Impact of quasi-periodic and steep-spectrum timing noise on the measurement of pulsar timing parameters

Michael J. Keith ⬤★ and Iuliana C. Niţu ⬤

Jodrell Bank Centre for Astrophysics, Department of Physics & Astronomy, The University of Manchester, M13 9PL Manchester, UK

ABSTRACT

Timing noise in pulsars is often modelled with a Fourier-basis Gaussian process that follows a power law with periodic boundary conditions on the observation time, T_{span}. However, the actual noise processes can extend well below $1/T_{span}$, and many pulsars are known to exhibit quasi-periodic timing noise. In this paper, we investigate several adaptions that try to account for these differences between the observed behaviour and the simple power-law model. First, we propose to include an additional term that models the quasi-periodic spin-down variations known to be present in many pulsars. Secondly, we show that a Fourier basis of $1/2T_{span}$ can be more suited for estimating long-term timing parameters such as the spin frequency second derivative (F2), and is required when the exponent of the power spectrum is greater than ∼4. We also implement a Bayesian version of the generalized least-squares 'Cholesky' method which has different limitations at low frequency, but find that there is little advantage over Fourier-basis methods. We apply our quasi-periodic spin-down model to a sample of pulsars with known spin-down variations and show that this improves parameter estimation of F2 and proper motion for the most pathological cases, but in general the results are consistent with a power-law model. The models are all made available through the RUN_ENTERPRISE software package.

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

1 INTRODUCTION

Pulsar timing relies on the comparison of the observed time of arrival (ToA) of a pulse from a pulsar with a parametric model of the pulsar's spin, astrometric, and other parameters. The difference between the observed ToA and the model for a given observation is termed the residual, and in an ideal case the post-fit residual would be white noise well described by the uncertainty on the ToA. However, it is well known that pulsar residuals often exhibit excess noise with both a white noise component (Osłowski et al. 2011; Parthasarathy et al. 2021), independent from observation to observation, and red noise processes which are correlated on time-scales much longer than the typical observing cadence (Cordes & Downs 1985; Hobbs, Lyne & Kramer 2010; Parthasarathy et al. 2019). For many pulsars, the red noise processes are very significant and can deviate from the model by many rotations.

Estimating the power-spectral density of the residuals reveals that the red noise typically appears well modelled by a power-law process in the Fourier domain (Reardon et al. 2016; Parthasarathy et al. 2019). The exponent of these power laws can be very steep, which hampers simple techniques for estimation of the power-spectral density, and typically requires using a method of pre-whitening the data using either a differential method or using the covariance matrix. Coles et al. (2011) showed that common fitting methods, such as weighted least-squares can be biased by this red noise, even when using simple methods to deal with the red noise by fitting polynomials

or sinusoids. The same paper also demonstrates a method to estimate the covariance matrix of the residuals from the estimate of the power-spectral density, which can then be used with generalized least-squares (GLS) to solve for the other parameters in the presence of red noise. Although this so-called 'Cholesky' method is effective, the iterative nature of requiring the covariance function of the data to properly estimate the power-spectral density used to model the covariance function means that it is possible to find a false peak in the overall likelihood. Bayesian techniques were proposed to solve for the data covariance function and the model parameters simultaneously, analytically integrating over less critical parameters in the timing model (van Haasteren & Levin 2013; Lentati et al. 2014).

The long-term fluctuations of the generally more stable millisecond pulsars are also expected to contain red noise terms caused by a stochastic background of gravitational waves. Detection of these waves in pulsar timing data is a key objective of several large international collaborations (Arzoumanian et al. 2020; Kerr et al. 2020; Chen et al. 2021; Miles et al. 2023; Tarafdar et al. 2022). This has necessitated the optimization of the Bayesian algorithms to reduce the very large computational complexity of solving a model containing perhaps 100 pulsars simultaneously. This led to the development of modelling the red noise as a Fourier-basis Gaussian process which can greatly reduce the number of computations required to fit the red noise models (Lentati et al. 2013). These methods have been widely applied to pulsar data for gravitational wave detection in codes such as TEMPONEST (Lentati et al. 2014) and ENTERPRISE (Ellis et al. 2019).

★ E-mail: michael.keith@manchester.ac.uk

These models are also recently being more widely applied to the timing of the larger population of pulsars (e.g. Parthasarathy et al. 2019). However, these algorithms are optimized for millisecond pulsars where the scale of the red noise means that only a small number of Fourier components are required to fully model the noise. Furthermore, one might question if the periodic nature of a Fourier-basis Gaussian process may impact the measurement of long-term spin parameters such as the spin frequency second derivative that is an important measurable when considering the long-term evolution of pulsars. More importantly, although the red noise in pulsars is often well fit by a power law, we have observed quasi-periodic spin evolution in a growing number of pulsars (e.g. Hobbs et al. 2010; Parthasarathy et al. 2019).

There have been some attempts to include these quasi-periodic variations in the timing noise model. For example studies of a sample of pulsars regularly timed with the Parkes radio telescope have been searched for periodic modulations in the timing residuals (Kerr et al. 2016; Parthasarathy et al. 2019). In these cases, the method employed was to add a sinusoidal component to the timing model with a free amplitude and periodicity, and use likelihood or Bayesian evidence ratios to test the favourability of a periodic model. Similarly, recent searches for planetary companions around pulsars on the Jodrell Bank pulsar timing archive demonstrated that periodic timing variations are detectable in a large number of pulsars (Niţu et al. 2022). However, although these methods are effective at detecting some quasi-periodic variability, especially in cases where the periodicity is relatively constant, they do not fully model the quasi-periodic variations and therefore may not be optimal for parameter estimation. In this work, we build on these ideas by developing a quasi-periodic term for the power-law timing model.

Young pulsars also often show a significant measurement $\ddot{\nu}$, the long-term frequency second derivative (F2). Estimation of F2 is important for the understanding of the long-term evolution of pulsars, and is important for understanding the pulsar braking index and in the recovery from pulsar glitches (Espinoza et al. 2011; Lyne et al. 2015; Lower et al. 2021), as well as having implications for high precision pulsar timing experiments (Liu et al. 2019). The F2 parameter appears as a cubic term in the residuals, and hence is most sensitive on the longest time-scales, which are also those where the timing noise dominates. Therefore the choice of timing noise model may be important for estimating the magnitude of F2, and particularly for quantifying the significance of any measurement. Particularly, the important question is if the observed F2 is consistent with being the low-frequency extension of the timing noise observed on shorter time-scales, or if it reflects an additional process such as glitch recovery or the long-term braking of the pulsar.

In this paper, we will attempt to address two questions regarding the application of the current Fourier-basis Gaussian process models on the canonical (i.e. non-millisecond) pulsar populations.

(i) Does the periodic boundary condition of Fourier basis models affect the ability to measure F2 for young pulsars? If so can this be mitigated by changing the lowest basis frequency?

(ii) Is there any advantage to modelling the pulsar timing noise with a model containing a quasi-periodic component, rather than the pure power-law model traditionally used?

We also take the opportunity to implement a Bayesian version of the Coles et al. (2011) model within the ENTERPRISE framework and confirm the results of Lentati et al. (2014) that the Fourier-domain Gaussian process model performs equally as well as the GLS method for the estimation of pulsar timing parameters.

2 TIMING NOISE MODELLING

We make use of the ENTERPRISE framework for the Fourier-basis Gaussian process models. A complete description of this method can be found in Lentati et al. (2013, 2014). In brief, we define a model of the Fourier-domain power-spectral density (PSD) and use this to constrain the amplitude of a harmonic series of sinusoids, i.e. a Fourier basis. The PSD is typically modelled as a power law, parametrized by the log-amplitude $\log_{10}(A)$ and the spectral exponent γ. For a stochastic Gaussian process, the underlying PSD is equal to the variance in the amplitude of the corresponding sinusoid, and so we can fit for harmonically related sinusoids with a Gaussian prior with variance defined by the model PSD. In practice this constrained fit is analytically marginalized over, making the method very computationally efficient.

The GLS approach assumes that the noise is drawn from a Normal distribution with zero mean and covariance given by a covariance matrix **C**. We typically estimate **C** by the Wiener–Khinchin theorem, which states that covariance as a function of lag is the Fourier transform of the PSD. Coles et al. (2011) propose fitting a power-law model to estimates of the PSD computed using a periodogram analysis of the pulsar data. However, this estimation of the PSD may depend on the choice of pulsar timing parameters, which in turn may depend on the choice of **C**. This can be somewhat addressed by an iterative approach, but this still does not allow for the uncertainty in the PSD model to be factored into the uncertainty on other fit parameters. The Bayesian approach is to fit the PSD hyperparameters (the amplitude, A and exponent, γ, of the power-law) at the same time as solving for other parameters of interest, whilst using the $\mathbf{C}(A, \gamma)$ to evaluate the likelihood, and GLS to analytically marginalize over any other linear parameters.

It is worth noting that for both cases, we cannot strictly model a pure power-law. For the GLS method, the integral to compute **C** is not finite for a pure power-law, and so a corner frequency, f_c is introduced, below which the PSD flattens, typically chosen to be of order $1/T_{span}$. The Fourier basis model must also make a choice of the set of harmonically related frequencies to use. Typically for a data set of length T_{span} the lowest frequency used is of order $f_{low} = 1/T_{span}$. The number of harmonics, N_{harm} used is also finite, and the natural choice is to ensure that the PSD at the frequency $f_{high} = N_{harm}f_{low}$ is dominated by white noise. Unlike the GLS approach, the Fourier basis model does not flatten below f_{low}, but rather the use of a Fourier basis imposes periodic boundary conditions. This means that the Fourier basis red noise model must be periodic over a window of $1/f_{low}$. Clearly the intrinsic timing noise has no knowledge of our observing span, and therefore any timing noise longer than this time-scale will leak into the fit parameters sensitive to the longest time-scales. Usually, it is assumed that this only has a small perturbation on the estimation of the long-term spin, F0, and spin derivative, F1, parameters. In effect it means that F0 and F1 represent the average spin frequency and spin-down frequency over the observing time-span rather than estimates of the long-term spin of the pulsar, but this usually has little impact on their interpretation.

2.1 Modelling quasi-periodic variations in spin down

The widely used timing noise models typically use a power-law model for the PSD model, with some use of multiple or broken power laws. However, there is good evidence that many pulsars show quasi-periodic behaviour, which may not be well modelled

by a power law. Here, we propose a modification to the existing PSD model to include a quasi-periodic term. We choose to do this in the Fourier domain to re-use the existing frameworks, but equivalent time-domain covariance kernels could also be constructed. We have chosen to model the quasi-periodic process with a Gaussian function centred on the fundamental frequency and harmonics of a periodic process. Motivated by the observed power spectra of pulsars exhibiting quasi-periodic timing noise (particularly PSRs B1828−11 and B1540-06), the Gaussian functions broaden and decrease in amplitude at higher harmonics. Particularly the integral of each Gaussian decays exponentially, and the Gaussians maintain a constant fractional bandwidth. This choice is largely arbitrary, but is similar in intent to the relatively common choice of a cosine or sine-squared kernel multiplied by a Gaussian envelope in the time domain.

The quasi-periodic process described in Lyne et al. (2010) appears in the spin-frequency derivative, $\dot{\nu}$, of the pulsar. Therefore we begin by considering a quasi-periodic Fourier-domain Gaussian process in $\dot{\nu}(t)$, with a PSD composed of a train of Gaussian components with integral of the kth harmonic decaying like $\exp(-(k-1)/\lambda)$, where λ is a free parameter that effectively describes the number of significant harmonics in the data. Specifically, we define the spectral shape of our model by

$$q(f, f_{\text{qp}}, \sigma, \lambda) = \sum_{k=1}^{N} \frac{\exp(-(k-1)/\lambda)}{k} \exp\left(\frac{-(f - kf_{\text{qp}})^2}{2k\sigma^2}\right), \quad (1)$$

where σ is the width of the Gaussian function and f_{qp} is the fundamental frequency ($k = 1$ harmonic) of the quasi-periodic process. Our analysis requires a model of the PSD in the residuals, $r(t)$, which is related to $\dot{\nu}$ by $\dot{\nu}(t) \propto \ddot{r}(t)$.

Hence, to convert our model of $\dot{\nu}(t)$ to the effect on the residuals we integrate twice with respect to time. In the Fourier domain, we only need to integrate the Fourier basis functions with respect to time, which has the effect of multiplying the power spectrum by f^{-4}, ignoring constant scaling factors. The resulting power spectrum is therefore proportional to $f^{-4}q(f, f_{\text{qp}}, \sigma, \lambda)$.

The choice of a Gaussian function for the spectral shape of our quasi-periodic process means that the power decays fairly quickly away from the peak of the quasi-periodic function. In practice, we find that this is not sufficient to model the observed spin variations and the observed PSD appears more like a quasi-periodic process on top of a power-law model. Some of this may be because the $\dot{\nu}$ variation process has intrinsically broader 'tails' than a Gaussian function. This might be observed if the $\dot{\nu}$ variation process is behaving more like a random switching of $\dot{\nu}$ and hence a random walk in ν. This would lead to a f^{-2} PSD in ν and hence a f^{-4} power-law process in $r(t)$. One approach for modelling this would be to replace the Gaussian function with a function with wider tails (e.g. the probability density function of the Cauchy distribution), but we find in practice it is simpler to add an additional power-law term to the model. This also has the advantage that it allows us to separate the 'purely' quasi-periodic component from the excess power-law noise and capture both the excess $\dot{\nu}$ noise as a process with spectral exponent of −4, but also allows us to model a wider range of noise processes with different spectral indices. Indeed, we will see that in practice γ is significantly divergent from −4 for some pulsars. The practical implementation of this model is described in Section 2.2.

2.2 Implementation of the noise models

2.2.1 Fourier basis Gaussian Process

For the Fourier-basis Gaussian process models we use the model of Lentati et al. (2014) as implemented in ENTERPRISE. The power-law noise model is defined by a one-sided power spectral density

$$P_{\text{pl}}(f) = \frac{A_{\text{pl}}^2}{12\pi^2} \left(\frac{f}{f_{\text{yr}}}\right)^{-\gamma} f_{\text{yr}}^{-3}, \quad (2)$$

where f_{yr} is a frequency of 1 per year. The hyperparameters are the spectral exponent γ, and log-amplitude $\log_{10}(A_{\text{pl}})$.

There are numerous ways to model quasi-periodic variations, and in this work we have chosen the model described in Section 2.1. Specifically our model for quasi-period variations in $\dot{\nu}$ has a power-spectral density given by

$$P_{\text{qp}}(f) = R_{\text{qp}} P_{\text{pl}}(f_{\text{qp}}) q(f, f_{\text{qp}}, \sigma, \lambda) \left(\frac{f}{f_{\text{yr}}}\right)^{-4}, \quad (3)$$

where R_{qp} is the ratio of quasi-periodic noise to red noise at the central frequency f_{qp}, and $q(f, f_{\text{qp}}, \sigma, \lambda)$ is the function describing the spectral shape of our quasi-periodic process, defined in equation (1).

This model consists of N harmonically related Gaussian functions of width σ, which decay with an exponential scale λ. The hyperparameters of the quasi-periodic model are $\log_{10}(R_{\text{qp}})$, f_{qp}, σ, and λ. The final red noise model including the quasi-periodic $\dot{\nu}$ term is $P_{\text{qp}}(f) + P_{\text{pl}}(f)$, and so contains the 6 hyperparameters from both red noise models.

In order to prevent P_{qp} growing at very low frequencies, where the Gaussian function decays slower than f^{-4}, a low-frequency cut-off is applied such that $P_{\text{qp}} = 0$ for $f < f_{\text{cut}}$. We choose f_{cut} to be the local minima value of P_{qp} for $N = 1$ (i.e. for only the lowest frequency harmonic) and hence can be determined by differentiation to be

$$f_{\text{cut}} = \frac{1}{2}\left(f_{\text{qp}} - \sqrt{(f_{\text{qp}}^2 - 16\sigma^2)}\right). $$

2.2.2 Generalized least squares

When directly computing the covariance function for GLS fitting we cannot use a pure power law as this diverges as the frequency approaches zero, so we adapt the model proposed by Coles et al. (2011), where the two-sided power-spectral density is given by

$$P_{\text{pl}}(f) = P_{\text{ref}} \left(1 + \left[\frac{f}{f_{\text{c}}}\right]^2\right)^{-\gamma/2} \left(\frac{f_{\text{c}}}{f_{\text{ref}}}\right)^{-\gamma}, \quad (4)$$

where f_{c} is the corner frequency below which the spectrum turns over and f_{ref} is the reference frequency at which P_{ref} is defined. This modification allows us to sample the power-spectral density at a meaningful scale, rather than at f_{c} which may correspond to a time-scale well beyond the observing span. This simplifies to the Coles et al. model for $f_{\text{c}} = f_{\text{ref}}$, and approaches equation (2) for $f_{\text{ref}} = f_{\text{yr}}$ and small values of f_{c}. Specifically, for f_{c} significantly smaller than $1/T_{\text{span}}$, conversion between Fourier basis Lentati et al. models (as output by ENTERPRISE or TEMPONEST) and Coles et al. GLS models (as used in TEMPO2) can be made by making use of the conversion

$$2P_{\text{ref}} = \frac{A_{\text{pl}}^2}{12\pi^2} \left(\frac{f_{\text{ref}}}{f_{\text{yr}}}\right)^{-\gamma} f_{\text{yr}}^{-3}, \quad (5)$$

where the factor of 2 converts from one-sided to two-sided power-spectral density.

Table 1. The four models used in this work.

Model	Method	Type	Equations
PF	Fourier basis	Power law	(2)
PG	GLS	Power law	(4)
QF	Fourier basis	Power law + Quasi-periodic	(2) + (3)
QG	GLS	Power law + Quasi-periodic	(4) + (3)

It is of course possible to also implement the quasi-periodic model within the GLS framework. The quasi-periodic model remains unchanged, except being scaled by $P_{pl}(f)$ from equation (4) rather than equation (2).

2.2.3 White noise

In addition to red noise, pulsars are also seen to exhibit excess white noise (Osłowski et al. 2011; Parthasarathy et al. 2021). Hence for the analysis in this paper we expand each of our noise models to also include the widely used 'EFAC' and 'EQUAD' white noise parameters that linearly scale and add in quadrature to the formal ToA uncertainty respectively. We follow the convention in TEMPONEST (Lentati et al. 2014) such that the output ToA error relates to the input ToA error, σ_{toa}, by

$$\sigma_{out} = \sqrt{(\text{EFAC} \times \sigma_{toa})^2 + \text{EQUAD}^2}. \qquad (6)$$

2.2.4 The four models

We therefore have four models, two using the Fourier basis and two using the GLS, which are summarized in Table 1. These have been implemented as part of a single pulsar Bayesian toolkit RUN_ENTERPRISE (Keith, Niţu & Liu 2022), which utilizes the ENTERPRISE framework and the pulsar timing package TEMPO2 as exposed by the python interface LIBSTEMPO. RUN_ENTERPRISE has been developed to allow fitting and comparison of a wide combination of pulsar noise models and timing parameters using a range of Bayesian samplers. In this work, we sample the hyperparameters using MULTINEST (Feroz & Hobson 2008), via the PYTHON interface PYMULTINEST (Buchner et al. 2014).

3 LOW FREQUENCY CUT-OFF AND PERIODIC BOUNDARY CONDITIONS

As we note in Section 2, even if the intrinsic spin-noise is a power-law over the observation time-span, we know that the power-law cannot extend to zero frequency as the PSD must remain finite whilst the power-law diverges. The two types of models, we use here behave differently at the lowest frequencies. The Fourier basis models define a lowest frequency that is both a cut-off for the power-law and also defines the time window for the periodic boundary conditions imposed by the Fourier basis on the red noise model. The GLS models define a corner frequency at which the model turns over and hence can have a finite integral. It is generally assumed that the effects of the frequencies below $1/T_{span}$ are absorbed in small changes in the F0 and F1 parameters, but this will also affect the estimation of F2 to some extent. Although these models have been tested extensively during their development, the impact on the F2 parameter was not the focus of these tests, yet F2 is an important parameter for studying the long-term spin evolution of pulsars. To investigate the impact of the choice of model on F2 estimation we perform a simple simulation. We generate 224 simulated ToAs over
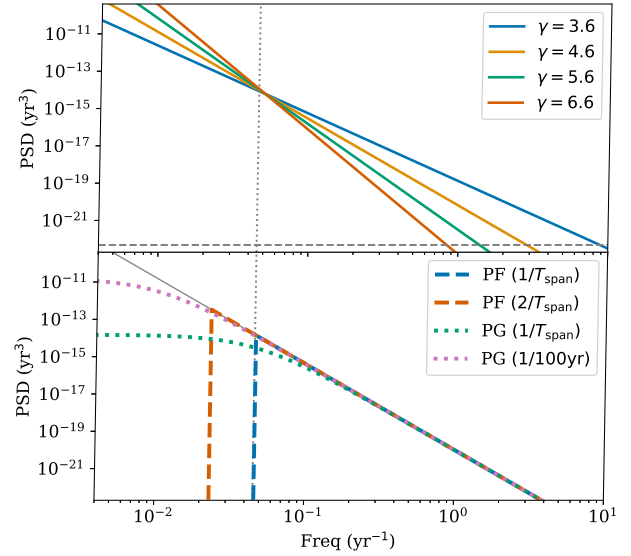
Figure 1. Upper panel: The injected power law red noise for $\gamma = 3.6$, 4.6, 5.6, and 6.6. The dotted vertical line is at $1/T_{span}$ and the dashed grey line shows the white noise PSD. Lower panel: The behaviour of the four different 'power-law' models used for $\gamma = 4.6$. Dashed curves are for the PF model with $f_{min} = 1/T_{span}$ and $f_{min} = 1/2T_{span}$ and dotted curves are for the PG model with $f_c = 1/T_{span}$ and $f_c = 1/100\text{yr}^{-1}$. The dotted vertical line is at $1/T_{span}$ and the solid grey line is a pure power law.

a timespan of 22 yr, with cadence and measurement errors and pulsar parameters based on the actual Lovell telescope observations of PSR J1909+0912. In addition to white noise, we inject a power-law spin noise that extends to a frequency of $1/200 \text{ yr}^{-1}$, i.e. the longest periodicity is ~ 9 times T_{span}. We run four sets of simulations, with red-noise spectral exponents of 3.6, 4.6, 5.6, and 6.6 to reflect the typical range of values we observe in the young pulsars where F2 is likely to be measured (Parthasarathy et al. 2019). The amplitude of the red noise is chosen such that it produces a root-mean-square (rms) residual similar to the real pulsar, which is dominated by the red noise, and scaled between the four exponents such that the power at $1/T_{span}$ is approximately the same. The four injected power laws are shown in the upper panel of Fig. 1. We compare four approaches to estimating the red noise and F2.

(i) PF model with the lowest frequency at $1/T_{span}$, as typically used for gravitational wave studies;

(ii) PF model with lowest frequency at $1/2T_{span}$ to reduce requirements for periodic boundary conditions;

(iii) PG model with $f_c = 1/T_{span}$, as suggested by Coles et al. (2011);

(iv) PG model $f_c = 1/100\text{yr}^{-1}$, a commonly used alternative for when it is felt that $f_c \ll 1/T_{span}$.

There are of course many other choices that could be made, but we feel that these four explore a reasonable range of commonly used models. An example of these four models is shown in the lower panel of Fig. 1. It is worth noting that the PG models diverge from a power-law well before reaching f_c, and hence the power at $1/T_{span}$ is very different for the $f_c = 1/T_{span}$ model compared to the other three models, which only significantly diverge from the power-law at frequencies below $1/T_{span}$.

Table 2. Statistics of parameter estimates from 40 realizations of simulated observations with power-law red noise. Parameters in the columns are defined in Section 3.1 and from left to right represent the scatter, error overestimation factor, and bias in the fitting.

γ	Model	$\sigma_p/\sigma_{p,\mathrm{min}}$	$\bar{e}_p/\sigma_p$	Δ_p/σ_p
3.6	PF $(1/T_{\mathrm{span}})$	1.0(1)	1.4(2)	0.0(2)
3.6	PF $(1/2T_{\mathrm{span}})$	1.0(1)	1.2(1)	0.0(2)
3.6	PG $(1/T_{\mathrm{span}})$	1.0(1)	0.7(1)	0.0(2)
3.6	PG (1/100yr)	1.0(1)	1.1(1)	0.0(2)
4.6	PF $(1/T_{\mathrm{span}})$	1.1(1)	1.0(1)	0.1(2)
4.6	PF $(1/2T_{\mathrm{span}})$	1.0(1)	1.2(2)	0.1(2)
4.6	PG $(1/T_{\mathrm{span}})$	1.1(1)	0.6(1)	0.1(2)
4.6	PG (1/100yr)	1.0(1)	1.0(1)	0.1(2)
5.6	PF $(1/T_{\mathrm{span}})$	1.0(1)	0.37(4)	−0.3(2)
5.6	PF $(1/2T_{\mathrm{span}})$	1.0(1)	1.1(1)	−0.3(2)
5.6	PG $(1/T_{\mathrm{span}})$	1.1(1)	0.33(4)	−0.3(2)
5.6	PG (1/100yr)	1.0(1)	0.61(7)	−0.3(2)
6.6	PF $(1/T_{\mathrm{span}})$	1.1(1)	0.09(1)	−0.1(2)
6.6	PF $(1/2T_{\mathrm{span}})$	1.0(1)	0.74(9)	−0.1(2)
6.6	PG $(1/T_{\mathrm{span}})$	1.2(1)	0.16(2)	−0.1(2)
6.6	PG (1/100yr)	1.1(1)	0.27(3)	−0.1(2)

3.1 Results from simulations

For each choice of exponent, we run 40 realizations of the simulation, and fit for the noise hyperparameters and F2, marginalizing over F0, F1, and the position of the pulsar. The initial parameters for the fitting were slightly perturbed from the simulated values. For a parameter p, in this case F2, we have a truth value p_{truth}, the best-fitting value, $\hat{p}_i$, and error, $e_{p,i}$, of the ith realization of the simulation. We compute the average deviation from the injected value,

$$\Delta_p = \frac{1}{N}\sum_{i=0}^{N}(\hat{p}_i - p_{\mathrm{truth}}),\qquad(7)$$

and σ_p^2 which is the variance of the $\hat{p}_i$ values. Note the distinction between $e_{p,i}$, the formal error returned by the fit algorithm, and σ_p which is derived from the observed variation between realizations of the simulation. Using these, we can derive metrics for the quality of the results obtained. First, we take $\sigma_p/\sigma_{p,\mathrm{min}}$, rms in the results normalized to the model with the lowest rms. This is the factor by which the actual scatter in the parameter estimates increases over the 'best' model. Next, we consider how well the fit errors model the scatter. Since the values of $e_{p,i}$ are consistent from realization to realization, we simply consider the mean error estimate $\bar{e}_p$. If the errors are well estimated, $\bar{e}_p/\sigma_p$, will be unity. A greater value indicates errors are typically overestimated, and a smaller value indicates errors are typically underestimated. Finally we consider an estimate of the normalized bias, Δ_p/σ_p, which should be consistent with zero. The results for fitting F2 in these models are given in Table 2.

For $\gamma = 3.6$, the PF models and the PG model with $f_c = 1/100$ yr all return reasonable values. The PG model with $f_c = 1/T_{\mathrm{span}}$ tends to underestimate the uncertainties, as might be expected since in this model the PSD starts to turn over well before the frequency range influenced by the F2 fit. As the spectral exponent increases, the PF model with lowest frequency at $1/T_{\mathrm{span}}$ and the PG model with $f_c = 1/T_{\mathrm{span}}$ begin to significantly underestimate the uncertainties. The PG model with $f_c = 1/100$ yr also starts to underestimate the uncertainty as the red noise becomes steeper. The PF model with lowest frequency at $1/2T_{\mathrm{span}}$ fairs better, consistently returning reliable uncertainties at all tested values of γ.

However, it is worth noting that we do not know what happens to the red-noise PSD at frequencies below our observing span, and with our longest data sets at ∼50 yr, we have no measurements for frequencies close to 1/100 yr. We should therefore always keep in mind that the measurements of F2 (and F1 and F0) depend on our choice of what happens to the red noise at frequencies below $1/T_{\mathrm{span}}$, and hence we may under (or over) estimate the significance of F2 even with a very small choice of f_c or f_{low}. Nevertheless we believe these simulations support our assertion that using the PF model with lowest frequency at $1/2T_{\mathrm{span}}$ significantly improves the uncertainty estimate on F2 under the influence of steep red noise. Using the PG model avoids any problems with the periodic boundary conditions, but the uncertainty in F2 depends strongly on the choice of f_c. Given the additional computational cost of the PG model, especially for small values of f_c, we therefore would recommend using the PF model (or QF model if appropriate) with lowest frequency at $1/2T_{\mathrm{span}}$ when studying the long-term evolution of pulsars.

4 QUASI-PERIODIC SPIN-DOWN

4.1 Simulations

In order to demonstrate the limitations of using power-law models for pulsars which exhibit quasi-periodic spin-down variations, we construct another simple simulation. The simulated pulsar switches between two spin-down states corresponding to a change of $\dot{\nu}$ by 1 per cent, with each 'high' state lasting for 162 ± 10 d and each 'low' state lasting for 440 ± 10 d. We again generate 224 ToAs over 22 yr, and based on actual Lovell telescope observations of PSR J1909+0912. This represents a highly, but not perfectly, periodic variation in the pulsar spin, similar to a pulsar such as PSR B1828−11 or B1540−06. In order to test the effect of the mis-match between model and data we consider two outputs – first, the estimates of the power-spectral density of the data, and secondly the estimates of the pulsar parameters.

We use the CHOLSPECTRA plugin of TEMPO2 to estimate the one-sided power-spectral density based on the maximum likelihood models from each realization of the simulation. The stochastic $\dot{\nu}$ switching model does not lend itself to an analytical model of the PSD, but we can estimate the expected PSD under ideal conditions by simulating a large number of uniformly sampled realizations of the injected signal. To avoid the spectral leakage prevalent when taking a periodogram of processes that fall faster than f^{-2}, we take the discrete Fourier transform of 2^{14} uniformly spaced samples of each realization of the injected $\dot{\nu}$ time-series, and multiply the resulting spectrum by f^{-4}. This should be equivalent to integrating twice in the time domain and hence recover the model spectrum for the residuals, though we caution that aliasing may lead to excess high-frequency noise that is then amplified if the spectrum does not decay rapidly enough at high frequencies.

Fig. 2 shows the average of 2^{10} realizations of the PSD estimated in this way, overlaid on the maximum likelihood model and an average of the PSD from CHOLSPECTRA. As might be expected, for such strongly periodic signals, the pure power-law models primarily capture the peaks in the spectra, and hence significantly overestimate the PSD at frequencies away from the harmonics of the quasi-periodic signal. In panel (b) of Fig. 2, we show the average PSD from the 20 realizations of the simulation. Here we see that on average the use of power-law models leads to a substantial overestimation of the PSD, especially at low frequencies where it can be out by more than 2 orders of magnitude. All models also overestimate the PSD at high frequencies, flattening well before reaching the injected white noise
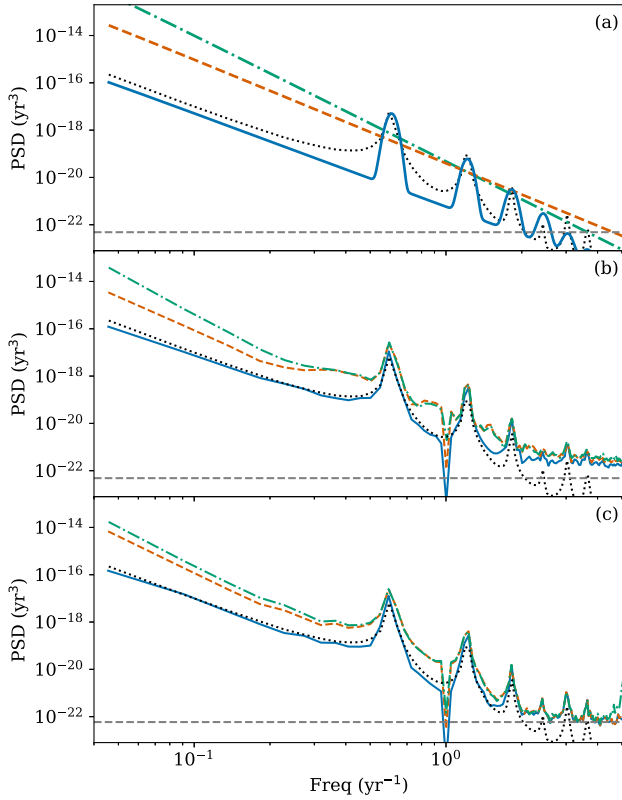
Figure 2. Power spectra derived from the simulations in Section 4.1. (a) The maximum-likelihood model power-spectrum for a single realization of ToAs from the simulation of quasi-periodic $\dot{\nu}$ variations on real pulsar sampling. (b) The mean value of the PSD estimated from 20 realizations of the simulation with irregular sampling. (c) The mean value of the PSD estimated from 20 realizations of the same simulation but with regular sampling. In each panel, the solid lines are for the QG, dotted lines are for the PG model and dotted–dashed lines are for the PF model. The QF model is not shown as it almost exactly overlaps with the QG model. The dotted spectrum is the average of the 1024 realizations of the simulation computed directly using the DFT on the generated $\dot{\nu}$ sequence. The dashed horizontal line is the mean PSD of the injected white noise only.

Table 3. Statistics of parameter estimates from 20 realizations of simulated observations with $\dot{\nu}$-switching. Columns are as in Table 2 and defined in Section 3.1.

Parameter	Model	$\sigma_p/\sigma_{p,\,\mathrm{min}}$	$\overline{e}_p/\sigma_p$	Δ_p/σ_p
F2	PF	5.0(5)	13(1)	0.3(2)
F2	PG	4.0(4)	3.4(4)	0.3(2)
F2	QF	1.0(1)	1.2(1)	0.2(2)
F2	QG	1.0(1)	1.1(1)	0.2(2)
PMRA	PF	3.6(4)	1.4(1)	0.3(2)
PMRA	PG	3.6(4)	1.3(1)	0.2(2)
PMRA	QF	1.0(1)	1.1(1)	0.1(2)
PMRA	QG	1.0(1)	1.1(1)	0.1(2)
PMDEC	PF	4.3(5)	1.0(1)	0.0(2)
PMDEC	PG	3.6(4)	1.1(1)	0.0(2)
PMDEC	QF	1.0(1)	0.9(1)	−0.1(2)
PMDEC	QG	1.0(1)	0.9(1)	−0.1(2)

In this simulation the PSD close to the periodicity is approximately 10^3 times the power-law component; however, it is important to note that this is only discernible when the PSD is formed using the quasi-periodic model. We also note that the simulation does not contain a separate power-law component, but rather the quasi-periodic $\dot{\nu}$ variations intrinsically cause a power-law red noise, at least below the fundamental frequency of the quasi-periodic process.

For parameter estimation, we focus on three parameters: the long-term frequency second derivative, F2, and the two proper-motion parameters PMRA and PMDEC. These parameters have interest to the pulsar astronomer, and have been shown they can be biased if incorrectly dealing with red noise (Coles et al. 2011). The results of fitting these parameters in 20 simulations are summarized in Table 3, using the same metrics as in Section 3.1. In all cases the quasi-periodic models are both a factor of 3–5 more precise, and have better error estimates. In practice the real pulsars have a wide range of observed periodicities, and the benefit of using a quasi-periodic model will be highly situational, depending on both the periodicity, amplitude, and purity of the observed quasi-periodic process.

4.2 Application to real data

Although the simulations can demonstrate that the quasi-periodic model can improve results for a particular case, it is perhaps more informative to see the effect when applied to a sample of pulsars with known quasi-periodic variations. Therefore we apply the quasi-periodic model fitting to the sample of pulsars from Lyne et al. (2010), which exhibit strong quasi-period variations in their spin-down rate. These pulsars have recently been revisited by Shaw et al. (2022; hereafter S22) using more recent data, and this provides an ideal data set with which we can test these models and investigate any effect on the pulsar parameters. We refer readers to S22 for details of the observations and preparation, but in brief the data set consists of observations made with the 76-m Lovell telescope at Jodrell Bank, supplemented with observations from the 25-m 'Mark-II' telescope also at Jodrell Bank. Most data prior to 2009 were centred at 1400 MHz and recorded using a 32-MHz analogue filterbank. Since 2009 most data are centred at 1520-MHz using a 384-MHz digital filterbank. These are supplemented with a small number of observations at 400, 610, and 925 MHz. For the digital filterbank data, radio frequency interference has been excised using a combination of median filtering and manual removal of affected channels or time intervals. ToAs are generated by cross correlation with a noise-free template using PSRCHIVE. A phase coherent timing solution is

level, giving the appearance of excess white noise. We attribute this high frequency excess to spectral leakage due to a combination of the irregular sampling of the data and the fact that the model does not perfectly capture the underlying PSD. This effectively limits the dynamic range available in the power-spectrum with irregular sampling. To test this hypothesis, we repeated the simulation with the same number of ToAs uniformly sampled across the same time-span, with the same white noise and quasi-periodic $\dot{\nu}$ switching. The average PSD from the 20 simulations with uniform sampling is shown in Fig. 2(c), and we can see that the estimated PSD correctly approaches the white noise level, but the low-frequency behaviour is largely unchanged.

This apparent excess white noise in irregular sampling may also be seen in real pulsar observations due to similar mismatches between the model and underlying noise process, though it is not clear how to distinguish this from excess white noise intrinsic to the pulsar. We note that in the case of irregular sampling, the spectral estimates at the high frequencies are highly correlated – indeed this must be the case since the effective Nyquist frequency is much higher than for the regularly sampled case yet the number of input data points has not increased.

Table 4. Results of fitting the QF quasi-periodic model to a sample of pulsars. $\ln(Z_{\mathrm{QF}}/Z_{\mathrm{PF}})$ is the log-Bayes factor in favour of a the quasi-periodic model. Columns 3–8 give model parameters are as defined in equations (2) and (3). $1/f_{\dot{\nu}}$ is the period associated with the peak of the $\dot{\nu}$ periodogram in S22 for comparison. Values in parenthesis indicate the uncertainty in the last given digit. Pulsars are ordered by the evidence for the quasi-periodic model.

PSR	$\ln\left(\dfrac{Z_{\mathrm{QF}}}{Z_{\mathrm{PF}}}\right)$	$\log_{10}\left(\dfrac{f_{\mathrm{qp}}}{\mathrm{yr}^{-1}}\right)$	$\log_{10}(R_{\mathrm{qp}})$	λ	σ	$\log_{10}(A_{\mathrm{pl}})$	γ	$1/f_{\mathrm{qp}}$ (d)	$1/f_{\dot{\nu}}$ (d)	$f_{\dot{\nu}}/f_{\mathrm{qp}}$
B1828−11	147.5	−0.111(2)	2.7(2)	0.7(1)	0.047(4)	−9.41(3)	4.3(1)	471(3)	490(50)	0.97(9)
B0740−28	25.5	0.42(1)	1.3(2)	3(2)	0.18(1)	−9.37(4)	4.9(2)	138(3)	130(5)	1.06(5)
B1540−06	23.8	−0.60(1)	2.7(3)	0.5(1)	0.04(2)	−11.2(3)	5.1(7)	1470(30)	1461(6)	1.01(2)
B1642−03	18.7	−0.82(5)	2.6(3)	1.6(3)	0.17(2)	−10.43(9)	4.0(3)	2400(300)	2000(600)	1.2(4)
B1826−17	16.7	−0.50(2)	1.9(3)	2.0(5)	0.14(3)	−10.3(3)	4.4(3)	1140(50)	1094(5)	1.05(5)
B1822−09	11.2	−0.41(4)	1.3(3)	1.6(5)	0.15(2)	−9.4(2)	5.5(4)	1000(100)	8900(200)	0.11(1)
B1929+20	10.1	−0.22(7)	1.1(2)	0.3(4)	0.13(4)	−10.3(1)	4.7(2)	600(200)	604(3)	1.0(3)
B0919+06	8.8	−0.19(2)	0.9(2)	2(1)	0.12(4)	−9.65(8)	4.3(2)	570(50)	600(300)	0.9(5)
B1818−04	6.5	−0.97(2)	1.7(3)	0.7(2)	0.06(4)	−10.27(4)	4.9(2)	3400(100)	4000(1000)	0.9(3)
B1714−34	6.4	−0.60(2)	1.4(3)	0.5(4)	0.05(3)	−10.0(1)	5.2(4)	1460(60)	1400(10)	1.04(4)
B2148+63	1.6	−0.43(9)	1.8(8)	5(3)	0.18(3)	−11.6(5)	4.8(7)	1000(200)	1300(200)	0.8(2)
B2035+36	1.1	−0.2(2)	0.5(6)	5(3)	0.13(4)	−9.5(1)	5.9(3)	700(300)	13200(200)	0.05(3)
B1839+09	0.8	−0.2(3)	0.3(9)	5(3)	0.11(6)	−10.0(1)	4.1(3)	800(800)	320(3)	2(2)
B1907+00	0.0	−0.7(3)	0(1)	3(3)	0.10(6)	−10.4(1)	5.1(3)	2100(700)	5900(100)	0.4(1)
B0950+08	−0.5	−0.8(3)	0(1)	2(3)	0.12(6)	−11.3(1)	5.5(3)	3000(1000)	5000(1000)	0.7(3)
B1903+07	−1.0	−0.5(3)	−0.6(8)	5(3)	0.10(6)	−9.06(7)	4.6(2)	1500(800)	1800(900)	0.8(6)
J2043+2740	−1.2	−0.2(4)	−0.9(7)	5(3)	0.10(5)	−9.25(5)	5.7(2)	800(600)	4010(900)	0.2(1)

obtained using TEMPO2, making use of the pulse numbering feature to track the rotation of the pulsar over the entire dataset. Because of the inhomogeneity of backend instruments, we fit for separate white noise parameters (EFAC and EQUAD) for the legacy data, analogue filterbank data and digital filterbank data.

We fit each of the models to the 17 pulsars in the data set of S22. The parameter estimates for the QF quasi-periodic noise model are given in Table 4. Results for the QG model are extremely similar and so are not repeated. Ten of the pulsars show significant preference for the quasi-periodic model, with a log-Bayes factor, $\ln Z_{\mathrm{QF}} - \ln Z_{\mathrm{PF}}$, greater than 6. Of these, nine find a periodicity that matches with the peak of the periodogram of $\dot{\nu}$ computed by S22. The remaining pulsar is PSR B1822−09, for which we find a much shorter period of 1000 d compared to a peak of 9000 d in S22. We attribute this difference to the large exponent of the power spectrum for the red noise in this pulsar. The $\dot{\nu}$ time series used in S22 for the periodogram is the second derivative of the pulsar residual. This process of taking derivatives will whiten the red noise (Coles et al. 2011), and hence reveal the quasi-periodic oscillations. However in the case where γ is significantly larger than 4 there will continue to be residual red noise in the $\dot{\nu}$ timeseries, which likely leads to the identification of a periodicity consistent with the data set length in S22 for PSR B1822−09.

Fig. 3 shows the PSD estimated from the residuals using the QG model as well as the maximum likelihood model PSD for each of the four models, for each of the pulsars with significant evidence in favour of a quasi-periodic model. The remaining pulsars are shown in Fig. 4. For the pulsars in Fig. 3, the quasi-periodic models are able to capture the shape of the estimated PSD more closely than the pure power-law models, with the power-law models consistently overestimating the power at the lowest frequencies. The PF model typically has a greater overestimation factor, in the most extreme case of B1828−11, overestimating the PSD at $1/T_{\mathrm{span}}$ by nearly 4 orders of magnitude. The post-fit F2, PMRA, and PMDEC parameters for the 10 pulsars in Fig. 3 are given in Table 5. As the main difference in the PSD models is at the lowest frequencies, the main impact is on the estimation of F2. In the case of PSR B1828−11 using the QF model results in a $>10\sigma$ measurement of F2, which is

otherwise not significant under the PF model. As we might expect, the proper motion measurements are most affected for pulsars with f_{qp} closest to 1 per year. For example, PSR B1828−11 ($f_{\mathrm{qp}} \approx 0.78\,\mathrm{yr}^{-1}$) sees a reduced uncertainty in proper motion parameters by a factor of around 4, whilst PSR B0740−28 ($f_{\mathrm{qp}} \approx 2.8\,\mathrm{yr}^{-1}$) reduces the uncertainty in proper motion by a factor of 2. These two pulsars also have the highest evidence in favour of the QF model. PSR B1929+20 ($f_{\mathrm{qp}} \approx 0.5\,\mathrm{yr}^{-1}$) sees a reduction in uncertainty of proper motion by a factor of 1.4. The remaining pulsars show little change in the value or uncertainty for proper motion, but all have a much longer periodicity or do not have strong evidence in favour of the quasi-periodic model (hence the quasi-periodic variations do not dominate the timing noise).

4.3 Spin-down variations

The prevailing model for the quasi-periodic variations is that the spin-down rate of the pulsar varies in a quasi-periodic manner, hence it is useful to visualize the $\dot{\nu}(t)$ time-series from our quasi-periodic model. The Fourier-domain Gaussian process model can be used to estimate $\dot{\nu}(t)$ by taking the second derivative of the Gaussian process model. This can be computed analytically once we have the Gaussian process in the form

$$r(t) = \sum_{i}^{N_{\mathrm{coef}}} A_i \cos \omega_i t + B_i \sin \omega_i t,$$

where A_i, B_i are the best-fitting amplitudes of the Fourier-domain Gaussian process components from TEMPO2 and N_{coef} is the number of Fourier coefficients. The spin-down rate can therefore be computed analytically, since by our definition of residual we have

$$\nu(t) = \mathrm{F0} + \mathrm{F1}t + \tfrac{1}{2}\mathrm{F2}t^2 - \mathrm{F0}\,\dot{r}(t),$$

and hence,

$$\dot{\nu}(t) = \mathrm{F1} + \mathrm{F2}t - \mathrm{F0}\,\ddot{r}(t)$$

$$= \mathrm{F1} + \mathrm{F2}t + \mathrm{F0} \sum_{i}^{N_{\mathrm{coef}}} A_i \omega_i^2 \cos \omega_i t + B_i \omega_i^2 \sin \omega_i t. \tag{8}$$
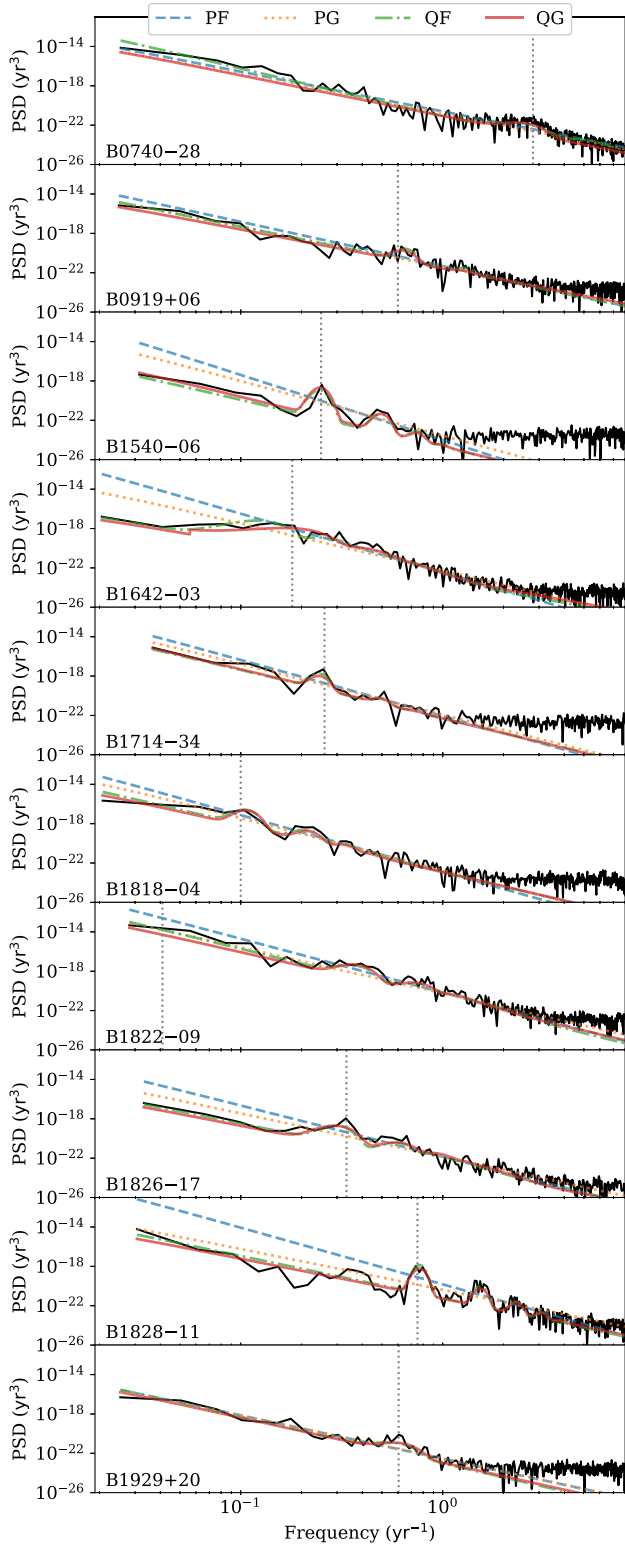
Figure 3. PSD of 10 pulsars estimated using the QG model overplotted with the model PSD from each of the four models. The dashed vertical line marks f_{qp}.
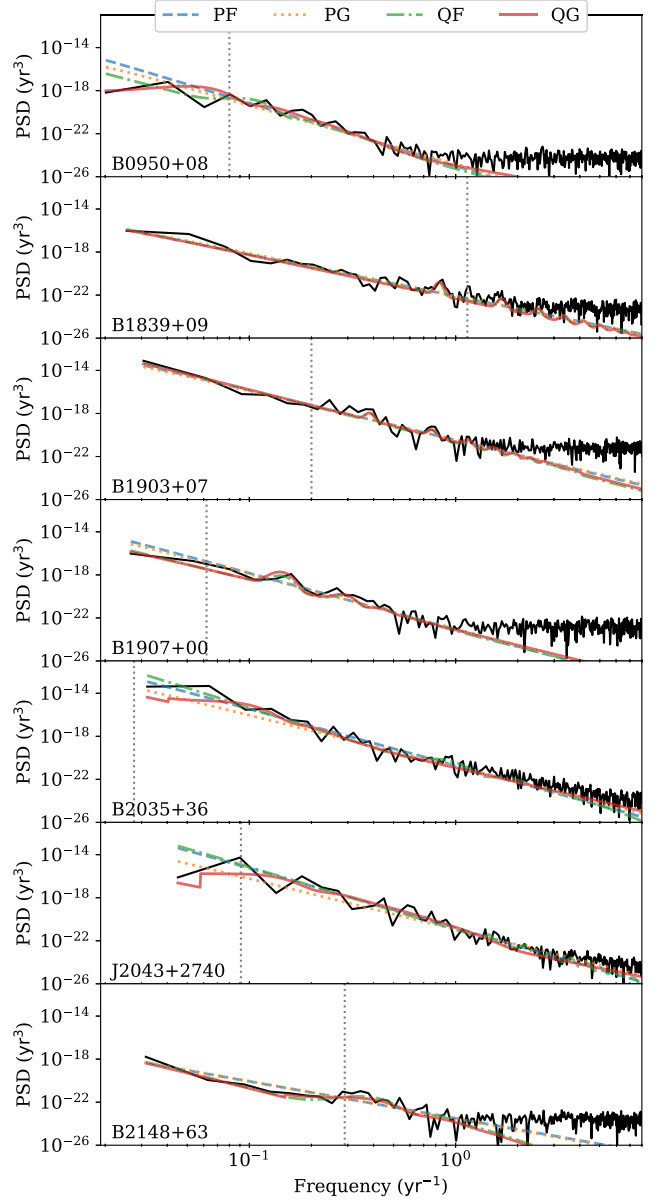

Figure 4. PSD of 7 pulsars without significant evidence for quasi-periodic variability, estimated using the QG model overploted with the model PSD from each of the four models.

The uncertainty on $\dot{\nu}(t)$ can be estimated from the parameter covariance matrix, and the ToA uncertainties as outlined in Appendix A.

Fig. 5 shows the $\dot{\nu}(t)$ derived directly from the QF model in this work, compared against the results presented in S22 which are generated from a time-domain Gaussian process applied to the post-fit residuals. There is generally a lot of similarity between the two methods for estimating $\dot{\nu}(t)$. Some differences appear around glitches, e.g. in PSR B0740−28, B0919+06 and B1828−11. The estimation of glitch parameters, particularly the change in $\dot{\nu}$ is difficult in the presence of quasi-periodic variations in $\dot{\nu}$ and hence the values obtained are sensitive to the choice of noise model used. We claim that the $\dot{\nu}$ timeseries around the glitch in the two pulsars with the highest evidence for the QF model, PSR B0740−28 and

Table 5. Post-fit timing parameters for a sample of pulsars when using the PF and QF models. The Bayes factor in favour of the QF model is also included comparison.

PSR	$\ln\left(\frac{Z_{QF}}{Z_{PF}}\right)$	F2 (10^{-25} Hz^3)		PMRA (mas yr^{-1})		PMDEC (mas yr^{-1})	
		PF	QF	PF	QF	PF	QF
B1828−11	147.5	12 ± 21	11.3 ± 0.8	-12 ± 40	2 ± 10	10 ± 178	2 ± 46
B0740−28	25.5	8.3 ± 3.2	10 ± 6	-28 ± 15	-26 ± 8	-16 ± 20	-12 ± 10
B1540−06	23.8	-1.2 ± 2.4	0.04 ± 0.07	-17.7 ± 0.9	-17.4 ± 0.8	-6.0 ± 2.8	-4.2 ± 2.4
B1642−03	18.7	0.3 ± 1.2	-0.01 ± 0.05	-2.6 ± 1.3	-2.8 ± 1.2	16 ± 4	16 ± 4
B1826−17	16.7	0 ± 4	0.45 ± 0.22	10 ± 5	10 ± 6	58 ± 49	58 ± 57
B1822−09	11.2	1 ± 14	5 ± 5	16 ± 28	18 ± 29	-100 ± 110	-90 ± 120
B1929+20	10.1	-0.45 ± 0.33	-0.48 ± 0.25	1.5 ± 2.0	2.1 ± 1.4	0.0 ± 3.1	0.4 ± 2.2
B0919+06	8.8	0.8 ± 1.0	1.4 ± 0.5	8 ± 14	4 ± 12	70 ± 40	60 ± 32
B1818−04	6.5	-0.2 ± 1.1	0.09 ± 0.18	-7.9 ± 1.0	-8.0 ± 0.9	15.5 ± 2.9	15.6 ± 2.9
B1714−34	6.4	-1 ± 4	0.1 ± 0.9	2 ± 6	2 ± 5	30 ± 28	27 ± 25

B1828−11, look qualitatively more plausible for our QF model than that from S22, however, PSR B0919+06 does not seem well modelled around the glitch for either case, perhaps reflecting other unmodeled transient glitch effects in this pulsar.

PSR B0950+08 shows a long-term deviation between our model and S22, a pulsar for which there is no evidence in favour of the QF model over the PF model. The deviation from S22 is because the QF model behaves like a power-law with a low frequency turn-over for this pulsar, and hence deviates from the power-law significantly at the lowest frequencies (see Fig. 4). This causes a change in the estimation of F2 in this pulsar leading to a linear deviation in the estimated $\dot{\nu}(t)$, however, neither measurement is significant. The maximum likelihood solution for F2 changes from $(1.4 \pm 1.1) \times 10^{-26}$ Hz^3 with the PF model to $(-0.4 \pm 1) \times 10^{-26}$ Hz^3 with the QF model. It is worth noting that this pulsar has variation time-scale of several thousand days, and so only about two cycles are seen within our observing window. It is possible that with a significantly longer observing span the evidence in favour of the quasi-periodic model would be more significant.

PSR B1822−09 and B2035+36 show rapid variations in the $\dot{\nu}(t)$ derived from our model compared to that in S22. For these two pulsars, the $\dot{\nu}$ timeseries do not appear to be well modelled by a Gaussian process, with evidence for non-stationarity in the statistics. In these cases, it appears the time-domain Gaussian process of S22 performs better than fitting directly to the residuals with the Fourier basis Gaussian process, though likely neither method really is optimal for this type of pulsar behaviour.

PSR B1818−04 shows high frequency oscillations in S22 that are not present in our model. A periodicity analysis shows that these oscillations are consistent with a period of 1 yr, which suggests an error in the position or proper motion in the timing model of S22. Although S22 used a Gaussian process model of $\dot{\nu}(t)$, the timing model was solved in a 'traditional' way, and is therefore susceptible to leakage from the spin noise into the pulsar parameters as described in Coles et al. (2011). This highlights the benefits of the noise modelling procedure for determining pulsar parameters and studying the rotation. A similar feature also appears in B1714−34, but to a lesser extent.

5 DISCUSSION

As may be naively expected, the results of this work show that the parameter estimates are generally improved when the noise model more closely matches the underlying noise process. However, the widely used power-law Fourier domain Gaussian process models are seen to be very robust and it takes fairly significant deviations to cause issues. Particularly the quasi-periodic model is only significantly advantageous in a small number of pulsars, even when we study a sample of pulsars known to exhibit significant quasi-periodic variations. Hence, we expect that it is not necessary to apply a quasi-periodic model to most pulsars, and it will be clear from inspection of the $\dot{\nu}$ time-series or power-spectra computed from a power-law model if the quasi-periodic model is likely to be of any benefit to the study of a given pulsar.

Based on the simulations in Section 3, we see that the Fourier basis model with low-frequency cut-off at $1/T_{\text{span}}$ is not suitable for estimating F2 for pulsars with steep timing noise. Indeed many pulsars do show timing noise with a power-law exponent at or above $\gamma = 4$ (e.g. Parthasarathy et al. 2019), hence some mitigation is needed for this. Here, we propose that a simple adjustment of the low-frequency cut-off to $1/2T_{\text{span}}$ significantly improves the results at large values of γ, though this does increase computation as twice as many frequency coefficients are needed to reach the same value of f_{high}.

Although this work focuses on the estimation of F2 and other pulsar parameters for the study of pulsars, this may also have implications for the ongoing gravitational wave pulsar timing array experiments that are searching for a steep spectrum signal in the pulsar data. Here, we have shown that quasi-periodic variations in the pulsar spin down can lead to a spurious overestimation of the modelled power at low frequencies. Although the actual spectral estimates are not typically overestimated, except perhaps in the most extreme cases, this would falsely increase the uncertainty from pulsar spin noise and potentially the total power estimated in the gravitational wave background (GWB). Of course, the pulsars forming the pulsar timing array do not typically show evidence for quasi-periodic oscillations in spin down, so this effect is unlikely to be significant, but it does demonstrate that the choice of model is important. Perhaps more importantly, the choice of how the model should behave at frequencies below $1/T_{\text{span}}$ clearly does have significant impact on the parameter estimates and it is unlikely that the pulsar timing noise really does cease or turn-over at frequencies around $1/T_{\text{span}}$. This is especially important for pulsars exhibiting steep red noise. We can use the estimation of F2 as a proxy for our ability to model the lowest frequencies in the data, since an error in the F2 estimate is caused by unmodelled low frequency red noise. We find that the critical power-law exponent for reliable measurements of F2 with modeled $f_{\text{low}} = 1/T_{\text{span}}$ is around $\gamma = 4$, which is also
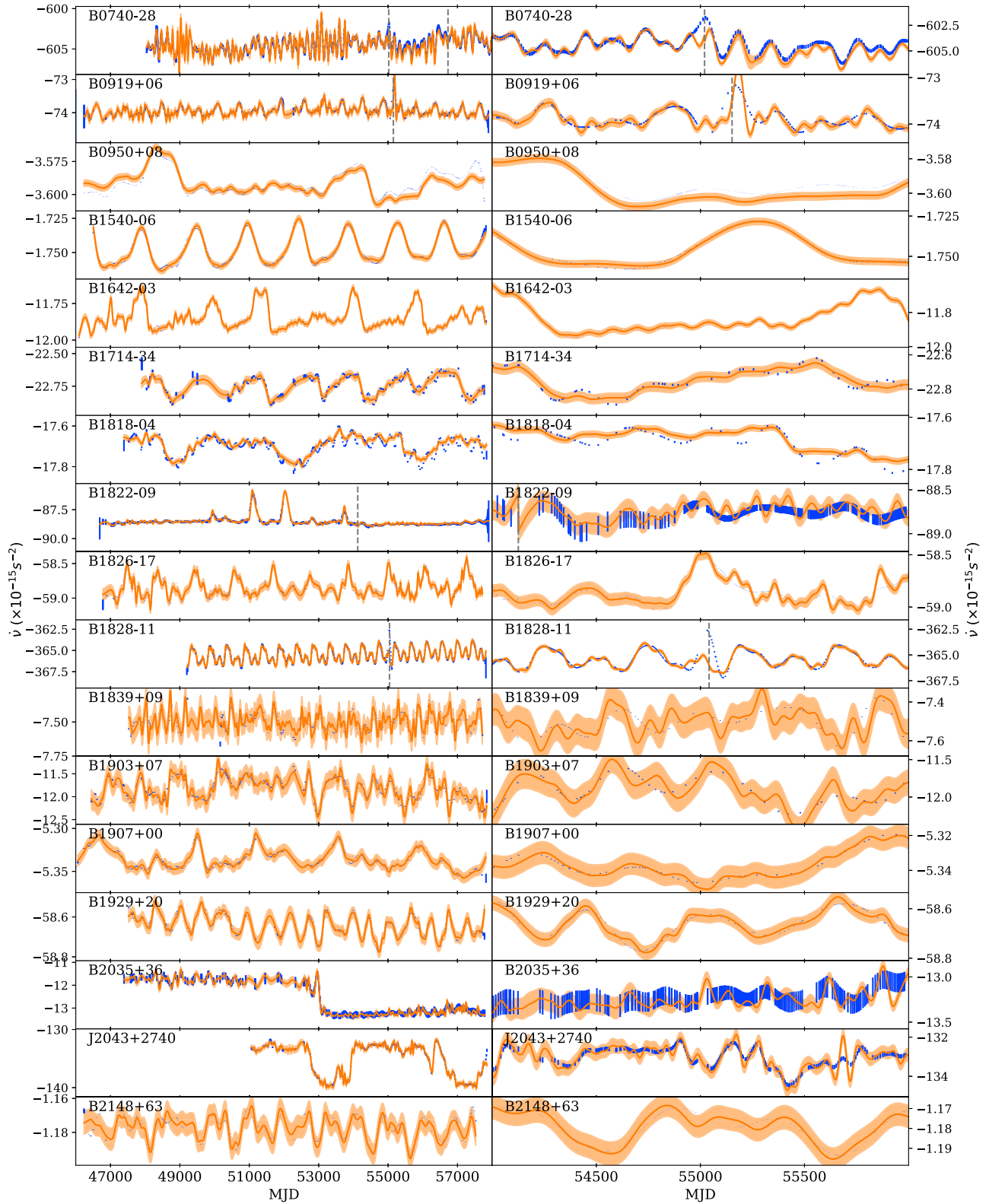
Figure 5. Frequency derivative of the 17 pulsars from S22. Points with errors are those presented in S22 and the smooth curve is that derived from the QF model in this work, with the filled band representing the 1σ uncertainty. Vertical dashed lines mark the epoch of glitches. The left column shows the full range of the data and the right column shows a zoom between MJD 54 000–56 000 to highlight some of the details of the time variation.

the exponent that would be produced by random variations in $\dot{\nu}$ as well as that expected from the GWB. This is consistent with the break in the uncertainty of measuring F2 against red noise power-law exponent observed by Liu et al. (2019). Therefore we argue that the choice of f_{low} can be important for the gravitational wave experiment, particularly as these analyses typically do not include F2 in the fitting. This means that any unmodelled power in the pulsar spin noise may be falsely attributed to another source, perhaps manifesting in the requirement of additional model terms, such as the recently observed common uncorrelated red noise process (Antoniadis et al. 2022). A full exploration of the impact of oversampling the spectrum for GWB searches is outside the scope of this paper, but there may be other benefits to oversampling the Fourier space by capturing second order statistics in the data, and we suggest further exploration of these effects should be included in confirming the robustness of a future GWB detection.

6 CONCLUSIONS

We set out to answer two questions regarding the choice of noise model for pulsar timing. For the case of the choice of low-frequency cut-off and the periodic boundary conditions, we found that below about $\gamma = 4$ all the models perform well even with the lowest frequency set to $1/T_{\text{span}}$. However, many pulsars show significantly steeper red processes, and hence the power must either be absorbed into F2 or another mitigation must be applied if F2 is a parameter of interest. Here, we propose a simple solution of setting the lowest frequency as $1/2T_{\text{span}}$, which performs reliably up to about $\gamma = 6$, beyond which the model again struggles.

For quasi-periodic timing variations, we find that a small number of pulsars are better modelled with our new noise model that includes a quasi-periodic term. However, the power-law models generally do well even without this quasi-periodic term for many of the pulsars, and it is only for the pulsars with the strongest quasi-periodic variations that there is a significant issue with measuring F2 with the power-law model. We find proper motion measurements can be affected where the quasi-periodic fluctuations have a period of the order of a year. We believe it is unlikely that quasi-periodic models will significantly improve pulsar timing if the quasi-periodic signal is not already clearly apparent in the power spectrum of the residuals.

We also confirm previous results that in general, the Fourier-basis Gaussian process performs largely as well as the equivalent GLS model. The new models and implementation of the GLS method are all made available for testing and real-world use through RUN_ENTERPRISE, which will allow for use, or further testing, on a wide range of data.

ACKNOWLEDGEMENTS

DATA AVAILABILITY

The raw data in this work are as used in Shaw et al. (2022), and requests to use these data should be made to the original authors of that work. The $\dot{\nu}$ timeseries from this work (i.e. those shown in Fig. 5) are available at doi:10.5281/zenodo.7664166.

REFERENCES

Antoniadis J. et al., 2022, MNRAS, 510, 4873
Arzoumanian Z. et al., 2020, ApJ, 905, L34
Buchner J. et al., 2014, A&A, 564, A125
Chen S. et al., 2021, MNRAS, 508, 4970
Coles W., Hobbs G., Champion D. J., Manchester R. N., Verbiest J. P. W., 2011, MNRAS, 418, 561
Cordes J. M., Downs G. S., 1985, ApJS, 59, 343
Ellis J. A., Vallisneri M., Taylor S. R., Baker P. T., 2019, Astrophysics Source Code Library, record ascl:1912.015
Espinoza C. M., Lyne A. G., Kramer M., Manchester R. N., Kaspi V. M., 2011, ApJ, 741, L13
Feroz F., Hobson M. P., 2008, MNRAS, 384, 449
Hobbs G., Lyne A. G., Kramer M., 2010, MNRAS, 402, 1027
Keith M. J., Niţu I. C., Liu Y., 2022, run_enterprise
Kerr M., Hobbs G., Johnston S., Shannon R. M., 2016, MNRAS, 455, 1845
Kerr M. et al., 2020, PASA, 37, e020
Lentati L., Alexander P., Hobson M. P., Taylor S., Gair J., Balan S. T., van Haasteren R., 2013, Phys. Rev. D, 87, 104021
Lentati L., Alexander P., Hobson M. P., Feroz F., van Haasteren R., Lee K. J., Shannon R. M., 2014, MNRAS, 437, 3004
Liu X. J., Keith M. J., Bassa C. G., Stappers B. W., 2019, MNRAS, 488, 2190
Lower M. E. et al., 2021, MNRAS, 508, 3251
Lyne A., Hobbs G., Kramer M., Stairs I., Stappers B., 2010, Science, 329, 408
Lyne A. G., Jordan C. A., Graham-Smith F., Espinoza C. M., Stappers B. W., Weltevrede P., 2015, MNRAS, 446, 857
Miles M. T. et al., 2023, MNRAS, 519, 3976
Niţu I. C., Keith M. J., Stappers B. W., Lyne A. G., Mickaliger M. B., 2022, MNRAS, 512, 2446
Osłowski S., van Straten W., Hobbs G. B., Bailes M., Demorest P., 2011, MNRAS, 418, 1258
Parthasarathy A. et al., 2019, MNRAS, 489, 3810
Parthasarathy A. et al., 2021, MNRAS, 502, 407
Reardon D. J. et al., 2016, MNRAS, 455, 1751
Shaw B. et al., 2022, MNRAS, 513, 5861 (S22)
Tarafdar P. et al., 2022, PASA, 39, e053
van Haasteren R., Levin Y., 2013, MNRAS, 428, 1147

APPENDIX: IMPLEMENTATION OF THE GAUSSIAN PROCESS MODEL IN TEMPO2 AND ESTIMATION OF THE UNCERTAINTIES

The TEMPO2 software performs the final step of evaluating the timing model and is used in this work to extract the Gaussian process time-series and the spin-down timeseries. TEMPO2 uses generalized least squares fitting, though for the Gaussian process model the data covariance matrix, $\mathbf{N}$, is assumed to be diagonal and contain the white noise uncertainties for each data point on the diagonal. A design matrix $\mathbf{M}$ maps the parameters $\boldsymbol{\beta}$ to a model of the residuals, $\boldsymbol{r}$ observed at at times $\boldsymbol{t}$. Following the standard least-squares methodology, we could compute an estimate of the parameters from the residuals,

$$\hat{\boldsymbol{\beta}} = (\mathbf{M}^{\mathsf{T}}\mathbf{N}^{-1}\mathbf{M})^{-1}\mathbf{M}\mathbf{N}^{-1}\boldsymbol{r}.$$

In practice this is performed using either the singular-value decomposition (SVD), or the QR decomposition to avoid the numerical precision issues with direct computation. Similarly, we can compute the covariance matrix of the parameter estimates,

$$\mathbf{C}_{\beta} = (\mathbf{M}^{\mathsf{T}}\mathbf{N}^{-1}\mathbf{M})^{-1},$$

which again is in practice computed via the SVD or QR decomposition.

To fit the Fourier domain Gaussian process with n coefficients to m data points, $\mathbf{M}$ will contain m rows and $2n$ columns with the form

$$M_{j,i} = \begin{cases} \sin\left(2\pi f_i t_j\right) \text{for } 0 < i < n, \\ \cos\left(2\pi f_{i-n} t_j\right) \text{for } n < i < 2n. \end{cases}$$

In order to implement the Gaussian priors on these parameters, we add $2n$ constraints to the least-squares problem. This is achieved by extending $\boldsymbol{r}$ with $2n$ zero valued elements, and extending $\mathbf{M}$ with $2n$ block diagonal elements that contain the Gaussian likelihood

$$M_{i+m,i+2n} = \begin{cases} P(f_i)^{-1/2} \, \delta f \text{ for } 0 < i < n, \\ P(f_{i-n})^{-1/2} \, \delta f \text{ for } n < i < 2n, \end{cases}$$

where $P(f_i)$ is the model PSD at frequency f_i and δf is the frequency resolution of the Fourier components, which scales from PSD to power. This constrained problem can then be solved using the least squares formalism. In practice, $\mathbf{M}$ will also contain columns associated with the pulsar timing model (e.g. the standard spin and astrometric parameters); however, for the remainder of this discussion it is assumed that we will extract only the relevant rows and columns related to the red noise model. The linear algebra from this point forward is implemented in the MAKE_PULSAR_PLOTS.PY script provided with RUN_ENTERPRISE.

Estimates of the $\dot{\nu}(t)$ at each epoch can be made by computing $\mathbf{D}$, the second time derivative of $\mathbf{M}$ scaled by F0 (cf. equation 8),

$$D_{j,i} = \begin{cases} 4\pi^2 f_i^2 \text{F0} \sin\left(2\pi f_i t_j\right) \text{for } i < n, \\ 4\pi^2 f_{i-n}^2 \text{F0} \cos\left(2\pi f_{i-n} t_j\right) \text{for } n < i < 2n, \end{cases}$$

In the same way, we can also compute a matrix $\mathbf{D}_m$ that maps the parameters onto an arbitrary vector of time samples $\boldsymbol{t}_m$, allowing us to interpolate the value of $\dot{\nu}(t)$ at any arbitrary time by

$$\dot{\nu}(\mathbf{t}_m) = \mathbf{D}_m \hat{\boldsymbol{\beta}}.$$

In order to compute the uncertainty on $\dot{\nu}(t)$ we first need the covariance matrix of Gaussian process in the absence of any observations,

$$\mathbf{C}_{\text{mm}} = \mathbf{D}_{\text{m}} \mathbf{C}_{\beta} \mathbf{D}_{\text{m}}^{\mathsf{T}}.$$

We also need the covariance between the $\dot{\nu}(t)$ and the observations,

$$\mathbf{C}_{\text{mo}} = \mathbf{D}_{\text{m}} \mathbf{C}_{\beta} \mathbf{M}^{\mathsf{T}},$$

and the covariance of the data with itself

$$\mathbf{C}_{\text{oo}} = \mathbf{M} \mathbf{C}_{\beta} \mathbf{M}^{\mathsf{T}} + \mathbf{N},$$

These can then be combined to estimate the covariance of the model $\dot{\nu}(t)$ in the presence of the observations,

$$\mathbf{C}_{\dot{\nu}} = \mathbf{C}_{\text{mm}} - \mathbf{C}_{\text{mo}} \mathbf{C}_{\text{oo}}^{-1} \mathbf{C}_{\text{mo}}^{\mathsf{T}}.$$

We take the diagonal elements of $\mathbf{C}_{\dot{\nu}}$ to draw representative uncertainties on the estimates of $\dot{\nu}(t)$. Note that this does not include any uncertainty related to the uncertainty on the model hyperparameters, but this can be estimated by repeating the process for a range of samples from the hyperparameter posteriors.

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