Universität Bielefeld

Outlier Rejection Scheme and basic improvements for EPTA Pulsar Timing

—-FAST/Future Pulsar Symposium 8

Jun Wang

Universität Bielefeld
https://www.AstronDog.com/

June 27, 2019

Schedule

Overview

European Pulsar Timing Array(EPTA) Telescopes

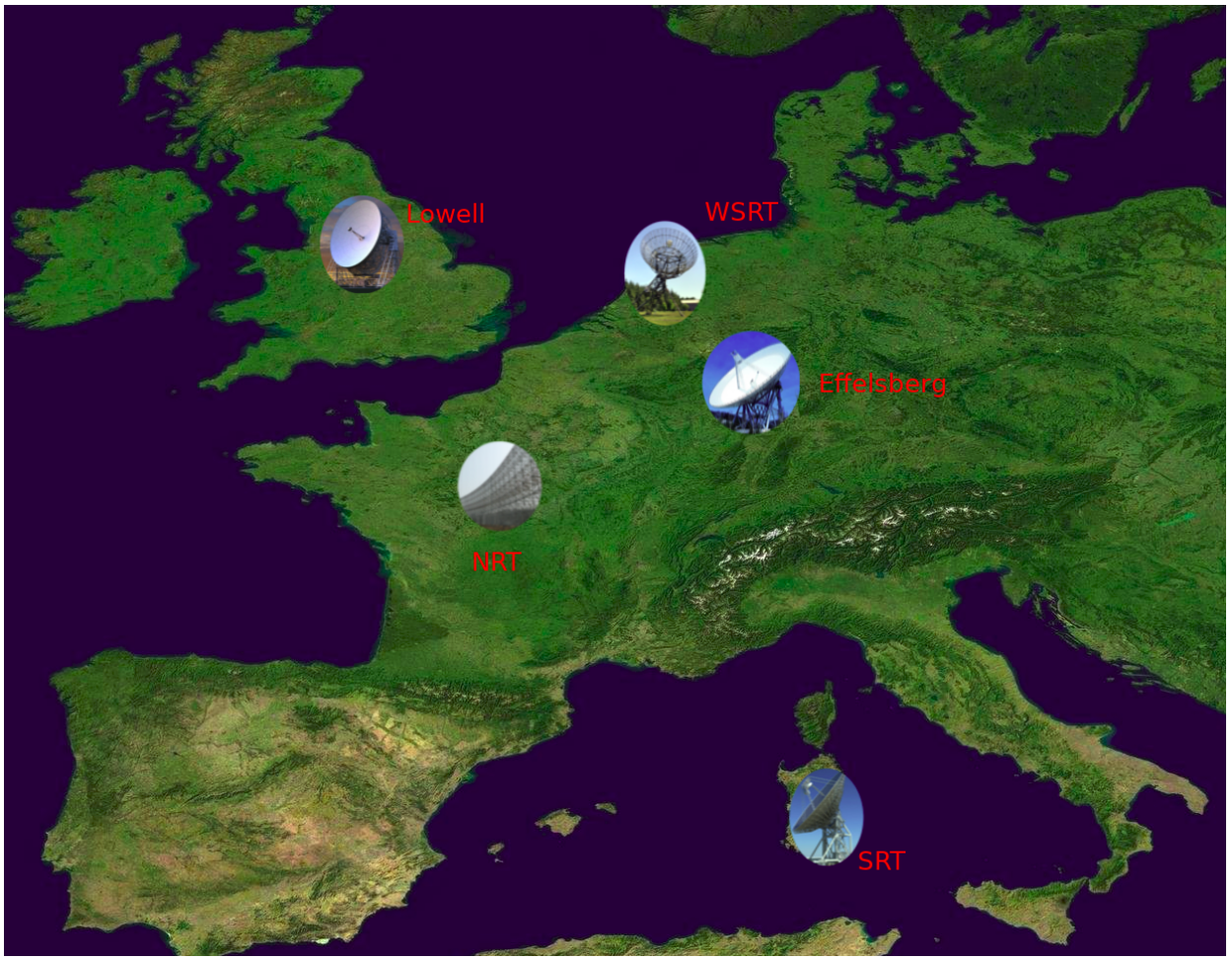


Figure: The telescopes of the EPTA.

Telescope List

- Effelsberg(100m)
- Westerbork(96m)
- Nançay(92m)
- Lovell(76m)
- Sardinia(64m)

Data-sets and Telescopes Parameters

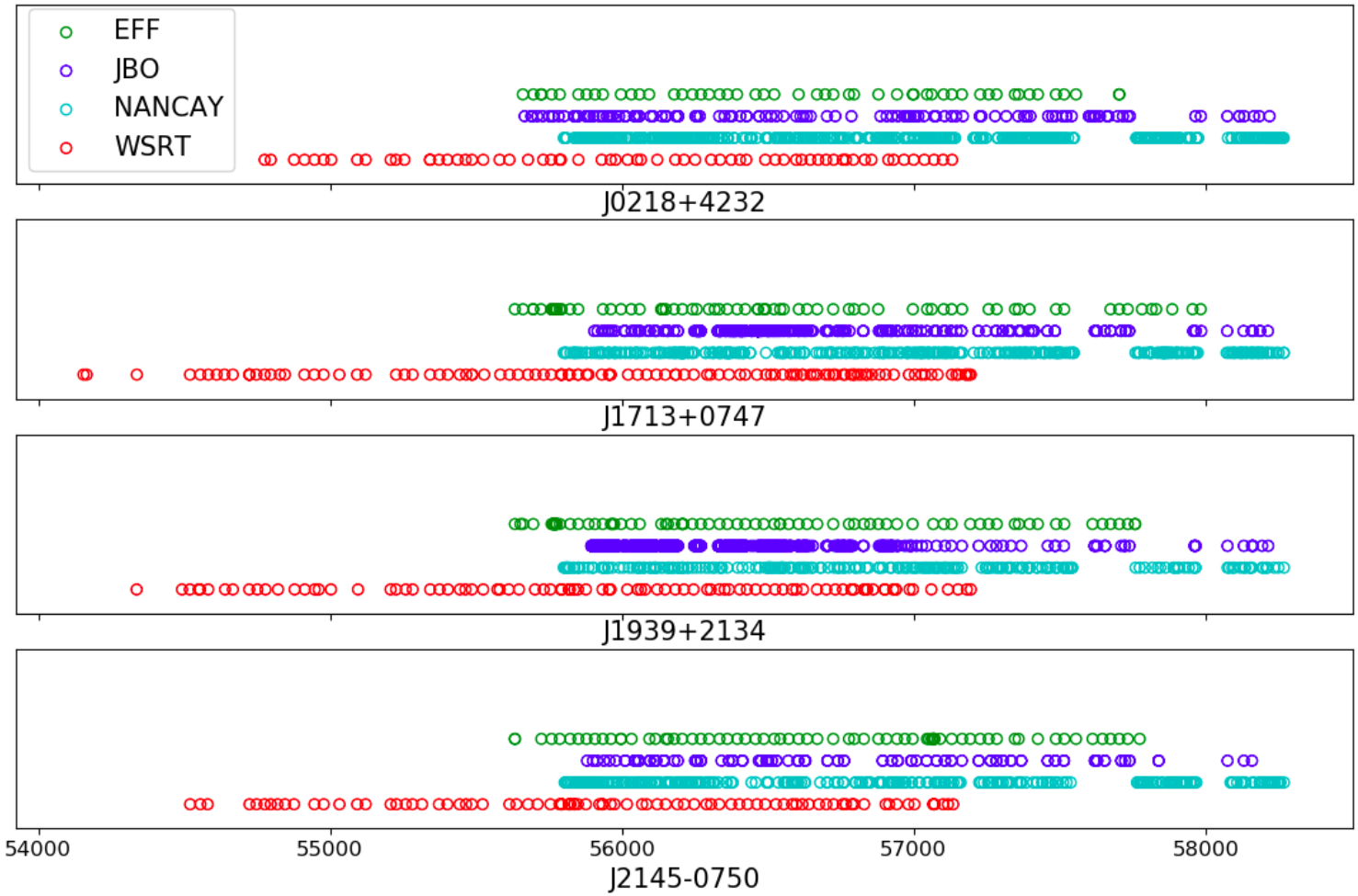
Pulsars

J0218+4232 J1713+0747
J1939+2134 J2145-0750

Telescopes Parameters

Telescope	**BW(MHz)**	**Freq(MHz)**	**nchan**	**nbin**
Effelsberg	200	1347.5	128	256
Westerbork	160	1380	512	256
Nançay	512	1484	128	256
Lovell	400	1532	1600	256

Epochs of all observations



Outliers influence the Frequency-resolved timing

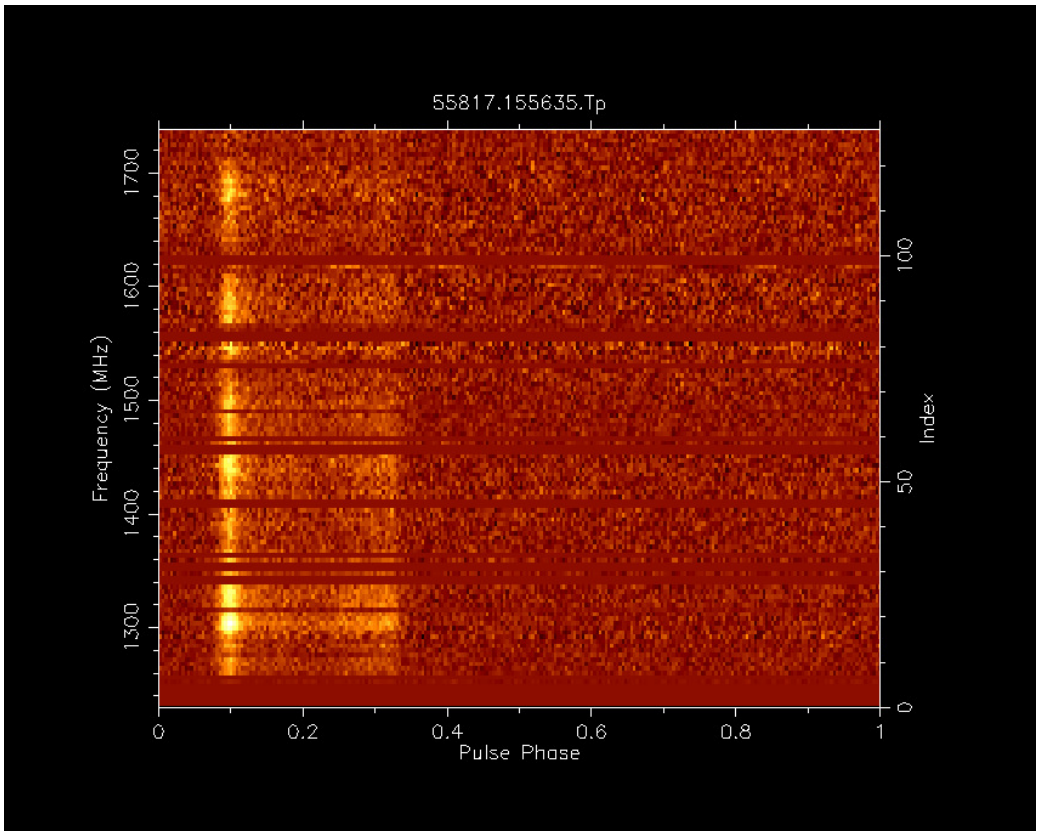


Figure: A frequency series of PSR J2145-0750, Nancay.

Outliers

1. Incomplete RFI Removal.
2. Low S/N.
3. Bad Channels.
4. Incorrect Calibration.
5. Instrumental failure.
6. ...

How to determine if the Scheme is optimal

Comprehensively analyze the following facors:

1. Minimise RMS.
2. Chisq is slightly larger than 1.
3. Maximize the Total Information Content $\frac{\sqrt{N}}{\sigma}$.
4. More ToAs are preserved.

Outlier Rejection Scheme

Scheme Evolution

Previous thoughts:

1. Goodness-of-fit criterion in the template fit process (0.5-1.5).
2. Median Standard Deviation(MAD).
3. Fitting the residual with Gaussian distribution model and remove non-Gaussian ToAs.
4. S/N criterion(S/N above 8).

Conclusion:

1. Median standard deviation.(k = 3)
2. Fitting the Normalized residual with Gaussian distribution model.(k = 3)

Subsection 1

1. Median Absolute Deviation(MAD)

What is MAD

The Commonly used Standard Deviation(SD) is the square root of the variance.
For a univariate data set X1, X2, ..., Xn, the MAD is defined as the median of the absolute deviations from the data's median:

$$MAD = median(|X_i - median(X_i)|) \tag{1}$$

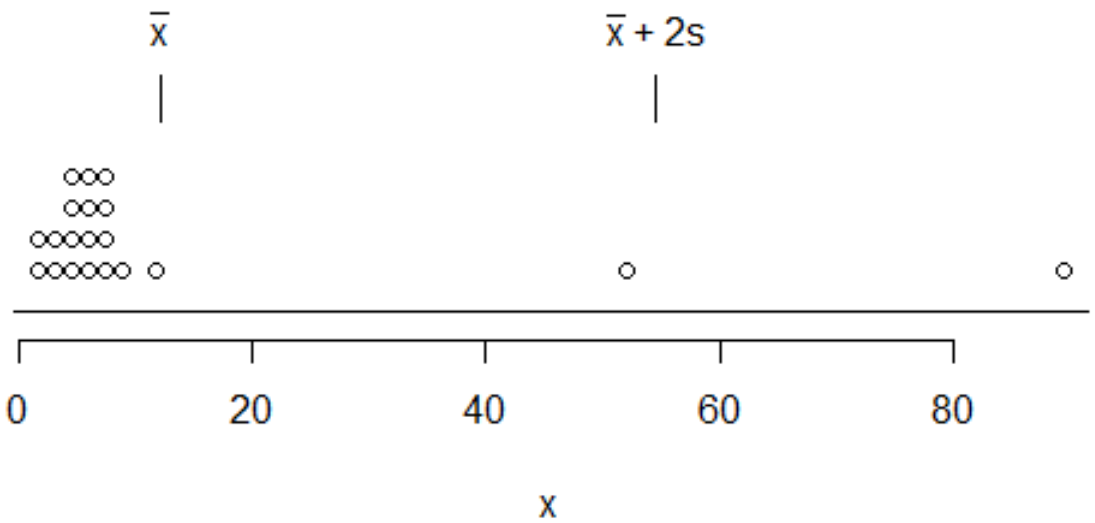


Figure: Weakness of SD in outlier detection.

Why MAD

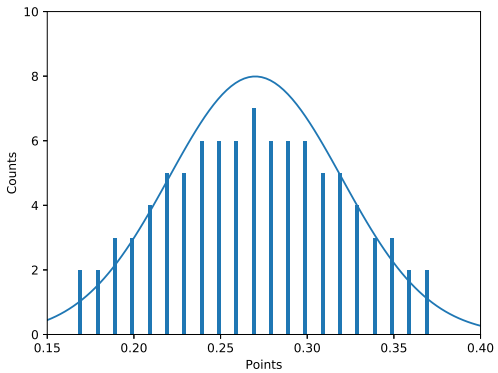


Figure: a) Normal distribution, n = 91, mean = 0.27, median = 0.27, standard deviation = 0.06.

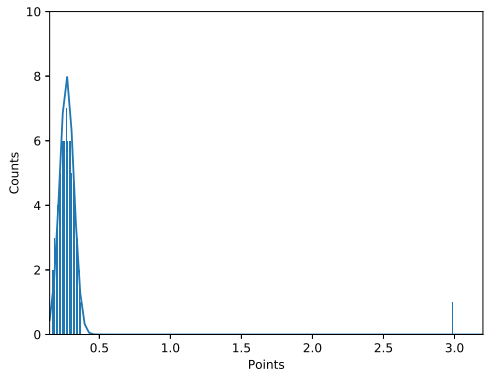
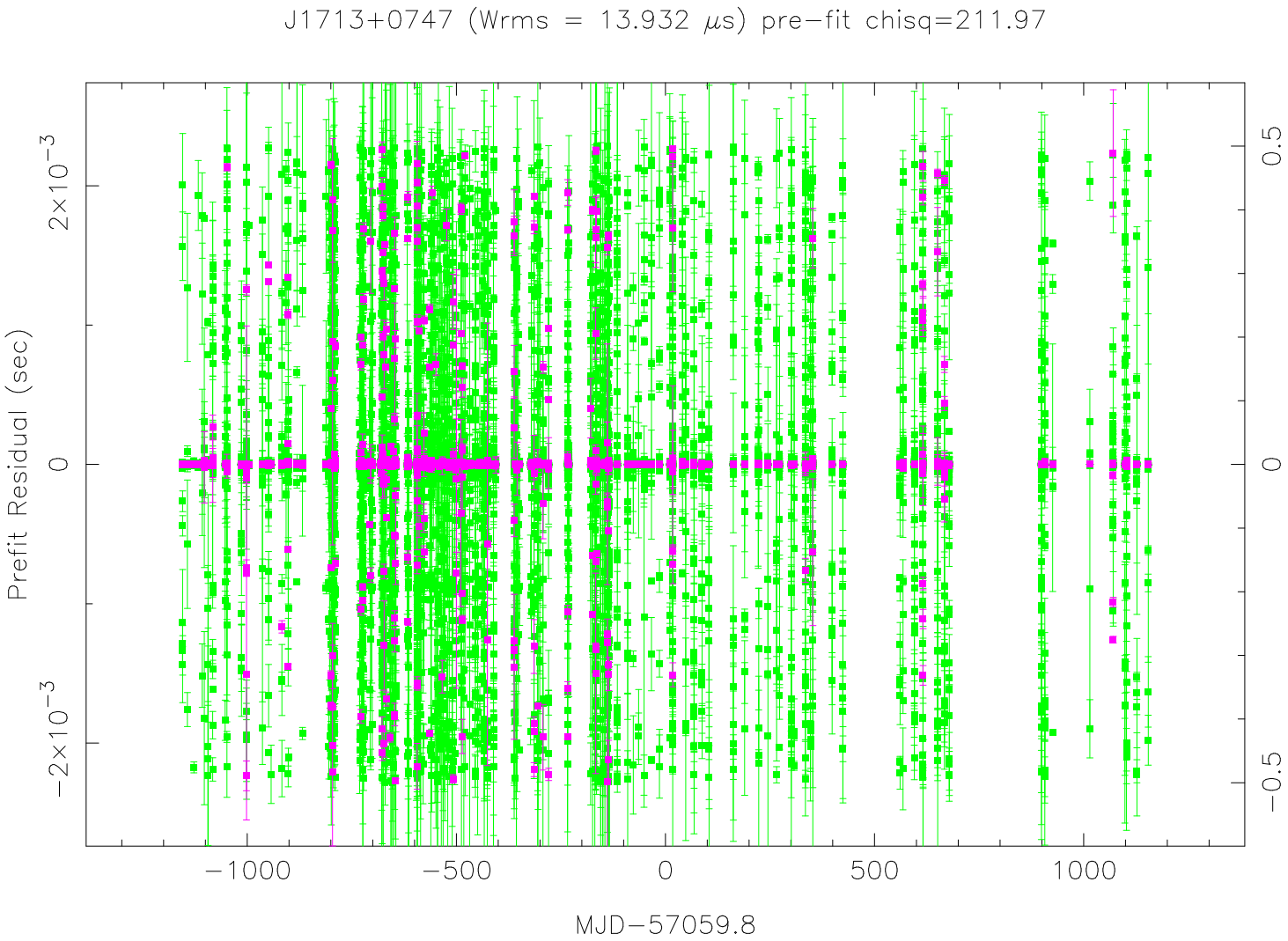


Figure: b) Asymmetry due to an outlier, n = 91, mean = 0.39, median = 0.27, standard deviation = 0.59.

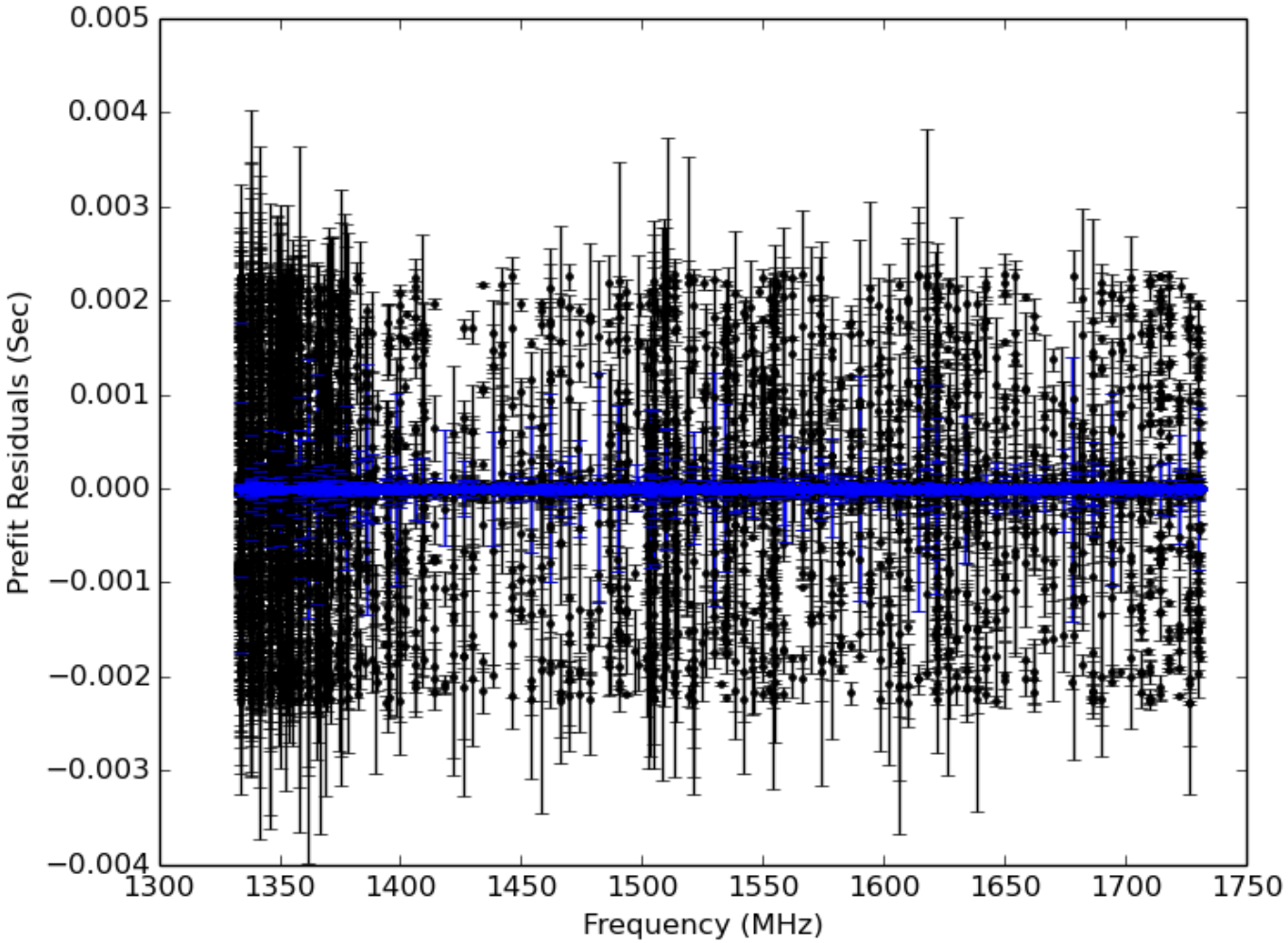
Advantage of Median Standard Deviation

1. Aimed at symmetric distribution.
2. MAD is more robust than Standard Deviation (SD).
3. MAD is totally immune to the sample size.

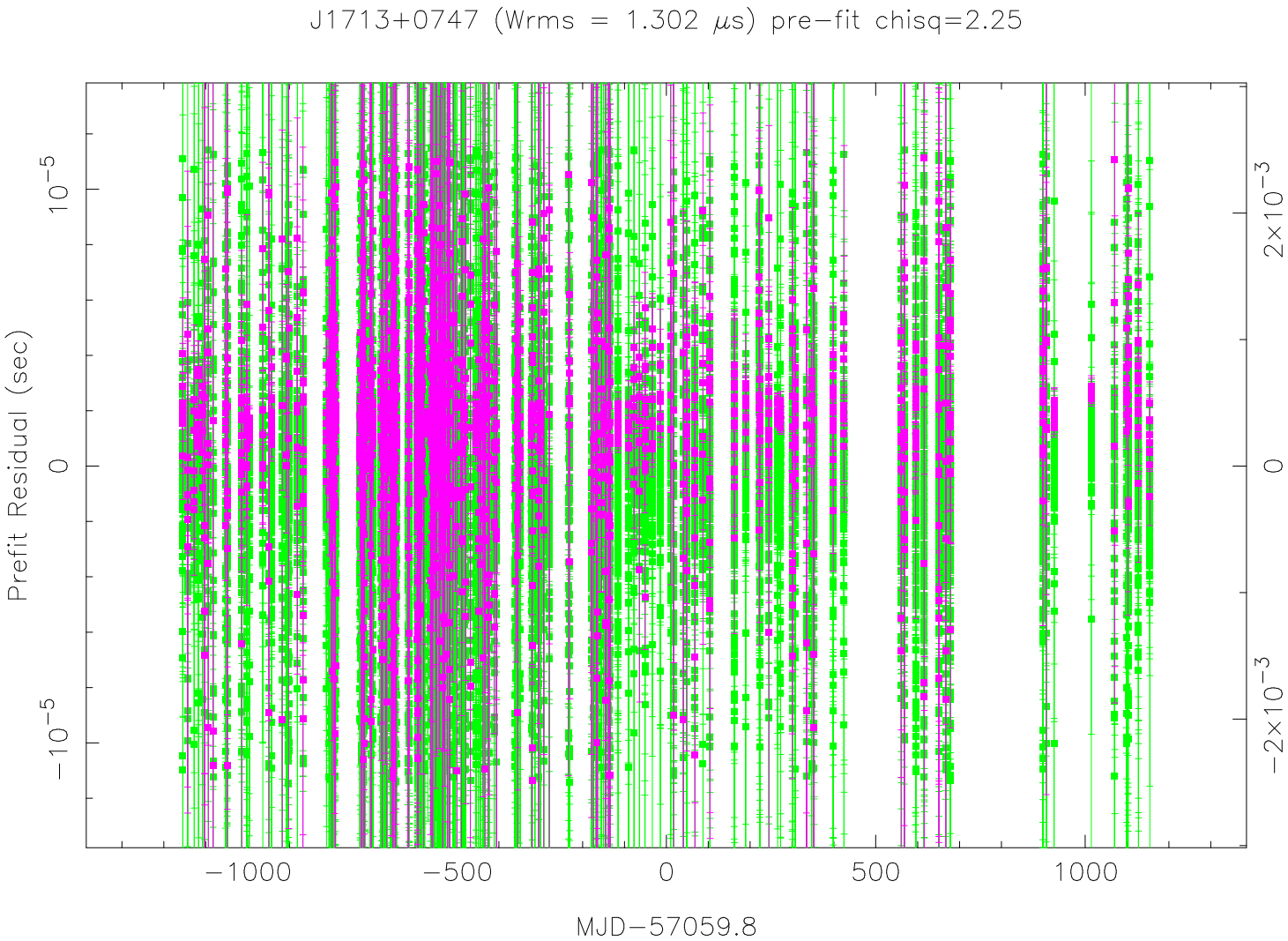
Raw Residuals from J1713+0747, JBO ($N_{TOA} = 28376$)



Example on MAD, MAD



Example on MAD,after MAD($N_{TOA} = 22588$)



Subsection 2

2. Normalized Residual Gaussian Distribution Fitting

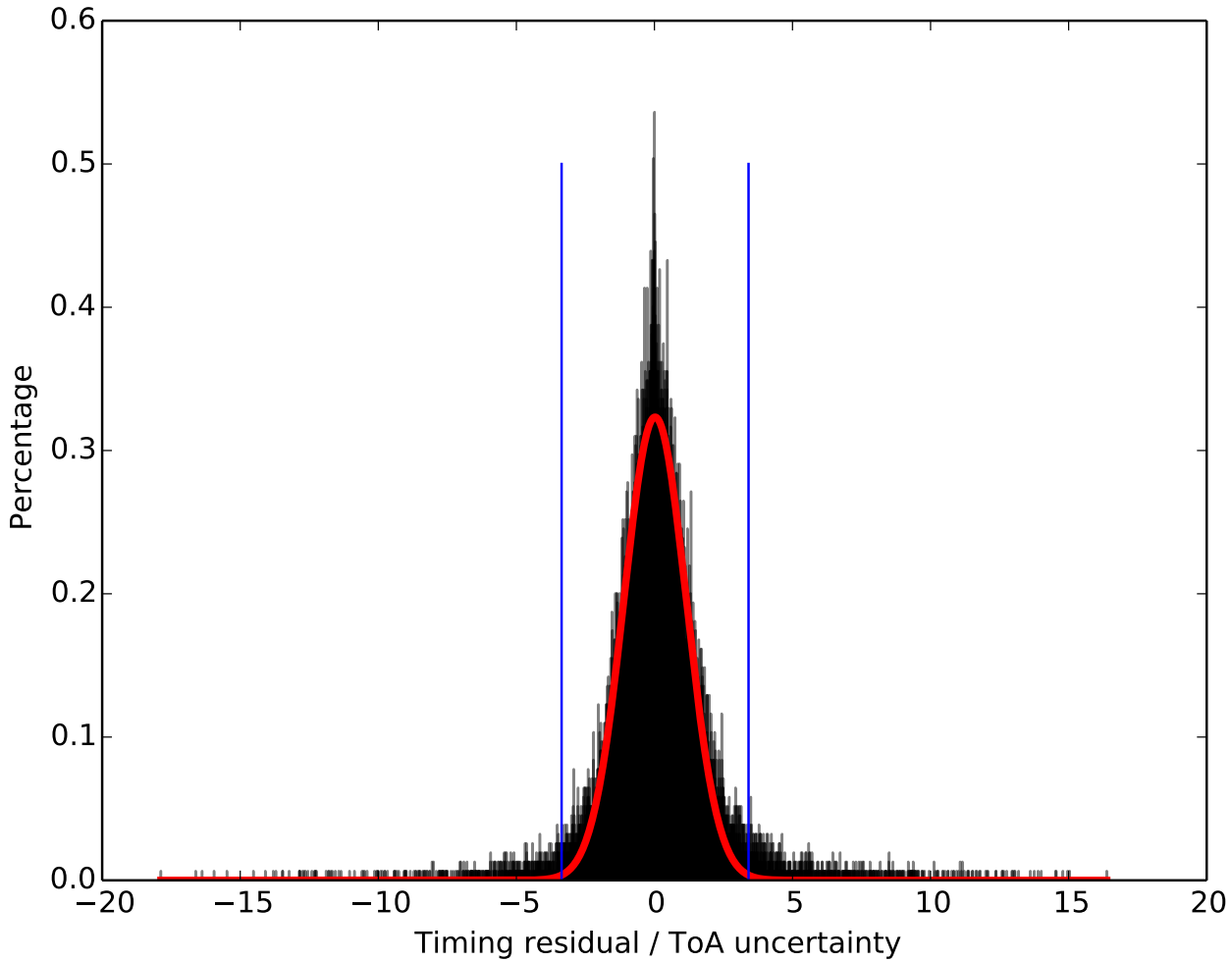
HOW

After employing the MAD scheme, the distribution of ToAs becomes more Gaussian. We then fit the histogram of Normalized Residual (residual divided by ToA uncertainty) with Gaussian Distribution:

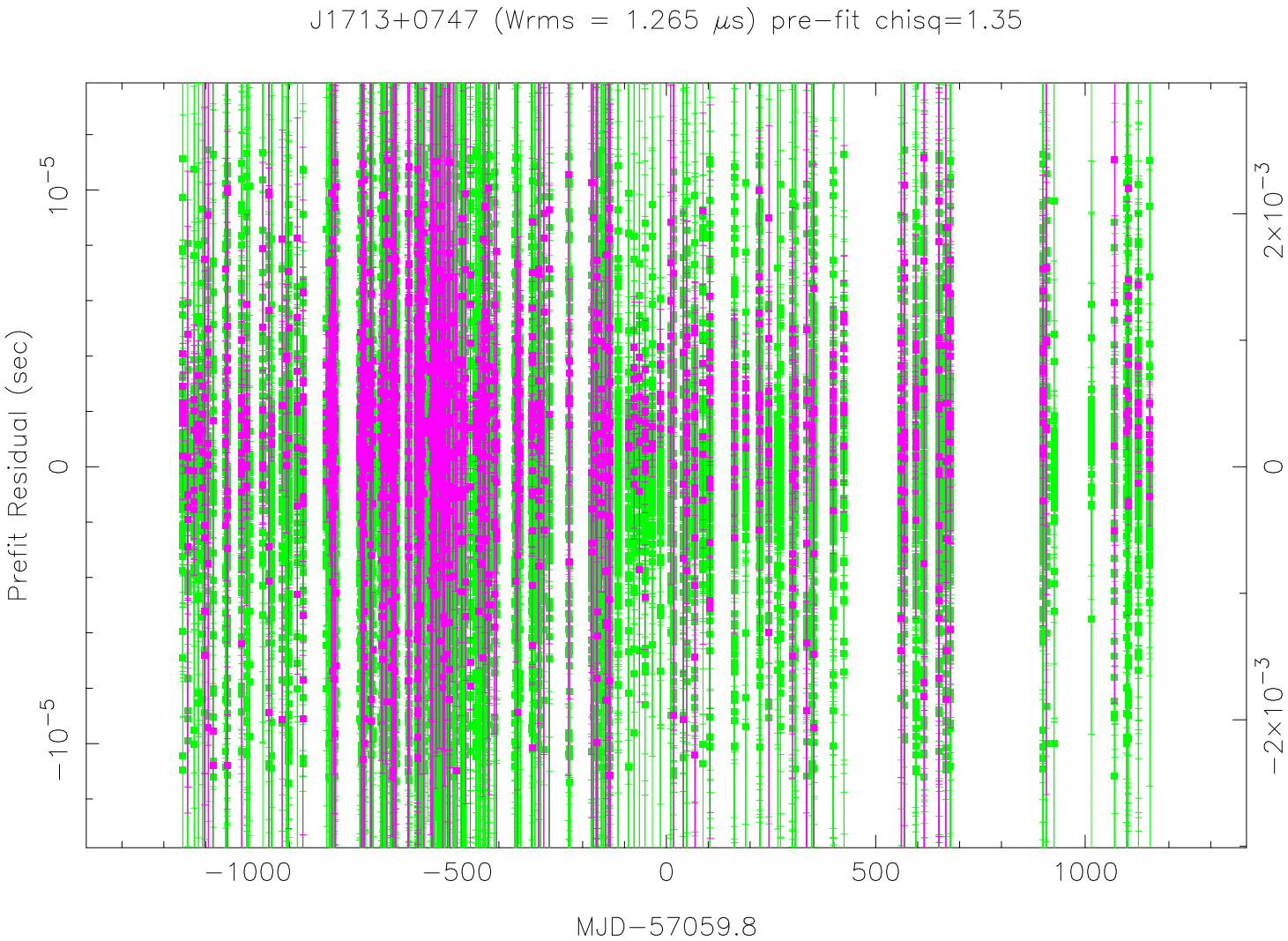
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{2}$$

and remove residuals which are 3σ away.

Examples on Gaussian Fitting



After Gaussian Fitting($N_{TOA} = 20978$)



Summary & Future Work

Summary & In Future

Summary

1. The Wrms is much lower than before, but it's still not great.
2. The reduced chi-squared is good and very comparable to the best results people typically get.
3. The present optimal outlier rejection scheme is first use MAD, then Normalised Residuals Fitting.

In Near Future

1. Decide which CCA is best.
2. Decide which template is best.
3. Check optimal ToA bandwidth.
4. Ask for more pulsars to be included.

The End

Thank you!