

# Effects of periodicity in observation scheduling on parameter estimation of pulsar glitches

L. Dunn <sup>1,2</sup>★ M. E. Lower <sup>3,4</sup> and A. Melatos<sup>1,2</sup>

<sup>1</sup>*School of Physics, University of Melbourne, Parkville, VIC 3010, Australia*

<sup>2</sup>*Australian Research Council Centre of Excellence for Gravitational Wave Discovery (OzGrav), University of Melbourne, Parkville, VIC 3010, Australia*

<sup>3</sup>*Centre for Astrophysics and Supercomputing, Swinburne University of Technology, PO Box 218, Hawthorn, VIC 3122, Australia*

<sup>4</sup>*CSIRO Astronomy and Space Science, Australia Telescope National Facility, Epping, NSW 1710, Australia*

Accepted 2021 April 14. Received 2021 March 24; in original form 2020 September 23

## ABSTRACT

In certain pulsar timing experiments, where observations are scheduled approximately periodically (e.g. daily), timing models with significantly different frequencies (including but not limited to glitch models with different frequency increments) return near-equivalent timing residuals. The average scheduling aperiodicity divided by the phase error due to time-of-arrival uncertainties is a useful indicator when the degeneracy is important. Synthetic data are used to explore the effect of this degeneracy systematically. It is found that phase-coherent TEMPO2 or TEMPONEST-based approaches are biased sometimes towards reporting small glitch sizes regardless of the true glitch size. Local estimates of the spin frequency alleviate this bias. A hidden Markov model is free from bias towards small glitches and announces explicitly the existence of multiple glitch solutions but sometimes fails to recover the correct glitch size. Two glitches in the UTMOST public data release are reassessed, one in PSR J1709–4429 at MJD 58178 and the other in PSR J1452–6036 at MJD 58600. The estimated fractional frequency jump in PSR J1709–4429 is revised upward from  $\Delta f/f = (54.6 \pm 1.0) \times 10^{-9}$  to  $(2432.2 \pm 0.1) \times 10^{-9}$  with the aid of additional data from the Parkes radio telescope. We find that the available UTMOST data for PSR J1452–6036 are consistent with  $\Delta f/f = 270 \times 10^{-9} + N/(fT)$  with  $N = 0, 1, \text{ and } 2$ , where  $T \approx 1$  sidereal day is the observation scheduling period. Data from the Parkes radio telescope can be included, and the  $N = 0$  case is selected unambiguously with a combined data set.

**Key words:** stars: neutron – pulsars: general – stars: rotation.

## 1 INTRODUCTION

The long-term spin-down of a radio pulsar may occasionally be interrupted by a glitch: an event in which the pulsar’s spin frequency suddenly increases. Glitches are typically recognized by their influence on the timing residuals (Espinoza et al. 2011; Yu et al. 2013), which are the deviations between the expected and measured pulse times of arrival (ToAs). The expected arrival times are predicted by a timing model, which parametrizes both the intrinsic evolution of the rotational phase of the pulsar (frequency and frequency derivatives, glitches) as well as astrometric effects (Roemer, Shapiro, and Einstein delays, dispersion, position, proper motion, and parallax; Lorimer & Kramer 2004; Edwards, Hobbs & Manchester 2006).

To estimate the parameters of a glitch, a model of the effect of the glitch on the rotational phase of the pulsar is assumed. A typical simple form based on a Taylor expansion is (Lower et al. 2020)

$$\Delta\phi_g(t) = \Delta\phi + \Delta f(t - t_g) + \frac{1}{2}\Delta\dot{f}(t - t_g)^2 + \dots, \quad (1)$$

where  $\Delta\phi_g(t)$  denotes the extra phase accumulated in response to the glitch relative to a no-glitch phase model. The free parameters

here are  $t_g$ , the glitch epoch;  $\Delta\phi$ , the permanent jump in rotational phase due to unmodelled effects or an uncertain glitch epoch;  $\Delta f$ , the permanent jump in spin frequency; and  $\Delta\dot{f}$ , the permanent jump in the first spin frequency derivative with respect to time. These parameters may be estimated in the same way as other parameters in the pulsar timing model: through a least-squares fit that minimizes the  $\chi^2$  of the post-fit residuals (Hobbs, Edwards & Manchester 2006). Alternatively, they may be estimated through Bayesian inference with a software package such as TEMPONEST (Lentati et al. 2014), which incorporates parameters describing the deterministic timing model and noise sources and calculates a posterior probability distribution for these parameters via nested sampling with MULTINEST (Feroz, Hobson & Bridges 2009). When reporting glitch parameters estimated in this way, it is tacitly assumed that the  $\chi^2$  of the post-fit residuals has a unique minimum.

In this paper we explore the validity of the single-minimum assumption and the consequences for glitch parameter estimation, when the scheduling of ToA measurements is periodic. By periodic we mean that the gap between consecutive ToAs is approximately equal to an integer multiple of some common period, e.g. if timing data are collected at the same local sidereal time for each observation. The role of observational scheduling on pulsar glitch measurement has received some attention previously. It is well understood that a higher density of observations is advantageous when trying to detect and characterize glitches (Wong, Backer & Lyne 2001; Dodson,

\* E-mail: liamd@student.unimelb.edu.au

McCulloch & Lewis 2002; Janssen & Stappers 2006; Ashton et al. 2019; Basu et al. 2020), but quantitative statements along these lines are rare, due to the large number of factors that may be considered. Espinoza et al. (2014) pointed out that an infrequent observing cadence, combined with a jump in first frequency derivative, can mask a glitch with sufficiently small permanent frequency jumps. They gave a quantitative lower bound on the detectable permanent frequency jump,  $\Delta f_{\text{lim}} = \Delta T |\Delta \dot{f}|/2$ , where  $\Delta T$  is the average time between observations. Similarly, Shannon et al. (2016) noted in their analysis of PSR J0835–4510 that glitches with  $\Delta f$  smaller than  $10^{-7}$  Hz are indistinguishable from timing noise, using timing data with a monthly observing cadence. Melatos et al. (2020) showed that the ability of a hidden Markov model (HMM) to detect glitches with  $\Delta f = 10^{-8}$  Hz in the presence of moderate timing noise is diminished, if the time between observations exceeds  $\sim 10$  d. In their timing of the frequently glitching pulsar PSR J0537–6910, Marshall et al. (2004) employed a strategy in which observations are spaced logarithmically, in order to keep phase uncertainty below 0.1 cycles without expending an inordinate amount of observing time. Finally, the Canadian Hydrogen Intensity Mapping Experiment (CHIME) collaboration has recently noted that their determinations of the pulse frequency of newly discovered pulsars may be in error by  $n/(1 \text{ sidereal day})$ , where  $n$  is a small integer, due to the transit nature of the instrument (Good et al. 2020). This is another manifestation of the underlying issue we explore here in detail. To our knowledge, the specific issue of the effect of periodicity in observation scheduling on the estimation of glitch parameters has not been considered previously.

The paper is structured as follows. In Section 2, we derive a practical condition for when periodic scheduling leads to errors in estimating the spin frequency of the pulsar, i.e. the time derivative of the phase to leading order in a Taylor expansion. We demonstrate that the condition is satisfied in some existing pulsar timing data sets, leading to a near-degeneracy between timing models with significantly different spin frequencies. In Section 3, we specialize to the case of degeneracy between glitch models with different permanent frequency jumps. In Section 4, we investigate how degeneracy in timing models affects specific glitch parameter estimation methods. Section 4.1 discusses glitch parameter estimation with TEMPO2 (Hobbs et al. 2006), both in the context of phase-coherent timing (Section 4.1.1) and local estimation of the pulsar’s spin frequency (Section 4.1.2). Section 4.2 discusses glitch parameter estimation with TEMONEST, and Section 4.3 discusses an HMM-based approach (Melatos et al. 2020). Finally, in Section 5 we discuss periodic scheduling in the context of the UTMOST public data release (Lower et al. 2020).

## 2 PHASE AMBIGUITY

In this section, we explore a simple, non-glitch case in which there may be ambiguity in the measurement of the evolution of the pulsar’s rotational phase. The ambiguity is ultimately due to near-periodicity in the scheduling of the ToA measurements used to determine the timing model parameters. The case without a glitch builds intuition for the more general case with a glitch, which is discussed in Section 3.

### 2.1 Quasi-periodic scheduling

We denote the measured ToAs by  $\{t_1, t_2, \dots, t_N\}$ , and consider the time gap between the  $i$ th and  $(i + 1)$ th ToA,  $\Delta t_i = t_{i+1} - t_i$ . We call a sequence of observations ‘periodic’, if we have the following

approximate equality for all  $i$ :

$$\Delta t_i \approx n_i T, \quad (2)$$

where  $n_i$  is an integer and  $T$  is a common, fundamental period independent of  $i$ .

Equation (2) expresses the periodicity condition intuitively as an approximate equality between the time gap between consecutive ToA measurements,  $\Delta t_i$ , and an integer multiple of some common period  $T$ . However, it is more accurate to think of this condition as a restriction on the fractional part of the factor that multiplies  $T$  in the exact version of equation (2), which is not strictly an integer. That is, we have the exact equality

$$\Delta t_i = (n_i + \epsilon_i)T, \quad (3)$$

where  $n_i$  is an integer, and the remainder  $\epsilon_i$  is a real number satisfying  $|\epsilon_i| < 0.5$  by construction.

### 2.2 Phase error

In the limiting case of perfect periodicity, there is a choice of  $T$  for which we have  $\epsilon_i = 0$  for all  $i$ . The degeneracy is exact: Any extra term in the timing model that contributes an integer number of pulsar rotations over the time-scale  $T$  predicts ToAs that coincide exactly with those measured, so timing models with and without such an extra term cannot be distinguished. The simplest extra term is of the form

$$\Delta \phi(t) = \frac{Nt}{T}, \quad (4)$$

where  $N$  is an integer. Equation (4) corresponds to a permanent change in the pulse frequency of size  $\Delta f = N/T$ .

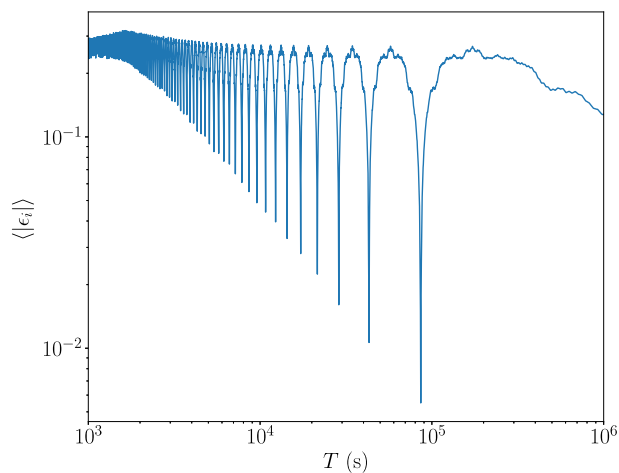
The limiting case  $\epsilon_i = 0$  is special. We now allow  $\epsilon_i \neq 0$ , start with the true phase model  $\phi(t)$ , and add a spurious term of the form equation (4). A term of this form produces indistinguishable timing models for  $\epsilon_i = 0$ , so one expects the degeneracy to break slightly, when  $\epsilon_i$  is small. The degeneracy can be quantified by asking: what is the effect of the spurious phase term on the timing residuals, as a function of  $\epsilon_i$ ? In general, the total phase residual at the  $i$ th ToA is given by  $R_i = \phi_i - n_i$ , where  $\phi_i$  is the predicted pulsar rotational phase at the  $i$ th ToA, and  $n_i$  is the closest integer to  $\phi_i$  (Taylor 1992). The phase residual contributed across a gap of length  $\Delta t_i = (n_i + \epsilon_i)T$  by  $\Delta \phi(t)$  is given by

$$\delta \phi_i = \text{frac} [\Delta \phi(\Delta t_i)] = \text{frac} \left[ \frac{N}{T} (n_i + \epsilon_i)T \right] = \text{frac}(N\epsilon_i), \quad (5)$$

where  $\text{frac}(\cdot)$  denotes the fractional part, and the last equality follows from the fact that  $N$  and  $n_i$  are both integers by construction, so their product has no fractional part. For  $N = 1$  we then simply have  $\delta \phi_i = \epsilon_i$ .

### 2.3 Worked example

We illustrate the arguments of this section with an example drawn from a real data set. The UTMOST pulsar timing programme (Bailes et al. 2017) uses data taken at the Molonglo Observatory Synthesis Telescope, and has released public data for a large number of pulsars (Lower et al. 2020). Here we consider the data set released for PSR J1452–6036, consisting of 287 ToAs measured between 2017 July and 2019 July. Our aim is to determine whether the observations of this pulsar are close enough to periodic that there is ambiguity in the phase evolution of the pulsar. We therefore seek to determine the value of  $T$  that minimizes the average value of  $|\epsilon_i|$  as defined in equation (3). If there is a choice of  $T$  which makes  $|\epsilon_i|$  particularly



**Figure 1.** Mean fractional scheduling aperiodicity  $\langle |\epsilon_i| \rangle$  as a function of scheduling quasi-period  $T$  for the UTMOST observations of PSR J1452–6036.

small, then the phase residuals induced by a phase term of the form in equation (4) will be correspondingly small.

Fig. 1 graphs the average of  $|\epsilon_i|$  as a function of  $T$  for this data set. A sequence of local minima is clearly visible, with the lowest value of  $\langle |\epsilon_i| \rangle$  occurring at  $T = 86158$  s, and  $\langle |\epsilon_i| \rangle = 4 \times 10^{-3}$ . We therefore expect that the magnitude of the timing residuals induced by an extra phase term of the form  $\Delta\phi(t) = t/(86158$  s) will be  $\sim 4 \times 10^{-3}$ .

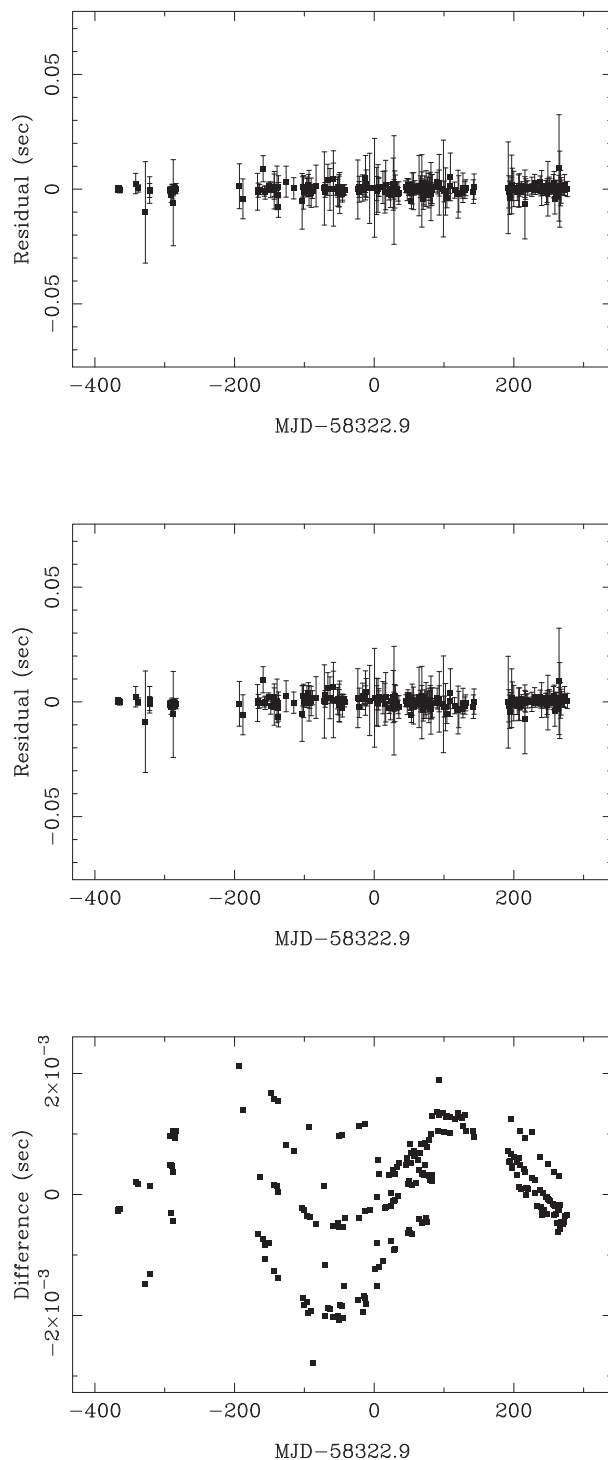
Once we know the magnitude of the residuals induced in a data set due to a spurious phase term of the form in equation (4), we can compare it against the magnitude of the stochastic residuals due to ToA measurement error. The quoted uncertainties on the ToA measurements in the PSR J1452–6036 data are typically 1 ms. The pulsar spins at roughly 6 Hz, so ToA uncertainties  $\sim 1$  ms correspond to phase residuals  $\sigma_{\text{ToA}} \sim 6 \times 10^{-3}$ . Hence, residuals induced by the extra phase term  $\Delta\phi(t)$  are of the same order as the residuals due to ToA measurement error. It is therefore reasonable to expect that the two sets of timing residuals look similar, with and without  $\Delta\phi(t)$ . Equivalently, we refer to the indicative ratio

$$R = \langle |\epsilon_i| \rangle / \sigma_{\text{ToA}}. \quad (6)$$

If the condition  $R \lesssim 1$  is satisfied, as it is in this example, we expect phase models with and without the  $\Delta\phi(t)$  term to give similar residuals. This is demonstrated in Fig. 2, where the top panel shows the timing residuals for the original UTMOST timing model and the middle panel shows the timing residuals for a timing model with an extra phase term  $\Delta\phi(t) = t/(86158$  s) added. The bottom panel shows the difference between the two sets of residuals. The difference is no more than a few milliseconds, of the same order as the typical ToA error. After re-fitting the spin frequency, the new timing model has a spin frequency  $1.1605 \times 10^{-5}$  Hz larger than the UTMOST timing model. Presented with the top or bottom panel in Fig. 2, an analyst would have no reason to doubt either timing model, even though they are different. We emphasize that in this case the degeneracy between the two timing models does not involve glitches.

## 2.4 Detecting residuals from frequency misestimation

If the frequency in the timing model is incorrect by a multiple of the inverse of the scheduling period  $T$ , the induced phase residual across a given ToA gap is given by equation (5). We may therefore



**Figure 2.** Timing residuals for the UTMOST observations of PSR J1452–6036 between MJD 57955 and MJD 58600, using the UTMOST timing model (*top*), and a model that matches the UTMOST timing model except for an overall frequency increment of  $1.1605 \times 10^{-5}$  Hz (*middle*). The bottom panel shows the difference between the two sets of residuals (error bars not shown). The uncertainty on the spin frequency reported by TEMPO2 is  $6 \times 10^{-10}$  Hz.

seek to detect the ‘signal vector’  $\mathbf{S}$ , the  $k$ th element of which is the cumulative sum of the induced phase residuals up to and including the  $k$ th ToA gap:

$$S_k = \sum_{i=1}^k \delta\phi_i. \quad (7)$$

In the absence of any other sources of timing residuals,  $\mathbf{S}$  is exactly what we expect to see if the frequency is misestimated due to periodic observation scheduling. Successful detection of this signal in the residuals enables disambiguation between timing models that produce timing residuals that appear by eye to be indistinguishable.

Our treatment of this detection problem follows Levy (2008). Since we know precisely the form of the expected signal, cross-correlating the vector  $\mathbf{S}$  with the residuals  $\mathbf{R} = (R_1, R_2, \dots, R_N)$  is an optimal means of detection if the noise in  $\mathbf{R}$  is additive, white, and Gaussian. We assume that the noise has the covariance matrix  $\mathbf{C} = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2)$ , where  $\sigma_i$  is the reported uncertainty on the  $i$ th ToA, measured in cycles. After calculating the expected signal  $\mathbf{S}$  according to equation (7), we may compute the cross-correlation test statistic

$$\gamma = \mathbf{R}^T \mathbf{C}^{-1} \mathbf{S} \quad (8)$$

and compare this test statistic to a threshold  $\gamma_{\text{th}}$  chosen to give a particular false alarm rate  $P_{\text{fa}}$ ,

$$\gamma_{\text{th}} = \sqrt{\mathbf{S}^T \mathbf{C}^{-1} \mathbf{S}} Q^{-1}(P_{\text{fa}}), \quad (9)$$

where  $Q^{-1}(x)$  is the inverse of

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} du \exp(-u^2/2). \quad (10)$$

The probability of detection  $P_d$  for a given probability of false alarm  $P_{\text{fa}}$  is then given by

$$P_d = 1 - Q \left[ \sqrt{\mathbf{S}^T \mathbf{C}^{-1} \mathbf{S}} - Q^{-1}(P_{\text{fa}}) \right]. \quad (11)$$

This gives a quantitative estimate of the reliability of detecting the residuals caused by misestimating the frequency when observations are periodically scheduled.

For concreteness, consider again the UTMOST data set for PSR J1452–6036. We calculate  $\epsilon_i$  according to (3) assuming  $T = 86\,158$  s, and subsequently calculate  $\mathbf{S}$  assuming  $N = 1$ . The latter quantity is the induced phase error signal due to the frequency in the timing model being  $1/(86\,158 \text{ s}) = 1.161 \times 10^{-5}$  Hz too large. For  $P_{\text{fa}}$  fixed at 0.01, we find  $P_d = 0.98$ . Hence, it is possible to reliably detect the presence of the signal  $\mathbf{S}$  due to an incorrect frequency measurement and therefore reject a timing model that contains such a signal, if one is aware of the effect and has a sufficiently long stretch of data. We caution that this is an idealized treatment that ignores the presence of other noise sources (e.g. timing noise). If significant timing noise is present in the data, the assumption of a diagonal noise covariance matrix breaks down, and a modified approach is recommended. We note also that for all of the results presented in this section,  $N$  can be either positive or negative – it makes no difference whether the spurious phase term  $\Delta\phi(t)$  corresponds to an increase or decrease in frequency.

### 3 GLITCHES

We now turn to the subject of degeneracy between glitch models. The idea is essentially the same: under the conditions described in Section 2, certain glitch sizes are difficult to distinguish from one another on the basis of ToA measurements, especially based on visual

inspection of the timing residuals. We begin by guiding the reader step by step through a worked example based on simulated data in Section 3.1 in order to illustrate the application of the key ideas. We then show in Section 3.2 that a jump in the frequency derivative during a glitch does not affect the argument in Section 2 nor the conclusions in the rest of the paper.

#### 3.1 Worked example

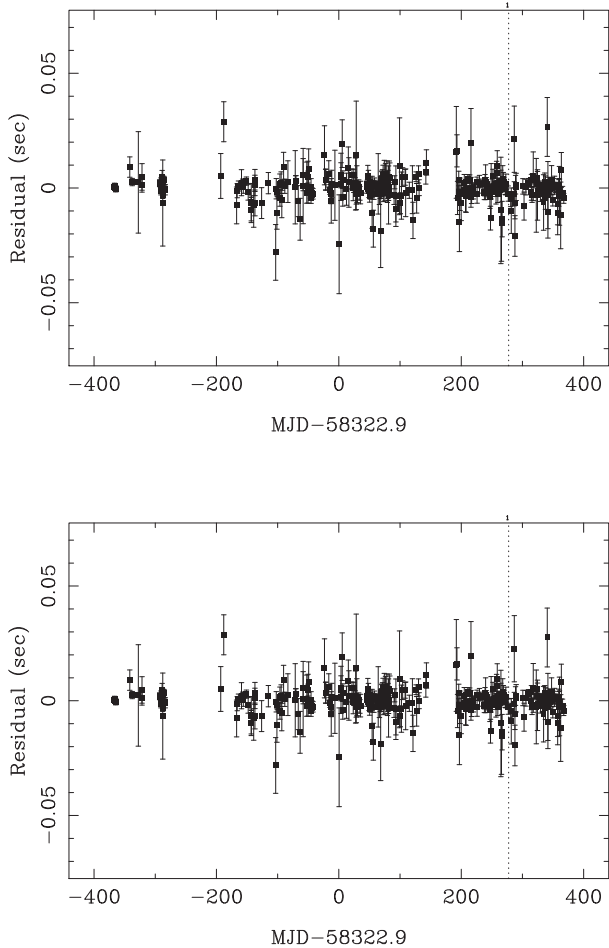
We examine a simulated data set that is generated as follows. First, a timing model is chosen. We choose to take as our starting point the timing model from the UTMOST data release for PSR J1452–6036, in which a glitch was detected at MJD 58600.29 with  $\Delta f = 1.745 \times 10^{-6}$  Hz,  $\Delta\phi = 0$ , and no reported  $\Delta\dot{f}$  or exponentially recovering component (Lower et al. 2020). The timing model used to generate the simulated data set matches the timing model for PSR J1452–6036 from the UTMOST data release, except that the glitch at MJD 58600.29 has size  $\Delta f = 1.2106 \times 10^{-5}$  Hz instead. All other glitch parameters are identical – there is no phase jump at the glitch epoch, change in frequency derivative, or exponential recovery included in the simulated data. We wish to generate a set of synthetic ToAs that are consistent with the chosen timing model. Starting with the real PSR J1452–6036 data set used in Section 2, we use LIBSTEMPO<sup>1</sup> to generate a new set of idealized ToAs. The synthetic ToAs begin as an exact copy of the PSR J1452–6036 ToAs but are shifted slightly so that they show zero residuals when analysed with the chosen timing model. White, Gaussian noise is then added to each ToA at a level commensurate with the reported uncertainty for that ToA. This synthetic data set serves two purposes. In the remainder of this section, it is used to illustrate the principle that periodicity in observation scheduling leads to a degeneracy between glitch models. In Sections 4.1, 4.2, and 4.3, it is used to examine how glitch analyses with TEMPO2, TEMPNEST, or an HMM-based approach may be confounded by this degeneracy between glitch models.

Fig. 3 compares two glitch models for the synthetic data set described above. The top panel of Fig. 3 shows the residuals for the true timing model (i.e. the chosen timing model used in generating the synthetic data set), while the bottom panel shows the residuals for a timing model in which the glitch size is set to  $5 \times 10^{-7}$  Hz. Why  $5 \times 10^{-7}$  Hz? Because the true glitch size of  $1.2107 \times 10^{-5}$  Hz may be written as  $1/(86\,158 \text{ s}) + 5 \times 10^{-7}$  Hz, and hence expresses the degeneracy noted in Section 2. The two sets of timing residuals both appear white by eye. The root mean square (rms) residuals for the original timing model with glitch size  $\Delta f = 1.2107 \times 10^{-5}$  Hz are 2729  $\mu\text{s}$ , while the rms residuals for the case with a glitch size of  $\Delta f = 5 \times 10^{-7}$  Hz are 2726  $\mu\text{s}$ . That is, the residuals are nearly equal, even though  $\Delta f$  is approximately 24 times larger in the former model. Note that in the glitch model with  $\Delta f = 5 \times 10^{-7}$  Hz, an unphysical phase jump of  $\Delta\phi = 0.326$  has also been included.

#### 3.2 Jumps in frequency derivative

Glitches are often accompanied by a jump in the frequency derivative, which can be a significant fraction of the pre-glitch frequency derivative. In Section 3.1, we do not include a jump in frequency derivative, to keep things as simple as possible. We now show that jumps in frequency derivative do not affect the arguments made in Section 2: The degeneracy persists even when a frequency derivative jump is present.

<sup>1</sup><https://vallis.github.io/libstempo/>



**Figure 3.** Timing residuals for the synthetic data set described in Section 3.1, using a timing model with (top)  $\Delta f = 1.2106 \times 10^{-5}$  Hz that matches the injected glitch size, and (bottom)  $\Delta f = 5 \times 10^{-7}$  Hz that is approximately 24 times smaller than the injected glitch.

To be explicit, we consider two glitch models:

$$\phi_{g,1}(t) = \Delta f(t - t_g) + \frac{1}{2} \Delta \dot{f}(t - t_g)^2 \quad (12)$$

$$\phi_{g,2}(t) = \left( \Delta f + \frac{1}{T} \right) (t - t_g) + \frac{1}{2} \Delta \dot{f}(t - t_g)^2, \quad (13)$$

where  $\Delta f$ ,  $\Delta \dot{f}$ , and  $t_g$  are defined as in equation (1) and  $T$  is the observing period. While these two glitch models contain a frequency derivative change, they none the less differ by  $t/T$  (plus a constant,  $t_g/T$ ), which has the form of equation (4). Therefore, the arguments of Section 2 apply, and these two glitch models show comparable residuals when the observations are periodic with period  $T$ .

The above argument extends to glitch models with exponentially decaying terms, or any other terms that may be appropriate, as long as those other terms are common to both glitch models being compared, and the only difference between them is a frequency increment as described in Section 2.2.

## 4 GLITCH PARAMETER ESTIMATION

In this section, we discuss the effects of periodic observation scheduling on specific glitch parameter estimation techniques using TEMPO2 (Section 4.1), TEMPNEST (Section 4.2), and an HMM (Section 4.3).

### 4.1 TEMPO2

#### 4.1.1 Phase-coherent timing

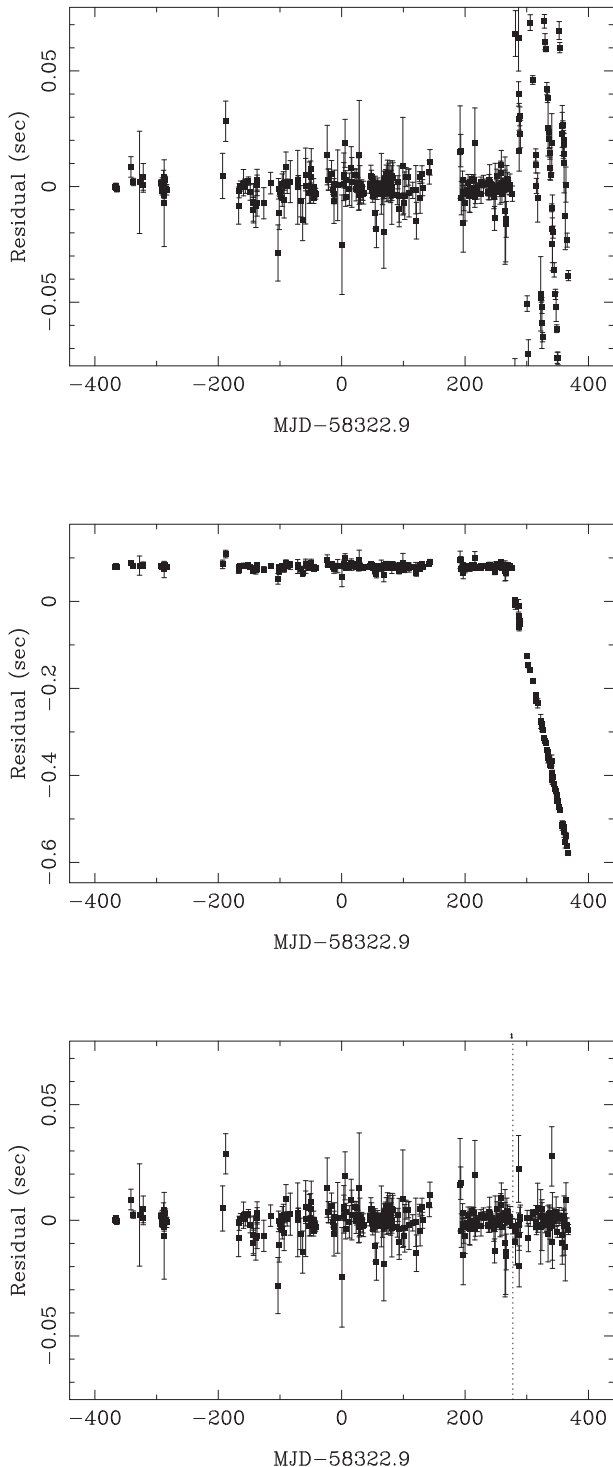
In Section 3 and Fig. 3, we demonstrate that when a glitch occurs in a pulsar, there may be multiple timing models that give similarly small timing residuals. Of course, the illustrative worked example in Section 3.1 is fashioned deliberately by exploiting our foreknowledge of the glitch parameters injected into a synthetic data set. Does the same outcome arise ‘naturally’, when the true glitch parameters are unknown, e.g. in real astronomical data or blind injections in synthetic data? In this section, we perform a phase-coherent timing analysis using TEMPO2 of the same synthetic data set. This is not a truly blind analysis, but the steps taken in the course of this analysis approximate one typical course of action for the parameter estimation of a previously unknown glitch using astronomical data.

We begin by inspecting the timing residuals for the synthetic data set used in Section 3.1 with a timing model that matches the injected parameters but without a glitch. The residuals are shown in the top panel of Fig. 4. The pre-glitch residuals are close to white, but there is a clear point at MJD 58600 where the timing model ‘fails’, and the residuals diverge and begin to wrap. At epochs where the phase residuals wrap around, we add phase jumps by hand to restore phase connection. At this stage we use TEMPO2 to tag each ToA with the number of pulses that have elapsed since the first ToA. We emphasize that this pulse numbering is model dependent. With the phase jumps added, we see a clear transition from flat phase residuals to a linear ramp in the residuals in the middle panel of Fig. 4 – a clean glitch signature. Using the pulse-numbered ToAs, it is straightforward to fit the glitch parameters with TEMPO2. For simplicity, we fit only the glitch frequency increment  $\Delta f$  and an unphysical glitch phase jump  $\Delta \phi$  to account for uncertainty in the glitch epoch. We set the glitch epoch to be the same as the injected glitch epoch, MJD 58600.29, for ease of comparison with the injected glitch parameters. The TEMPO2 fit returns  $\Delta f = (5.008 \pm 0.006) \times 10^{-7}$  Hz, and  $\Delta \phi = 0.320 \pm 0.003$ . The associated residuals are plotted in the bottom panel of Fig. 4. By eye, they appear close to white, with variance consistent with the ToA error bars, which are of the order of 1 ms. This timing model is a good fit: The reduced  $\chi^2$  returned by TEMPO2 is 0.98. It is not, however, an accurate recovery of the injected glitch parameters, which are  $\Delta f = 1.2106 \times 10^{-5}$  Hz and  $\Delta \phi = 0$ .

The above analysis exemplifies a general principle: Given a set of ToAs containing a glitch, if the ToAs are consistent with multiple glitch models due to periodicity in the observation schedule, then the glitch model with the smallest  $\Delta f$  is more likely to be recovered by a phase-coherent timing analysis. This occurs because ToAs that are consistent with a small glitch display a relatively gentle linear ramp in the post-glitch residuals. This linear ramp will be picked out by eye, and subsequently used to number the post-glitch pulses. A larger glitch (e.g.  $1.2106 \times 10^{-5}$  Hz compared to  $5 \times 10^{-7}$  Hz) has a different phase model, which assigns a different numbering to the post-glitch pulses. Once the analyst restores phase connection (whether correctly or incorrectly) and numbers the pulses according to this phase connected solution, the range of possible glitch models is restricted.

#### 4.1.2 Local frequency estimation

Rather than constructing a phase-connected timing solution and subsequently estimating glitch parameters, it is also possible to estimate the frequency evolution *locally* by fitting for the frequency using small sets of ToAs closely spaced in time that are derived from



**Figure 4.** Timing residuals for the synthetic data set described in Section 3.1 before (*top panel*) and after (*middle panel*) adding phase jumps to account for wraps in pulse phase. The bottom panel shows the residuals obtained after accounting for a glitch with  $\Delta f = (5.006 \pm 0.015) \times 10^{-7}$  Hz and  $\Delta\phi = 0.323 \pm 0.008$ .

sub-integrations of a longer observation. In this case, the inter-ToA spacing is much smaller than  $T$ , the fundamental observing period. We stipulate that the ToAs are closely spaced in time to sidestep the issue of distinguishing between phase models with frequencies differing by  $1/T$ , which arises once ToAs separated by more than  $T$

are included in the fit. In order to distinguish between timing models that differ in spin frequency by  $1/T$ , the time  $\tau$  between the first and last ToAs in each local fit must be long enough that the accumulated phase error exceeds the phase error  $\sigma_{\text{ToA}}$  due to the ToA uncertainty:

$$\tau/T \gg \sigma_{\text{ToA}}. \quad (14)$$

We consider three applications of local frequency estimation to synthetic data. For each application, we generate a synthetic data set with the same timing model parameters and observation schedule as described in Section 3.1 (so the observation period  $T$  is still roughly 86 158 s). At each session when a synthetic observation is made, we generate four ToAs spaced closely, with  $\tau \ll T$ . In the first of the three applications, we take  $\tau = 1100$  s,<sup>2</sup> giving  $\tau/T \approx 2\sigma_{\text{ToA}}$ . The results of the local spin frequency estimation for this data set are shown in the top panel of Fig. 5. The scatter in the post-glitch frequency estimates is of the order of 10  $\mu\text{Hz}$ , prohibiting a determination of the post-glitch spin frequency that distinguishes between glitch models with sizes separated by  $1/T \approx 1.2 \times 10^{-5}$  Hz. However, if  $\tau$  is increased by a factor of 5, we instead have the result shown in the middle panel of Fig. 5. The scatter in the frequency estimation is much reduced to  $\sim 1$   $\mu\text{Hz}$ , and it is clear that a glitch with  $\Delta f \approx 1.2 \times 10^{-5}$  Hz is preferred over the value recovered by the phase-coherent analysis in Section 4.1.1 ( $\Delta f \approx 5 \times 10^{-7}$  Hz). Finally, we can analyse the same synthetic data set but use eight ToAs spanning two consecutive observing sessions in each local frequency estimate, so that  $\tau \approx T$ . The results for this case are shown in the bottom panel of Fig. 5. The scatter in post-glitch frequency estimations is again reduced, down to roughly 0.1  $\mu\text{Hz}$ . However, the local estimates are centred around  $5 \times 10^{-7}$  Hz – the same incorrect estimate recovered by the phase-coherent analysis in Section 4.1.1. In both the phase-coherent case and the  $\tau \approx T$  case, the assumption that the increase in frequency due to the glitch contributes less than one full rotation between two observation sessions leads to the incorrect estimate.

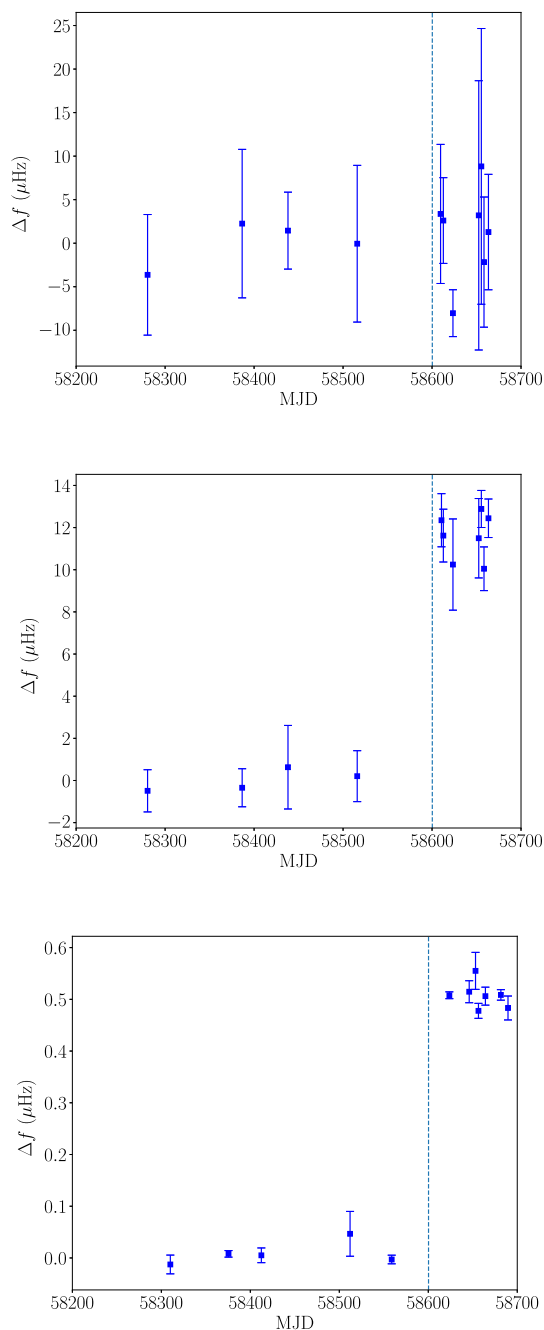
## 4.2 TEMPONEST

The Bayesian pulsar timing package TEMPONEST is sometimes used to estimate glitch parameters (Shannon et al. 2016; Yu & Liu 2017; Lower et al. 2020; Parthasarathy et al. 2020). If the glitch parameters are not already well constrained, it is common to first use TEMPO2 to obtain a set of pulse-numbered ToAs, following the procedure described in Section 4.1.1. Note that it is impossible to run TEMPONEST without pulse-numbered ToAs, unless the timing model parameters are already well constrained. As noted in Section 4.1.1, pulse numbering restricts the range of viable glitch models. Even if one chooses an uninformative prior for  $\Delta f$ , TEMPONEST only assigns high posterior probability to glitch models that are consistent with the pulse numbering used. As such, glitch sizes estimated with TEMPONEST exhibit the same bias towards small glitch sizes as those estimated with a phase-coherent TEMPO2 analysis.

## 4.3 HMM

Recently, Melatos et al. (2020) presented a new approach to pulsar glitch detection that models the rotational evolution of the pulsar with an HMM and selects between models with and without glitches present. Like TEMPO2-based approaches, the HMM operates on ToAs. Therefore, the arguments in Sections 2 and 3 suggest that one

<sup>2</sup>This corresponds to one ToA every 4.5 min, which is quite typical for pulsars observed by UTMOST.



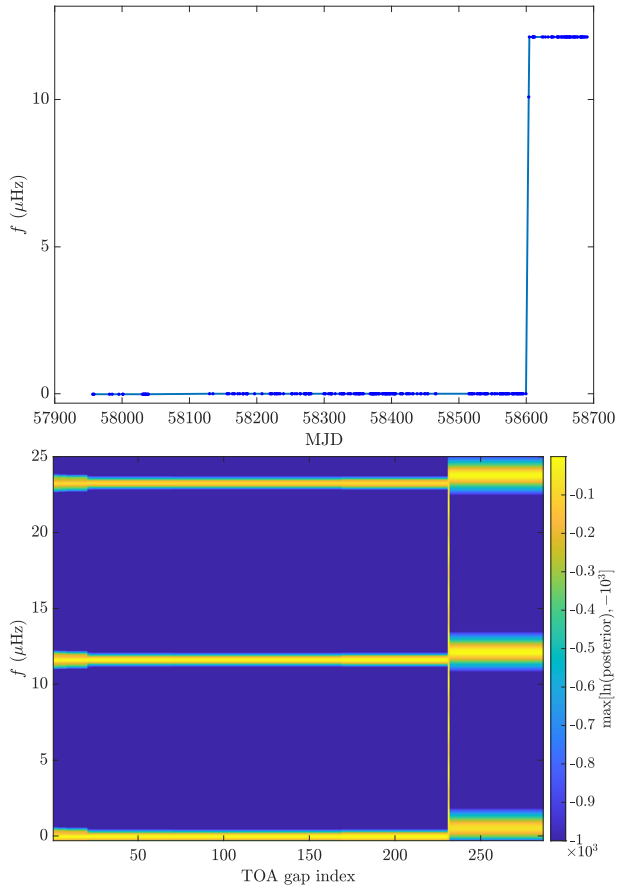
**Figure 5.** Results of local spin frequency estimation described in Section 4.1.2. The top panel shows the case  $\tau/T \approx 2\sigma_{\text{ToA}}$ , the middle panel shows the case  $\tau/T \approx 10\sigma_{\text{ToA}}$ , and the bottom panel shows the case where multiple observing sessions are used in each spin frequency fit. The location of the glitch is indicated by the vertical dotted line. In all cases, the true size of the glitch is  $\Delta f = 1.21 \times 10^{-5}$  Hz. Note the  $10^2$ -fold decrease in the vertical scale from the top to the bottom panels.

ought to be careful when estimating glitch parameters with an HMM, if the observations are scheduled periodically. It turns out that the HMM estimates  $\Delta f$  more accurately than TEMPO2 and TEMPONEST for the synthetic data set from Section 3.1, as reported in Section 4.3.1. However, it is still somewhat prone to the same ambiguities arising from periodic scheduling, as demonstrated in Section 4.3.2 with the aid of specific examples.

The HMM consists of three essential components, which we briefly describe here. We refer the reader to Appendix A and Melatos et al. (2020) for further details. In the HMM, the state of the pulsar is described by its pulse frequency  $f$ , and the time derivative of the pulse frequency  $\dot{f}$ . The values of  $f$  and  $\dot{f}$  are measured relative to a fiducial frequency evolution that is taken from a Taylor expansion of the phase model calculated by TEMPO2. For example, the state  $(f, \dot{f}) = (0, 0)$  indicates that the state of the pulsar is exactly as predicted by the Taylor expansion at that time-step. The state  $(f, \dot{f})$  is ‘hidden’: It is not observed directly, since the HMM operates only on ToAs. Instead, we define an ‘emission probability’  $L(\Delta t_i; f, \dot{f})$  that gives the probability of observing a given ToA gap  $\Delta t_i$  if the pulsar’s state is  $(f, \dot{f})$  during this gap. The expression for  $L(\Delta t_i; f, \dot{f})$  used in this paper is given by equation (A2). Finally, it is necessary to specify the ‘transition probabilities’ that determine the probability that one state transitions to another state after each ToA gap. The HMM tracks the spin wandering directly, as a realization of a Markov chain. This is in contrast to TEMPONEST, which estimates the ensemble characteristics (power spectral density) of the residuals due to timing noise (Lentati et al. 2014). In this work, we use transition probabilities that assume a random walk in the second frequency derivative of the pulsar. This matches the prescription adopted by Melatos et al. (2020). Other, qualitatively similar forms of the transition probabilities are also possible and produce qualitatively similar results. Once detected, the parameters of the glitch may be estimated by constructing the sequence of a posteriori most likely  $(f, \dot{f})$  states of the pulsar using the forward–backward HMM algorithm, and subsequently reading off  $\Delta f$  at the most probable glitch epoch. With the parameter choices adopted in this work, the HMM is computationally cheap – the search for glitches described in Section 4.3.1 takes roughly 10 min to run on a modern desktop CPU (the quoted time is measured on an Intel Core i5-9300H CPU running at 2.40 GHz.)

#### 4.3.1 Worked example

To illustrate the effect of periodic scheduling on glitch measurement with an HMM, we re-analyse the same synthetic data set presented in Section 3.1 and analysed with TEMPO2 in Section 4.1. The details of the parameter choices in the HMM analysis can be found in Appendix A. Most relevant to the issue at hand is the range of  $f$ , which is taken to be  $[-3 \times 10^{-7}, 2.5 \times 10^{-5}]$  Hz, bracketing the TEMPO2 timing solution. This range of frequencies is chosen to encompass a large enough range of glitch sizes to demonstrate the essential point. In particular, it covers  $\Delta f \pm 1/(1 \text{ sidereal day})$ , where  $\Delta f = 1.21 \times 10^{-5}$  Hz is the injected glitch size. A wider frequency range is possible, but would significantly increase the computation time required without adding anything new [see section 4.4 of Melatos et al. (2020)]. A glitch is detected at the 231st ToA gap, between MJDs 58599 and 58603, with log Bayes factor 250. We denote this model by  $M_1(231)$ . While the epoch matches the injected glitch, which occurs at MJD 58600.29, we mimic a realistic analysis by persevering and searching ‘blindly’ for a second glitch by comparing  $M_1(231)$  to a set of two-glitch models  $M_2(231, k)$  that contain glitches during the 231st gap and the  $k$ th gap ( $k \neq 231$ ). This procedure is the greedy hierarchical algorithm described in section 4.2 in Melatos et al. (2020). We find that the two-glitch model  $M_2(231, 232)$  is favoured over  $M_1(231)$  with a log Bayes factor of 53. We then compare the two-glitch model  $M_2(231, 232)$  to a set of three-glitch models  $M_3(231, 232, l)$ . None of the three-glitch models are favoured over  $M_2(231, 232)$  with log Bayes factor less than 0 for all  $l \neq 231, 232$ , so we terminate the search. Note that we do



**Figure 6.** (Top) Frequency path recovered by the HMM forward-backward algorithm for the analysis of the synthetic data set described in Section 3.1. The recovered glitch size is  $\Delta f = 1.2078 \times 10^{-5}$  Hz. (Bottom) Heatmap showing the evolution of the logarithm of the posterior probability for  $f$  from the HMM analysis of the same synthetic data set. Time increases from left to right, and at each time-step the posterior has been marginalized over  $\dot{f}$ . Values in the heatmap are clipped below  $-10^3$  to aid readability. The favoured model,  $M_2(231, 232)$ , includes glitches during ToA gaps 231 and 232.

not interpret  $M_2(231, 232)$  as a model with two truly distinct spin-up events, because the ToA gaps are adjacent.  $M_2(231, 232)$  is favoured over  $M_1(231)$  because the one-glitch models are constructed such that the glitch occurs at the beginning of the ToA gap, whereas in these data the glitch occurs nearly 1 d into a 4-d-long gap. The two-glitch model contains enough freedom to mitigate the phase error caused by the simple form of the glitch model assumed by the HMM, which only allows for a glitch to occur at the beginning of a ToA gap, and includes no phase jumps.

Given  $M_2(231, 232)$ , the HMM forward-backward algorithm computes the posterior distribution of  $f$  and  $\dot{f}$  during each ToA gap. From this posterior distribution, we construct a sequence of frequency and frequency derivative states by choosing the a posteriori most likely states during each ToA gap (Rabiner 1989; Melatos et al. 2020). The frequency sequence  $f(t_i)$  constructed in this way is shown in the top panel of Fig. 6. The location of the glitch is clear. Reading off the size of the glitch gives  $\Delta f = 1.2078 \times 10^{-5}$  Hz, compared to the true glitch size of  $1.2107 \times 10^{-5}$  Hz. While the glitch model allows for a change in  $\dot{f}$  (which can be positive or negative), we do not observe such a change in the recovered  $\dot{f}(t_i)$  sequence. This is in agreement with the injected glitch parameters, which include  $\Delta \dot{f} = 0$ .

While this appears to be a relatively successful recovery of the glitch parameters – more so than the phase-coherent analysis of Section 4.1.1 – we can further inspect the posterior distribution of spin states to determine what effect the observation schedule has on this mode of analysis. A heatmap of the logarithm of the posterior distribution of  $f$  (marginalized over  $\dot{f}$ ) as a function of ToA gap index is shown in the bottom panel of Fig. 6. It exhibits a multiply peaked structure both before and after the glitch, with peaks separated by  $1.1618 \times 10^{-5}$  Hz. This spacing is significant: It is close to  $1/(86\,158\text{ s}) = 1.1607 \times 10^{-5}$  Hz, where 86 158 s is the observation scheduling period for this data set. These multiple peaks in the posterior  $f$  distribution are indicative of the degeneracy caused by periodic observation scheduling. While the peaks appear to be equal in height based on the logarithmic heatmap, after the glitch the peak at  $\Delta f = 1.2078 \times 10^{-5}$  Hz is systematically higher (by a factor of at least 1.5) than the other two, and therefore features in the recovered sequence.

#### 4.3.2 Robustness of parameter estimation against uncertainty in glitch epoch

The HMM succeeds in recovering the correct glitch size in Section 4.3.1 despite the periodic scheduling. However, small variations in the epoch of the glitch can significantly perturb the recovered frequency path when the scheduling is periodic.

Fig. 7 shows the marginalized post-glitch posterior distributions of  $f(t_{250})$  ( $t_{250}$  is chosen to show the posterior distribution of  $f$  well after the glitch, which occurs at  $t_{231}$ ) for 12 injected glitches that have the same size as the previous example,  $\Delta f = 1.2107 \times 10^{-5}$  Hz, but epochs distributed uniformly between MJD 58599 and MJD 58603, i.e. anywhere within the 231st ToA gap. The posterior in each case has the same multiply peaked structure as in the bottom panel of Fig. 6, with the largest peak located randomly near one of three values:  $5 \times 10^{-7}$ ,  $1.2 \times 10^{-5}$ , and  $2.4 \times 10^{-5}$  Hz. In all but one of the panels of Fig. 7, only one or two peaks of the three-peak structure seen in the bottom panel of Fig. 6 are high enough to be seen by eye.

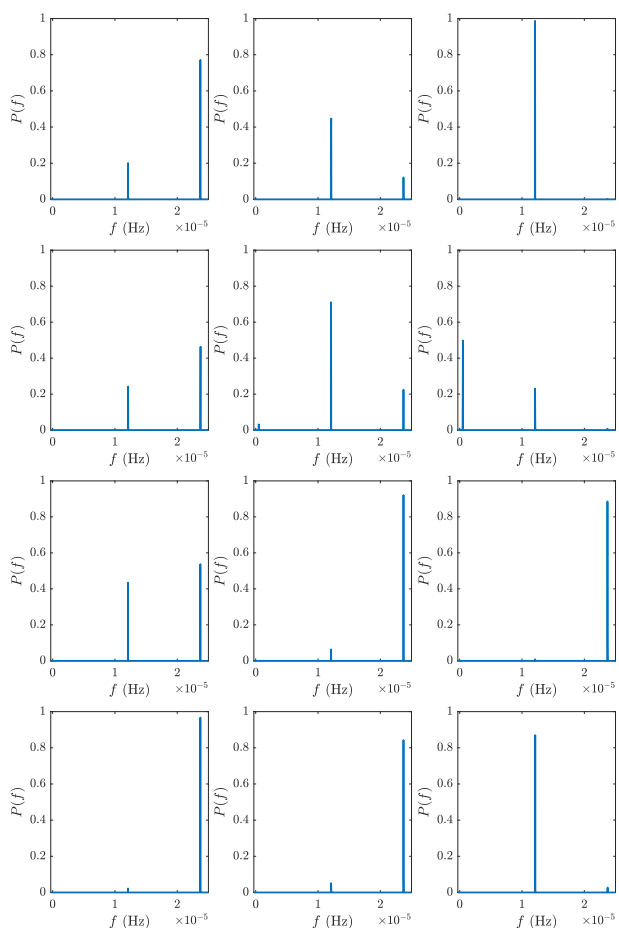
Thus, HMM-based analyses of data sets with periodic scheduling do not necessarily recover the correct glitch size: Of the 12 posterior  $f$  distributions shown in Fig. 7, only 4 peak at the correct location of  $1.2107 \times 10^{-5}$  Hz. The rest have peaks displaced from the true glitch size by approximately  $1/(86\,158\text{ s})$ , indicating that the ambiguity due to periodic scheduling is responsible for the failures to recover the correct glitch size in the other realizations.

Phase-coherent timing analyses of similar data sets are biased towards recovering the smallest plausible glitch size. The HMM shows no systematic bias. None the less, it is not guaranteed to return the true glitch size either. Therefore, the methods complement one another and are safest to use in tandem. If periodic scheduling is unavoidable for some reason, a chance discrepancy between the methods is one way to catch errors in the estimate of  $\Delta f$ , as the example in Sections 4.1.1, 4.1.2, and 4.3.1 demonstrates in practice. Interestingly, the HMM posterior distribution readily reveals the existence of multiple high-likelihood frequency tracks, which are not seen so easily in a timing analysis.

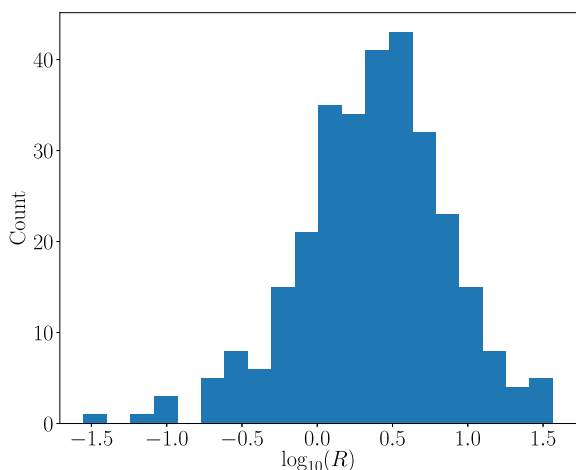
## 5 REASSESSING UTMOST GLITCHES

From mid-2017 onwards, it was decided to use the Molonglo Observatory Synthesis Telescope in transit-only mode (Venkatraman Krishnan et al. 2020). In this mode, the transit of astronomical objects through the primary beam occurs at the same time each

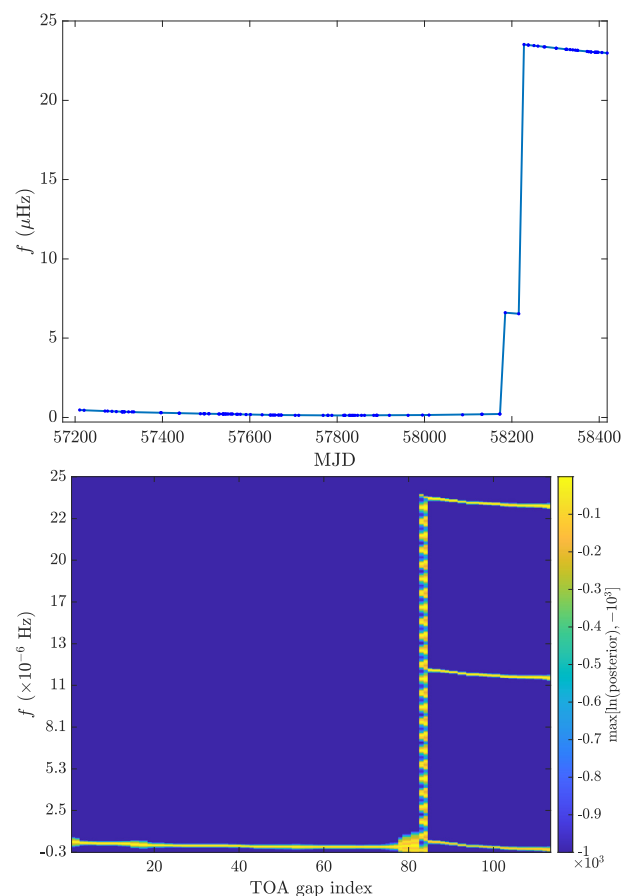




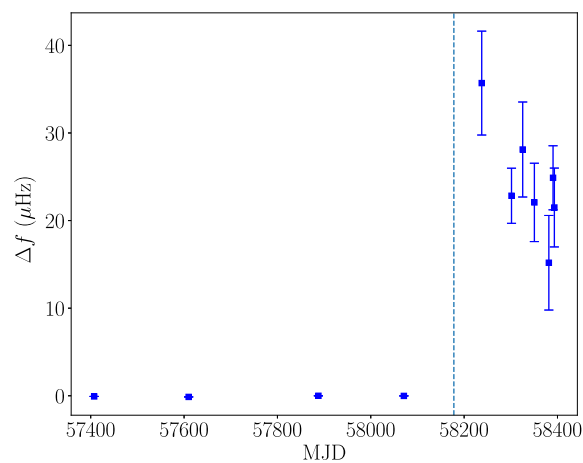
**Figure 7.** Posterior probability distributions  $P(f)$  of  $f$  during the 250th ToA gap (well after the glitch at  $t_{231}$ ) for 12 synthetic data sets with randomized glitch epochs in the interval  $\text{MJD } 58599 < t_g < \text{MJD } 58603$  as described in Section 4.3. In all cases, the true frequency deviation is  $1.2107 \times 10^{-5}$  Hz and the glitch occurs during the 231st ToA gap.



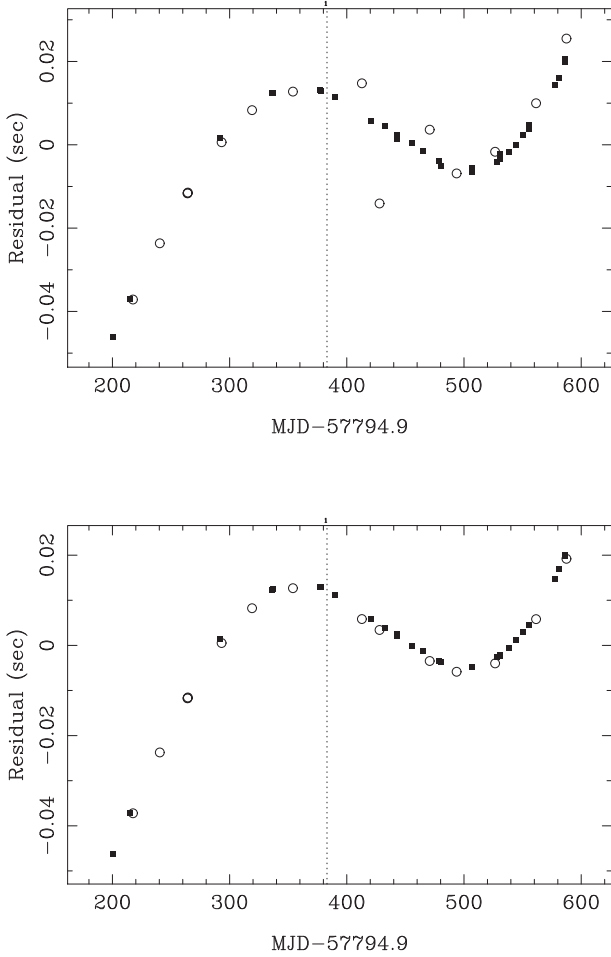
**Figure 8.** Histogram of  $R$  for all 300 pulsars in the UTMOST data release.



**Figure 9.** Recovered frequency path (*top*) and posterior distribution of  $f$  (*bottom*) for the HMM analysis of the UTMOST observations of PSR J1709–4429, laid out as in Fig. 6. Before the glitch, which occurs during the 81st ToA gap, the posterior distribution of  $f$  is well constrained, showing only a narrow band of support near  $f = 0$  Hz. After the glitch, the posterior distribution of  $f$  has support in three distinct  $f$  regions, separated by  $1/(1 \text{ sidereal day})$ . The parameters used in this analysis are reported in Table A1.



**Figure 10.** Local spin frequency estimation of the glitch size  $\Delta f$  for the glitch in PSR J1709–4429 at MJD 58178 using UTMOST observations, laid out as in Fig. 5. The error bars on the pre-glitch frequency estimates are too small to be seen.



**Figure 11.** Timing residuals for combined UTMOST and Parkes observations of PSR J1709–4429 for two timing models with glitch sizes  $\Delta f/f = 54.6 \times 10^{-9}$  (top) and  $\Delta f/f = 2429.7 \times 10^{-9}$  (bottom). In each plot, the unphysical phase jump  $\Delta\phi$  has also been adjusted to minimize the jump in residuals before and after the glitch. UTMOST observations are shown as squares, and the Parkes observations are shown as circles. In both plots, all other timing model parameters besides  $\Delta f$  and  $\Delta\phi$  are identical to those published in the UTMOST data release.

sidereal day, with a typical dwell time of 5–20 min (Lower et al. 2020). Thus, the ToAs used in the UTMOST timing programme are collected in clusters separated by integer multiples of 1 sidereal day (86 164 s), closely matching the observing period of 86 158 s seen in the UTMOST data set for PSR J1452–6036. In most cases, the cadence of observations is 3–7 d.

With this in mind, we check to see if periodic scheduling compromises the phase reconstruction of UTMOST pulsars. We calculate  $\langle |\epsilon_i| \rangle$  [where  $\epsilon_i$  is defined in equation (3)] for each of the 300 pulsars in the UTMOST data release, using only ToAs measured after 2017 September in transit mode, and assuming  $T = 86\,164$  s. Recall that  $\langle |\epsilon_i| \rangle$  measures the magnitude of timing residuals (measured in terms of cycles), which can be expected if one chooses a timing model that displaces  $f$  by  $1/T$  from its true value, where  $T$  is the scheduling period. Fig. 8 shows a histogram of  $R$  [defined in equation (6)], for the 300 targets. Data sets with  $R \gtrsim 1$  are relatively unlikely to support multiple glitch models with comparable residuals. The residuals induced by the wrong choice of  $\Delta f$  are so large, that the error is obvious. However, there is a sizeable population of pulsars with  $\langle |\epsilon_i| \rangle / \langle \sigma_{\text{ToA}} \rangle < 1$ , including two pulsars that are

known to have glitched since 2017 September: PSR J1709–4429 ( $\langle |\epsilon_i| \rangle / \langle \sigma_{\text{ToA}} \rangle = 0.81$ ) and PSR J1452–6036 ( $\langle |\epsilon_i| \rangle / \langle \sigma_{\text{ToA}} \rangle = 0.12$ ). The glitch in PSR J1709–4429 has not been reported by any timing programmes other than UTMOST. Given the periodic observation schedule, the possibility that the reported glitch sizes are in error by  $1.16 \times 10^{-5}$  Hz or more must be taken seriously. We now examine the two objects in turn.

### 5.1 PSR J1709–4429

Lower et al. (2018) reported a glitch in PSR J1709–4429 at MJD  $58178 \pm 6$  of size  $\Delta f/f = (52.4 \pm 0.1) \times 10^{-9}$ , based on UTMOST data analysed with TEMPO2 and TEMONEST. An updated parameter estimate was published by Lower et al. (2020), who gave a glitch size of  $\Delta f/f = (54.6 \pm 1.0) \times 10^{-9}$ , again based only on the UTMOST data. An HMM-based analysis of these data recovers a glitch size of  $\Delta f/f = 2405 \times 10^{-9}$ , as shown in Fig. 9. The glitch size recovered by the HMM is larger than the Lower et al. (2020) result by  $2.29 \times 10^{-5}$  Hz – roughly 2/(1 sidereal day). This suggests that at least one of these methods is confounded by the periodicity of the observations, but without further information it is difficult to decide which glitch size is closer to the truth.

Re-processing the UTMOST data to produce multiple ToAs per observation session allows us to estimate the local spin frequency post-glitch, as described in Section 4.1.2. The results of this exercise are shown in Fig. 10. The post-glitch frequency measurements appear to be centred around roughly  $2 \times 10^{-5}$  Hz, consistent with the HMM estimate. Note that preceding the glitch, this pulsar was not observed with sufficiently long observations and sufficient sensitivity to extract enough high-quality ToAs per observing session for a useful local frequency fit. Each of the pre-glitch frequency estimates therefore incorporates ToAs from observing sessions separated by multiple days, and the error bars are correspondingly much smaller than the post-glitch frequency estimates. The use of widely separated ToAs is not a concern for the pre-glitch frequency estimates, as the pre-glitch frequency is not in question, having been well measured by UTMOST before the switch to a periodic observing schedule, and this is the first glitch since the switch.

Fortunately, the pulsar timing programme carried out at the Parkes radio telescope has released public data covering the period during which the glitch was reported (Hobbs et al. 2011), and the Parkes pulsar timing programme does not schedule observations with the same regularity as UTMOST. Hence, the combined UTMOST and Parkes data can be expected to estimate the glitch parameters better than the UTMOST data alone. Fig. 11 shows timing residuals for two glitch models with the combined UTMOST and Parkes data. With the combined data, the glitch model with  $\Delta f/f = 2429.7 \times 10^{-9}$  is clearly preferred, close to what was recovered in the HMM and local frequency estimation analyses.

As a final step, we re-estimate the glitch parameters with the combined UTMOST and Parkes data using TEMONEST to fit for the timing model parameters and the noise parameters. We start with a pulse numbering derived from a timing model with glitch size  $\Delta f/f = 2429.7 \times 10^{-9}$ . The glitch model used follows Lower et al. (2020), and is given by

$$\begin{aligned} \Delta\phi_g(t) = & \Delta\phi + \Delta f(t - t_g) \\ & + \frac{1}{2} \Delta \dot{f} (t - t_g)^2 - \Delta f_d \tau_d e^{-(t-t_g)/\tau_d}. \end{aligned} \quad (15)$$

The obtained glitch parameters are shown in Table 1. The unphysical phase jump  $\Delta\phi = -0.342 \pm 0.005$  arises because we choose to fix the glitch epoch at MJD 58178, as in the analysis of Lower et al.

**Table 1.** Estimated glitch parameters for the glitch in PSR J1709–4429 at MJD 58178. The glitch model parameters are defined in equation (15).

	$t_g$ MJD	$\Delta\phi$	$(\Delta f + \Delta f_d)/f$ $\times 10^{-9}$	$\Delta \dot{f}/f$ $\times 10^{-3}$	$\Delta f/f$ $\times 10^{-9}$	$\tau_d$ d
This work	$58178 \pm 6$	$-0.342 \pm 0.005$	$2432.2 \pm 0.1$	$4.7 \pm 0.3$	$<0.1$	–
Lower et al. (2020)	$58178 \pm 6$	0.372	$54.6 \pm 1.0$	$1.06_{-0.43}^{+0.36}$	$54.3 \pm 1.0$	$99.1_{-9.6}^{+11.3}$

(2020). We report only an upper limit on the size of the exponentially decaying term,  $\Delta f_d$ . The posterior distribution for  $\Delta f_d$  has support between the lower end of the prior range,  $10^{-18}$  Hz, and  $10^{-9}$  Hz, but no support above  $10^{-9}$  Hz. The decay time-scale  $\tau_d$  is unconstrained – the posterior distribution has significant support across the entire prior range of (1, 1000) d. Previous glitches of this pulsar have been measured with an exponentially decaying component roughly 1 per cent as large as the permanent frequency jump, with a decay time-scale of approximately 100 d (Yu et al. 2013).

## 5.2 PSR J1452–6036

Lower et al. (2020) also reported a glitch in PSR J1452–6036 at MJD  $58600.29 \pm 0.05$ , with a glitch size of  $\Delta f/f = (270.7_{-0.4}^{+0.3}) \times 10^{-9}$  and no measured  $\Delta \dot{f}$ . In the following, we keep  $\Delta \dot{f}$  fixed at zero for simplicity.

Local frequency estimation with the UTMOST data does not constrain the glitch size – the typical ToA uncertainty is high, roughly 1 ms, and so the corresponding phase error is  $\sim 6 \times 10^{-3}$ . Based on the arguments of Section 4.1.2, an observation session long enough to break the degeneracy between glitch models would be roughly 1 h. This is much longer than the actual  $\sim 5$  min observation sessions. The maximum possible observation time in a single transit with UTMOST for this pulsar is somewhat longer, roughly 20 min, but not long enough to break the degeneracy. In general, the maximum time that UTMOST can observe a given pulsar depends strongly on its declination: Pulsars near the ecliptic transit the primary beam in approximately 10 min, while pulsars near the south celestial pole remain in the primary beam for hours. However, most pulsars are not routinely observed by UTMOST for more than 10 min at a time (Jankowski et al. 2019). Alternatively, the per-ToA uncertainty that would allow a 5 min observation session to break the degeneracy is roughly 0.1 ms.

We may apply the arguments of Section 2.4 to search for excess post-glitch phase residuals due to misestimation of the glitch size. We set  $P_{fa} = 0.01$ , and find that the threshold calculated according to equation (9) is  $\gamma_{th} = 1.99$ . The probability of detection given by equation (11) is only  $P_d = 0.07$ . We search for the induced residual signal defined by equation (7) in the residuals of three glitch models: the originally reported model, a model with  $\Delta f$  increased by  $1/(1$  sidereal day) to give  $\Delta f/f = 2069 \times 10^{-9}$ , and a model with  $\Delta f$  increased by  $2/(1$  sidereal day) to give  $\Delta f/f = 3868 \times 10^{-9}$ . The residuals for each glitch model are shown in Fig. 12. We find in each case that the test statistic  $\gamma = \mathbf{R}^T \mathbf{C}^{-1} \mathbf{S}$  does not exceed  $\gamma_{th}$ ; i.e. we do not detect an induced phase error.  $\gamma$  is largest for the glitch model with  $\Delta f/f = 3868 \times 10^{-9}$ , where we calculate  $\gamma = 0.62$ . As  $P_d$  is low, this non-detection does not allow us to constrain the true parameters of the glitch.

Jankowski, Keane & Stappers (2021) recently reported on wide-band observations of this pulsar at the Parkes radio telescope, which by chance happen to lie on either side of this glitch. Combining these observations with the UTMOST data, they report a glitch size  $\Delta f/f = 270.52(3) \times 10^{-9}$ , consistent with the value found by Lower

et al. (2020). As with the PSR J1709–4429 observations, the Parkes observations of PSR J1452–6036 are not on the same schedule as the UTMOST observations, and so the estimate of  $\Delta f$  from the combined data is not confounded by the periodicity of the UTMOST observations.

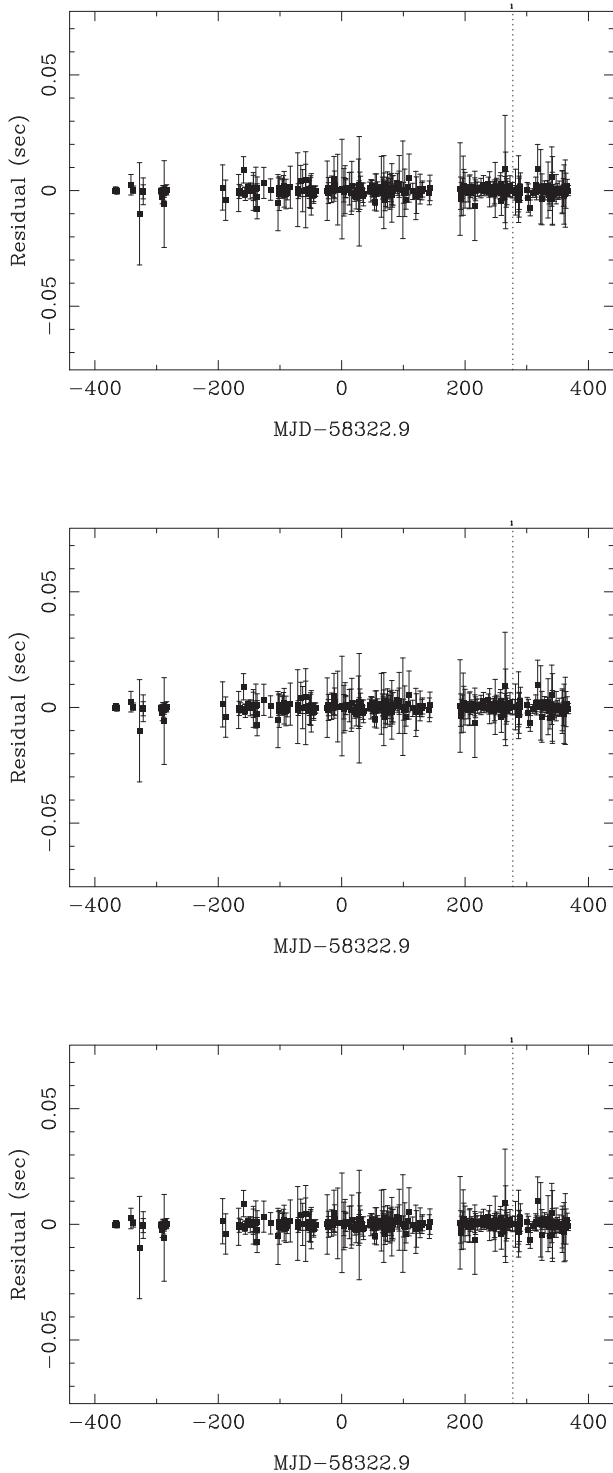
## 6 CONCLUSION

In this paper, we show that periodic scheduling of pulsar observations can lead to erroneous estimates of frequency in pulsar timing models if the frequency is not well measured a priori. We examine in detail the effect this has on the estimation of pulsar glitch parameters. Specifically, the estimated permanent frequency jump  $\Delta f$  may be displaced from its true value by an integer multiple of  $1/T$ , where  $T$  is the scheduling period. We find that in certain existing data sets the excess timing residuals induced by misestimating the spin frequency of the pulsar due to periodic scheduling are comparable to the stochastic residuals induced by ToA measurement error.

We find that ‘by-eye’ attempts to restore phase connection through the use of pulse numbering can bias the recovered glitch size towards smaller values. When the true value of  $\Delta f$  is larger than  $1/T$ , this bias yields incorrect estimates of  $\Delta f$ . Local frequency estimation can mitigate this bias, as long as the ToAs used in the fits are sufficiently accurate. An HMM-based approach may also fail to recover the correct  $\Delta f$ . However, the HMM does not appear to be biased in the same way as the phase-coherent timing-based method, and the existence of multiple solutions is readily apparent from a brief inspection of the products of the analysis, e.g. the posteriors of  $f(t_1), \dots, f(t_N)$ .

We re-evaluate two glitches detected by UTMOST, in PSR J1709–4429 and PSR J1452–6036. In the case of PSR J1709–4429, additional public data from the Parkes radio telescope break the degeneracy in glitch models, and we recover a new glitch size  $\Delta f/f = (2432.2 \pm 0.1) \times 10^{-9}$ , much larger than the previously reported value  $\Delta f/f = (54.6 \pm 1.0) \times 10^{-9}$ . For PSR J1452–6036, we attempt to detect the presence of phase residuals induced by an incorrect choice of glitch model by cross-correlating the observed residuals with the expected induced signal. However, we are unable to definitively confirm or exclude the previously reported glitch model using the UTMOST data alone, because the noise is relatively high, and there are relatively few post-glitch observations. Recently released complementary observations from the Parkes radio telescope break the degeneracy and confirm that the glitch size is  $\Delta f/f = 270.52(3) \times 10^{-9}$  (Jankowski et al. 2021).

In view of the issues raised here, we recommend that wherever possible, future observing campaigns aimed at glitch measurement should avoid periodic observation scheduling. In cases where periodic observations are unavoidable, such as the CHIME/Pulsar (Ng 2018; Good et al. 2020) and future UTMOST-2D (Venkatraman Krishnan et al. 2020) programmes, care should be taken when making inferences about the frequency of a pulsar, particularly after a glitch has occurred. A small number of complementary observations from



**Figure 12.** Timing residuals for UTMOST observations of PSR J1452–6036. Glitches are indicated by the vertical dotted line, with sizes  $\Delta f/f = 270 \times 10^{-9}$  (top),  $2069 \times 10^{-9}$  (middle), and  $3868 \times 10^{-9}$  (bottom). The post-glitch residuals appear to be identical, despite the significant differences in glitch sizes in the three timing models. Other timing model parameters match those released in the UTMOST data release.

another observatory can also help break the degeneracy between glitch models.

## ACKNOWLEDGEMENTS

Parts of this research are supported by the Australian Research Council Centre of Excellence for Gravitational Wave Discovery (OzGrav) (project number CE170100004) and Australian Research Council Discovery Project DP170103625. LD is supported by an Australian Government Research Training Program Scholarship and by the Rowden White Scholarship. MEL receives support from the Australian Research Council Laureate Fellowship FL15010014 and Commonwealth Scientific and Industrial Research Organisation Astronomy and Space Science. The authors are grateful to Bill Moran, Rob Evans, and especially Sofia Suvorova for work on the HMM and useful discussions. The authors are also grateful to the anonymous referee for their helpful comments.

## DATA AVAILABILITY

The public UTMOST data underlying this work are available at <https://github.com/Molonglo/TimingDataRelease1/>. The other data underlying this work will be shared on reasonable request to the corresponding author.

## REFERENCES

- Ashton G., Lasky P. D., Graber V., Palfreyman J., 2019, *Nat. Astron.*, 3, 1143  
 Bailes M. et al., 2017, *Publ. Astron. Soc. Aust.*, 34, E045  
 Basu A., Joshi B. C., Krishnakumar M. A., Bhattacharya D., Nandi R., Bandhopadhyay D., Char P., Manoharan P. K., 2020, *MNRAS*, 491, 3182  
 Dodson R. G., McCulloch P. M., Lewis D. R., 2002, *ApJ*, 564, L85  
 Edwards R. T., Hobbs G. B., Manchester R. N., 2006, *MNRAS*, 372, 1549  
 Espinoza C. M., Lyne A. G., Stappers B. W., Kramer M., 2011, *MNRAS*, 414, 1679  
 Espinoza C. M., Antonopoulou D., Stappers B. W., Watts A., Lyne A. G., 2014, *MNRAS*, 440, 2755  
 Feroz F., Hobson M. P., Bridges M., 2009, *MNRAS*, 398, 1601  
 Good D. C. et al., 2020, preprint ([arXiv:2012.02320](https://arxiv.org/abs/2012.02320))  
 Hobbs G. et al., 2011, *Publ. Astron. Soc. Aust.*, 28, 202  
 Hobbs G. B., Edwards R. T., Manchester R. N., 2006, *MNRAS*, 369, 655  
 Jankowski F. et al., 2019, *MNRAS*, 484, 3691  
 Jankowski F., Keane E. F., Stappers B. W., 2021, *MNRAS*, 504, 406  
 Janssen G. H., Stappers B. W., 2006, *A&A*, 457, 611  
 Lentati L., Alexander P., Hobson M. P., Feroz F., van Haasteren R., Lee K. J., Shannon R. M., 2014, *MNRAS*, 437, 3004  
 Levy B. C., 2008, *Principles of Signal Detection and Parameter Estimation*. Springer US, New York  
 Lorimer D. R., Kramer M., 2004, *Handbook of Pulsar Astronomy*. Cambridge Univ. Press, Cambridge  
 Lower M. E. et al., 2018, *Res. Notes Am. Astron. Soc.*, 2, 139  
 Lower M. E. et al., 2020, *MNRAS*, 494, 228  
 Marshall F. E., Gotthelf E. V., Middleditch J., Wang Q. D., Zhang W., 2004, *ApJ*, 603, 682  
 Melatos A., Dunn L. M., Suvorova S., Moran W., Evans R. J., 2020, *ApJ*, 896, 78  
 Ng C., 2018, *Proc. IAU Symp. 337, Pulsar Astrophysics the Next Fifty Years*. Kluwer, Dordrecht, p. 179  
 Parthasarathy A. et al., 2020, *MNRAS*, 494, 2012  
 Rabiner L. R., 1989, *Proc. IEEE*, 77, 257  
 Shannon R. M., Lentati L. T., Kerr M., Johnston S., Hobbs G., Manchester R. N., 2016, *MNRAS*, 459, 3104  
 Taylor J. H., 1992, *Phil. Trans. R. Soc. Ser. A*, 341, 117  
 Venkatraman Krishnan V. et al., 2020, *MNRAS*, 492, 4752  
 Wong T., Backer D. C., Lyne A. G., 2001, *ApJ*, 548, 447

Yu M., Liu Q., 2017, *MNRAS*, 3041, 3031  
 Yu M. et al., 2013, *MNRAS*, 429, 688

## APPENDIX A: HMM RECIPE AND PARAMETERS

The HMM-based analysis described in Section 4.3 involves choosing a number of input parameters. In this appendix, we briefly describe these choices. For a more detailed discussion of the considerations involved in choosing the analysis parameters, see Melatos et al. (2020). There are broadly three classes of parameters involved: those that specify the  $(f, \dot{f})$  pairs under consideration, those that specify the connection between observations (ToAs) and  $(f, \dot{f})$  pairs, and those that specify the probabilities of transitions between  $(f, \dot{f})$  pairs. A complete list of parameters is given in Table A1.

We specify the allowed range of frequencies and frequency derivatives in two stages. First, we specify a fiducial phase evolution by fixing the frequency  $f_0$  and frequency derivative  $\dot{f}_0$  at a reference epoch  $T_0$ , derived from a Taylor expansion computed by TEMPO2. The HMM tracks deviations away from this fiducial model on a discrete grid in the  $f$ - $\dot{f}$  plane. The range of allowed deviations is specified by lower ( $f_-$ ,  $\dot{f}_-$ ) and upper ( $f_+$ ,  $\dot{f}_+$ ) bounds. The discretization is specified by bin sizes  $\eta_f$  and  $\eta_{\dot{f}}$ .

To incorporate timing noise, we adopt a simple prescription that drives the second frequency derivative with a white noise term  $\xi(t)$  satisfying  $\langle \xi(t)\xi(t') \rangle = \sigma^2 \delta(t-t')$  [see section 3.4 in Melatos et al. (2020)]. Other, similar prescriptions yield similar results (Melatos et al. 2020). The free parameter  $\sigma$  controls the magnitude of the timing noise in the model. The discrete nature of the  $f$ - $\dot{f}$  grid sets a lower bound on  $\sigma$ : Errors in the estimation of  $\dot{f}$ , caused by the finite bin width  $\eta_{\dot{f}}$ , cause the frequency to spuriously wander across a ToA gap by an amount  $\delta f_{\text{bin}} \sim \eta_{\dot{f}} \Delta t_i$ . It is desirable that the HMM ‘correct’ this spurious wandering, through the freedom allowed in the timing noise model. With  $\xi(t)$  defined as above, the frequency wandering across  $\Delta t_i$  is given by  $\delta f_{\text{TN}} = \sigma (\Delta t_i)^{3/2}$ . Equating  $\delta f_{\text{bin}}$  and  $\delta f_{\text{TN}}$  with  $\Delta t_i$  replaced by its average over all ToAs,  $\langle \Delta t_i \rangle$ , suggests the following approximate lower bound:

$$\sigma \geq \eta_{\dot{f}} \langle \Delta t_i \rangle^{-1/2}. \quad (\text{A1})$$

In the analyses presented here, we choose  $\sigma = \eta_{\dot{f}} \langle \Delta t_i \rangle^{-1/2}$ .

Given a particular ToA gap  $\Delta t_i$ , the probability of a particular  $(f, \dot{f})$  state is calculated as a von Mises distribution (Melatos et al. 2020)

$$L(\Delta t_i; f, \dot{f}) = [2\pi I_0(\kappa)]^{-1} \times \exp\{\kappa \cos[2\pi(\Delta t_i f - \Delta t_i^2 \dot{f}/2 + \Delta \Phi_i)]\}, \quad (\text{A2})$$

where  $I_0(x)$  is the modified Bessel function of the first kind,  $\kappa$  is a free parameter known as the concentration, and  $\Delta \Phi_i$  is the phase contribution over the ToA gap from the fiducial phase model specified by  $f_0$ ,  $\dot{f}_0$ , and the reference epoch  $T_0$ . This distribution is peaked when the number of cycles accumulated across the ToA gap,  $\Delta t_i f - \Delta t_i^2 \dot{f}/2 + \Delta \Phi_i$ , is an integer (the minus sign appears because we employ a backwards Taylor expansion). To a good approximation, we may identify  $\kappa$  as the reciprocal of the squared uncertainty of the phase. There are two important contributions to  $\kappa$ . One is the uncertainty on individual ToAs. For each ToA gap, the phase uncertainties on the two ToAs that bracket the gap,  $\sigma_{\text{ToA},1}$  and  $\sigma_{\text{ToA},2}$ , contribute independently to the total phase uncertainty across the gap. The second contribution comes from the discrete  $f$ - $\dot{f}$  grid. The binning in frequency contributes a phase uncertainty  $\eta_f \Delta t_i$ , and the binning in frequency derivative contributes a phase uncertainty  $\eta_{\dot{f}} \Delta t_i^2/2$ . Combining all of these uncertainties in quadrature, and recalling the identification of  $\kappa$  with the reciprocal squared phase uncertainty, we arrive at the expression for  $\kappa$  used in these analyses:

$$\kappa = [\sigma_{\text{ToA},1}^2 + \sigma_{\text{ToA},2}^2 + (\eta_f \Delta t_i)^2 + (\eta_{\dot{f}} \Delta t_i^2/2)^2]^{-1}. \quad (\text{A3})$$

Note that  $\kappa$  depends strongly on the length of the ToA gap, and therefore is recalculated for each gap in the data set. More details on the theoretical underpinnings of (A2) and (A3) appear in section 3.3 and appendix C in Melatos et al. (2020).

Finally, we choose a Bayes factor threshold to be used in model selection via the greedy hierarchical algorithm described in section 4.2 of Melatos et al. (2020), when models containing glitches are compared against models with no or fewer glitches. Informed by synthetic data tests described by Melatos et al. (2020), we choose here a threshold of  $K_{\text{th}} = 10^{1/2}$ .

**Table A1.** HMM parameters for the analysis (described in Section 4.3) of the synthetic data set described in Section 3.1 and the PSR J1709–4429 analysis described in Section 5.1. Parameters marked with an asterisk are from the UTMOST data release.

Parameter	Symbol	Units	Synthetic data	PSR J1709–4429
Timing model reference epoch*	$T_0$	MJD	57 600	57 600
Fiducial frequency*	$f_0$	Hz	6.451 939 727 51	9.754 290 04
Fiducial frequency derivative*	$\dot{f}_0$	Hz s <sup>-1</sup>	$-6.038 24 \times 10^{-14}$	$-8.847 \times 10^{-12}$
Frequency deviation	$[f_-, f_+]$	Hz	$[-3, 250] \times 10^{-7}$	$[-3, 250] \times 10^{-7}$
Frequency derivative deviation	$[\dot{f}_-, \dot{f}_+]$	Hz s <sup>-1</sup>	$[-6.04, 6.04] \times 10^{-15}$	$[-4.9, 5.1] \times 10^{-14}$
Frequency bin size	$\eta_f$	Hz	$1.7 \times 10^{-8}$	$1.7 \times 10^{-8}$
Frequency derivative bin size	$\eta_{\dot{f}}$	Hz s <sup>-1</sup>	$1.21 \times 10^{-15}$	$2 \times 10^{-15}$
Timing noise strength	$\sigma$	Hz s <sup>-3/2</sup>	$2.56 \times 10^{-18}$	$2.05 \times 10^{-18}$
Mean ToA uncertainty*	$\langle \sigma_{\text{ToA}} \rangle$	ms	5.327	0.506
Bayes factor threshold	$K_{\text{th}}$	None	$10^{1/2}$	$10^{1/2}$

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