Bayesian strategies for the generation of pulsar time-of-arrival measurements

C. Messenger[1], A. Lommen[2], P. Demorest[3], S. Ransom[3]
[1]Albert Einstein Institüt, Hannover, Germany
[2]Franklin & Marshall College, Lancaster, Pennsylvania, USA
[3]National Radio Astronomy Observatory, Charlottesville, Virginia, USA

The increasing sensitivities of pulsar timing arrays to ultra-low frequency (nHz) gravitational waves promises to make direct gravitational wave detection within the next 5-10 years. While there are many parallel efforts being made in the improvement of telescope sensitivity, in the detection of stable millisecond pulsars and the improvement of the timing software there are reasons to believe that the methods used to accurately determine the time-of-arrival (TOA) of pulses from radio pulsars can be improved upon. More specifically, the determination of the uncertainties on these TOAs, which strongly effect the ability to detect GWs through pulsar timing, may be unreliable. We propose two Bayesian methods for the generation of pulsar TOAs starting from pulsar search-mode data and pre-folded data.

▶ Introduction

Pulsar timing arrays could well be used to detect ultra-low frequency gravitational waves (GWs) within the next 5-10 years [1]. This is an especially exciting prospect given the concurrent efforts of the LIGO-Virgo Scientific collaboration (LVC) who's aim is to make direct detection of GWs (in the $\sim$10-1000 Hz regime) using the 2^{nd} generation of ground based interferometric detectors within the same timescale [2].

In this work we outline the beginnings of a Bayesian approach to the detection of GWs with pulsar timing using simplistic signal and noise models onto which can be built further levels of sophistication in the future. We also focus on a *single piece* of the complete pulsar timing analysis, the generation of time-of-arrival (TOA) measurements. Given a single pulsar observation, this is the arrival time of the average pulse at the telescope where in this context "average" means the sum of pulses produced by "folding" the data with a periodicity equal to the assumed pulse period. It is from these TOAs that pulsar astronomers then model the spin evolution of pulsars taking into account the motion of the radio telescope relative to the pulsar [3]. The presence of GWs in the field between the telescope and the pulsar will result in small shifts in the arrival times of pulses.

Since the definition of the TOA is defined at the telescope and the GW timescale $\gg$ the timescale of a single observation, we are able to neglect any GW effect in the generation of TOAs. We will discuss two different strategies for the estimation of parameters (including the TOA) from two separate starting points, what we will call "search-mode" data and "pre-folded" data. In both cases we perform the analysis using a standard Bayesian integration algorithm in order to obtain posterior probability distributions on the signal parameters.

▶ The signal model

We begin with a dataset defined on a discrete 2-dimensional grid of time t_j versus radio-frequency f_k. In practice this is referred to as "search-mode" data for which we assume the following signal model

$$x(t_j, f_k) = s(t_j, f_k) + n(t_j, f_k), \tag{1}$$

where $x(t_j, f_k)$ represents the data-set, $s(t_j, f_k)$ is the signal and $n(t_j, f_k)$ is the noise which for simplicity we assume is Gaussian distributed with zero mean and unit variance. The signal itself we define as

$$s(t_j, f_k) = \sum_{\alpha=0}^{n-1} A \exp\left[-\left(t_j - \mu_{\alpha k}\right)^2 / 2w^2\right], \tag{2}$$

where α sums over all $n \approx T/P$ Gaussian profile pulses present in the time-series for each frequency channel. We use A as the pulse peak amplitude, w is the pulse width, and the centre of the α'th pulse in the k'th frequency channel is defined as $\mu_{\alpha k} = (\alpha + \phi_k)P$ with P as the pulse period and ϕ,defined on the range $[0, 1)$, is the phase of the first pulse with reference to the observation start $t = 0$. It is related to the more commonly used phase ϕ_0 defined as the phase of the pulse at the midpoint frequency channel f_{mid} and with reference to the midpoint of the observation $t = T/2$ by

$$\phi_k = \text{mod}\left(T/2 + \phi_0 P + \Delta t_k, P\right), \tag{3}$$

where the relative delay due to dispersion in the k'th frequency channel Δt_k is given by

$$\Delta t_k = 4.148808 \times 10^3 \left(f_k^{-2} - f_{\text{mid}}^{-2}\right) D \ \text{ sec} \tag{4}$$

where D is the dispersion measure in cm^{-3}pc and the units of the frequencies are MHz.

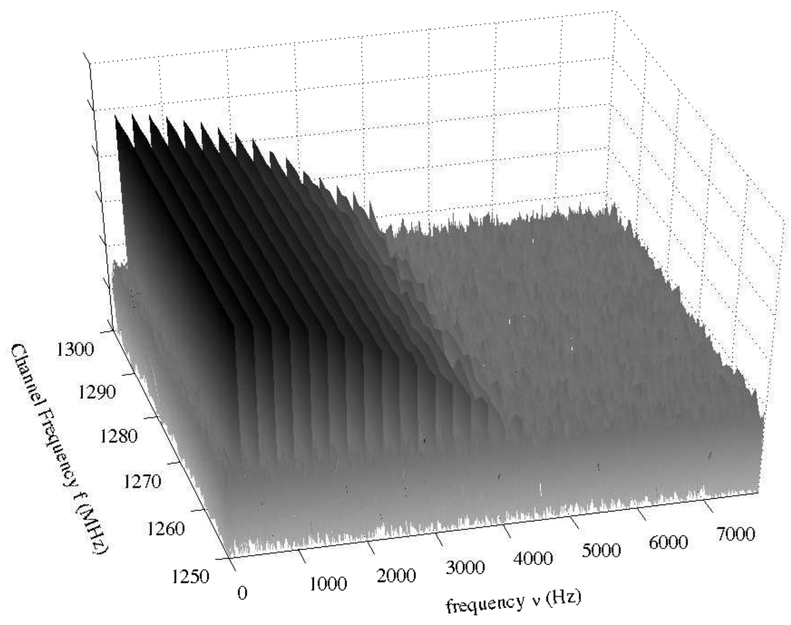

FIG. 1: An example frequency-frequency map computed for a small block of time across 50 frequency channels. The z-axis scale is proportional to the signal power. Note that the 250.003294761 Hz simulated signal is clearly visible as a series of diminishing amplitude harmonics. Not plotted here is the phase, which for a given harmonic as a function of channel frequency, is determine by the dispersion measure.

▶ Method 1 : A frequency domain approach

The signals recieved from pulsars are periodic and their frequency evolution is slow. By Fourier transforming each channel's time-series we find that a pulsar signal can be represented as a series of narrow-band harmonics (see Fig. 1) allowing us to be economic with the data samples used since we will have good prior information on the pulse period. If we now apply the Fourier transform to a complete noise-free signal pulse train we obtain

$$\tilde{s}(\nu_l, f_k) = \frac{AwT}{P}\sqrt{2\pi}\exp\left\{-2(\pi w \nu_l)^2 - 2\pi i \mu_{0k}\nu_l\right\}\tilde{W}_l, \tag{5}$$

where the complex function $\tilde{W}_l$ is given by

$$\tilde{W}_l = \sum_{\alpha=0}^{n-1}\left\{\frac{\sin(2\pi\Delta\nu_{l\alpha}T)}{2\pi\Delta\nu_{l\alpha}T} + i\left[\frac{\cos(2\pi\Delta\nu_{l\alpha}T) - 1}{2\pi\Delta\nu_{l\alpha}T}\right]\right\} \tag{6}$$

and $\Delta\nu_{l\alpha} = \nu_l - \alpha/P$. Note that in this frequency-frequency plane where l indexes the Fourier frequency (as opposed to the channel frequency), we see that each pulse harmonic will be identical in profile for each channel. It will however be rotated in phase by a quantity dependent upon the dispersion measure and the precise frequency value of the discrete Fourier frequency bin in question.

▶ Method 2 : Using pre-folded data

If we consider a dataset that has been already folded at a specific (non-exact) pulse period $P' \neq P$ (see Fig. 2) we can model the signal profile as

$$X(\phi', P') = \sum_{\beta=0}^{n'-1} s\left((\beta + \phi')P', f_k\right), \tag{7}$$

where β indexes each fold up to $n' \approx T/P'$. Substituting in our signal model (Eq. 2) we can accurately approximate the discretely summed pulse profile as

$$X(\phi', P') \approx \frac{A}{2}\sqrt{\frac{\pi}{a}}e^{\left\{\frac{b^2}{4a} - c\right\}}\left[\text{erf}\left(\frac{2an' - b}{2\sqrt{a}}\right) + \text{erf}\left(\frac{b}{2\sqrt{a}}\right)\right] \tag{8}$$

where we have used

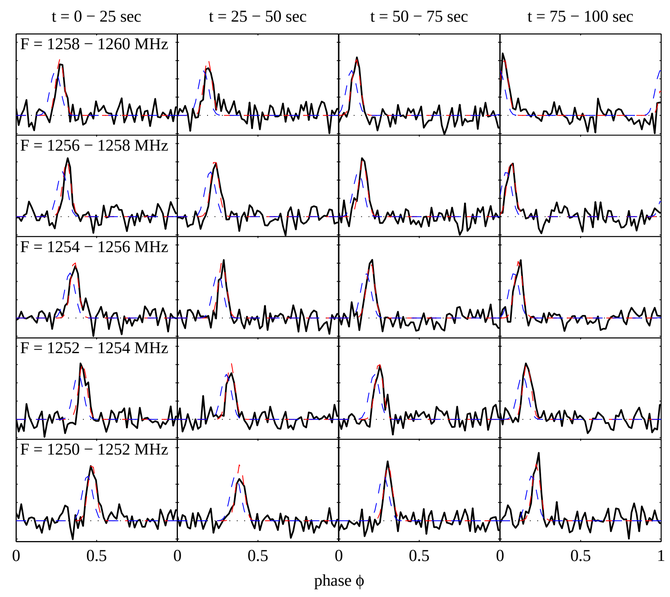

FIG. 2: The folded pulse profiles for a 250.003294761 Hz signal constructed in 25 sec sub-intervals within a 100 sec observation and in 5, 2 MHz wide frequency bands. Shown in solid black are the curves for a simulated signal of pulse amplitude 0.1, dispersion measure 500 cm^{-3}pc and folded with the true signal period. The dashed red curve is the noise free, profile model generated using Eq. 8 with no error in folding period. The dashed blue curve is for identical parameters using a 10^{-5} fractional period error.

$$a = (P' - P)^2/2w^2 \tag{9}$$
$$b = -(P' - P)\left(\phi' P' - \phi P\right)/2w^2 \tag{10}$$
$$c = -\left(\phi' P' - \phi P\right)^2/2w^2 \tag{11}$$

▶ Conclusions and future work

We show in Fig. 3 the results of a Bayesian nested-sampling integration technique implemented with `MultiNest` [4]. The results from this simplistic model indicate that the data reduction procedure of folding the data does not affect our ability to estimate parameters. One does see however, that this is not the case when using an incorrect folding period, even when one accounts for this.

Here we have focused on the parameter estimation of a single TOA. In future work we hope to include more of the physically relevant effects including scattering, polarisation calibration, and detector motion. We also intend to apply this method to some real pulsar data, to allow for more realistic noise behaviour and to extend the analysis to span multiple observations in order to sensibly compare these approaches to existing methods.

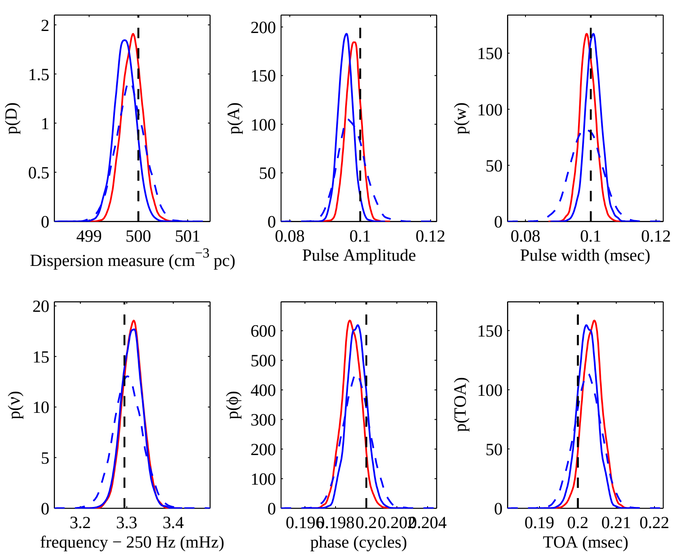

FIG. 3: The posterior probability distributions of the signal parameters. The solid red curves correspond to posteriors generated using the frequency domain approach, and both solid and dashed blue curves correspond to the pre-folded profile approach. The solid blue curve represents the case where the folding period was equal to P. The dashed blue curve represents the case where the folding period was $P' = P(1 + 10^{-5})$. The vertical black dashed lines are the true simulated signal parameter values.

References

[1] G. Hobbs. Pulsar timing array projects. In *IAU Symposium*, volume 261 of *IAU Symposium*, pages 228–233, January 2010.

[2] The LIGO Scientific Collaboration. LIGO: the Laser Interferometer Gravitational-Wave Observatory. *Reports on Progress in Physics*, 72(7):076901–+, July 2009.

[3] G. B. Hobbs, R. T. Edwards, and R. N. Manchester. TEMPO2, a new pulsar-timing package - I. An overview. *MNRAS*, 369:655–672, June 2006.

[4] F. Feroz and M. P. Hobson. Multimodal nested sampling: an efficient and robust alternative to MCMC methods ... *MNRAS*, 384:449–463, February 2008.