Intern report first complete draft

Robert Ni (supervised by Prof. Matthew Bailes and Dr. Emma Carli)

December 22, 2025

Contents

Abstract

This report summarises three analyses on MeerKAT PTA timing data. Project 1 applies subband-level cleaning and selective epoch cuts to unfiltered timing files and compares results to the `first_pass` dataset; combined exclusion of early (2019) and late (2025) epochs reduces dipolar contamination while preserving the Hellings–Downs significance. Project 2 compares evidence for noise components between FD-removed templates and the original `third_pass` using pBilby, finding FD-removed templates often lead to better-constrained parameters but can under-estimate TOA uncertainties for some pulsars. Project 3 studies how removing observations near the Sun affects HD and dipolar SNRs using real and simulated datasets; the results highlight the sensitivity of correlation statistics to solar-wind modelling and ephemeris/systematics. Conclusions and open questions are presented.

1 Introduction

1.1 Pulsars

Pulsars are rapidly rotating neutron stars with strong magnetic fields. In the classic 'lighthouse' model, they emit narrow beams of radio waves from their magnetic poles. Each pulse is observed when the beam sweeps across our line of sight from Earth. The pulse period equals the star's rotation period, which lengthens gradually as the neutron star loses rotational energy to magnetic dipole radiation. Pulsars thus act like huge flywheels (moment of inertia $\gtrsim 10^{38}$ kg m^2) whose spin slows over time. This spin-down has been directly measured – for example, the Crab pulsar (PSR B0531+21) shows the expected slow increase in its period (Richards, D.W., and Comella, J.M., "The period of pulsar NP 0532", *Nature*, 222, 551–552, 1969).

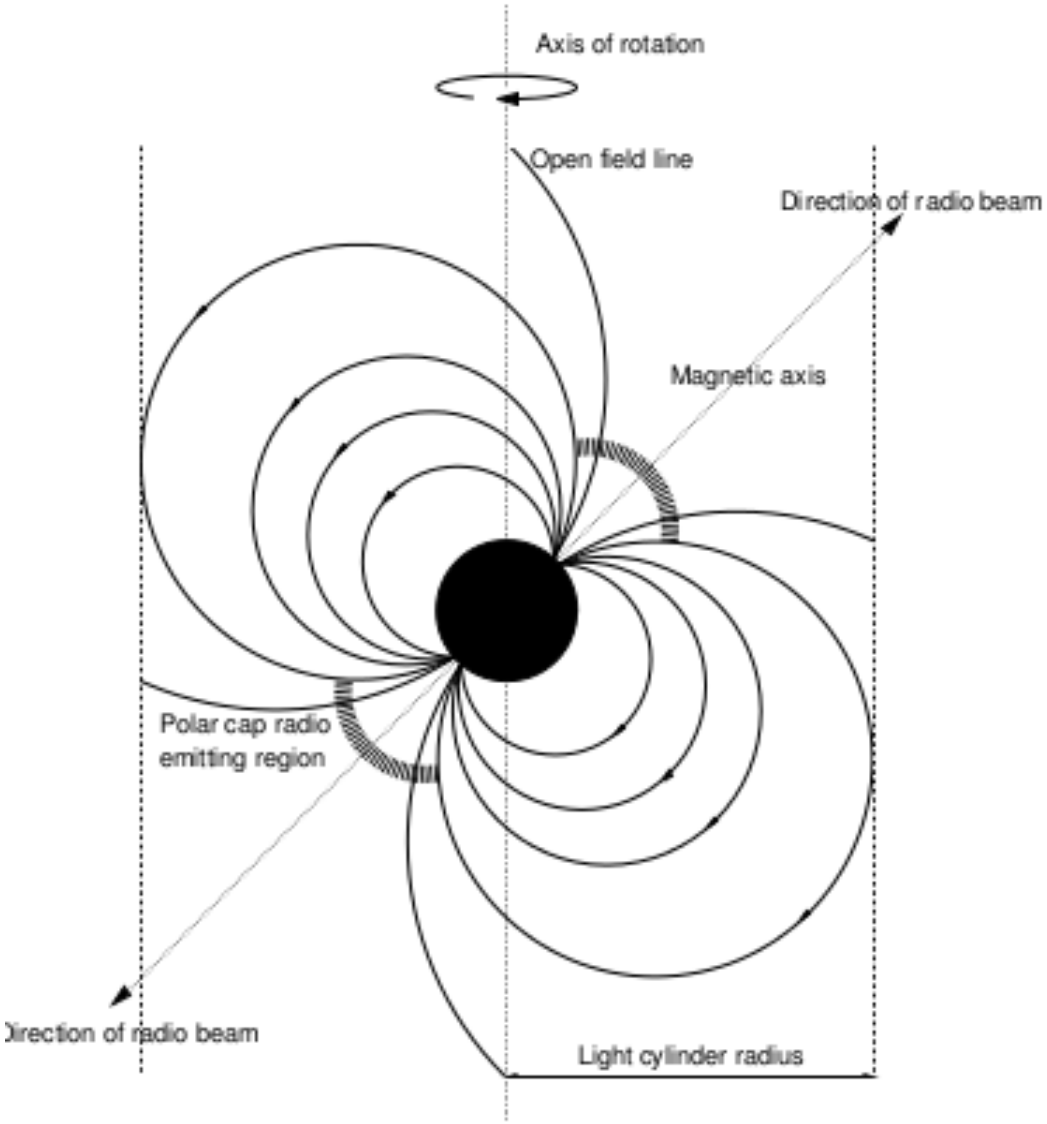


Figure 1: Image 2.

Most known pulsar exhibits pulse periods $P \sim 0.5$ s, which increase at rates $\dot{P} \sim 10^{-15}$ s/s. The relationship is as follow, and is commonly analysed via the P-$\dot{P}$ diagram

(Figure **??**).

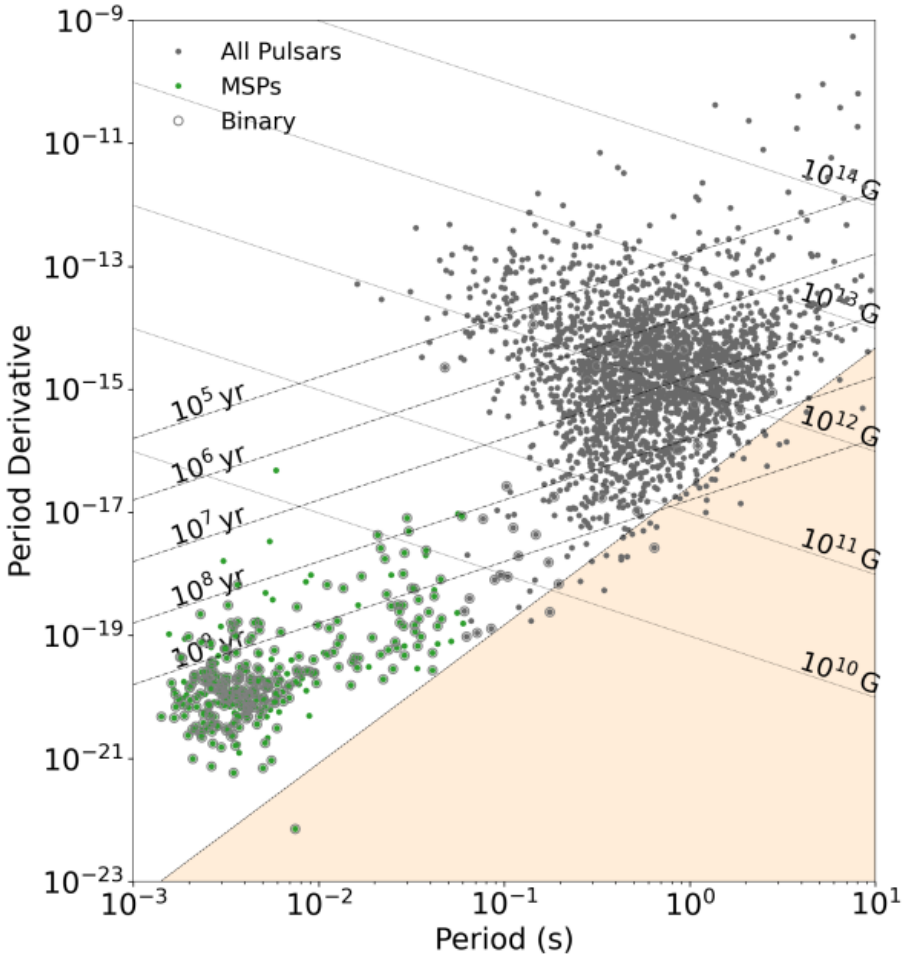


Figure 2: Scatter plot of noise correlations. (Manchester R. N., Hobbs G. B., Teoh A., Hobbs M., 2005)

Marked green in Figure **??** are the population of millisecond pulsars (MSPs), first discovered by Alpar M. A. (1982). Most MSPs are in binary systems (marked by open grey circles). Typical companions are white dwarfs, main-sequence stars, or other neutron stars. A standard formation scenario (see Figure **??**) of millisecond pulsars is the following: the primary star explodes into a neutron star, often separating the binary system or imparting a 'kick'. For those few binaries that remain bound, and in which the companion is sufficiently massive to evolve into a giant and overflow its Roche lobe, the old spun-down neutron star can 'spun-up' as a pulsar by accreting matter and angular momentum at the expense of the orbital angular momentum of the binary system.

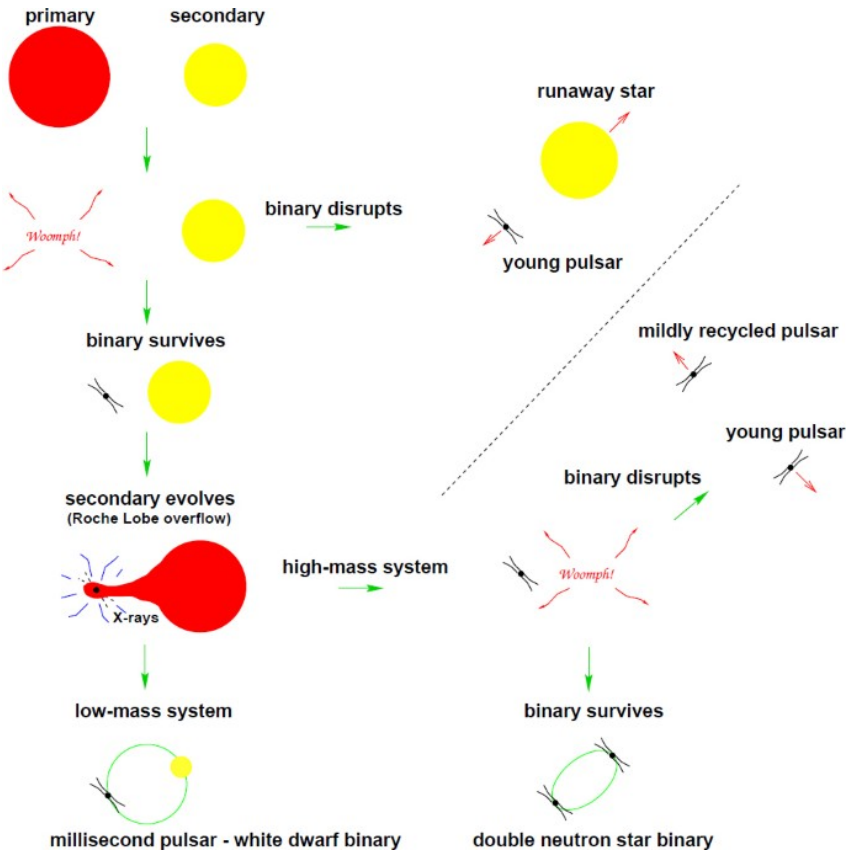


Figure 3: Image 1. (https://link.springer.com/article/10.12942/lrr-2008-8)

1.2 Pulsar timing

Pulsar timing is the precise measurement of pulse times of arrival (TOAs) in order to track every rotation of the neutron star. We obtain the observed pulse profile by averaging 8-second subintegrations over the full observation and 32 frequency channels. This profile is then cross-correlated with a high-SNR standard template to determine the pulse phase. The TOAs (initially referenced to an observatory clock) are then corrected for known delays and transformed into a common reference frame. For example, clock offsets are applied (e.g. UTC→TDB), and the TOAs are converted to barycentric dynamical time at the Solar-System barycentre.

Propagation delays, such as dispersion in the interstellar medium, gravitational propagation effects, and tropospheric delays, alongside the pulsar's own motion (binary orbit, proper motion) are included in the timing model. In practice, one fits a deterministic timing model (including spin frequency, spin-down, position, DM, and orbital parameters if present) to the full set of TOAs.

The observations are written in `psrfits` (Hotan, van Straten & Manchester 2004) format, containing 256s((?)) subintegrations of the pulsar observation at a phase resolution of 1024 bins, with four polarization products, with the early data being recorded with 928 channels3 and the latter data with 1024 frequency channels.

A *subintegration* is the averaging of folded pulse profiles over a short time interval (here typically 256 s) to produce a single profile per subintegration while preserving time resolution across the observation. Subintegrating reduces data volume and increases SNR for each profile used in template matching, but choices in subintegration length (for example

`mode` vs `max`) trade time resolution against per-subintegration SNR and can change TOA uncertainties and error estimation. The `max` files adopt the largest allowable subintegration such that each subintegration meets an SNR threshold; inconsistent subintegration strategies hence affect downstream noise estimates and combining bands (UHF/L-band).

The timing residuals are defined as the differences between the observed TOAs and those predicted by the timing model. These residuals contain the imprints of any unmodeled effects or signals, including white and red noise processes, as well as low-frequency gravitational waves.

In the MeerKAT PTA analysis, TOAs are extracted using a 32-subband frequency resolution scheme, in which each observation is divided into 32 contiguous frequency channels before template matching and TOA estimation (Spiewak et al., 2023). This method preserves chromatic timing information, improves the modelling of dispersion measure variations, and enhances sensitivity to frequency-dependent effects such as scattering and pulse profile evolution.

2 Noise models & timing residuals

Timing residuals include a variety of stochastic noise processes. These are conventionally split into white noise (high-frequency, uncorrelated) and red noise (low-frequency, correlated) components. White noise arises from random processes including instrumental uncertainties and pulse-phase jitter, and is often parametrized phenomenologically by three parameters: EFAC (a multiplication factor on TOA uncertainties), EQUAD (an added white-noise variance, added in quadrature), and ECORR (an epoch-correlated "jitter" term). Red noise consists of timing noise that can mimic the long-timescale signals of interest. It includes achromatic components (frequency-independent), such as intrinsic "spin noise" or irregularities in the pulsar's rotation, and chromatic components (frequency-dependent). In practice, one commonly models one achromatic red process (a power-law spin/noise term) plus chromatic processes for DM variations, scattering delays, solar wind fluctuations, etc.. These noise models are fitted to the timing residuals to isolate deterministic trends, mainly the GWB, from stochastic fluctuations that could smear the signal.

The methods used to determine these noise processes can all be classified as Monte Carlo Markov Chains (MCMC), two different packages with slightly different functionalities will be later introduced and used in projects 2 and 3, namely `tempered_nest` and `ptmcmc`.

3 GWB & PTA

First proposed by Albert Einstein in his infamous general theory of relativity (Einstein, A. (1915) Die Feldgleichungen der Gravitation. Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften, 48, 844-847), gravitational waves have been indirectly detected by Hulse, Taylor, and collaborators by observing binary pulsars in 1976. A century after its proposal, the LIGO experiment provided the first direct detection of gravitational waves.

The MPTA collaboration have detected ? sigma significance GWB in Data Release 2

4 Project 1: Cuts on unfiltered data and comparison to the 'first pass' dataset

4.1 Methods

4.1.1 Obtaining the data

The unfiltered data was obtained in the form of .par(parameter) and .tim(timing) files.

The .par files were obtained directly from: `/fred/oz002/users/mmiles/MPTA_GW/partim_frank/pp`

The .tim files were obtained by concatenating all observation .tim files into a single .tim file for a pulsar: `/fred/oz005/timing_processed/{psrname}/2*/?/timing/PTA/J*_chopped.32ch_1`

As seen in the naming convention, the tim files chosen will possess the following properties:

- 32 sub-channels (explained in intro)
- 1 total polarisation product – only the total intensity kept, polarisation information is discarded
- 'mode' refers to subintegrating once over one usual observation period (256s).
  - Sometimes these files were empty, 'max' files were used instead, which means the observation was subintegrated the maximum amount of time for the SNR of each subintegration to be higher than 12.

4.1.2 Cleaning the data

The method for cleaning the data imitated the procedure in the ASTRAL report, with slight modifications in the code used and in the specific cutoff thresholds applied.

All cuts were applied at the subband level, meaning that individual sub-channels failing the filtering criteria could be removed without discarding the entire observation. However, this creates the issue that some observations retain only a small number of subbands. In such cases, a single subband may disproportionately represent the entire observation, introducing artifacts such as systematic offsets and complications in modeling chromatic effects. To mitigate this, all observations with fewer than five surviving subbands were removed.

The signal-to-noise ratio (SNR) cutoff was set at 10. All observations with TOA uncertainties greater than 10 microseconds were discarded.

The minimum observation length was set to 128 seconds, and long observations made with the UHF band were removed because they are difficult to combine consistently with L-band observations due to differing frequency coverage and noise characteristics. (This issue was later addressed by others to make use of the UHF band data, but was not included in this analysis)

Further, all observations with reduced chi-squared values (per subband) greater than 3.6 were excluded, based on the analysis performed in the ASTRAL report.

For some pulsars, the problem arose that most of their observations failed a particular criterion (e.g., large TOA uncertainty), leaving very little usable data after cleaning. To address this, a safeguard was implemented: if a single criterion rejected more than 30% of a pulsar's data, that criterion was relaxed, and the cutoff for that pulsar was determined manually (see appendix for specific cutoffs).

Two pulsars, J1843-1448 and J0900-3144, were removed from the dataset entirely, as consistent removal criteria could not be determined for them.

4.2 Results and analysis

In this project and project 3, the optimal statistic package was used to estimate the significance of gravitational waves in the data. It is a fast, frequentist estimator that computes a weighted sum of cross-correlations between pulsar timing residuals to estimate the amplitude and SNR of spatial correlation templates (e.g. the Hellings–Downs quadrupole or a dipole). It is useful for quick scans and for comparing template SNRs across many data cuts.

As shown in Figure 1, it is straightforward to overplot a dipolar correlation on the same axes as the quadrupolar Hellings–Downs curve, and by eye the two can be difficult to distinguish—raising the concern that an apparent HD signal could instead be a dipolar signature arising from ephemeris errors.

Quantitatively, although the optimal-statistic returns a high signal-to-noise ratio (SNR) for the HD template, it also yields a comparably high SNR for the dipole template. This suggests that the putative HD detection may be contaminated by solar-system ephemeris errors or other systematics that mimic a dipolar pattern.

To investigate further, I applied a set of data cuts based on several criteria: Sun–pulsar separation (removing observations below various elevation angles), dispersion measure thresholds, exclusion of all data from 2019 and 2025, and manual excision of outliers. The strongest HD SNR overall (15.52) came from the DM$\leq$60 pc cm^{-3} cut. However, when simultaneously excluding both 2019 and 2025 data, the HD SNR remains healthy (8.75) while the dipole SNR falls to only 3.83—yielding the largest disparity between quadrupole and dipole significance.

This outcome is plausible: early MeerKAT observations in 2019 were affected by non-optimal hardware and software settings, and post-2024 timing may suffer from unmodeled solar-wind or ephemeris changes (since our noise model was calibrated on data through mid-2024). Removing both epochs thus minimizes dipolar systematics while preserving a robust HD correlation.

Directory	SNR_HD	SNR_dipolar
GWB_search_2	12.237026420376978	10.582305696757597
GWB_search_2_angle10	12.036049434027658	10.791268727744255
GWB_search_2_angle20	11.457201595506715	10.82832353432228
GWB_search_2_angle30	11.554110414653067	10.879401568394524
GWB_search_2_angle50	7.631584657002278	8.27166492385602
GWB_search_2_dm100	14.276981264457598	13.43022838088923
GWB_search_2_dm20	0.6281049361049306	2.0616303541216223
GWB_search_2_dm40	11.46340934603799	6.658734626707023
GWB_search_2_dm60	15.523151321473124	13.6080501815082
GWB_search_2_dm80	15.388922036086159	13.891987552255918
GWB_search_2_no2019	15.144103974568402	11.902882940519328
GWB_search_2_no2025	6.952835904140578	4.935448839925726
GWB_search_2_no2025_2019	8.74795707430896	3.8307057216285947
GWB_search_2_psrcut	14.111231021468354	10.587572649245462
GWB_search_2_psrcut2	13.636544993177322	10.288101047436156
GWB_search_2_psrcut2_no_pattern	11.066674346944119	9.502086467225705
GWB_search_2_psrcut2_pattern	14.703701066347413	11.360942952496437

Table 1: Summary of optimal-statistic SNRs across different data cuts.

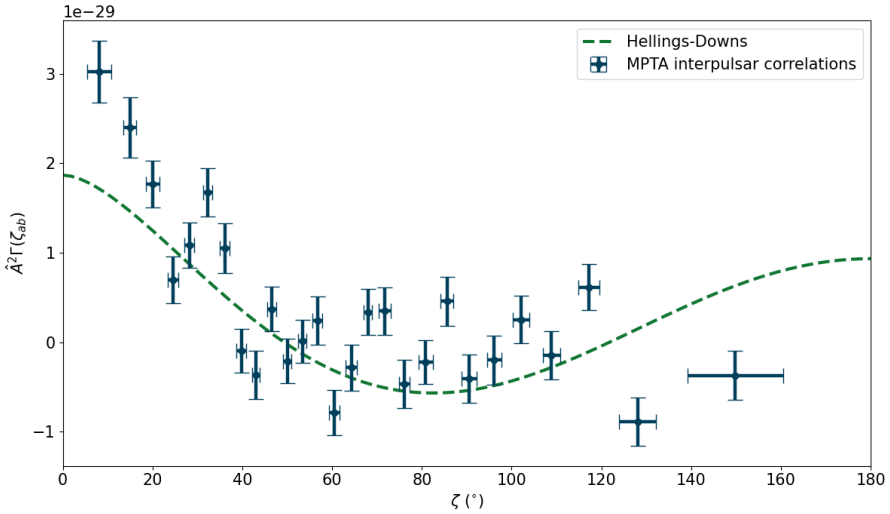


Figure 4: Residuals after cleaning (Project 1).

5 Project 2: Noise analysis with Hendrik's FD removed files

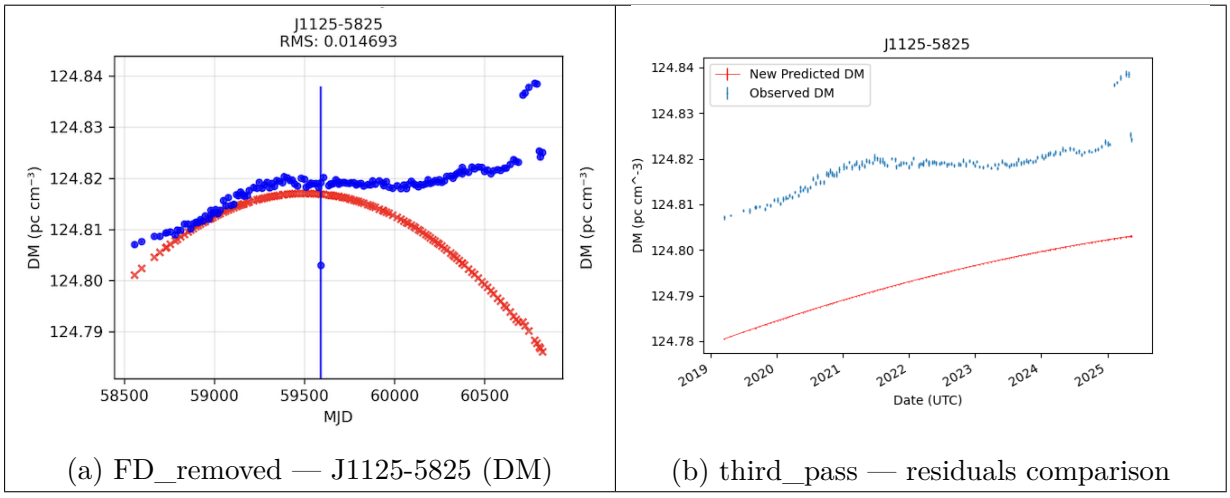
5.1 Method

pBilby (parallel Bilby) is the parallelised interface to the Bilby Bayesian-inference framework. It is well-suited to evidence-based model comparison on individual pulsars because it can compute marginal likelihoods (evidences) efficiently in high-dimensional spaces and supports samplers that handle multimodal posteriors. In this work pBilby was used to compare evidence for white and red noise components between FD-removed and `third_pass` templates, which helps identify where template changes systematically alter parameter constraints or evidence ratios.

In this analysis, I used pbilby to compare the evidence levels of various noise parameters for the same pulsar but created from different template, namely the FD-removed dataset created by Hendrik and the original third_pass dataset.

The pulsars are generally classified as "well behaved", taken from a list given by Hendrik, or "poorly behaving". The criteria used to determine this is the plots of DM_observed vs DM_predicted (using the Taylor series in the .par file). By removing the FD parameters, the new template should recover a DM_predicted aligning closer with DM_observed by removing Freuquency dependent parameter which absorb some of the DM noise. Poorly behaving pulsars generally have a large offset DM_observed and DM_predicted, and/or having a DM_predictred taylor polynomial which doesn't mimic the pattern of DM_observed well.

Here, J1125-5825 is provided as an example as a poorly behaving pulsar, though the pattern resembles decently, there is a large offset of ~ 0.3 DM units remaining even after the FD removal.



(a) FD_removed — J1125-5825 (DM)

(b) third_pass — residuals comparison

Table 2: Comparison of FD_removed and third_pass results.

The first test case is a "well-behaving" pulsar J1911-1114
Evident in the plot below, notice the offset is around 0.002 DM units

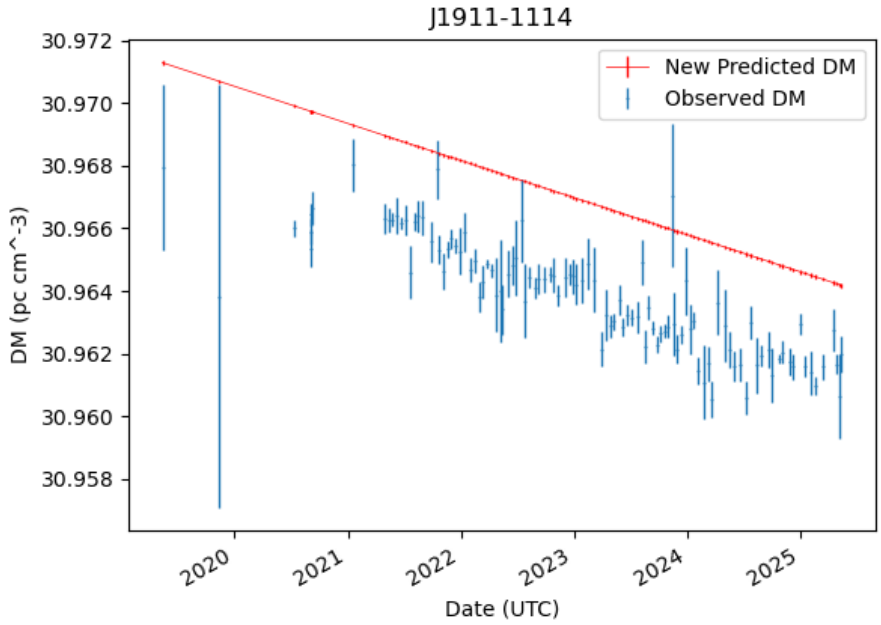


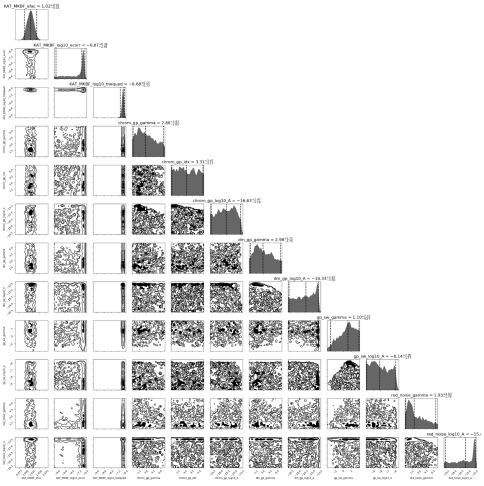
Figure 5: Evidence plot: FD-removed vs third_pass.

When modelling the FD-removed data, significantly more parameters are well-constrained compared to the third_pass data. This is reasonable, as only the DM terms are active in the .par file for the FD-removed data, leading to residual trends being absorbed into the noise model.

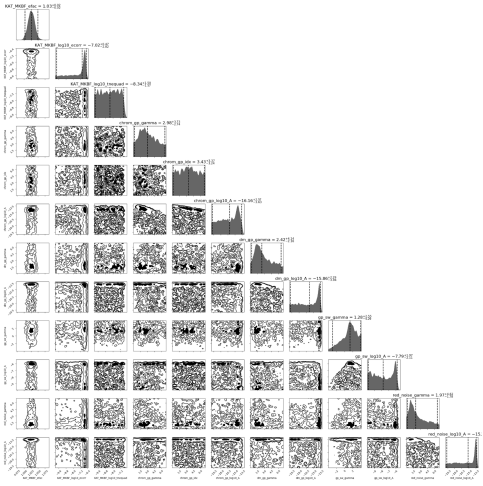
To enable a fairer comparison, we re-fit all timing model parameters in the .par file for the test pulsars. In the case of J1911-1114, most of the peaks in the corner plots appear similar between the two datasets. However, the FD-removed data shows a prominent peak for the TNEQuad parameter, suggesting systematic effects introduced by the new template. This was later determined to be likely caused by an underestimation of TOA uncertainties in the new template.

(For a better full picture, I decided to use the everything model (red + DM + CHROM + SW) instead of the predetermined choice for analysing the pulsars)

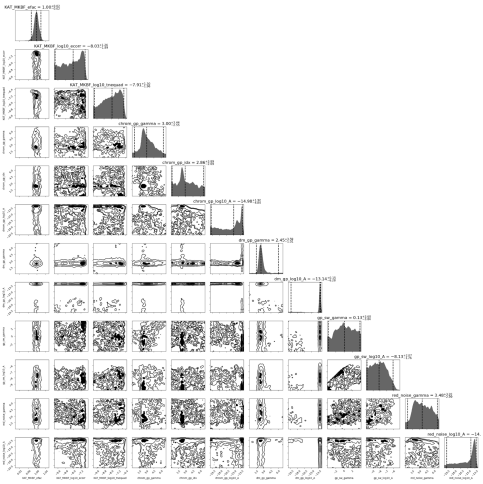
5.1.1 Well behaving pulsars



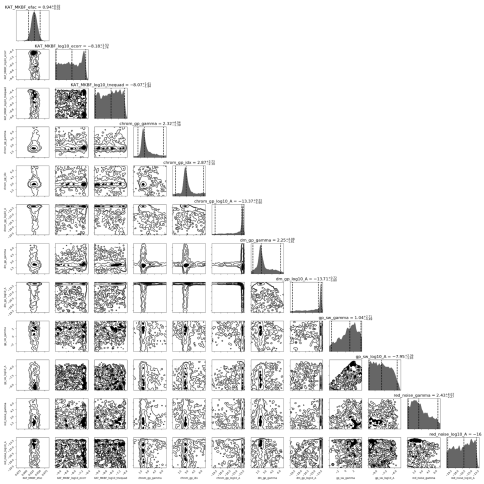
(a) J0711-6830 (FD_removed)



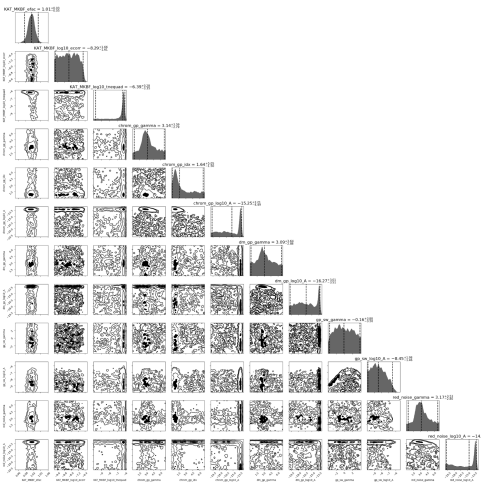
(b) J0711-6830 (third_pass)



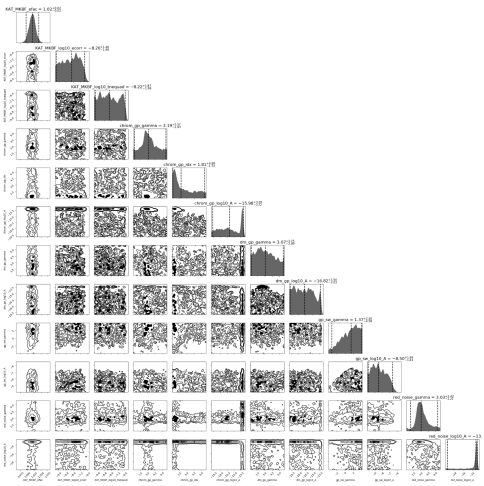
(c) J1216-6410 (FD_removed)



(d) J1216-6410 (third_pass)



(e) J1719-1438 (FD_removed)

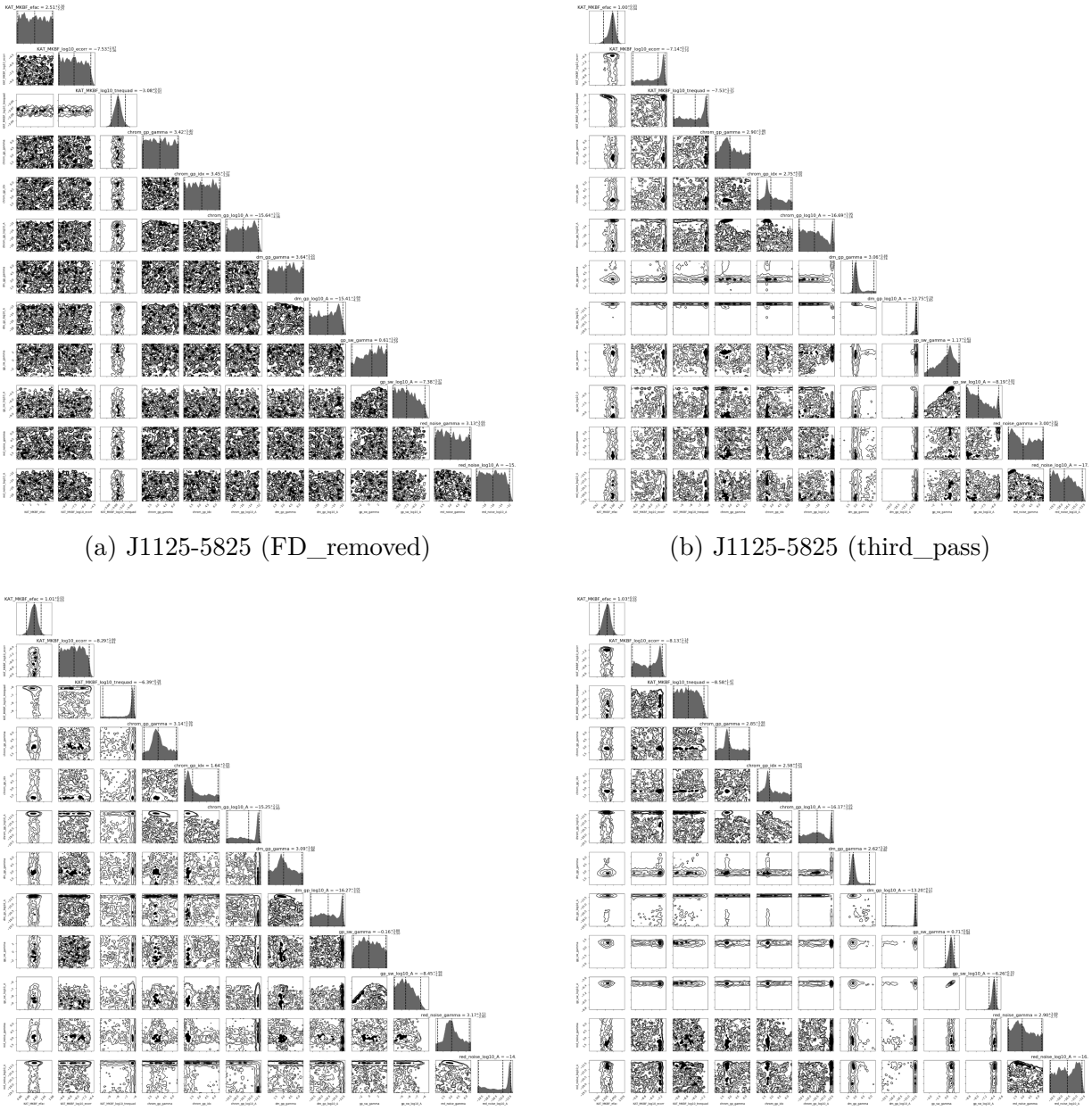


(f) J1719-1438 (third_pass)

Figure 6: Global caption describing the 3×2 panel.

By eye, there is not much of a difference in all the noise parameter evidence except for equad, similar to J1911−1450 (expect for J1216−6410, for this there seems to be some evidence, but not that clear). As above, the reason predicted for this is from underestimation of error bars in the new templates.

5.1.2 Poorly behaving pulsars



(a) J1125-5825 (FD_removed)

(b) J1125-5825 (third_pass)

(c) J1614-2230 (FD_removed)

(d) J1614-2230 (third_pass)

Figure 7: Global caption describing the 2×2 panel.

For the "pooly behaving" pulsars, we see a completely different comparison. It's hard to summarize any specific comparative trends for these pulsars. But we can note the following worrying questions:

- Why is the white noise so different? Is it absorbing any red noise which it shouldn't do?
- Intuitively, we would expect the DM or SW evidence to increase in these pulsars, as they should account for the unmodeled discrepancy between DM_observed and DM_predicted. However, in reality, the opposite occurs—the third_pass dataset shows stronger evidence in almost all red noise posteriors.

6 Project 3: How does removing data in proximity to the sun affect the signal of the Gravitational Wave detected

6.1 Short intro: HD and Dipolar

The Hellings–Downs (HD) curve is the quadrupolar angular correlation expected between pulsar timing residuals from an isotropic stochastic gravitational-wave background. A dipolar correlation instead produces a $\cos(\theta)$-like pattern and is characteristic of systematics that project onto the solar-system frame (for example ephemeris errors or poorly modelled solar wind). Extending the optimal-statistic discussion: comparing HD and dipole SNRs helps diagnose whether observed correlations follow the astrophysical quadrupole expectation or are dominated by systematics. However, as seen later in this analysis, Optimal statistics can conflate these especially at low angular seperations where they look similar, resulting in spuriously high SNRs.

(Maybe add a diagram?? didn't really find a good one)

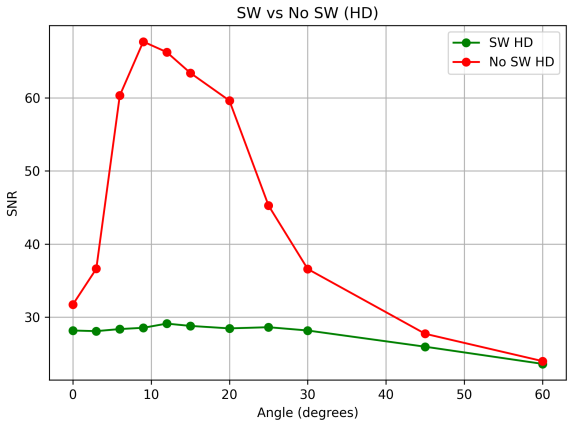
6.2 Aim and datasets

This analysis aimed at testing how removing pulsar data within a certain angle of the Sun will affect the signal of the Gravitational Wave detected.

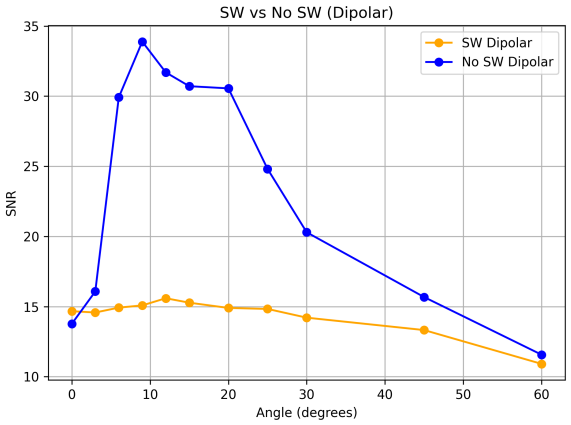
It presents 2 datasets, the real dataset (`third_pass`), and the simulated dataset (created by Saurav, with only efac, sw, and gwb injected). For each of the datasets, I chose the noise models in 2 different ways: with and without Solar wind parameters.

6.3 Simulated dataset results

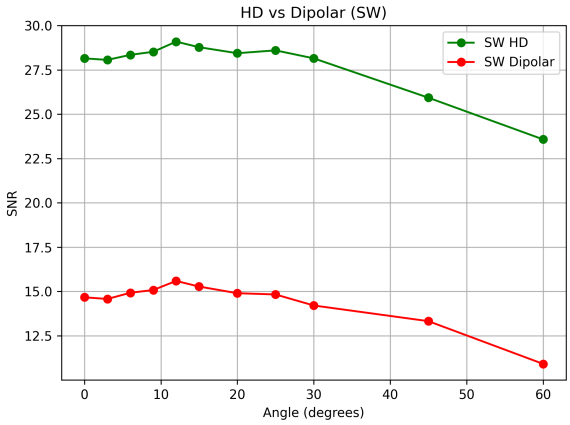
There are a few trends worth noting:



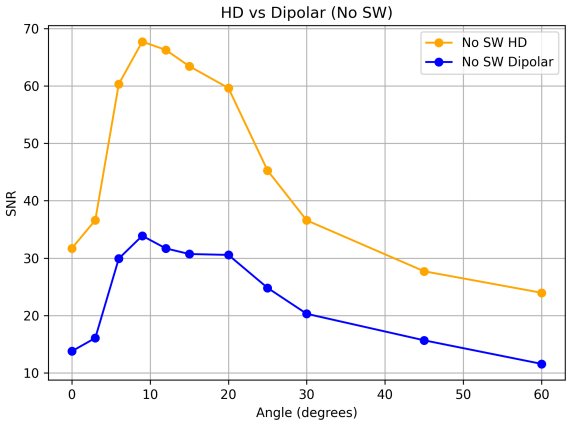
(a) SNR_HD with different noise models



(b) SNR_dipolar with different noise models



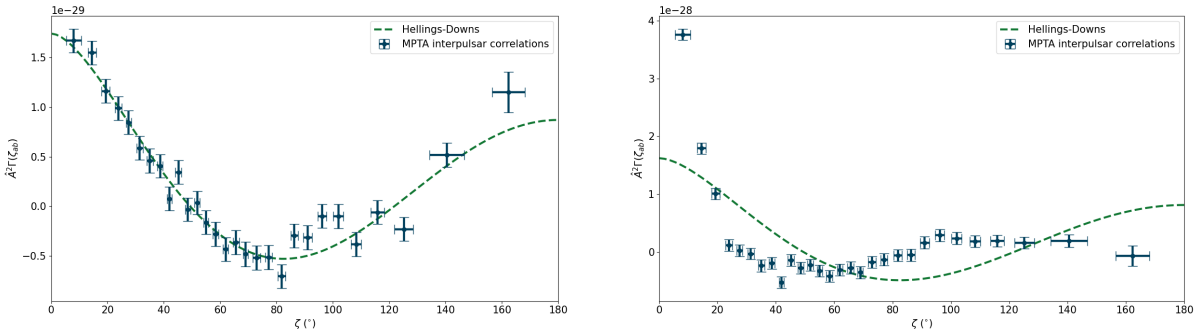
(c) SNR_HD and SNR_dipolar with SW model



(d) SNR_HD and SNR_dipolar with no_SW model

Figure 8: Global caption describing the 2×2 panel.

- For the SW dataset, the SNR of the Hellings-Downs (HD) correlation pattern remains steady until around 30 degrees (the peak is at 12 degrees cut, though not by much). This pattern is rather straightforward to explain
  - At sufficiently large angles, not enough data remains, decreasing the sensitivity of the simulated array, hence a decrease in SNR
  - (I'm not sure about this) There is no significant increase in SNR for any cuts, as the noise injected is well understood and well modelled by our current solar wind modelling, hence we don't expect deleting data close to the sun to improve SNR significantly.
- For the no_SW dataset, SNR(HD) seems to be always higher than the SW dataset. Though the numbers might be surprising, the observation can be explained by visually examining the plots.
  - The pattern of the datapoints(bins) deviates significantly from the theoretical HD curve. This can be seen as a consequence of unmodelled solar wind effects, producing a stronger dipolar signature (evident in the trend later discussed), which becomes absorbed by the HD SNR. (But it's not a perfect Dipolar pattern as well? Perhaps it is a mixture of both?)



(a) HD plot with normal noise modeling

(b) HD plot with no SW noise modelling

Figure 9: Global caption describing the 1×2 comparison.

The above trends are rather predicted/well explained. A more interesting observation is the correlation between the HD and Dipolar SNR for both the SW and no_SW datasets. The similarity of the 2 patterns is easily seen in the plots shown above. So I decided to compute the Pearson correlation coefficient for them. The following results were obtained:

The SNR of HD and Dipolar for the same dataset almost perfectly imitates a linear relationship.

It is totally expected that they will be absorbed somewhat by each other, but is it really expected that there is such a perfectly linear correlation?

Is this a natural consequence of the Opstat algorithm when the dataset is ideal (why is it not reflected in the real dataset, is it only because the dataset is ideal and well modelled?), or does it suggest other issues?

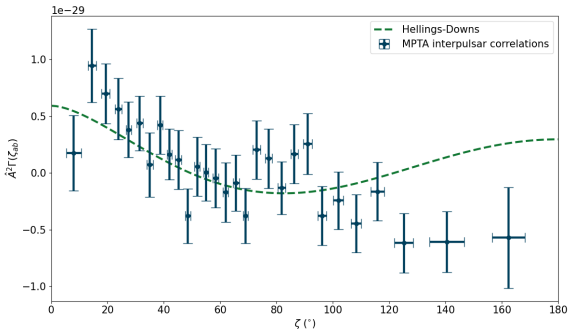
Pearson r (SW HD vs SW Dipolar) = 0.985, p = 1.08e-16

Pearson r (No SW HD vs No SW Dipolar) = 0.982, p = 7.14e-16

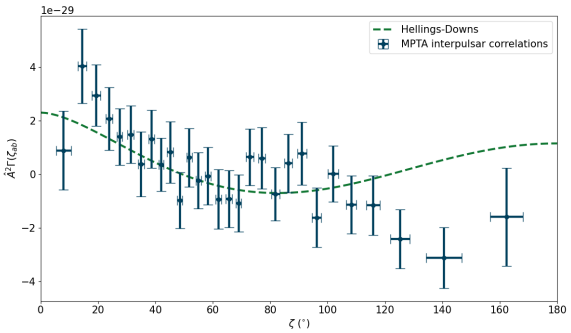
6.4 Real dataset results

For the real data:

- The same trends are not observed with the real dataset. However, the value of the statistics produced for the real dataset is questionable. The datapoints for the normal (SW) dataset look strongly dipolar, and are not much different from the no_SW version. This may suggest that this current preliminary model doesn't model solar wind that well or there are other processing required (e.g. removal of points etc.)
- Correlation values for the real data:
  - Pearson r (SW HD vs No SW HD) = 0.812, $p = 4.46 \times 10^{-6}$
  - Pearson r (SW Dipolar vs No SW Dipolar) = 0.566, $p = 6.05 \times 10^{-3}$
  - Pearson r (SW HD vs SW Dipolar) = 0.154, $p = 0.495$
  - Pearson r (No SW HD vs No SW Dipolar) = 0.959, $p = 2.29 \times 10^{-12}$



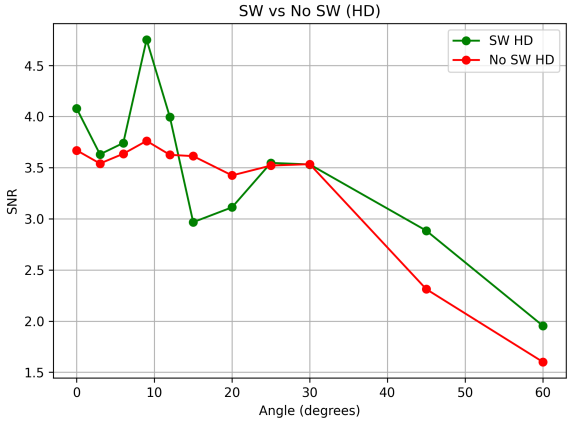
(a) HD plot with normal noise modeling



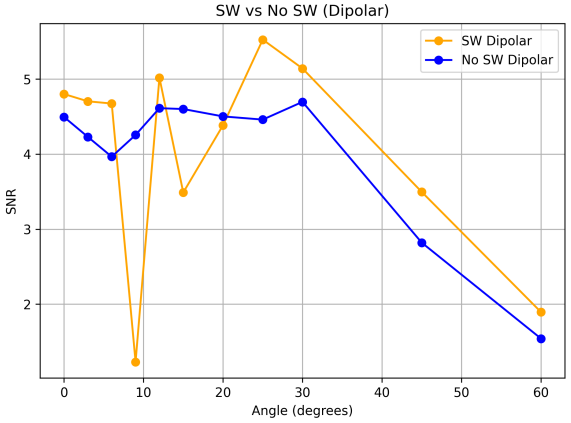
(b) HD plot with no SW noise modelling

Figure 10: Global caption describing the 1×2 comparison.

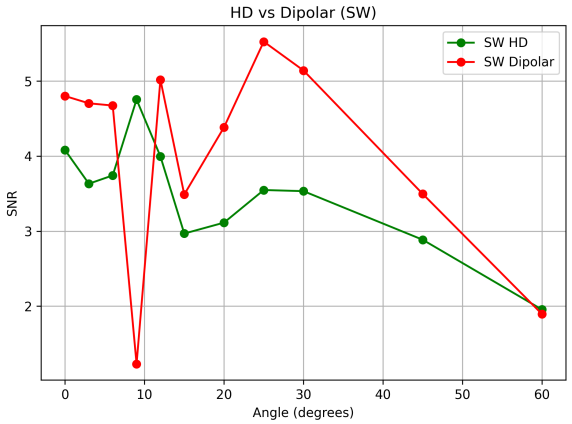
I haven't analysed the trends for the real data that much due to the questionable reliability of the statistics mentioned above. The only interesting plots that I looked at were the 9 degree cut with SW modelling, the HD SNR was the highest and the Dipolar SNR was the lowest. However, examining the plot the datapoints just seem like a mess and look like random jitter rather than any identifiable patterns.



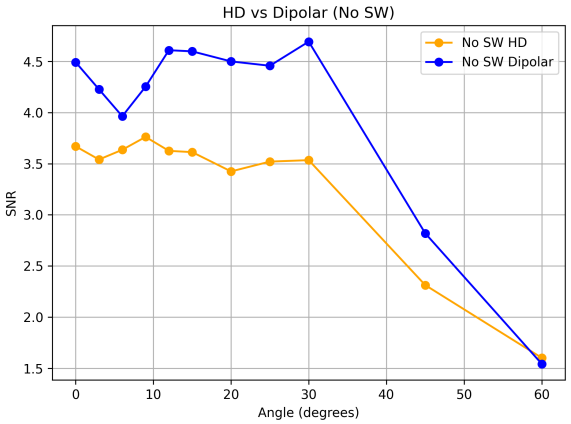
(a) SNR_HD with different noise models



(b) SNR_dipolar with different noise models



(c) SNR_HD and SNR_dipolar with SW model



(d) SNR_HD and SNR_dipolar with no_SW model

Figure 11: Global caption describing the 2×2 panel.

7 Discussion and openings

This work highlights several consistent themes: data selection strongly affects dipolar vs quadrupolar SNRs, FD removal can improve parameter constraints for some pulsars but may reduce TOA uncertainties artificially in others, and solar-proximity cuts have dataset-dependent impacts that interact with solar-wind modelling choices. .....

Acknowledgements

I thank Hendrik and Saurav for providing templates and simulated datasets, and the MPTA team for guidance and feedback during the internship.

Appendix