the true parameter values and specific noise realisations are unknown.

TK: do we want an extended discussion re Wasserstein distance over many noise realisations analogous to what we did in paper 1? I think it is a bit distracting to main structure of paper which is: (i) show that including pulsar terms works generally (Section 4) and (ii) show the improvement over the Earth terms (Section 5). Relegating WD stuff to an appendix for now.

4.3 Extension to broader parameter space

Section 4.2 focuses on a single representative system, parameterised in Table 1. In this section we test the method for various systems, varying the SMBHB source parameters $\boldsymbol{\theta}_{\mathrm{gw}}$ through astrophysically relevant ranges. The aim of this section is to confirm that the analysis scheme can be applied across an astrophysically relevant parameter space and that the arbitrary choice of $\boldsymbol{\theta}_{\mathrm{gw}}$ in Section 4.2 is not particularly advantageous by accident.

We analyse X injections constructed by fixing $h_0 = $ Y and $\iota = 1.0$ rad and drawing the remaining five elements of $\boldsymbol{\theta}_{\mathrm{gw}}$ randomly from the prior distributions defined in Table 1. We fix h_0 and ι in order to maintain an approximately constant signal-to-noise ratio (SNR) across the X injections. For each injection we compute the posterior distribution of $\boldsymbol{\theta}_{\mathrm{gw}}$.

The results are displayed in Figure 3. Since it is prohibitive to produce corner plots analogous to Figure 1 for X injections, in order to summarise the results we use a parameter-parameter (PP) plot (Cook et al. 2006). A PP plot displays the fraction of injections included within a given credible interval of the estimated posterior, plotted as a function of the credible interval itself. In the ideal case of perfect recovery, the PP plot should be a diagonal line of unit slope. In Figure 3 the shaded grey contours enclose the 1σ, 2σ, and 3σ significance levels for X injections. We see that... TBD once simulations complete

5 EARTH TERMS VS PULSAR TERMS

In Section 4 we have validated the state-space analysis scheme when the pulsar terms are included in the inference model. In this section we compare the results obtained using the pulsar-term model (i.e. Equation (17)) to those obtained using the Earth-term model (i.e. Equation (15)) for the specific representative case of the individual quasi-monochromatic SMBHB source described in Table 1. The pulsar-term model and the Earth-terms model are applied to identical realisations of the data, generated using the procedure as described in Section 4.1. The priors on the static parameters in each model are also identical, with the addition of a uniform prior on $\chi^{(n)}$ for the pulsar terms model, as described in Appendix C.

In Section 5.1 we compare the accuracy of the estimates of $\boldsymbol{\theta}'$ returned by the two models. We consider two separate cases: a "high-SNR" case with $h_0 = 10^{-12}$ and a "low-SNR" case with $h_0 = $

5×10^{-15}. In Section 5.2 we compare the minimal detectable strain for the two models.

5.1 Accuracy of parameter estimation

In this section apply we apply the Kalman filter in conjunction with nested sampling in order to infer the joint posterior distribution $p(\boldsymbol{\theta}'|\boldsymbol{Y})$. For the Earth-term model we apply the Kalman filter using Equation (15) to return $\mathcal{L}(\boldsymbol{Y}|\boldsymbol{\theta}')$. For the pulsar-term model we we apply the Kalman filter using Equation (17). The likelihood returned by the Kalman filter is then used by nested sampling in order to estimate $p(\boldsymbol{\theta}'|\boldsymbol{Y})$. The settings of the nested sampler (for example the number of live points, see Appendix B) are identical for both models.

We consider two representative systems. The first is a "low-SNR" system with $h_0 = 5 \times 10^{-15}$, i.e. the exact system described in Table 1. The second is a "high-SNR" system which has the same static injection parameters, but with an increased $h_0 = 10^{-12}$. The "high-SNR" system is considered in order to evaluate the parameter estimation accuracy, and in particular any potential bias, without confusion from the measurement noise. In order to enable a clear comparison between the posterior probability distributions calculated using the two models, in this section we consider a single noise realisation of the synthetic data $\boldsymbol{Y}$. The results quoted in the below paragraph are generally consistent across different realisations, but plotting multiple realisations simultaneously overcrowds the figure and makes it harder to draw conclusions.

The results for the seven parameters of $\boldsymbol{\theta}_{\mathrm{gw}}$ are shown in Figure 4 for the high-SNR system and in Figure 5 for the low-SNR system. The corner plot is arranged identically to Figure 1, however the different coloured curved now correspond to different inference models, rather than different realisations of $\boldsymbol{Y}$. The green curves are the results derived using the Earth-terms measurement equation, Equation 15. The blue curves are the results derived using the full measurement equation, including the pulsar terms, Equation (8).

For the high-SNR case, Figure 4, the one-dimensional posteriors for the Earth-term model are evidently biased, as was observed in K24, as well as Zhu et al. (2016) and Chen & Wang (2022). Particular biases are observed in ψ, ι and α, with the blue posterior displaced with respect to the orange injection line. For example, for ι, whilst the injected value is contained within the 90% credible interval, the median of the posterior of the Earth-term model is shifted from the injected value by 0.35 radians. The inclusion of the pulsar terms corrects for this bias; the green one-dimensional exhibit no bias and are generally symmetric about the injected value. Moreover, for every pair of parameters the injection values are contained within the 2-sigma equivalent contours for the pulsar-term model, which is not true in the Earth-terms case.

We define a relative error metric to quantify the accuracy of the one-dimensional posteriors with respect to the injection value as

$$\Delta_{\mathrm{M}}(\theta) = \frac{1}{K} \sum_{k=1}^{K} \frac{p_k(\theta) - \hat{\theta}}{\hat{\theta}}, \qquad (21)$$

where $p_k(\theta)$ is the k-th one-dimensional posterior sample for parameter θ, K is the total number of samples, $\hat{\theta}$ is the true injection value, and the subscript $M \in \{\mathrm{earth, pulsar}\}$ indicates whether the posterior is estimated using Equation (15) or Equation (8) respectively. The dependence of $p_k(\theta)$ on M is implied. $\Delta_{\mathrm{M}}(\theta)$
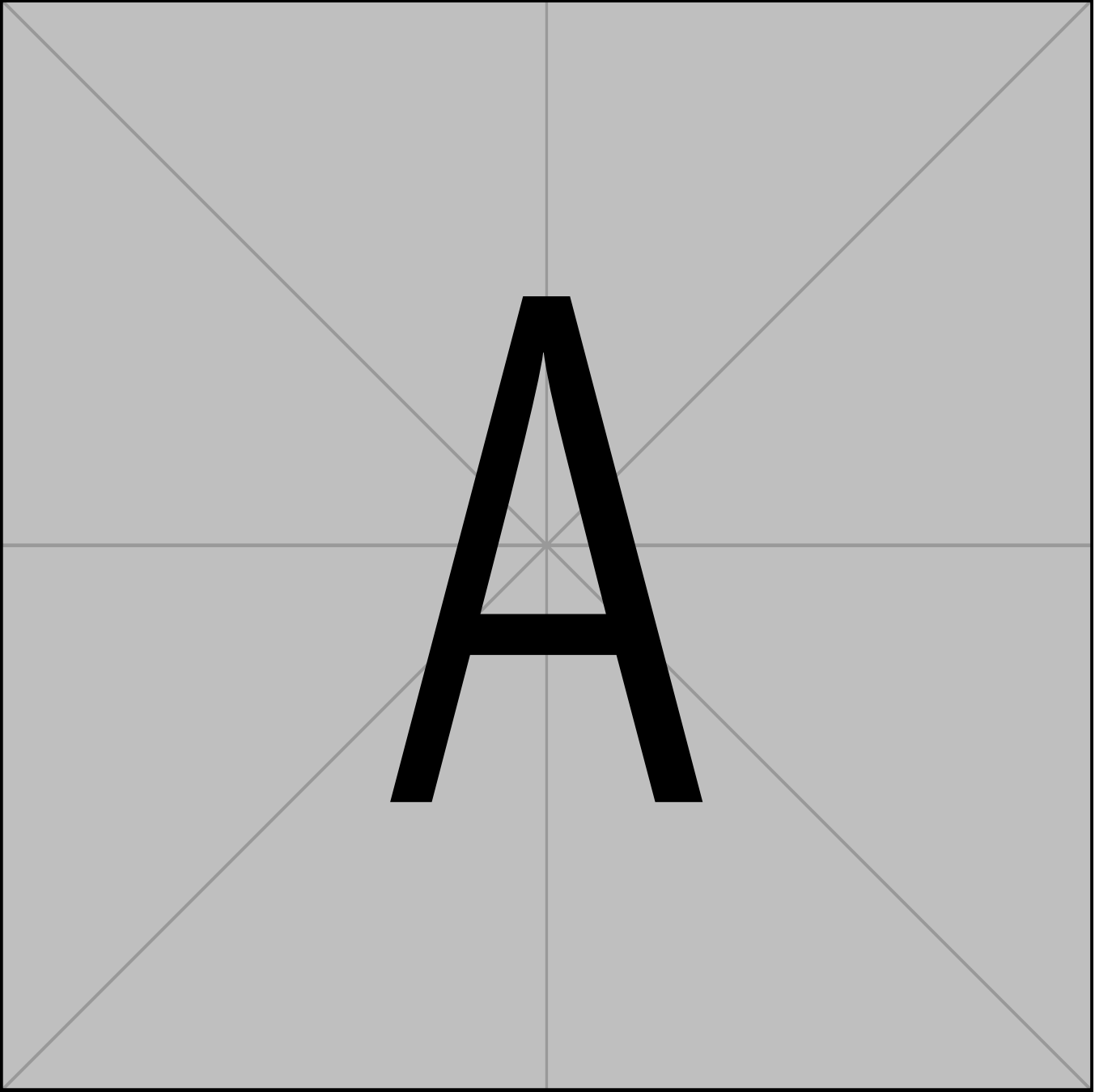
Figure 1. Posterior distribution of the GW source parameters $\boldsymbol{\theta}_{\mathrm{gw}}$ for the representative system in Table 1, for nine realisations of the noise processes, with curves coloured uniquely. The horizontal and vertical orange lines indicate the true injected values. The contours in the two-dimensional histograms mark the (0.5, 1, 1.5, 2)-σ levels after marginalizing over all but two parameters. The one-dimensional histograms correspond to the joint posterior distribution marginalized over all but one parameter. The supertitles of the one-dimensional histograms record the posterior median and the 0.16 and 0.84 quantiles of the median realisation. We plot the scaled variables $10^9\Omega$ (units: $\mathrm{rad\,s}^{-1}$) and $10^{15}h_0$. The Kalman filter and nested sampler estimate accurately all seven parameters in $\boldsymbol{\theta}_{\mathrm{gw}}$. The horizontal axes span a subset of the prior domain for all seven parameters. The known, injected value lies within the 90% credible interval for X out of the 63 combinations of seven parameters and nine noise realizations. There is an appreciable dispersion among the peaks of the one-dimensional posteriors, with coefficient of variation X across the 63 combinations.

represents the mean relative error in the one-dimensional posterior. The values of $\Delta_{\mathrm{M}}(\theta)$ for each parameter in Figure 4 are summarised in Table 2. The difference in the mean relative error between the two models, $\Delta_{\mathrm{pulsar}}(\theta) - \Delta_{\mathrm{earth}}(\theta)$ is also presented. For every parameter except Ω, $\Delta_{\mathrm{pulsar}}(\theta) - \Delta_{\mathrm{earth}}(\theta)$ is negative, indicating that the parameter estimates are more accurate using the pulsar

terms (i.e. the mean relative error when using the pulsar model is smaller). The exception for Ω is due to the fact that Ω is recovered with high accuracy in both cases; the the mean relative error is small $\Delta_{\mathrm{M}}(\Omega) \sim 10^{-5}$. The improvement from using the pulsar terms is largest for ι, with $\Delta_{\mathrm{pulsar}}(\iota) - \Delta_{\mathrm{earth}}(\iota) = -0.25$.

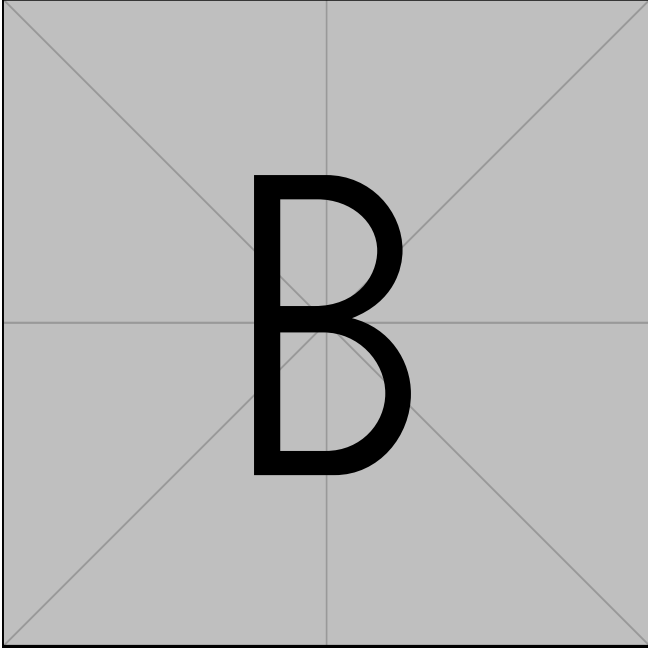**Figure 2.** Same as Figure 1 but for 5 static parameters of $\chi^{(n)}$. The known, injected value lies within the 90% credible interval for X out of the X combinations of five parameters and vX noise realizations.
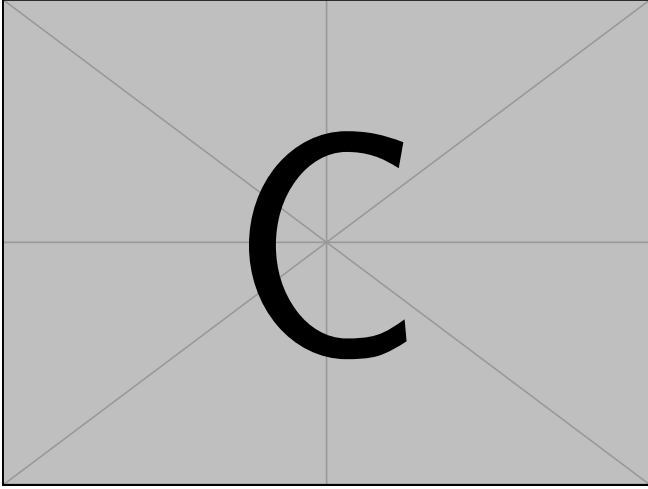

Figure 3. Fraction of injections included within a given credible interval of the estimated posterior, as a function of the credible interval itself (i.e. PP plot). The injections are 200 simulated GW sources generated by drawing randomly five parameters in θ_{gw} from the prior distributions in Table 1. Each coloured curve corresponds to a different parameter (see legend). The parameters h_0 and ι are fixed at 5×10^{-15} and 1.0 rad respectively in order to maintain an approximately constant SNR. The grey shaded contours label the $1\sigma, 2\sigma$ and 3σ confidence intervals. For parameters with well estimated posteriors, the PP curve should fall along the diagonal of unit slope. Ω is generally well-estimated (i.e. it lies close to the unit diagonal) but the four other parameters show evidence of being over-constrained (i.e. the curves lie above the unit diagonal for low credible intervals, and below the unit diagonal for high credible intervals). This is due to a modelling bias whose origin is discussed in Sections **??** and **??**.

	θ	$\Delta_{\text{earth}}(\theta)$	$\Delta_{\text{pulsar}}(\theta)$	$\Delta_{\text{pulsar}}(\theta) - \Delta_{\text{earth}}(\theta)$
High SNR	Ω	1.23e-05	1.49e-05	2.54e-06
	Φ_0	0.0179	0.0085	-0.00935
	ψ	0.0386	0.000254	-0.0384
	ι	0.374	0.126	-0.248
	δ	0.00223	0.000242	-0.00198
	α	0.0397	0.000582	-0.0391
	h_0	0.168	0.0846	-0.0839
Low SNR	Ω	1.23e-05	1.49e-05	2.54e-06
	Φ_0	0.0179	0.0085	-0.00935
	ψ	0.0386	0.000254	-0.0384
	ι	0.374	0.126	-0.248
	δ	0.00223	0.000242	-0.00198
	α	0.0397	0.000582	-0.0391
	h_0	0.168	0.0846	-0.0839

Table 2. Mean relative error $\Delta_M(\Theta)$ (Equation (21)) in the one-dimensional posteriors calculated using the Earth-term model (M = earth, Equation (15)) and the pulsar-term model (M=pulsar, Equation (8)) for the elements $\theta \in \theta_{gw}$. The injection values are as summarised in Table 1. The top and bottom halves of the table contain the high-SNR ($h_0 = 10^{-12}$) and low-SNR ($h_0 = 5 \times 10^{-15}$) cases respectively. The rightmost column quotes the difference in the mean relative error between the models for each parameter; a negative value indicates that the pulsar-term model is more accurate. TK numbers to be revised and double checked, but probably representative

For the low-SNR case, Figure 5, there is less of a notable improvement in the parameter estimates. Qualitatively, the green and blue one-dimensional posteriors are generally overlaid. The two-dimensional contours suggest a slight improvement through using the pulsar terms, but the difference is modest. The mean relative errors $\Delta_M(\theta)$ for the low SNR case are also reported in Table 2. We see that.... TK: numbers to be inserted here. We conclude that the inclusion of the pulsar terms does not significantly improve the parameter estimates in the low-SNR case. This is intuitive since the system is highly noise dominated such that an increasing model complexity does not aid parameter estimates. Whilst the inclusion of the pulsar terms only modestly improve the accuracy of the parameter estimates in the case of low-SNR, we will show in Section 5.2 that they are important for increasing the detection probability for a given signal strength.

5.2 Detectability vs h_0

The evidence integral $\mathcal{Z}$ returned by nested sampling is the probability of the data Y given a model $\mathcal{M}_i$. We compare competing models via a Bayes factor,

$$\beta = \frac{\mathcal{Z}(Y|\mathcal{M}_1)}{\mathcal{Z}(Y|\mathcal{M}_0)}. \tag{22}$$

Throughout this work we take $\mathcal{M}_1$ to be the state-space model that includes the presence of a GW. $\mathcal{M}_0$ is the null model, which assumes no GW exists in the data. This is equivalent to setting $g^{(n)}(t) = 1$ in Equations (**??**) and (8). The Bayes factors we quote in this work therefore quantify whether the data supports evidence for a GW compared to there being no GW signal present.

In Section **??** we calculate the detectability of the source as a function of h_0.

We frame the problem of detecting a GW in noisy PTA data in terms of the Bayesian model selection procedure described in
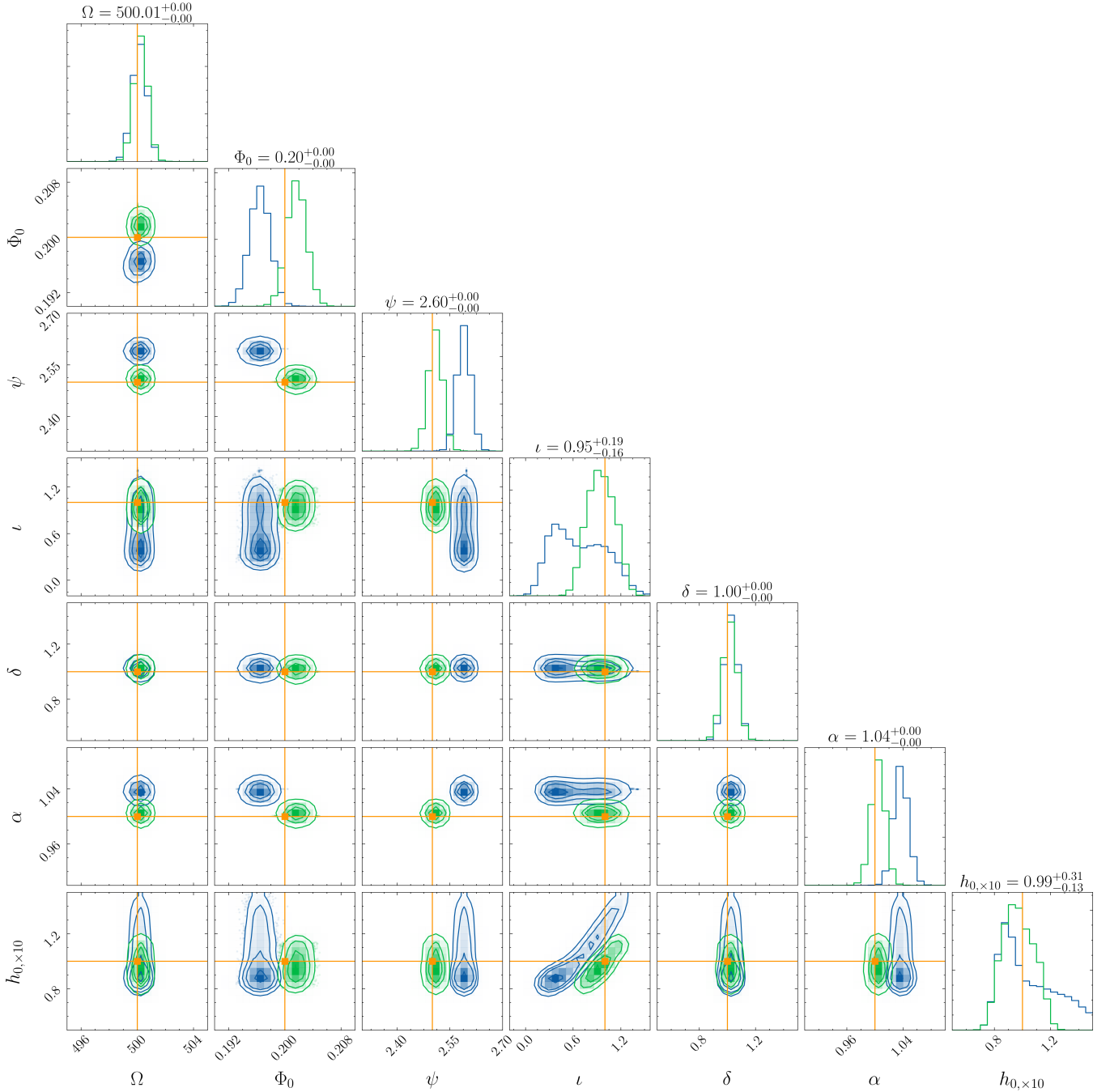
Figure 4. Posterior distribution of the GW source parameters $\boldsymbol{\theta}_{\mathrm{gw}}$ for the representative system described in Table 1, with $h_0 = 10^{-12}$, for a single realisation of the system noise. The blue curves are the posteriors derived using the Earth-term model, Equation (15), the green curves using the pulsar-term model, Equation (8). The vertical orange lines indicate the true injected values. The contours in the 2D histograms denote the (0.5, 1, 1.5, 2)-σ levels. We plot the scaled variables $10^9\Omega$ (units: rad s^{-1}) and $10^{15}h_0$. Some parameters (e.g. ψ, ι) exhibit a bias when using the Earth-term model which is absent when using the pulsar-term model. TK: need to remove these subtitles and correct the labels for omega and h. The actual results here may also be updated (same for Figure 5)

Section F2. In Equation (F1) $\mathcal{M}_1$ is the Earth-term-only model with a GW present, i.e. the state-space model with a Kalman filter based on Equation (14). The Bayes factor, β, as defined in Equation (F1) is presented in Figure **??** for the representative source in Table 1, except that we now vary the source amplitude, h_0, from 10^{-15} (undetectable) to 10^{-12} (easily detectable). To control the test, the noise processes in the synthetic data are identical realisations for each value of h_0; the only change from one h_0 value to the next is

h_0 itself. [2]

We see in Figure **??** an approximate quadratic relationship $\beta \propto h_0^2$ for $h_0 \gtrsim 10^{-14}$. The GW source is detectable with decisive evidence ($\beta \geqslant 10$) for $h_0 \gtrsim 4 \times 10^{-15}$. Of course, the minimum detectable

---

[2] Changing the noise realizations as well, from one value of h_0 to the next, adds uninformative scatter to the trend in Figure **??**.
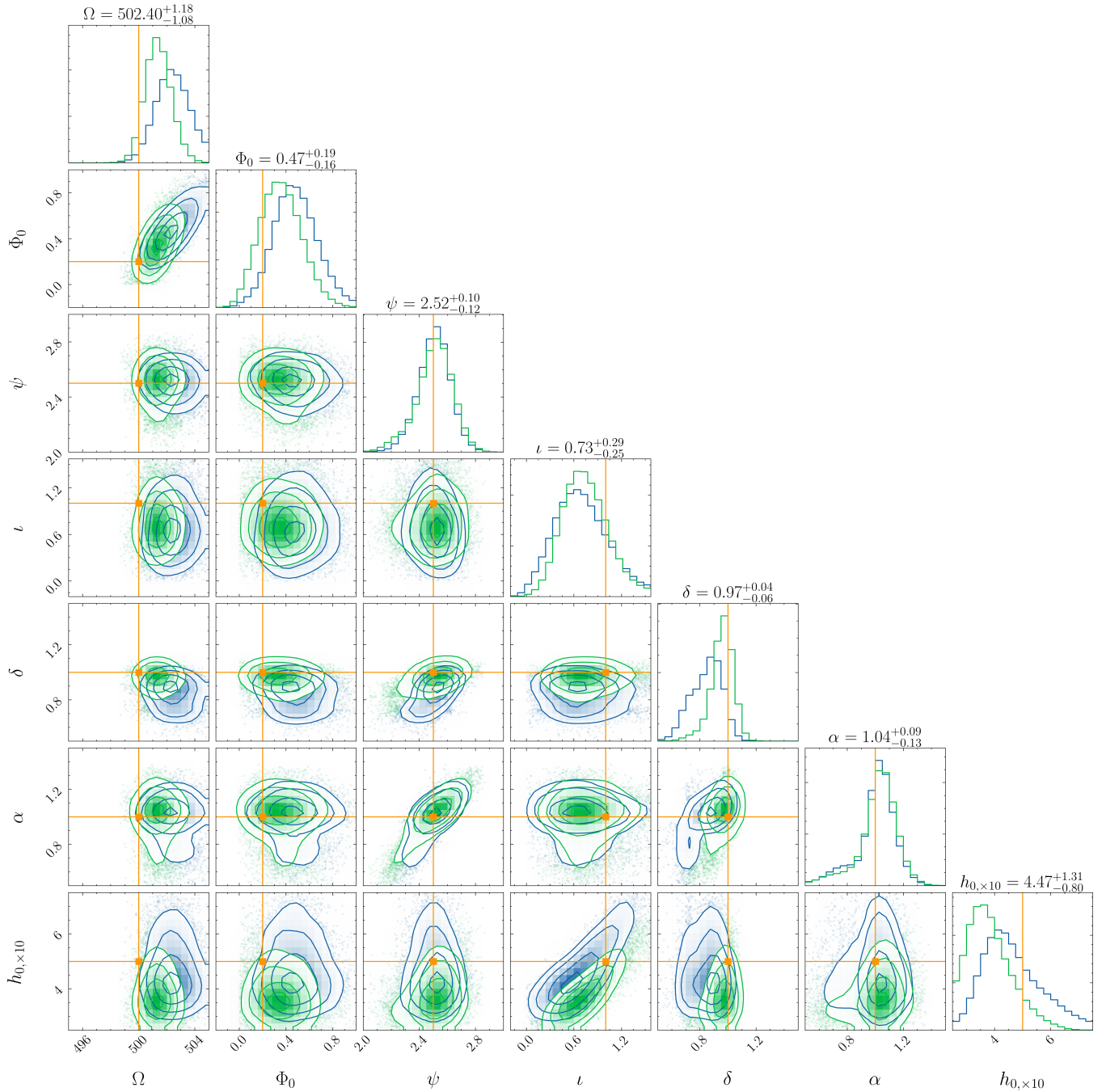
Figure 5. Same as Figure 4, but for a low-SNR system with $h_0 = 5 \times 10^{-15}$. The posterior distributions from the Earth-term and pulsar-term models are generally overlaid

strain is particular to the system in Table 1. It is influenced in general by $T_{\rm obs}$, $\boldsymbol{\theta}_{\rm gw}$, and $\boldsymbol{\theta}_{\rm psr}$, as discussed in Section 4.3. Adjusting $\sigma_{\rm m}$ we find that an approximate quadratic relationship $\beta \propto h_0^2/\sigma_{\rm m}^2$ also exists, where $h_0/\sigma_{\rm m}$ is the effective signal-to-noise ratio.

The odds ratio β drops off for $h_0 \lesssim 4 \times 10^{-15}$. This happens because the two competing models becoming increasingly indistinguishable once the measurement noise dominates the GW signal. Moreover, the points in Figure **??** become sparser for $h_0 \lesssim 4 \times 10^{-15}$. This happens because we obtain $\beta < 0$ for the missing, intermediate points. This is a noise artefact of the nested sampler. When the sam-

pler converges sub-optimally, the hierarchical relationship between $\mathcal{M}_0$ and $\mathcal{M}_1$ fails, and $\mathcal{Z}(\boldsymbol{Y}|\mathcal{M}_1) > \mathcal{Z}(\boldsymbol{Y}|\mathcal{M}_0)$ no longer holds. There is nothing special about the particular missing points; if one recreates the $\beta(h_0)$ curve by rerunning the nested sampler with another random seed, a different set of points are missing. Increasing the number of live points $n_{\rm live}$ used by the nested sampler improves the performance at small values of h_0, because the uncertainties in $\mathcal{Z}$ scale as $\mathcal{O}(n_{\rm live}^{-1/2})$ (see Section B). In this section we set $n_{\rm live} = 2000$.

We frame the problem of claiming a detection of a GW in the noisy data in terms of a Bayesian model selection procedure as described in Section F2. $\mathcal{M}_1$ is the Earth-terms only model, i.e. the
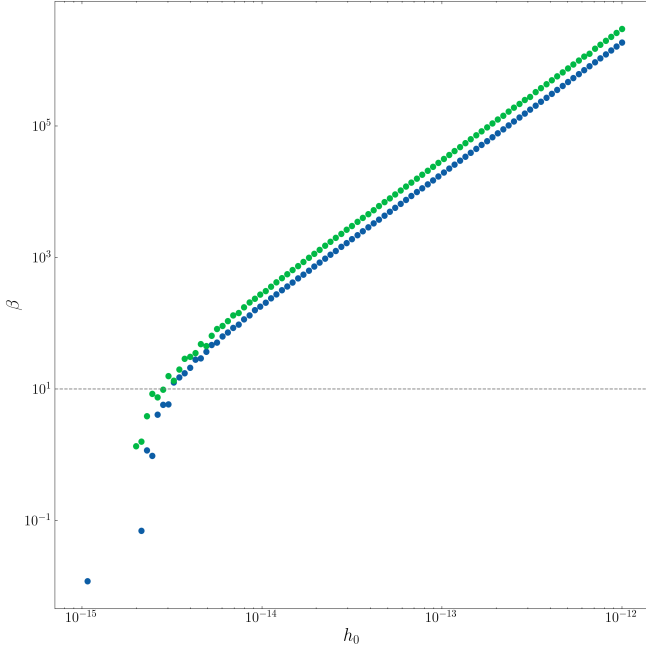
Figure 6. Bayes factor (odds ratio) β between the competing models $\mathcal{M}_1$ (GW present in data) and $\mathcal{M}_0$ (GW not present in data) at different GW strains, h_0, for the representative example in Table 1. The horizontal grey dashed line labels an arbitrary detection threshold, $\beta = 10$. The minimum detectable strain, for $\beta < 10$, equals 4×10^{-15}. Missing points for $\beta \lesssim 10$ occur when noise subverts the hierarchical relationship between $\mathcal{M}_0$ and $\mathcal{M}_1$ (see Section **??**).

state-space model with a Kalman filter using Equation (14). The Bayes factor, β, as defined in Equation (F1) is presented in Figure **??** for our representative system, where we now vary the magnitude of the GW strain, h_0. The noise processes in the synthetic data are identical realisations for each value of h_0; the only change in the data is the change in h_0. We can see that for $\beta > 10$ there is an approximate log-linear relationship between β and h_0 where the GW source is easily detectable with decisive evidence ($\beta \sim 100$) for $h_0 \gtrsim 10^{-14}$. We take $\beta = 10$ as a tolerance cut-off above which we can accept $\mathcal{M}_1$ and claim a detection of the GW. For this system, given this realisation of the noise, the minimal detectable strain is $\sim 3.9 \times 10^{-15}$. This minimal detectable strain is of course particular to this system and will be influenced in general by e.g. T_{obs}, σ_{m} as well as the parameters of the GW source such as its location on the sky.

There is an evident sparsity and steep drop-off of β for values of $h_0 \lesssim 3.9 \times 10^{-15}$. The steep drop off is due to the two competing models becoming increasingly indistinguishable as the strain signal decreases and the measurement noise starts to dominate. The sparsity is due some of the β value becoming negative at these points. This is a noise artefact of the nested sampling method. For these points the noise in the sampler starts to dominate, the sampler converges sub-optimally, and the hierarchical relationship between $\mathcal{M}_0$ and $\mathcal{M}_1$ fails. That is, the requirement that $\mathcal{Z}(Y|\mathcal{M}_1) > \mathcal{Z}(Y|\mathcal{M}_0)$ no longer holds. There is nothing special about these particular points; if one recreates the $\beta - h_0$ curve by rerunning the nested sampler these sparse locations appear at different values of h_0, for $\beta < 10$. For $\beta > 10$ the sampling noise is sufficiently small with respect to the signal and the model hierarchy holds. Increasing the number of live points n_{live} used by the nested sampler improves the performance at

small values of h_0 since as discussed in Section B the uncertainties in $\mathcal{Z}$ scale as $\mathcal{O}(1/\sqrt{n_{\text{live}}})$. For the results presented in this subsection we set $n_{\text{live}} = 2000$.

6 CONCLUSION

In this paper we demonstrate a new method for the detection and parameter estimation of GWs from individual, monochromatic SMBHBs in PTA data. The new method is complementary to traditional approaches. We track the evolution of the intrinsic pulsar timing noise explicitly via a state-space method, rather than fitting for the ensemble-averaged power spectral density of the noise. That is, we disentangle statistically the specific time-ordered realisation of the timing noise from the GW-induced modulations, and thereby infer the GW source parameters conditional on the specific observed realisation of the noisy data. We implement a Kalman filter in order to track the intrinsic rotational state evolution of the pulsar and combine it with a Bayesian sampling framework to estimate the posterior distributions of each static parameter, as well as the associated Bayesian evidence (marginalised likelihood) of the model with and without a GW.

We test the new method on synthetic data and find that it detects injected signals successfully and estimates their static parameters accurately with relatively low computational cost. We initially focus on a single, astrophysically representative, SMBHB GW source observed synthetically by the 12.5-year NANOGrav pulsars with $T_{\text{obs}} = 10$ years. The minimum detectable strain is estimated to be $\min(h_0) \approx 4 \times 10^{-15}$ for $\iota = 1.0$ rad. We then repeat the parameter estimation exercise for 1000 noise realisations to compute the natural dispersion ("cosmic variance") in the recovered values and hence quantify the statistical accuracy of the method in a real astrophysical application, when the true parameter values are unknown. Consistent posteriors are obtained for most realisations. The median WD is limited to $\lesssim 3\%$ of the width of the prior domain (i.e. negligible railing). The median WD divided by the injected value, an approximate measure of the natural dispersion, ranges from 0.2% for Ω to 67% for Φ_0.

Exploration of a broader SMBHB parameter domain at fixed ι and h_0 via 200 randomly sampled parameter vectors reveals a bias in the estimates of the static parameters δ and α, and to a lesser extent ψ and Φ_0. The bias is examined for a specific SMBHB GW source in the limit of high-SNR. It is greatest for ι, amounting to ≈ 0.3 rad. Smaller biases of $\lesssim 0.1$ rad are also observed in Φ_0, ψ, δ and α. The bias is shown to result from dropping the pulsar terms from the measurement equation in the Kalman filter, consistent with the work of previous authors (Zhu et al. 2016; Chen & Wang 2022). The method also exhibits a trade-off between h_0 and ι, which follows formally from the weak identifiability of h_0 and ι.

We emphasize that the Kalman tracker and nested sampler in this paper do not supplant traditional PTA analysis approaches; they complement traditional approaches and are most powerful when used in tandem. Relatedly, it is misleading to ask whether the Kalman tracker and nested sampler are more or less sensitive than traditional approaches for two reasons. First, one must still generalize the Kalman tracker and nested sampler to ingest TOAs directly, as happens traditionally, instead of ingesting a frequency time series, as in this paper. Generalizing the algorithm is a subtle task and is postponed to a forthcoming paper. Second, the Kalman tracker and nested sampler are conditional on a noise model that is

related but different to the noise model in traditional approaches. The analysis in this paper assumes a specific, time-ordered realization of a mean-reverting Ornstein-Uhlenbeck process satisfying Equations (4)–(3), whereas traditional analyses assume a stationary Gaussian process described by an ensemble-averaged, power-law power spectral density, whose amplitude and exponent are adjustable. Hence the sensitivities cannot be compared directly. Clarifying the similarities and differences between various approaches promises to be a fruitful avenue of future work. It is also a subject of attention in audio-band GW data analysis involving hidden Markov models applied to data from terrestrial long-baseline interferometers (Middleton et al. 2020; Abbott et al. 2022a,b; Abbott et al. 2022c).

The approach in this paper can be extended in at least five ways, enumerated as (i)–(v) below.

(i) A natural first extension is to retain the pulsar terms in Equation (5). Alternatively, if the method continues to be used with solely the Earth term, in line with standard practice in published PTA analyses, it would be desirable to evaluate systematically the incurred biases in the model parameters across an astrophysically representative parameter domain, supplementing the results obtained by other authors (Zhu et al. 2016; Chen & Wang 2022).

(ii) In this paper, we consider one specific configuration of the synthetic PTA, namely the same pulsars that make up the 12.5-year NANOGrav (Section **??**). It is interesting to compare the performance of the method using different pulsar configurations, e.g. those in the PPTA and EPTA. Adding pulsars to PTAs increases the computational cost so it may be advantageous to select a subset of pulsars by exploiting formal optimization techniques from electrical engineering (Speri et al. 2023), although the Kalman filter and nested sampler are already cheaper computationally than some other methods.

(iii) In practice different PTA pulsars are observed with different cadences at different times. Extending the Kalman filter to non-uniform time sampling is straightforward (Zarchan & Musoff 2000).

(iv) The assumption of a monochromatic source is well-justified astrophysically in various regimes (see Section **??**) and is an appropriate starting point for this introductory paper. Nevertheless, SMBHBs are not strictly monochromatic. It is interesting to extend the state-space framework such that $f_{\rm gw}$ evolves in time. For $\Delta f_{\rm gw} < 1/T_{\rm obs}$ (see Equation (9)) evolution adds noise incoherently to the pulsar terms, whilst for $\Delta f_{\rm gw} > 1/T_{\rm obs}$ the pulsar terms induce phase shifts that affect the overall phase coherence (Sesana & Vecchio 2010; Perrodin & Sesana 2018). Careful consideration of the evolution of $f_{\rm gw}$ will be needed when including the pulsar terms in the inference model, as in point (i).

(v) Finally, we assume in this paper that there is only one GW source. However, it may be possible to resolve multiple continuous GW sources concurrently (Babak & Sesana 2012b). The Kalman filter extends naturally to multiple sources; one can modify Equation (5) easily to accommodate a linear superposition of GWs. Taking the logic further, the stochastic background itself is arguably an incoherent sum of many individual GW sources. As long as a way can be found to summarize economically the many static parameters associated with the background sources, it should be possible for a Kalman filter and nested sampler to operate together to detect the stochastic background by generalizing the model selection procedure

in Sections F2 and **??**. Summarizing the parameter set economically, while respecting the mathematical structure of the Kalman filter, is a subtle challenge, which we postpone to a forthcoming paper. If successful, it will complement the traditional approach of cross-correlating pulsar residuals to uncover the Hellings-Downs curve (Hellings & Downs 1983; Agazie et al. 2023a).

DATA AVAILABILITY

No new data were generated or analysed in support of this research.

REFERENCES

Abbott R., et al., 2022a, Phys. Rev. D, 105, 022002
Abbott R., et al., 2022b, Physical Review D, 106
Abbott R., et al., 2022c, Phys. Rev. D, 106, 062002
Agazie G., et al., 2023a, ApJ, 951, L8
Agazie G., et al., 2023b, ApJ, 951, L9
Allen B., 1997, in Marck J.-A., Lasota J.-P., eds, Relativistic Gravitation and Gravitational Radiation. pp 373–417 (arXiv:gr-qc/9604033), doi:10.48550/arXiv.gr-qc/9604033
Antoniadis J., et al., 2023a, arXiv e-prints, p. arXiv:2306.16214
Antoniadis J., et al., 2023b, arXiv e-prints, p. arXiv:2306.16224
Antoniadis J., et al., 2023c, arXiv e-prints, p. arXiv:2306.16226
Arjovsky M., Chintala S., Bottou L., 2017, arXiv e-prints, p. arXiv:1701.07875
Arzoumanian Z., et al., 2020, ApJ, 905, L34
Arzoumanian Z., et al., 2023, arXiv e-prints, p. arXiv:2301.03608
Ashton G., Talbot C., 2021, MNRAS, 507, 2037
Ashton G., et al., 2022, Nature Reviews Methods Primers, 2, 39
Babak S., Sesana A., 2012a, Phys. Rev. D, 85, 044034
Babak S., Sesana A., 2012b, Phys. Rev. D, 85, 044034
Babak S., et al., 2016, MNRAS, 455, 1665
Baym G., Pethick C., Pines D., Ruderman M., 1969, Nature, 224, 872
Bhagwat S., De Luca V., Franciolini G., Pani P., Riotto A., 2021, J. Cosmology Astropart. Phys., 2021, 037
Buchner J., 2021a, arXiv e-prints, p. arXiv:2101.09675
Buchner J., 2021b, The Journal of Open Source Software, 6, 3001
Burke-Spolaor S., et al., 2019, A&ARv, 27, 5
Caballero R. N., et al., 2016, MNRAS, 457, 4421
Charisi M., Taylor S. R., Witt C. A., Runnoe J., 2023, arXiv e-prints, p. arXiv:2304.03786
Chen J.-W., Wang Y., 2022, ApJ, 929, 168
Christensen N., 2019, Reports on Progress in Physics, 82, 016903
Cook S. R., Gelman A., Rubin D. B., 2006, Journal of Computational and Graphical Statistics, 15, 675
Cordes J. M., Lazio T. J. W., 2002, arXiv e-prints, pp astro-ph/0207156
Desvignes G., et al., 2016, Monthly Notices of the Royal Astronomical Society, 458, 3341
Dunn L., et al., 2022, MNRAS, 512, 1469
Dunn L., Melatos A., Espinoza C. M., Antonopoulou D., Dodson R., 2023, MNRAS, 522, 5469

Ellis J. A., Siemens X., Creighton J. D. E., 2012, ApJ, 756, 175
Feroz F., Hobson M. P., 2008, MNRAS, 384, 449
Feroz F., Hobson M. P., Bridges M., 2009, Monthly Notices of the Royal Astronomical Society, 398, 1601
Gardiner C., 2009, Stochastic Methods: A Handbook for the Natural and Social Sciences. Springer Series in Synergetics, Springer Berlin Heidelberg, https://books.google.com.au/books?id=otg3PQAACAAJ
Gelb A., Kasper J. F., Nash R. A., Price C. F., Sutherland A. A., eds, 1974, Applied Optimal Estimation. MIT Press, Cambridge, MA
Goldreich P., Julian W. H., 1969, ApJ, 157, 869
Goldstein J. M., Veitch J., Sesana A., Vecchio A., 2018, MNRAS, 477, 5447
Goncharov B., et al., 2021, MNRAS, 502, 478
González-Delgado J., Sagar A., Zanon C., Lindorff-Larsen K., Bernadó P., Neuvial P., Cortés J., 2023, Journal of Molecular Biology, 435, 168053
Gügercinoğlu E., Alpar M. A., 2017, MNRAS, 471, 4827
Handley W. J., Hobson M. P., Lasenby A. N., 2015, MNRAS, 450, L61
Hellings R. W., Downs G. S., 1983, ApJ, 265, L39
Hobbs G. B., Edwards R. T., Manchester R. N., 2006, Monthly Notices of the Royal Astronomical Society, 369, 655
Jaffe A. H., Backer D. C., 2003, The Astrophysical Journal, 583, 616
Jankowski F., et al., 2018, Monthly Notices of the Royal Astronomical Society, 484, 3691
Jenet F. A., Lommen A., Larson S. L., Wen L., 2004, ApJ, 606, 799
Kalman R. E., 1960, Journal of Basic Engineering, 82, 35
Kimpson T., Melatos M., O'Leary J., Carlin J., et al.,, Submitted to MNRAS
Kimpson T., Paxton E. A., Chantry M., Palmer T., 2023, Quarterly Journal of the Royal Meteorological Society, 149, 843
Lasky P. D., Melatos A., Ravi V., Hobbs G., 2015, MNRAS, 449, 3293
Lower M. E., et al., 2021, MNRAS, 508, 3251
Manchester R. N., Hobbs G. B., Teoh A., Hobbs M., 2005, AJ, 129, 1993
McWilliams S. T., Ostriker J. P., Pretorius F., 2014, The Astrophysical Journal, 789, 156
Melatos A., Dunn L. M., Suvorova S., Moran W., Evans R. J., 2020, ApJ, 896, 78
Melatos A., O'Neill N. J., Meyers P. M., O'Leary J., 2023, ApJ, 944, 64
Meyers P. M., Melatos A., O'Neill N. J., 2021a, MNRAS, 502, 3113
Meyers P. M., O'Neill N. J., Melatos A., Evans R. J., 2021b, MNRAS, 506, 3349
Middleton H., Clearwater P., Melatos A., Dunn L., 2020, Phys. Rev. D, 102, 023006
Morris M. J., Lipp A. G., Roberts G. G., 2023, Geophys. Res. Lett., 50, e2023GL103880
Mukherjee P., Parkinson D., Liddle A. R., 2006, ApJ, 638, L51
Pártay L. B., Bartók A. P., Csányi G., 2009, arXiv e-prints, p. arXiv:0906.3544
Paxton E. A., Chantry M., Klöwer M., Saffin L., Palmer T., 2022, Journal of Climate, 35, 1215
Perrodin D., Sesana A., 2018, in Rezzolla L., Pizzochero P., Jones D. I., Rea N., Vidaña I., eds, Astrophysics and Space Science Library Vol. 457, Astrophysics and Space Science Library. p. 95 (arXiv:1709.02816), doi:10.1007/978-3-319-97616-7_3
Petiteau A., Babak S., Sesana A., de Araújo M., 2013, Phys. Rev. D, 87, 064036
Pitkin M., 2018, Journal of Open Source Software, 3, 538
Price S., Link B., Shore S. N., Nice D. J., 2012, MNRAS, 426, 2507
Rajagopal M., Romani R. W., 1995, ApJ, 446, 543
Ravi V., Wyithe J. S. B., Shannon R. M., Hobbs G., 2015, MNRAS, 447, 2772
Reardon D. J., et al., 2023, ApJ, 951, L6
Renzini A. I., Goncharov B., Jenkins A. C., Meyers P. M., 2022, Galaxies, 10, 34
Rosado P. A., Sesana A., Gair J., 2015, Monthly Notices of the Royal Astronomical Society, 451, 2417
Sesana A., 2013, Classical and Quantum Gravity, 30, 224014
Sesana A., Vecchio A., 2010, Phys. Rev. D, 81, 104008
Sesana A., Vecchio A., Colacino C. N., 2008, Monthly Notices of the Royal Astronomical Society, 390, 192
Shannon R. M., Cordes J. M., 2010a, ApJ, 725, 1607
Shannon R. M., Cordes J. M., 2010b, ApJ, 725, 1607

Simon D., 2006, Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches. Wiley-Interscience, USA
Skilling J., 2006, Bayesian Analysis, 1, 833
Sornette D., 2004, Critical phenomena in natural sciences : chaos, fractals selforganization and disorder : concepts and tools
Speagle J. S., 2020, Monthly Notices of the Royal Astronomical Society, 493, 3132
Speri L., Porayko N. K., Falxa M., Chen S., Gair J. R., Sesana A., Taylor S. R., 2023, MNRAS, 518, 1802
Sykes B., Middleton H., Melatos A., Di Matteo T., DeGraf C., Bhowmick A., 2022, MNRAS, 511, 5241
Taylor S. R., Huerta E. A., Gair J. R., McWilliams S. T., 2016, ApJ, 817, 70
Tiburzi C., 2018, Publ. Astron. Soc. Australia, 35, e013
Trassinelli M., 2019, Proceedings, 33
Vargas A., Melatos A., 2023, TBD, 1, 1
Vaserstein L., 1969, Probl. Peredachi Inf., 5
Verbiest J. P. W., Weisberg J. M., Chael A. A., Lee K. J., Lorimer D. R., 2012, ApJ, 755, 39
Verbiest J. P. W., Osłowski S., Burke-Spolaor S., 2021, in , Handbook of Gravitational Wave Astronomy. p. 4, doi:10.1007/978-981-15-4702-7_4-1
Villani C., 2009, The Wasserstein distances. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 93–111, doi:10.1007/978-3-540-71050-9_6, https://doi.org/10.1007/978-3-540-71050-9_6
Wan E., Van Der Merwe R., 2000, in Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373). pp 153–158, doi:10.1109/ASSPCC.2000.882463
Wyithe J. S. B., Loeb A., 2003, ApJ, 590, 691
Xu H., et al., 2023, Research in Astronomy and Astrophysics, 23, 075024
Yao J. M., Manchester R. N., Wang N., 2017, ApJ, 835, 29
Zarchan P., Musoff H., 2000, Fundamentals of Kalman Filtering: A Practical Approach. Progress in astronautics and aeronautics, American Institute of Aeronautics and Astronautics, https://books.google.com.au/books?id=AQxRAAAAMAAJ
Zhu X. J., et al., 2014, MNRAS, 444, 3709
Zhu X.-J., et al., 2015, Monthly Notices of the Royal Astronomical Society, 449, 1650
Zhu X.-J., Wen L., Xiong J., Xu Y., Wang Y., Mohanty S. D., Hobbs G., Manchester R. N., 2016, Monthly Notices of the Royal Astronomical Society, 461, 1317
Zic A., et al., 2023, arXiv e-prints, p. arXiv:2306.16230
van Eysden C. A., Melatos A., 2010, MNRAS, 409, 1253

APPENDIX A: KALMAN FILTER

In this appendix, we describe the Kalman filter algorithm used in this paper. General recursion relations for the discrete-time Kalman filter are written down for an arbitrary linear dynamical system in Section A1. The mapping onto the specific continuous-time state-space model in Section **??** is written down in Section A2.

The Kalman filter (Kalman 1960) is a Gauss-Markov model used to algorithmically recover a temporal sequence of stochastically evolving system state variables, $\boldsymbol{X}(t)$, which are not observed directly, given a temporal sequence of noisy measurements, $\boldsymbol{Y}(t)$. It finds common use in engineering applications and has been applied successfully in neutron star astrophysics (e.g. Meyers et al. 2021a,b; Melatos et al. 2023). In this work we use the linear Kalman filter, which assumes a linear relation between $d\boldsymbol{X}/dt$ and $\boldsymbol{X}(t)$ (dynamics) and between $\boldsymbol{Y}(t)$ and $\boldsymbol{X}(t)$ (measurement), with $\boldsymbol{X}(t) = \{f_{\mathrm{p}}^{(n)}(t)\}$ and $\boldsymbol{Y}(t) = \{f_{\mathrm{m}}^{(n)}(t)\}$. Extension to non-linear problems is straightforward using either an extended Kalman filter (Zarchan & Musoff 2000), unscented Kalman filter (Wan & Van Der Merwe 2000) or particle filter (Simon 2006). Equations (5) and (**??**) are non-linear in the static parameters (e.g. $d^{(n)}$), even though they are linear in $\boldsymbol{X}(t)$ and $\boldsymbol{Y}(t)$. Hence inferring the static parameters is a non-linear exercise, to be tackled separately after the linear Kalman filter operates on the data for a fixed set of static parameters. Inference of the static parameters in Equations (5) and (**??**) by nested sampling is discussed in Section B.

Implementation of the linear Kalman filter for the PTA state-space model in Section **??**, including the full set of Kalman recursion relations is presented in Appendix A. At each discrete timestep indexed by $1 \leqslant i \leqslant M$, the Kalman filter returns an estimate of the state variables, $\hat{\boldsymbol{X}}_i = \hat{\boldsymbol{X}}(t_i)$, and the covariance of those estimates, $\boldsymbol{P}_i = \langle \hat{\boldsymbol{X}}_i \hat{\boldsymbol{X}}_i^{\mathsf{T}} \rangle$, where the superscript T denotes the matrix transpose. The filter tracks the error in its predictions of $\boldsymbol{X}_i$ by converting $\hat{\boldsymbol{X}}_i$ into predicted measurements $\hat{\boldsymbol{Y}}_i$ via Equations (5) and (**??**) and comparing with the actual, noisy measurements $\hat{\boldsymbol{Y}}_i$. This defines a residual $\boldsymbol{\epsilon}_i = \boldsymbol{Y}_i - \hat{\boldsymbol{Y}}_i$, which is sometimes termed the innovation. The Kalman filter also calculates the uncertainty in $\boldsymbol{\epsilon}_i$ via the innovation covariance $\boldsymbol{S}_i = \langle \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^T \rangle$. The innovation and the innovation covariance are then used to correct the state estimate $\hat{\boldsymbol{X}}_i$ according to Equation (A10). For a fixed set of static parameters, the Kalman filter returns an estimate of the state sequence $\hat{\boldsymbol{X}}_1, \ldots, \hat{\boldsymbol{X}}_M$ which minimizes the mean square error. We explain how to use this intermediate output to infer the optimal values of the static parameters $\boldsymbol{\theta}$ in Section B

The Gaussian log-likelihood of obtaining $\boldsymbol{Y}_i$ given $\hat{\boldsymbol{X}}_i$ can then be calculated at each timestep from the Kalman filter output according to

$$\log \mathcal{L}_i = -\frac{1}{2} \left(N \log 2\pi + \log |\boldsymbol{S}_i| + \boldsymbol{\epsilon}_i^{\mathsf{T}} \boldsymbol{S}_i^{-1} \boldsymbol{\epsilon}_i \right) . \tag{A1}$$

The total log-likelihood for the entire sequence is

$$\log \mathcal{L} = \sum_{i=1}^{M} \log \mathcal{L}_i . \tag{A2}$$

Given $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_M$, $\mathcal{L}$ is a function of the estimates $\hat{\boldsymbol{\theta}}$ of the static parameters passed to the Kalman filter, i.e. $\mathcal{L} = \mathcal{L}(\boldsymbol{Y}|\hat{\boldsymbol{\theta}})$. Similarly the estimates of the state and measurement variables, $\hat{\boldsymbol{X}}$ and $\hat{\boldsymbol{Y}}$, are functions of $\hat{\boldsymbol{\theta}}$. If $\hat{\boldsymbol{\theta}}$ is close to the true underlying parameters $\boldsymbol{\theta}$, then the errors in $\hat{\boldsymbol{X}}$ and $\hat{\boldsymbol{Y}}$ are minimized and $\mathcal{L}$ is maximised. This is illustrated with synthetic data in Figure **??**. In the left column, a time

series of $f_{\mathrm{m}}^{(1)}(t)$ including Gaussian noise (middle panel, red curve) is generated from Equations (4)–(3), (5), and (**??**) for a single pulsar and fed into the Kalman filter along with the true static parameters $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}$. The Kalman filter recovers the evolution of $f_{\mathrm{p}}^{(1)}(t)$ with high fidelity; the estimate of $\hat{f}_{\mathrm{p}}^{(1)}(t)$ (left top panel, blue curve) overlaps almost perfectly with the true $f_{\mathrm{p}}^{(1)}(t)$ (left top panel, green curve). The predicted state $\hat{f}_{\mathrm{p}}^{(1)}(t)$ is converted into a predicted measurement $\hat{f}_{\mathrm{m}}^{(1)}(t)$ (middle panel, magenta curve), which again overlaps almost perfectly with the true measurement. The residuals $\epsilon(t)$ between the true and predicted measurements are small ($\lesssim 0.1\%$) and normally distributed (left bottom panel). By contrast, in the right column, the exercise is repeated while passing slightly incorrect static parameters ($\hat{\boldsymbol{\theta}} \neq \boldsymbol{\theta}$) to the Kalman filter, where Ω is displaced from its true value by 20%. In this case the Kalman filter fails to track $f_{\mathrm{p}}^{(1)}(t)$ accurately, as the discrepancy between the blue and green curves in the top panel of the right-hand column indicates. It similarly fails to predict $f_{\mathrm{m}}^{(1)}(t)$ accurately, as shown by the discrepancy between the red and magenta curves in the middle panel, and the residuals are no longer distributed normally (right bottom panel). In Section B, we explain how to combine the Kalman filter with a nested sampler to iteratively guide $\hat{\boldsymbol{\theta}}$ towards the true value of $\boldsymbol{\theta}$.

A1 Recursion equations

The linear Kalman filter operates on temporally discrete, noisy measurements $\boldsymbol{Y}_k$, which are related to a set of unobservable discrete system states $\boldsymbol{X}_k$, via a linear transformation

$$\boldsymbol{Y}_k = \boldsymbol{H}_k \boldsymbol{X}_k + \boldsymbol{v}_k , \tag{A3}$$

where $\boldsymbol{H}_k$ is the measurement matrix or observation model, $\boldsymbol{v}_k$ is a zero-mean Gaussian measurement noise, $\mathcal{N} \sim (0, \boldsymbol{R}_k)$ with covariance $\boldsymbol{R}_k$, and the subscript k labels the time-step. The Kalman filter evolves the underlying states according to

$$\boldsymbol{X}_k = \boldsymbol{F}_k \boldsymbol{X}_{k-1} + \boldsymbol{G}_k \boldsymbol{u}_k + \boldsymbol{w}_k , \tag{A4}$$

where $\boldsymbol{F}_k$ is the system dynamics matrix, $\boldsymbol{G}_k$ is the control matrix. $\boldsymbol{u}_k$ is the control vector, and $\boldsymbol{w}_k$ is a zero-mean Gaussian process noise, $\mathcal{N} \sim (0, \boldsymbol{Q}_k)$ with covariance $\boldsymbol{Q}_k$

The Kalman filter is a recursive estimator with two distinct stages: a "predict" stage and an "update" stage. The predict stage predicts $\hat{\boldsymbol{X}}_{k|k-1}$, the estimate of the state at discrete step k, given the state estimates from step $k-1$. Specifically, the predict step proceeds as

$$\hat{\boldsymbol{X}}_{k|k-1} = \boldsymbol{F}_k \hat{\boldsymbol{X}}_{k-1|k-1} + \boldsymbol{G}_k \boldsymbol{u}_k , \tag{A5}$$

$$\hat{\boldsymbol{P}}_{k|k-1} = \boldsymbol{F}_k \hat{\boldsymbol{P}}_{k-1|k-1} \boldsymbol{F}_k^{\mathsf{T}} + \boldsymbol{Q}_k , \tag{A6}$$

where $\hat{\boldsymbol{P}}_{k|k-1}$ is the covariance of the prediction. Note that the predict stage is independent of the measurements. The measurement $\boldsymbol{Y}_k$ is included to update the prediction during the update stage as follows:

$$\boldsymbol{\epsilon}_k = \boldsymbol{Y}_k - \boldsymbol{H}_k \hat{\boldsymbol{X}}_{k|k-1} , \tag{A7}$$

$$\boldsymbol{S}_k = \boldsymbol{H}_k \hat{\boldsymbol{P}}_{k|k-1} \boldsymbol{H}_k^{\mathsf{T}} + \boldsymbol{R}_k , \tag{A8}$$

$$\boldsymbol{K}_k = \hat{\boldsymbol{P}}_{k|k-1} \boldsymbol{H}_k^{\mathsf{T}} \boldsymbol{S}_k^{-1} , \tag{A9}$$

$$\hat{\boldsymbol{X}}_{k|k} = \hat{\boldsymbol{X}}_{k|k-1} + \boldsymbol{K}_k \boldsymbol{\epsilon}_k , \tag{A10}$$

$$\hat{\boldsymbol{P}}_{k|k} = (\boldsymbol{I} - \boldsymbol{K}_k \boldsymbol{H}_k) \hat{\boldsymbol{P}}_{k|k-1} . \tag{A11}$$

Equation (A9) defines the "Kalman gain" $\boldsymbol{K}_k$ which is defined so as to minimise the mean squared error in the state estimate, i.e.

$K_k = \mathrm{argmin}\left\{ \boldsymbol{E}[(\boldsymbol{X}_k - \hat{\boldsymbol{X}}_k)^2] \right\}$. For a full review of the Kalman filter, including its derivation, we refer the reader to Gelb et al. (1974) and Zarchan & Musoff (2000).

To apply the Kalman filter in practice means specifying the eight component matrices that make up the "machinery" of the filter: $\boldsymbol{X}_k$, $\boldsymbol{Y}_k$, $\boldsymbol{F}_k$, $\boldsymbol{G}_k$, $\boldsymbol{u}_k$, $\boldsymbol{H}_k$, $\boldsymbol{Q}_k$ and $\boldsymbol{R}_k$. In Section A2 we describe how the machinery is determined for the state-space model in Section **??**.

A2 State space representation of a PTA analysis

We apply the Kalman recursion relations in Section A1 to the state-space model of a PTA with N pulsars described in Section **??** as follows.

We identify $\boldsymbol{X}(t)$ with a vector of length N composed of the intrinsic pulsar frequency states, i.e.

$$\boldsymbol{X}(t) = \left(f_{\mathrm{p}}^{(1)}(t), f_{\mathrm{p}}^{(2)}(t), ..., f_{\mathrm{p}}^{(N)}(t) \right) . \tag{A12}$$

Analogously, we package the measured pulsar frequencies as

$$\boldsymbol{Y}(t) = \left(f_{\mathrm{m}}^{(1)}(t), f_{\mathrm{m}}^{(2)}(t), ..., f_{\mathrm{m}}^{(N)}(t) \right) . \tag{A13}$$

The states evolve according to the continuous stochastic differential equation (c.f. Equation (1))

$$d\boldsymbol{X} = \boldsymbol{A}\boldsymbol{X}\,dt + \boldsymbol{C}(t)dt + \boldsymbol{\Sigma}\,d\boldsymbol{B}(t) , \tag{A14}$$

where $\boldsymbol{A}$ is a diagonal $N \times N$ matrix,

$$\boldsymbol{A} = \mathrm{diag}\left(-\gamma^{(1)}, -\gamma^{(2)}, ..., -\gamma^{(N)} \right) , \tag{A15}$$

and $\boldsymbol{C}(t)$ is a time-dependent vector with n-th component

$$C^{(n)} = \gamma^{(n)} \left[f_{\mathrm{em}}^{(n)}(t_1) + \dot{f}_{\mathrm{em}}^{(n)}(t_1)\,t \right] + \dot{f}_{\mathrm{em}}(t_1)^{(n)} . \tag{A16}$$

The $N \times N$ square matrix $\boldsymbol{\Sigma}$ governs the magnitude of the increments of Brownian motion (Wiener process) $d\boldsymbol{B}(t)$, with

$$\boldsymbol{\Sigma} = \mathrm{diag}\left(\sigma^{(1)}, \sigma^{(2)}, ..., \sigma^{(N)} \right) . \tag{A17}$$

In the idealized model above, each pulsar's rotational state evolves phenomenologically according to a mean-reverting Ornstein-Uhlenbeck process, described by a Langevin equation, Equation (A14), whose general solution is given by (Gardiner 2009),

$$\boldsymbol{X}(t) = e^{\boldsymbol{A}t}\boldsymbol{X}(0) + \int_0^t e^{\boldsymbol{A}(t-t')}\boldsymbol{C}(t')dt' + \int_0^t e^{\boldsymbol{A}(t-t')}\boldsymbol{\Sigma}\,d\boldsymbol{B}(t') . \tag{A18}$$

From Equation (A18) we construct the discrete, recursive solution for $\boldsymbol{X}(t_k) = \boldsymbol{X}_k$ in the form of Equation (A4), with

$$\boldsymbol{F}_k = e^{\boldsymbol{A}\Delta t} \tag{A19}$$

$$= \mathrm{diag}\left(e^{-\gamma^{(1)}\Delta t}, e^{-\gamma^{(2)}\Delta t}, ..., e^{-\gamma^{(N)}\Delta t} \right) , \tag{A20}$$

$$\boldsymbol{G}_k \boldsymbol{u}_k = \int_{t_k}^{t_{k+1}} e^{\boldsymbol{A}(t_{k+1}-t')}\boldsymbol{C}(t')dt' , \tag{A21}$$

$$= \left(G_k^{(1)}, G_k^{(2)}, ..., G_k^{(N)} \right), \tag{A22}$$

$$\boldsymbol{w}_k = \int_{t_k}^{t_{k+1}} e^{\boldsymbol{A}(t_{k+1}-t')}\boldsymbol{\Sigma}\,d\boldsymbol{B}(t') , \tag{A23}$$

$$\begin{aligned} G_k^{(n)} = &f_{\mathrm{em}}^{(n)}(t_1) + \dot{f}_{\mathrm{em}}^{(n)}(t_1)\,(\Delta t + t_k) \\ &- e^{-\gamma\Delta t}\left[f_{\mathrm{em}}^{(n)}(t_1) + \dot{f}_{\mathrm{em}}^{(n)}(t_1)t_k \right] , \end{aligned} \tag{A24}$$

and $\Delta t = t_{k+1} - t_k$. From Equation (A23) the process noise covariance matrix is

$$\boldsymbol{Q}_k \delta_{kj} = \langle \boldsymbol{\eta}_k \boldsymbol{\eta}_j^{\mathsf{T}} \rangle = \mathrm{diag}\left(Q^{(1)}, Q^{(2)}, ..., Q^{(N)} \right) , \tag{A25}$$

with

$$Q^{(n)} = \frac{[\sigma^n]^2}{2\gamma^{(n)}}\left[1 - e^{-2\gamma^{(n)}\Delta t} \right] . \tag{A26}$$

The two remaining unspecified component matrices of the Kalman filter are the measurement matrix $\boldsymbol{H}_k$ and the measurement covariance matrix $\boldsymbol{R}_k$. These are defined straightforwardly from Equations (5)–(8). Specifically, $\boldsymbol{H}_k$ is a diagonal matrix where the n-th component of the diagonal is given by $g^{(n)}(t_k)$ from Equation (**??**). The measurement covariance satisfies $\boldsymbol{R}_k = E\left[\boldsymbol{v}\boldsymbol{v}^{\mathsf{T}} \right] = \sigma_{\mathrm{m}}^2$ for all k.

APPENDIX B: NESTED SAMPLING

We can use the likelihood returned by the Kalman filter, Equation (A2), in conjunction with likelihood-based inference methods to estimate the posterior distribution of $\boldsymbol{\theta}$ by Bayes' Rule,

$$p(\boldsymbol{\theta}|\boldsymbol{Y}) = \frac{\mathcal{L}(\boldsymbol{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\mathcal{Z}} , \tag{B1}$$

where $\pi(\boldsymbol{\theta})$ is the prior distribution on $\boldsymbol{\theta}$ and $\mathcal{Z}$ is the marginalised likelihood, or evidence

$$\mathcal{Z} = \int d\boldsymbol{\theta}\,\mathcal{L}(\boldsymbol{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) . \tag{B2}$$

We estimate the posterior distribution and the model evidence through nested sampling (Skilling 2006) in this paper. Nested sampling evaluates marginalised likelihood integrals, of the form given by Equation (B2). It also approximates the posterior by returning samples from $p(\boldsymbol{\theta}|\boldsymbol{Y})$. It does so by drawing a set of n_{live} live points from $\pi(\boldsymbol{\theta})$ and iteratively replacing the live point with the lowest likelihood with a new live point drawn from $\pi(\boldsymbol{\theta})$, where the new live point is required to have a higher likelihood than the discarded point. The primary advantage of nested sampling is its ability to compute $\mathcal{Z}$, on which model selection relies. Nested sampling is also computationally efficient and can handle multi-modal problems (Ashton et al. 2022). For these reasons, it has enjoyed widespread adoption in the physical sciences, particularly within the cosmological community (Mukherjee et al. 2006; Feroz & Hobson 2008; Handley et al. 2015), neutron star astrophysics (Meyers et al. 2021a,b; Melatos et al. 2023), particle physics (Trassinelli 2019) and materials science (Pártay et al. 2009). For reviews of nested sampling we refer the reader to Buchner (2021a) and Ashton et al. (2022). Multiple nested sampling algorithms and computational libraries exist. (e.g. Feroz & Hobson 2008; Feroz et al. 2009; Handley et al. 2015; Speagle 2020; Buchner 2021b). In gravitational wave research it is common to use the dynesty sampler (Speagle 2020) via the Bilby (Ashton & Talbot 2021) front-end library. We follow this precedent and use Bilby for all nested sampling Bayesian inference in this work.

The primary tunable parameter in nested sampling is n_{live}. More live points assists with large parameter spaces and multi-modal problems, whilst the uncertainties in the evidence and the posterior scale as $O\left(n_{\mathrm{live}}^{-1/2} \right)$. However the computational runtime scales as

$O(n_{\text{live}})$. Ashton et al. (2022) offered a rule-of-thumb trade-off, where the minimum number of live points should be greater than the number of static parameters. Informal empirical tests conducted as part of this paper support the trade-off suggested by Ashton et al. (2022); we find typically that the true $\boldsymbol{\theta}$ is contained within the 90% credible interval of the one-dimensional marginalised posteriors of $\hat{\boldsymbol{\theta}}$ for $n_{\text{live}} > 7 + 5N$ with $N \leqslant 50$. Unless stated otherwise we take $n_{\text{live}} = 1000$ for all results presented in this work.

APPENDIX C: CHOICE OF PRIORS

To deploy the analysis scheme of K24 it is necessary to specify a Bayesian prior $\boldsymbol{\theta}$ on the static parameters.

For $\pi(\boldsymbol{\theta}_{\text{gw}})$ we choose standard non-informative priors (e.g. Bhagwat et al. 2021). For $\pi(\boldsymbol{\theta}_{\text{psr}})$ *a priori* information from electromagnetic observations does exist. We adopt constrained uniform priors on $f_{\text{em}}^{(n)}(t_1)$ and $\dot{f}_{\text{em}}^{(n)}(t_1)$, which extend $\pm 10^3 \eta_f^{(n)}$ and $\pm 10^3 \eta_{\dot{f}}^{(n)}$ respectively about the central, injected values, where $\eta_f^{(n)}$ and $\eta_{\dot{f}}^{(n)}$ denote the errors quoted in the ATNF Pulsar Database. By using wider-than-necessary priors we expose the analysis scheme to a more stringent test. We set an uninformative broad prior $\pi[\sigma^{(n)}/(1\,\text{s}^{-3/2})] \sim \text{LogUniform}(10^{-23}, 10^{-19})$ which covers the range of values inferred from Equation 19. We set an non-informative prior on the new static parameter $\chi^{(n)} \sim \text{Uniform}(0, 2\pi)$. We do not set a prior on $\gamma^{(n)}$, because $\gamma^{(n)} T_{\text{obs}} \sim 10^{-5}$, and consequently $\gamma^{(n)}$ is effectively "unobservable" over T_{obs}. For validation purposes it is therefore sufficient to carry $\gamma^{(n)}$ through the analysis at its injected value. This reduces the total dimension of the parameter space to $7 + 4N$. The chosen prior $\boldsymbol{\theta}$ is the same as used in K24, with the addition of a prior on $\chi^{(n)}$ due to the inclusion of the pulsar term.

The chosen priors for the representative analysis in Section 4 are summarised in **??**

APPENDIX D: COMPLETE WORKFLOW SUMMARY

For the reader's convenience we now summarise the workflow for a representative PTA analysis using the Kalman filter and nested sampler for parameter estimation and model selection:

   (i) Specify a PTA composed of N pulsars
   (ii) Obtain N data inputs $f_{\text{m}}^{(n)}(t)$, collectively labelled $\boldsymbol{Y}$
   (iii) Specify a state-space model $\mathcal{M}$, with static parameters $\boldsymbol{\theta}$
   (iv) Specify prior distribution $\pi(\boldsymbol{\theta})$
   (v) Sample n_{live} points from $\pi(\boldsymbol{\theta})$
   (vi) For each live point:

      (a) Pass the sample $\boldsymbol{\theta}_{\text{sample}}$ to the Kalman filter
      (b) Iterate over the input data using the Kalman filter and obtain a single log $\mathcal{L}$ value, Equation (A2)

   (vii) Remove the live point with the lowest likelihood value, log $\mathcal{L}_{\text{lowest}}$
   (viii) Sample a new live point from $\pi(\boldsymbol{\theta})$, subject to the requirement that the new likelihood obeys $\mathcal{L}_{\text{new}} > \mathcal{L}_{\text{lowest}}$, where log $\mathcal{L}_{\text{new}}$ is calculated via steps (vi)(a)–(vi)(b).
   (ix) Update $p(\boldsymbol{\theta}|\boldsymbol{Y})$ and $\mathcal{Z}$ with nested sampler
   (x) Repeat steps (vii)–(ix) until convergence criteria are satisfied.

In order to compute β the above workflow is repeated for a different $\mathcal{M}$. The resulting $\mathcal{Z}$ values can then be compared. We remind the reader that the above workflow differs from a realistic PTA analysis in one important respect, namely that the data are input as frequency time series $f_{\text{m}}^{(n)}(t)$ instead of pulse TOAs. The generalization to TOAs is subtle and will be tackled in a forthcoming paper.

APPENDIX E: MULTIPLE NOISE REALISAITONS AND THE WASSERSTEIN DISTANCE

We can extend the analysis of these two points and consider 1000 noise realisations of $\boldsymbol{Y}$, beyond the X plotted in Figure 1. As in Figure 1 we independently estimate the static parameters, $\boldsymbol{\theta}$ for each realisation and calculate $p(\boldsymbol{\theta}'|\boldsymbol{Y})$.

To this end we start with the representative example in Table 1 and generate 1000 realisations of the process noise $\xi^{(n)}(t)$ and measurement noise $\varepsilon^{(n)}(t)$. For each realisation we

We now consider the complete set of 10^3 noise realisations, going beyond the subset of nine realisations in the preceding discussion. We ask the question: how "similar" are the 10^3 marginalized, one-dimensional posteriors computed for each of the seven parameters in $\boldsymbol{\theta}_{\text{gw}}$? There is no unique way to answer this question. In this paper we appeal to the Wasserstein distance (WD; Vaserstein 1969; Villani 2009) from optimal transport theory, which defines an intuitive notion of similarity between probability distributions. The WD is a popular metric in machine learning (Arjovsky et al. 2017), climate modelling (Paxton et al. 2022; Kimpson et al. 2023), computational biology (González-Delgado et al. 2023) and geophysics (Morris et al. 2023); a short review is presented in Appendix **??**. The WD measures the cost of an optimal strategy for moving probability mass between two distributions from position x to position y, with respect to some cost function $c(x, y)$. In this paper we use the first WD moment, $W_1(\mu, \nu)$, which is defined and interpreted in Appendix **??**. Intuitively W_1 bounds the difference in the expectation value of a parameter selected from $\boldsymbol{\theta}_{\text{gw}}$ with respect to the PDFs μ and ν. Taking a concrete example, suppose that we infer two one-dimensional posterior distributions $\mu(\iota)$ and $\nu(\iota)$ for ι, for two different realisations of the noise, and calculate $W_1(\mu, \nu) = 0.5$ rad. Then we can infer $|\langle\iota\rangle_\mu - \langle\iota\rangle_\nu| \leqslant 0.5$ rad.

Table E1 summarizes the WD between the 5×10^5 pairs of one-dimensional posteriors across the 10^3 realizations, for each of the seven parameters in $\boldsymbol{\theta}_{\text{gw}}$. The median W_1 for each parameter is tabulated in the rightmost column and denoted by $W_{1,\text{median}}$. As a percentage of the known, injected value, $W_{1,\text{median}}$ ranges from 0.2% for Ω to 67% for Φ_0. These values quantify approximately the natural dispersion in parameter estimates when the analysis scheme is applied to real, astronomical data, where the true parameter values are unknown. Moreover, as a percentage of the prior domain, $W_{1,\text{median}}$ ranges from 8.1×10^{-4} % for h_0 to 2.7% for ψ. These values confirm that the nested sampler converges reliably to a single, narrow peak without railing against the prior bounds for 10^3 noise realizations. Further analysis of the WD is performed in Appendix **??**.

APPENDIX F: CHALLENGES OF THE PULSAR TERM

In this section....

To elucidate the complications introduced by including the pulsar term, we generate some noisy synthetic data $\boldsymbol{Y}$ using the methods outlined in K23. $\boldsymbol{\theta}_{\text{gw}}$ are chosen arbitrarily, with $\alpha = 1.0$ rad.
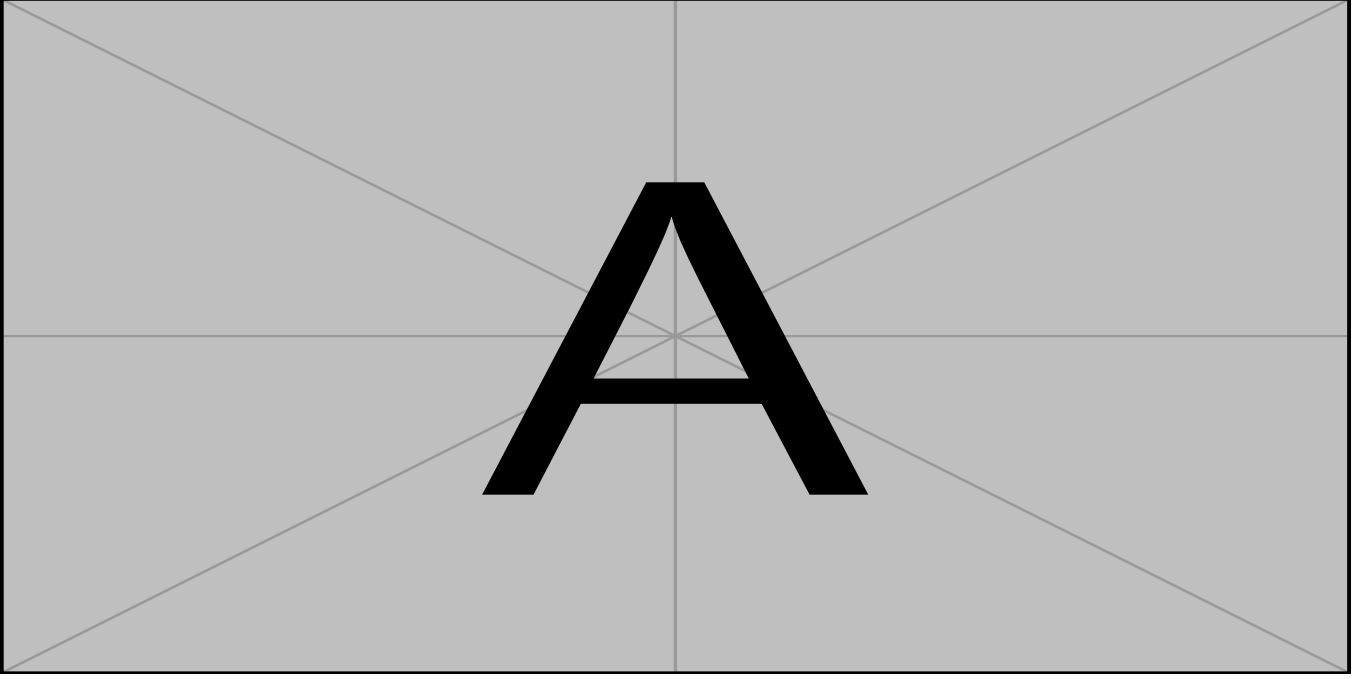
Figure F1. Logarithm of the one-dimensional likelihood $\mathcal{L}(\alpha)$ returned by the Kalman filter acting on synthetic data $\boldsymbol{Y}$. All other static parameters of $\boldsymbol{\theta}$ are held constant at their true injected values. In the left-hand panel the Kalman filter uses Equation (8) to calculate $\mathcal{L}(\alpha)$, i.e. inclusive of the pulsar terms. In the right-hand panel the Kalman filter uses Equation 15, i.e. just the Earth terms. The vertical dashed lines indicate the true injected value of α (= 1.0 rad) used to generate $\boldsymbol{Y}$. The inset plot covers the region $1.0 - 10^{-4} \leqslant \alpha \leqslant 1.0 + 10^{-4}$ rad. In the left hand panel $\mathcal{L}(\alpha)$ is very noisy, though a smooth concave optima exists on scales $O(10^{-4})$ rad. This noise hampers the ability of sampling algorithms to converge to a posterior distribution. Conversely, in the right hand panel $\mathcal{L}(\alpha)$ is globally smooth.

Parameter	Injected value	Units	$W_{1,\text{median}}$
Ω	5×10^{-7}	Hz	10^{-9}
Φ_0	0.20	rad	0.13
ψ	2.50	rad	0.16
ι	1.0	rad	0.14
δ	1.0	rad	0.09
α	1.0	rad	0.17
h_0	5×10^{-15}	—	8×10^{-16}

Table E1. Median value of the Wasserstein distance $W_{1,\text{median}}$, for each parameter in $\boldsymbol{\theta}_{\text{gw}}$, calculated across the $\binom{10^3}{2}$ pairs of probability posteriors, for the 10^3 noise realisations in Figure **??**. $W_{1,\text{median}}$ is generally smaller than the prior domain.

The choice of $\boldsymbol{\theta}_{\text{psr}}$ corresponds to the 47 pulsars in the 12.5 year NANOGrav data set (Arzoumanian et al. 2020). We run the Kalman filter using the measurement equation inclusive of the pulsar terms, Equation (8) on $\boldsymbol{Y}$ for different values of α, $0.0 \leqslant \alpha \leqslant \pi$ rad. We obtain a curve $\mathcal{L}(\alpha)$, which is plotted in the left hand panel of Figure F1 (blue curve). The dashed grey line indicates the injected value of α (= 1.0 rad). The inset plot shows a zoomed-in section of $\mathcal{L}(\alpha)$ within the region $1.0 - 10^{-4} \leqslant \alpha \leqslant 1.0 + 10^{-4}$ rad.

There are two important features to discuss in the left-hand panel of Figure F1. The first is that on scales of order the width of the prior $\pi(\alpha)$ (i.e radians), the curve of $\mathcal{L}(\alpha)$ is very noisy. Similar noisy curves are obtained for $\mathcal{L}(\Omega)$ and $\mathcal{L}(\delta)$. For all other static parameters, the $\mathcal{L}(\Theta)$ curves are smooth and globally concave. The second is that despite being noisy, a true likelihood maximum coincident with the injected value does exist and is locally

concave on scales $O\left(10^{-4}\right)$ radians. This means that whilst it is theoretically possible to use likelihood-based methods for Bayesian estimation of α, in practice the computation is intractable since the noise in the likelihood curve inhibits sampling algorithms to reliably converge to a definitive posterior. Moreover, $\mathcal{L}(\alpha)$ is only a one-dimensional function. The challenge for sampling algorithms is further exacerbated when we consider the three-dimensional parameter space in $\alpha - \delta - \Omega$, or the full $7+5N$ dimensions of Section 2.3.

Why do α, δ and Ω exhibit noisy $\mathcal{L}(\Theta)$ curves whereas other static parameters do not? These parameters appear in the pulsar term as a product in the pulsar-dependent phase correction $\Omega\left(1 + \boldsymbol{n} \cdot \boldsymbol{q}^{(n)}\right) d^{(n)}$, where we remind the reader that $\boldsymbol{n} = \boldsymbol{n}(\delta, \alpha)$. Over the typical pulsar distances $d^{(n)} \sim 1$ kpc, a 10 nHz GW can accumulate 10^3 cycles. This creates a confusion problem when taking the cosine of $\Omega\left(1 + \boldsymbol{n} \cdot \boldsymbol{q}^{(n)}\right) d^{(n)}$ since we do not know how many cycles the wave has gone through, modulo 2π. TK: this explanation is rubbish! How can we say this better? Conversely, if we drop the pulsar term, the $\mathcal{L}(\alpha)$ curve (as well as $\mathcal{L}(\Omega)$, $\mathcal{L}(\delta)$) This is shown by the orange curve in the right-hand panel of Figure F1 where we plot $\mathcal{L}(\alpha)$, analogous to the left-hand panel, but now use Equation 15 in our Kalman filter (i.e. just the Earth terms) rather than the full Equation 8. In this case there is no confusion regarding the number of cycles the wave has gone through over a distance $d^{(n)}$ and the curve is smooth.

In this section we present a new method to infer $\boldsymbol{\theta}$ and calculate the marginal likelihood (i.e. the model evidence). In Section F1 we outline how noisy measurements of the pulsar frequency, $f_{\text{m}}^{(n)}(t)$, can be used to estimate the hidden state sequence, $f_{\text{p}}^{(n)}(t)$, using

a Kalman filter. In Section B we demonstrate how to deploy the Kalman filter in conjunction with a nested sampling technique to perform Bayesian inference. Model selection and the specification of the null model are described in Section F2. A complete summary of the workflow is presented in Section **??**. The method complements traditional PTA analyses because (i) it does not assume a specific functional form (e.g. power law) for the power spectral density of the TOA fluctuations; and (ii) it infers $\boldsymbol{\theta}$ conditional on the specific, time-ordered, random realization of the noise corresponding to the observed sequence of TOAs rather than on ensemble-averaged quantities like the power spectral density.

F1 Kalman filter and likelihood

F2 Model selection

The evidence integral $\mathcal{Z}$ returned by nested sampling is the probability of the data $\boldsymbol{Y}$ given a model $\mathcal{M}_i$. We compare competing models via a Bayes factor,

$$\beta = \frac{\mathcal{Z}(\boldsymbol{Y}|\mathcal{M}_1)}{\mathcal{Z}(\boldsymbol{Y}|\mathcal{M}_0)} \ . \tag{F1}$$

Throughout this work we take $\mathcal{M}_1$ to be the state-space model that includes the presence of a GW. $\mathcal{M}_0$ is the null model, which assumes no GW exists in the data, and is equivalent to setting $g^{(n)}(t) = 1$ in Equations (**??**) and (8). The Bayes factors we quote in this work therefore quantify whether the data supports evidence for a GW signal compared to there being no GW signal present.

This paper has been typeset from a TeX/LaTeX file prepared by the author.