



Robust fitting for pulsar timing analysis

DOI:

[10.1093/mnras/stx579](https://doi.org/10.1093/mnras/stx579)

Document Version

Final published version

[Link to publication record in Manchester Research Explorer](#)

Citation for published version (APA):

Wang, Y., Keith, M., Stappers, B., & Zheng, W. (2017). Robust fitting for pulsar timing analysis. *Monthly Notices of the Royal Astronomical Society*, 468(3), 2637-2644. <https://doi.org/10.1093/mnras/stx579>

Published in:

Monthly Notices of the Royal Astronomical Society

Citing this paper

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.



Robust fitting for pulsar timing analysis

Yidi Wang,^{1,2★} Michael J. Keith,^{2★} Benjamin Stappers^{2★} and Wei Zheng¹

¹College of Aerospace Science and Engineering, National University of Defense Technology, Deya Road, Changsha 410073, China

²Jodrell Bank Centre for Astrophysics, School of Physics and Astronomy, The University of Manchester, Manchester M13 9PL, UK

Accepted 2017 March 3. Received 2017 March 1; in original form 2016 August 9

ABSTRACT

We introduce a robust fitting method into pulsar timing analysis to cope with the non-Gaussian noise. The general maximum likelihood estimator (M-estimator) can resist the impact of non-Gaussian noise by employing convex and bounded loss functions. Three loss functions, including the Huber function, the Bisquare function and the Welsch function, are investigated. The Shapiro–Wilk test is employed to test whether the uncertainty in the observed times of arrival is drawn from a non-Gaussian distribution. Two simulations, where the non-Gaussian distribution is modelled as contaminated Gaussian distributions, are performed. It is found that M-estimators are unbiased and could achieve a root-mean-square error smaller than that obtained by the least square (LS) at the cost of a slightly higher computation complexity in a non-Gaussian environment. M-estimators are also applied to the real timing data of PSR J1713+0747. The results have shown that the fitting results of M-estimators are more accurate than those of LS and are closer to the result of very long baseline interferometry.

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

1 INTRODUCTION

Pulsars are rapidly rotating neutron stars that are useful because of their exceptional rotational stability. With decades of timing measurements, pulsars, especially millisecond pulsars, exhibit a rotational stability similar to the current atomic clocks (Matsakis, Taylor & Eubanks 1997). They can be employed in the pursuit of many scientific goals, including tests of general relativity and for the direct detection of gravitational waves (e.g. Taylor & Weisberg 1989; Jenet et al. 2005).

Pulsar timing is a powerful tool to accomplish the above goals. In brief, a series of pulse times-of-arrival (ToAs) of a pulsar is recorded by an observatory and is transferred to the inertial frame of the Solar system barycentre in order to remove the impact caused by the motion of the Earth. A timing model that characterizes the properties of a pulsar’s orbital motion as well as its timing properties is established by fitting the transferred ToAs. A detailed overview of the pulsar timing method can be found in Hobbs, Edwards & Manchester (2006).

The fitting problem can be well addressed by using the pulsar timing package TEMPO2, which employs the least squares (LS) algorithm (Hobbs et al. 2009). Initially, TEMPO2 worked with the assumption that the uncertainty within ToAs is Gaussian white noise. The generalized least square algorithm (GLS) was introduced into TEMPO2 to cope with time-correlated noise (Coles et al. 2011). On the other hand, the noise in realistic data sets does not behave as

nically as Gaussian noise. GLS can whiten the correlated noise but cannot guarantee that the resulting white noise is Gaussian.

More recently, a Bayesian timing package TEMPONEST was released (Lentati et al. 2014a). TEMPONEST employs the Bayesian inference to fit the timing model and can solve correlated and non-Gaussian noise problem in theory (Lentati, Hobson & Alexander 2014b). However, its fitting performance depends heavily on how precisely the derived model could describe the distribution of noise. In addition, it needs intensive computation, which grows dramatically as the number of fitting parameters increases.

In practice, we usually encounter a type of non-Gaussian noise that is not drawn from a completely new distribution but can be viewed as a small departure from the presumed Gaussian one. The departure might be caused by unknown uncertainties within the actual ToAs or instrumentation problems. In addition, the non-negligible influence of outliers in the data could cause a distribution with a tail longer than that of a Gaussian one. In this case, the robust fitting can be a better option. The robust fitting treats the small departure as a perturbation, resisting its impact by modifying the loss function of the parameter fitting problem as a convex and bounded one (Huber 1981). In 1964, Huber proposed the concept of the robust statistics that laid the foundation for robust fitting (Huber 1964). Over the last 50 yr, several approaches to robust fitting have been proposed, including the R-estimator and the L-estimator (Hampel et al. 1986). However, the general maximum likelihood estimator (MLE; M-estimator) now appears to dominate the field because of its generality, high breakdown point and efficiency. The M-estimator was first proposed for the location and scale parameter estimation in Huber (1972) and was extended to the linear regression problem in Huber (1973). The M-estimator is

* E-mail: wangyidi_nav@163.com (YW); michael.keith@manchester.ac.uk (MJK); ben.stappers@manchester.ac.uk (BS)

computationally efficient, requiring a computation cost slightly more expensive than LS. In addition, it is practical, as it does not need a precise description of the non-Gaussian distribution.

In this paper, we introduce the robust fitting into pulsar timing to cope with non-Gaussian noise. By modelling the timing process as a linear regression problem, we analyse the relation and difference between the M-estimator and LS methods, providing a robust estimation procedure. Three loss functions of M-estimator (e.g. Huber function, Bisquare function and Welsch function) are investigated. The Shapiro–Wilk (SW) test is performed on the post-fit residuals to detect the non-Gaussianity of the recorded ToAs. M-estimators are tested by simulating two non-Gaussian cases and also by applying them to real data.

The remainder of this paper is organized as follows. A brief review of pulsar timing via LS is provided in Section 2. Section 3 shows how to handle non-Gaussian noise using robust fitting. Section 4 provides the method of detecting the non-Gaussianity of ToAs. M-estimators are applied to simulated and real data in Sections 5 and 6, respectively.

2 BRIEF REVIEW OF PULSAR TIMING VIA LS

Given that the parameters of the timing model are linear, at least for all small perturbations, the pulsar timing process can be modelled as a linear system (Coles et al. 2011)

$$\mathbf{R} = \mathbf{M}\mathbf{P} + \mathbf{E}, \quad (1)$$

where $\mathbf{R} \in \mathbb{R}^{m \times 1}$ denotes the pre-fit timing residuals, $\mathbf{M} \in \mathbb{R}^{m \times n}$ represents the pulsar timing model, $\mathbf{P} \in \mathbb{R}^{n \times 1}$ describes the fitted parameters and $\mathbf{E} \in \mathbb{R}^{m \times 1}$ denotes the post-fit residuals with an assumed covariance matrix, \mathbf{C} .

If \mathbf{C} is unknown, it can be estimated by analysing the spectrum of \mathbf{E} . When the off-diagonal elements of \mathbf{C} are non-zero, \mathbf{E} is called correlated noise. In this case, GLS can use Cholesky factorization to whiten \mathbf{E} . As the result of Cholesky factorization, \mathbf{C} can be expressed as $\mathbf{L}\mathbf{L}^T$, where \mathbf{L} is the lower triangular matrix. And then, by multiplying \mathbf{L}^{-1} to each side of equation (1), we have

$$\mathbf{R}_w = \mathbf{M}_w\mathbf{P} + \mathbf{E}_w, \quad (2)$$

in which $\mathbf{R}_w = \mathbf{L}^{-1}\mathbf{R}$, $\mathbf{M}_w = \mathbf{L}^{-1}\mathbf{M}$ and $\mathbf{E}_w = \mathbf{L}^{-1}\mathbf{E}$ are the result of the whitening. In this manner, \mathbf{E}_w contains only a white noise.

And then, \mathbf{P} can be estimated by LS by solving the minimization problem,

$$J_{LS} = (\mathbf{R}_w - \mathbf{M}_w\mathbf{P})^T(\mathbf{R}_w - \mathbf{M}_w\mathbf{P}). \quad (3)$$

Finally, the estimate of \mathbf{P} , $\hat{\mathbf{P}}_{LS}$, is

$$\hat{\mathbf{P}}_{LS} = (\mathbf{M}_w^T\mathbf{M}_w)^{-1}\mathbf{M}_w^T\mathbf{R}_w, \quad (4)$$

and its covariance matrix is

$$\text{cov}(\hat{\mathbf{P}}_{LS}) = (\mathbf{M}_w^T\mathbf{M}_w)^{-1}, \quad (5)$$

where $\mathbf{r} = \mathbf{R}_w - \mathbf{M}_w\hat{\mathbf{P}}_{LS}$ is the fitting residual.

Note that equation (3), which gives the mean squared error, is exactly the log likelihood function for a Gaussian process. It means that the estimate result of LS coincides with that of an MLE in a Gaussian environment. However, when LS is applied to a non-Gaussian case, its estimate result would depart from the result of corresponding MLE, as it discards the information of the moments higher than the second order.

3 COPING WITH NON-GAUSSIAN NOISE VIA ROBUST FITTING

A natural idea for parameter estimation in the presence of non-Gaussian noise would be to derive a new MLE based on the underlying probability density function (PDF) of the noise. Of course, this is typically difficult if one has to estimate the PDF from the data itself. A great deal of attention has been paid to accurately modelling non-Gaussian distributions, using for instance, Gaussian mixture models (Bishop 2006). Alternatively, when the PDF shows only small deviations from a Gaussian distribution, for instance a Gaussian contaminated by outliers, then it is possible to derive a general purpose estimator for which the contamination of the distribution is a tuning parameter that can more easily be estimated from the data. The robust estimator we propose for the pulsar timing problem is the M-estimator (Huber 1964), and in the remainder of this paper we will show how this is derived and then applied to pulsar timing.

3.1 Description of a non-Gaussian distribution

In robust statistics, a non-Gaussian distribution, F , is generally modelled as a contaminated Gaussian distribution (Huber 1981), i.e.,

$$F = (1 - \epsilon)G + \epsilon H, \quad (6)$$

where ϵ is the contamination ratio, $0 \leq \epsilon < 1$, G is the modelling Gaussian distribution and H is a contamination distribution. For the purposes of robust statistics, the exact nature of the contamination distribution does not need to be known, it may be caused by our imperfect understanding of the intrinsic PDF of the investigated noise source, or from some other unmodelled error.

As mentioned earlier, robust fitting only works well in the case when the non-Gaussian distribution can be viewed as being within the vicinity of the modelling Gaussian distribution, the scale of which is quantified by the contamination ratio ϵ . Typically, when ϵ is less than 0.5, the modelling Gaussian distribution remains the dominant distribution and the robust fitting is effective.

For pulsar timing, we recommend that the robust fitting should be employed when the uncertainty within the observed ToAs can be viewed as a slight departure from the Gaussian distribution expected from e.g. radiometer noise.

3.2 Derivation of M-estimator

Assume the post-fit residual, \mathbf{r} , satisfies the multiplicative model

$$\mathbf{r} = s\mathbf{u}, \quad (7)$$

where \mathbf{u} is independently distributed with a PDF of $f(\mathbf{u})$ and $s > 0$ is an unknown scaling parameter.

An MLE can be determined by maximizing (Holland & Welsch 1977)

$$\sum_{i=1}^n \ln f\left(\frac{r_i}{s}\right), \quad (8)$$

where r_i is the i th component of \mathbf{r} .

When there is modelling error within the expression of $f(\frac{r_i}{s})$, the solution of MLE is sub-optimal. In this case, Huber (1964) proposed a general MLE (M-estimator), which substitutes $-\ln f(r_i)$ by $\rho(r_i)$, to pursue a *near-optimal* estimate result.

Table 1. Predominant ρ functions, and the corresponding ϕ functions and Ψ functions.

	LS	M-estimator		
		Huber function	Bisquare function	Welsch function
ρ function	η^2	$\begin{cases} \frac{1}{2}\eta^2 & \text{if } \eta \leq a \\ a \eta - \frac{1}{2}a^2 & \text{else} \end{cases}$	$\begin{cases} \frac{1}{6}a^2[1 - (1 - (\frac{\eta}{a})^2)^3] & \text{if } \eta \leq a \\ \frac{1}{6}a^2 & \text{else} \end{cases}$	$a^2[1 - \exp(-(\frac{\eta}{a})^2)]$
ϕ function	2η	$\begin{cases} \eta & \text{if } \eta \leq a \\ \text{sgn}(\eta)a & \text{else} \end{cases}$	$\begin{cases} \eta(1 - (\frac{\eta}{a})^2)^2 & \text{if } \eta \leq a \\ 0 & \text{else} \end{cases}$	$2a\eta \exp(-(\frac{\eta}{a})^2)$
Ψ function	1	$\begin{cases} 1 & \text{if } \eta \leq a \\ \frac{\text{sgn}(\eta)}{\eta}a & \text{else} \end{cases}$	$\begin{cases} (1 - (\frac{\eta}{a})^2)^2 & \text{if } \eta \leq a \\ 0 & \text{else} \end{cases}$	$2a \exp(-(\frac{\eta}{a})^2)$

The M-estimator can be determined by minimizing (Huber 1981)

$$J_M = \sum_{i=1}^m \rho\left(\frac{r_i}{s}\right). \quad (9)$$

For further comparison, equation (3) can also be expressed in a form similar to equation (9) by setting $\rho(\frac{r_i}{s})$ to be $(\frac{r_i}{s})^2$. On the premise that the ρ function is convex, the minimum of equation (9) can be obtained by solving the equation

$$\sum_{i=1}^m \frac{\partial \rho(r_i/s)}{\partial r_i} \frac{\partial r_i}{\partial \mathbf{P}} = \mathbf{0}. \quad (10)$$

By defining functions $\phi_i = \frac{\partial \rho(r_i/s)}{\partial r_i}$ and $\Psi_i = \phi(r_i)/r_i$ and matrix $\Psi = \text{diag}[\Psi_i]$, equation (10) can be rewritten in a matrix form as

$$\mathbf{M}_w^T \Psi \mathbf{r} = \mathbf{0}. \quad (11)$$

Substituting the expression of \mathbf{r} into equation (11), we have

$$\mathbf{M}_w^T \Psi (\mathbf{R}_w - \mathbf{M}_w \mathbf{P}) = \mathbf{0}, \quad (12)$$

which is the derivative of

$$J_M = (\mathbf{R}_w - \mathbf{M}_w \mathbf{P})^T \Psi (\mathbf{R}_w - \mathbf{M}_w \mathbf{P}). \quad (13)$$

Thus, the estimate of \mathbf{P} via an M-estimator is

$$\hat{\mathbf{P}}_M = (\mathbf{M}_w^T \Psi \mathbf{M}_w)^{-1} \mathbf{M}_w^T \Psi \mathbf{R}_w, \quad (14)$$

which would coincide with equation (4) if Ψ is a unit matrix.

However, the covariance matrix of $\hat{\mathbf{P}}_M$ cannot be directly calculated in a manner similar to equation (5), as the robust weight matrix Ψ is random and its value depends on the value of \mathbf{r} .

The asymptotic and unbiased estimate of $\text{cov}(\hat{\mathbf{P}}_M)$ in a finite-sample case is (Hampel et al. 1986)

$$\text{cov}(\hat{\mathbf{P}}_M) = s^2 \frac{m^2 \sum_{i=1}^m \phi_i^2(r_i/s)}{(m-n) \left[\sum_{i=1}^m \frac{\partial \phi(r_i/s)}{\partial r_i} \right]^2} (\mathbf{M}_w^T \mathbf{M}_w)^{-1}, \quad (15)$$

Huber (1981) has demonstrated that when $\frac{n^3}{m} \rightarrow 0$, equation (15) asymptotically approaches to the true covariance of \mathbf{P}_M .

If $\phi(\mathbf{r})$ is set to be \mathbf{r} , equation (15) becomes

$$\text{cov}(\hat{\mathbf{P}}_M) = \frac{\mathbf{r}^T \mathbf{r}}{m-n} (\mathbf{M}_w^T \mathbf{M}_w)^{-1}, \quad (16)$$

which is exactly the alternative form of $\text{cov}(\hat{\mathbf{P}}_{LS})$.

Given that the scale s is usually unknown for most cases, it needs to be estimated with \mathbf{P} simultaneously.

In order to have a robust scale estimate result, the median absolute deviation (MAD) of the fitting residual is employed, i.e.,

$$\hat{s} = \text{MAD} = \text{median}|\mathbf{r}|/0.6745. \quad (17)$$

3.3 Numerical solution of M-estimator

It follows from equations (3) and (13), compared with LS, that the M-estimator has an extra weight matrix depending on the fitting residual. Thus, the M-estimator can be solved by using an iterative solution as (Street, Carroll & Ruppert 1988)

$$\hat{\mathbf{P}}_M^{(j+1)} = (\mathbf{M}_w^T \Psi^{(j)} \mathbf{M}_w)^{-1} \mathbf{M}_w^T \Psi^{(j)} \mathbf{R}_w, \quad (18)$$

where the superscript (j) denotes the j th iteration. The iteration can be initialized using the solution of LS, i.e., equation (4).

The above iterative algorithm is termed iteratively reweighted LSs, which will converge if the Ψ function is non-increasing (for $r_i > 0$).

Meanwhile, the scaling parameter s can also be iteratively calculated by

$$\hat{s}^{(j+1)} = \text{median}|\mathbf{r}^{(j)}|/0.6745. \quad (19)$$

Note that there is not a rigorous demonstration that equation (19) will converge except for the Huber loss function, which will be introduced in Section 3.4. Thus, in practice, we recommend the Huber loss function is employed first to estimate s and then the other robust loss functions can be used to fulfil various M-estimators.

3.4 Loss functions of M-estimator

There are three principles for designing a ρ function of the M-estimator: (1) the function should be convex to ensure the presence of a minimum; (2) the function should be bounded to resist the impact of non-Gaussian noise and (3) the function should guarantee the corresponding M-estimator has a fitting result similar to LS in a Gaussian environment. Based on the above principles, there are various ρ functions that have been proposed. For a univariate estimation problem, three predominant ρ functions (i.e., Huber function, Bisquare function and Welsch function) and the corresponding ϕ functions and Ψ functions are shown in Table 1 (Llanos, Sanchez & Maronna 2015). For comparison, the counterpart of ρ function, ϕ function as well as the Ψ function of LS is also provided.

In Table 1, η represents the normalized residual (i.e. residual scaled by the expected error) and a is a tuning parameter crucial to guarantee a satisfactory robust fitting result. If a approaches infinity, the ρ function of an M-estimator approaches that of LS, reducing its effectiveness at resisting the impact of non-Gaussian noise. If a approaches zero, the ρ function of the M-estimator approaches zero, indicating that it rejects all the residuals even in the case of a purely Gaussian environment. Thus, a sensible choice a is needed. One approach to set a is to evaluate the efficiency of the

Table 2. Optimal tuning parameters.

Robust function	a
Huber function	1.345
Bisquare function	4.685
Welsch function	2.11

M-estimator in a Gaussian environment (Llanos et al. 2015), which is defined as

$$\text{eff} = 100 \frac{\text{STD}_M}{\text{STD}_{LS}} \text{ per cent}, \quad (20)$$

with STD_M of the standard deviation of M-estimator calculated in a Gaussian environment and STD_{LS} of the standard deviation of LS calculated in the same case. When the efficiency is 95 per cent, Table 2 shows the optimal tuning parameters of the above three ρ functions of M-estimators.

Fig. 1 shows the graphic representations of the ρ functions and Ψ s. The growth of the ρ functions of the M-estimators in the central regions ($\eta < a$) is generally slower than that of the LSs. Outside of the central region, the estimators become linear (Huber), negatively exponential (Welsch) or flat (Bisquare), which severely reduces the impact of data points with large η to the parameter estimates.

In this way, the Ψ functions behave like ‘windows’ that are symmetrical about zero and adjust the weights of data points according to their η value. Roughly speaking, the data within the ‘windows’ are given much greater weight than those outside. In addition, the weights continue to reduce as the absolute values of η grows. The main difference among the three Ψ functions of M-estimators is the shape of ‘window’. The ‘window’ of Huber function is flat within the threshold and becomes hyperbolas outside the threshold. The ‘windows’ of Bisquare function and of Welsch function are like

a bell that emphasizes the impact of data with η close to zero and gradually reduces the weights of those further from zero.

4 IDENTIFICATION OF NON-GAUSSIAN DISTRIBUTION

The most common formal normality tests available in literature are the SW test, the Kolmogorov–Smirnov (KS) test, the Anderson–Darling (AD) test and the Lilliefors (LF) test. Razali & Wah (2011) utilized Monte Carlo simulation to compare the power of the above four tests and found that the SW test was the most powerful normality test, followed by the AS test, the LF test and the KS test although the power of all four tests dropped for small sample size. Therefore, we employ the SW test to identify the presence of non-Gaussian noise within the recorded ToAs.

Assuming that there is an ordered random sample, $y_1 < y_2 < \dots < y_n$, the statistic of SW test is (Shapiro & Wilk 1965)

$$W = \frac{(\mathbf{b}^T \mathbf{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (21)$$

where \bar{y} is the sample mean, and

$$\mathbf{b} = \frac{\mathbf{m}^T \mathbf{V}^{-1}}{(\mathbf{m}^T \mathbf{V}^{-1} \mathbf{V}^{-1} \mathbf{m}^T)^{1/2}}. \quad (22)$$

In equation (21), $\mathbf{m} = (m_1, m_2, \dots, m_n)^T$ is a vector consisted of the expected values of the order statistics of independent and identically distributed variables sampled from the standard Gaussian distribution and \mathbf{V} is the covariance matrix of the order statistics.

The statistic W can be interpreted as the squared correlation coefficient of \mathbf{y} and \mathbf{b} , the higher the value of which, the more likely the samples are normally distributed.

In practice, the identification of the non-Gaussianity of ToAs proceeds by two steps: (1) obtaining the post-fit residuals via LS

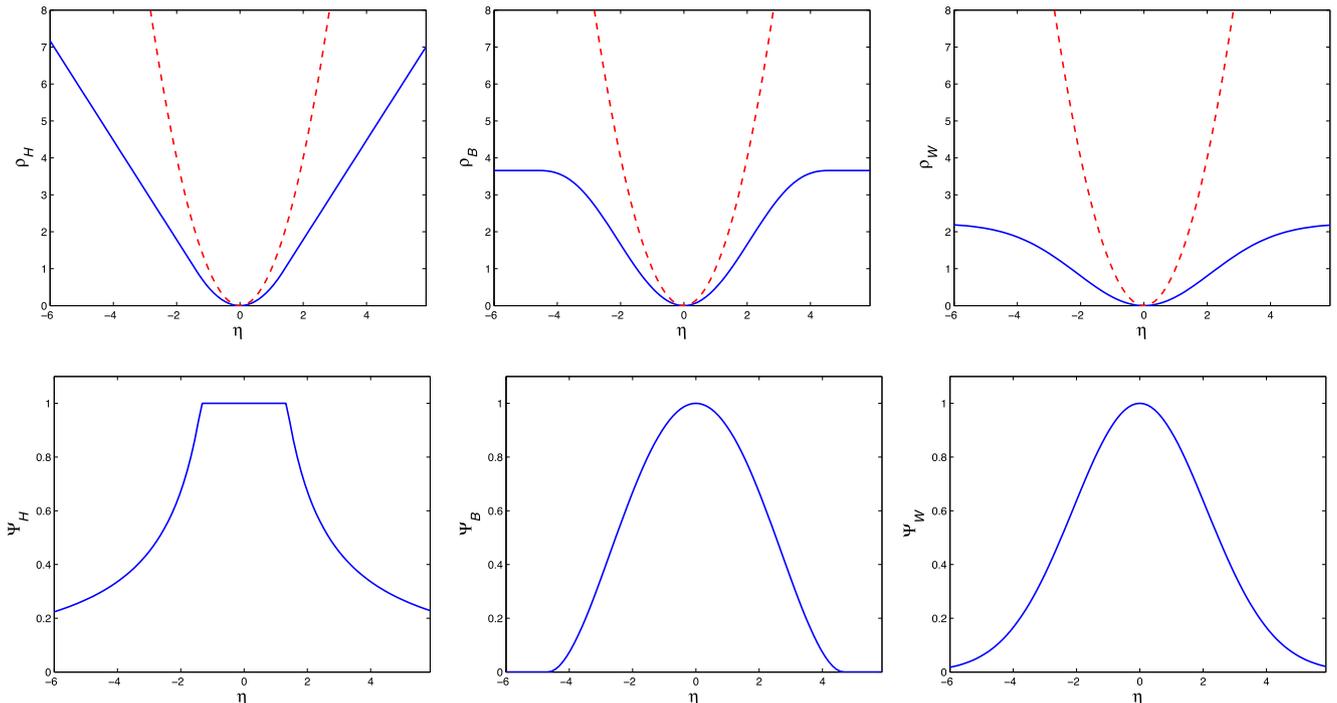


Figure 1. Graphic representations of the Huber function(ρ_H), Bisquare function(ρ_B) and Welsch function(ρ_W), Ψ_H , Ψ_B , Ψ_W . The top row shows the ρ functions, where the dashed line represents the ρ function of LS and the solid lines represent the ρ functions of M-estimator. The bottom row shows the Ψ s.

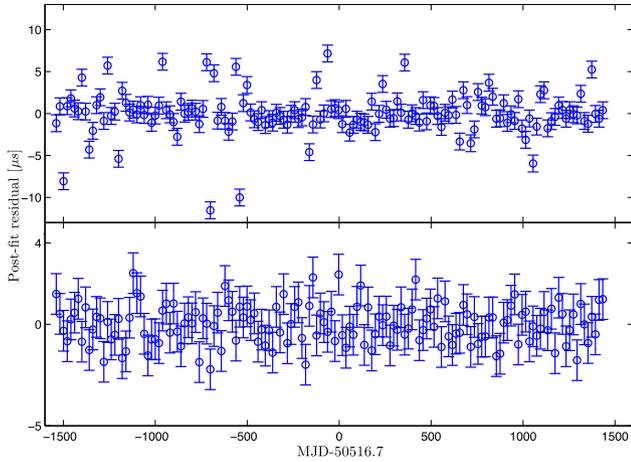


Figure 2. One realization of non-Gaussian post-fit residuals (upper) and Gaussian post-fit residuals (bottom) after subtracting the timing model under the condition that β is 0.3.

and (2) applying the *SW* test to the resultant post-fit residuals. The ToAs would be viewed as being of a non-Gaussian distribution, if the p value of the *SW* test is less than 0.05.

5 APPLICATION TO SIMULATION

In this section, we investigate the estimate performance of robust fitting by simulating a series of ToAs from a pulsar. Simulations are performed using `TEMPO2`. More specifically, the true observations are simulated by the plug-in `FORMIDEAL`, and a new plug-in `ADDDGAUSSIANMIXTURE` was developed to simulate the non-Gaussian noise by employing Gaussian mixture model. The plug-in `CREATEREALISATION` is employed to generate ToAs. In the following simulations, the error bars of the generated ToAs are set to the standard deviation of the modelling distribution, simulating a practical situation where we have no idea whether there is a contamination source.

The root-mean-square (RMS) error of the fitted parameter with respect to the true value is a good metric to evaluate the performance of the estimator. It includes contributions from both bias and the variance of the estimator. For each simulation, we compute the mean and RMS of the 1000 realizations and also the square root of sample variance.

5.1 Simulation for contaminated Gaussian distribution

In this case, the modelling and contamination distributions are assumed to be $g(x|0, 1\mu\text{s})$ and $g(x|0, 5\mu\text{s})$, where $g(x|\mu, \sigma)$ denotes the PDF of a Gaussian distribution with the mean of μ and the standard deviation of σ .

And then, the PDF of the investigated distribution is

$$p_1(x) = (1 - \beta)g(x|0, 1\mu\text{s}) + \beta g(x|0, 5\mu\text{s}), \quad (23)$$

where β is the unknown occurrence probability of $g(x|0, 5\mu\text{s})$.

Fig. 2 provides the simulated non-Gaussian post-fit residuals as well as the Gaussian post-fit residuals for comparison. Compared with Gaussian residuals, the non-Gaussian residuals are less central and have ‘pseudo-outliers’. Those points having values of more than $3\mu\text{s}$ would be viewed as outliers if $g(x|0, 1\mu\text{s})$ is assumed to be the true distribution. In fact, those points are produced by the contamination distribution and thus it is not recommended that they

are simply deleted. Therefore, we call them ‘pseudo-outliers’. It should be noted that if the contamination distribution is not caused by instrument problems, it may lead to some new discoveries.

On the other hand, given that equation (23) provides the explicit expression of investigated non-Gaussian PDF, a log likelihood function can be derived as

$$\ln p_1 = \sum_{i=1}^m \ln[(1 - \beta)g(\bar{r}_i|0, 1\mu\text{s}) + \beta g(\bar{r}_i|0, 5\mu\text{s})], \quad (24)$$

where \bar{r}_i is the i th component of $\mathbf{R} - \mathbf{MP}$. And then, the parameters of pulsar timing model can be estimated by maximizing equation (24).

In order to completely investigate the performance of M-estimator, Table 3 shows the fitting results of LS, M-estimators and MLE when β is 0.3 and 300 observation ToAs are employed. The statistics, including sample mean and RMS, are provided. The right ascension, declination and pulse frequency are fitted. The tuning parameters required by M-estimators are set according to Table 2. The sample means are all less than 1 per cent of RMSs, which indicates that the listed methods are unbiased and $\sqrt{\text{variance}}$ is close to RMS. Thus, the sample variance is not provided in the table. In order to facilitate the comparison between M-estimators and LS, an index defined as

$$\lambda = \frac{1}{n} \sum_{i=1}^n \frac{\text{RMS}_E}{\text{RMS}_{\text{MLE}}} \quad (25)$$

is employed. In equation (25), RMS_E denotes the RMSs calculated by LS and M-estimators and RMS_{MLE} is the RMS of MLE. The λ quantifies the divergence from the result of M-estimator or LS to the best possible, i.e. MLE. Thus, the λ s for M-estimators are 3.76 (Huber), 3.51 (Bisquare) and 3.33 (Welsch), respectively, whereas that for LS is 6.48. It means that M-estimators could provide a fitting result more accurate than LS. The differences among M-estimators are insignificant, indicating that the exact expression of ρ function puts little impact on the fitting result. Considering that the pulsar timing analysis keeps pursuing increasingly high-fitting accuracy, the robust fitting is therefore of importance.

Next, we investigate the consistency of the covariance estimated by M-estimators. First of all, an index that can quantify the consistency is defined as

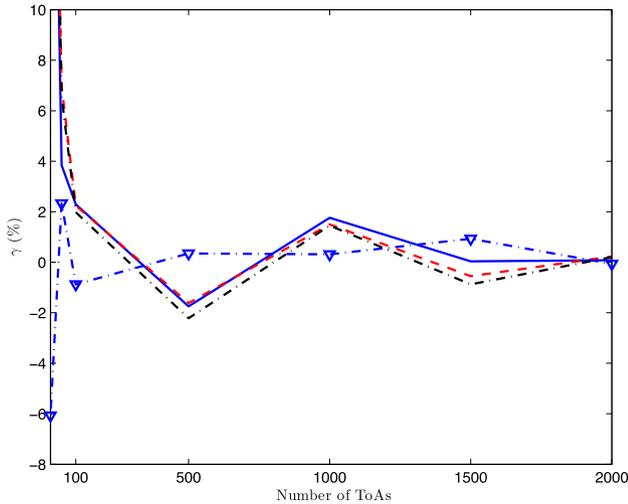
$$\gamma = \frac{1}{n} \sum_{i=1}^n \frac{\text{Var}_S - \text{STD}_E}{\text{Var}_S} \times 100 \text{ per cent}, \quad (26)$$

where Var_S is the square root of sample variance and STD_E is the analytic standard deviation of estimators, which is calculated as the square root of the diagonal of covariance matrix of estimator (i.e., equation 5 for LS and equation 15 for M-estimator).

Fig. 3 shows how γ varies with an increasing number of ToAs, i.e., m . The γ s for all methods gradually converge to zero as m increases, which indicates that equation (15) could provide a consistent and asymptotic standard deviation for M-estimators. Then, regarding M-estimators as unbiased, the performance of M-estimator could be assessed by the diagonal of the resultant covariance matrix of M-estimator. The γ s for M-estimators converge at a speed slower than LS. It is because that the speed of M-estimators’ convergence depends on the speed at which $\frac{n^2}{m}$ approaches 0 whereas the convergence speed of LS is up to the speed at which $\frac{n}{m}$ approaches 0 (Huber 1973).

Table 3. Fitting results for the case where the non-Gaussian distribution is a contaminated Gaussian distribution with β of 0.3.

		LS	Huber	Bisquare	Welsch	ML
Declination (10^{-5} arcs)	Mean	-1.32	-0.49	-0.24	-0.19	-0.18
	RMS	19.05	11.07	10.27	9.74	8.57
Right ascension (10^{-6} s)	Mean	-1	-0.39	-0.18	-0.16	-0.51
	RMS	17.44	9.93	9.24	8.81	7.6
Pulse frequency (10^{-14} s $^{-1}$)	Mean	2.89	0.62	-0.04	-0.087	0.0285
	RMS	62	36.76	34.64	32.89	31.65

**Figure 3.** γ varying with the number of ToAs. The dot-dashed line with the marker ‘ ∇ ’ represents the γ for LS. The solid, dashed, dot-dashed lines provide the γ s for the M-estimator (Huber), the M-estimator (Bisquare) and the M-estimator (Welsch), respectively.

Now, we analyse how the fitting performance of M-estimator will change as β grows from 0 to 0.5. In order to facilitate the investigation, a new index, θ , is defined as

$$\theta = \frac{1}{n} \sum_{i=1}^n \frac{\text{RMS}_E - \text{RMS}_{\text{MLE}}}{\text{RMS}_{\text{MLE}}}. \quad (27)$$

Fig. 4 shows how θ varies with an increasing β . When β is 0, the investigated distribution becomes Gaussian, the fitting performance of the estimators is similar to each other. When β is positive, the distribution becomes non-Gaussian. θ for LS rapidly increases as β grows, whereas M-estimators are less affected. When β is less than 0.3, θ s for M-estimators maintain a level, which indicates that M-estimators could provide results close to MLE then. At last, there is insignificant difference among different M-estimators.

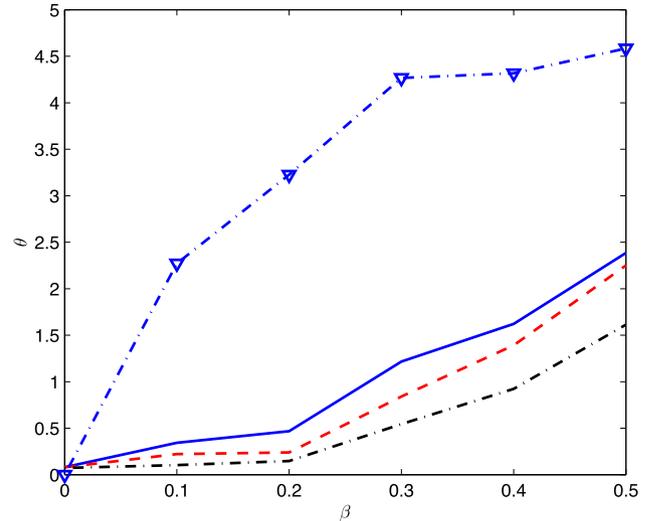
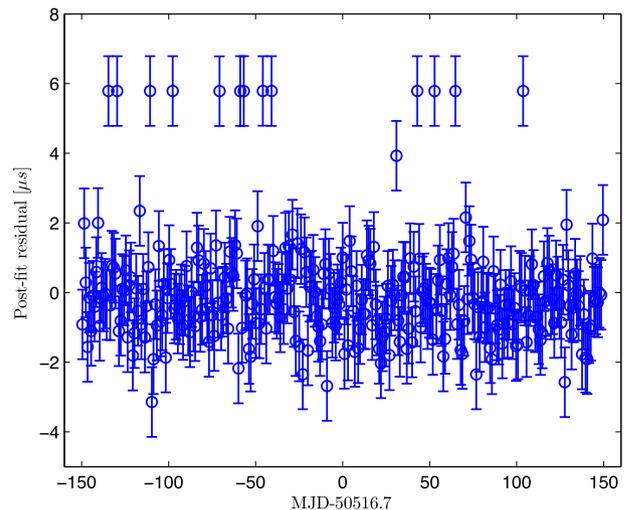
5.2 Simulation for outliers

In this section, we model the non-Gaussian noise as impulsive outliers. In this way, the investigated non-Gaussian distribution has a PDF of

$$p_2(x) = (1 - \beta)g(x|0, 1\mu\text{s}) + \beta h, \quad (28)$$

where h is the impulsive noise that occurs at $6\mu\text{s}$ and β is set to be 0.05 according to the fact that there are not many outliers in a data set. Fig. 5 shows the post-residual in this case.

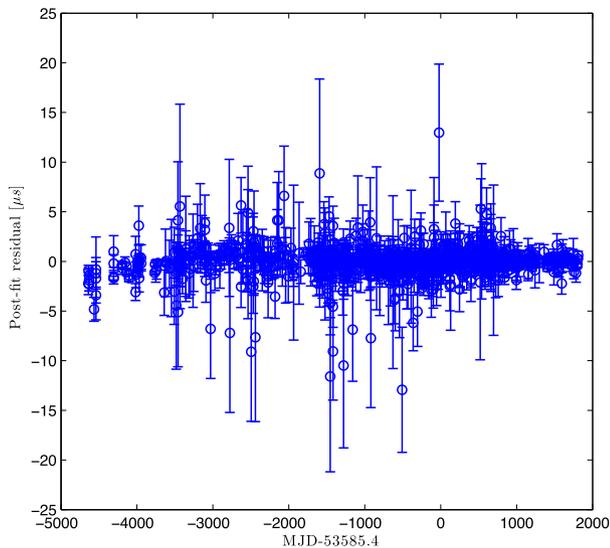
Table 4 shows the fitting result. As the MLE is difficult to be derived, we provide the result of LS handling ToAs with no outliers as

**Figure 4.** θ varying with the contamination ratio. The dot-dashed line with the marker ‘ ∇ ’ represents the θ for LS. The solid, dashed, dot-dashed lines provide the θ s for the M-estimator (Huber), the M-estimator (Bisquare) and the M-estimator (Welsch), respectively.**Figure 5.** One realization of the post-fit residuals for the case where outliers occur. There are some outliers with values of $6\mu\text{s}$, which is caused by the impulsive component of $p_2(x)$.

a reference to analyse the improvement of M-estimators over LS in this case. By replacing the RMS_{MLE} in equation (25) to be the RMS calculated by the reference case, we could quantify the improvement of M-estimators over LS. Specifically, the modified λ s for M-estimators are 3.6 (Huber), 3.15 (Bisquare) and 3.26 (Welsch),

Table 4. Fitting results for the case where the outliers occur.

		LS	Huber	Bisquare	Welsch	LS (no outliers)
Declination (10^{-5} arcs)	Mean	-0.09	-0.049	-0.046	-0.039	-0.061
	RMS	11.22	7.92	6.95	7.15	6.53
Right ascension (10^{-6} s)	Mean	-0.34	-0.067	-0.041	0.016	-0.012
	RMS	10.19	7.18	6.31	6.55	6.03
Pulse frequency (10^{-14} s $^{-1}$)	Mean	1.12	0.55	0.14	0.14	0.16
	RMS	37.58	26.61	23.3	24.13	22.22

**Figure 6.** Post-fit residuals of the PSR J1713+0747. The ToAs are from EPTA and fitted by LS.

respectively and that for LS is 5.1. Thus, although all the methods are affected by the presence of outliers, the M-estimators could still provide a fitting result more accurate than LS.

6 APPLICATION TO REAL DATA

In this section, we make use of radio observations of PSR J1737+0747 taken by the European Pulsar Timing Array (EPTA), including the Effelsberg Radio Telescope in Germany, the Lovell Radio Telescope at the Jodrell Bank Observatory, the Nançay Radio Telescope in France, and the Westerbork Synthesis Radio Telescope in the Netherlands. The observation spans from MJD 50360 to 56810, and includes 1188 observed ToAs. The detailed description of the observation can be found in Desvignes et al. (2016).

Fig. 6 shows the post-fit residuals after subtracting the timing model. The ToAs are viewed to be drawn from a non-Gaussian distribution, as the p value of SW test is 0.042 652. Since the exact expression of ρ function has little impact on the final fitting result, only the results of M-estimator(Huber) and M-estimator(Bisquare) are provided in Table 5. The Solar system ephemeris DE 421, binary model T2 and clock correction procedure of TT(BIPM2011) are employed. The M-estimators could provide fitting results with uncertainties less than LS, which indicates that the results of M-estimators are more accurate than LS. It is because that, as illustrated in Section 5, all the listed methods are unbiased and the fitting performance can be well assessed by the uncertainties within the fitting results.

PSR J1737+0747 has been well observed by various means, including the timing and the very long baseline interferometry (VLBI). In Table 6, we list the parameters for PSR J1737+0747 determined by timing (this work) and by VLBI (Chatterjee et al. 2009). All parameters are for equinox J2000, results of timing are shifted to match the VLBI data epoch of MJD 52275, and positions are listed as offsets from right ascension 17h13min49s ($\Delta\alpha$) and declination $7^{\circ}37'47''$ ($\Delta\delta$). All timing observations agree with the VLBI within 2σ of the VLBI error bars and the timing M-estimators agree with each other within 2σ . However, the timing observations by LS do not agree with those by M-estimator within 2σ . Compared with LS, the positions provided by M-estimators are closer to VLBI. Regarding that the difference between the pulsar positions acquired by timing and VLBI reflects the ties between the dynamical Solar system ephemeris and the extragalactic international celestial reference frame, which is derived to be around 0.2 mas (Folkner & Border 2015), M-estimator can also be employed to realize the reference frame ties to high accuracy. In addition, the parallax is a good test as it is not affected by the frame tie, the ephemeris version or the VLBI reference source, and it is equally well determined by VLBI and timing. In this case, the M-estimators are consistent with the VLBI estimate but the LS estimator is not.

The computational times for the J1713+0747 data set were 1.89, 2.07 and 2.1 s for LS, M-estimator (Huber) and M-estimator (Bisquare), respectively. Since the M-estimators are at worst 11 per cent more expensive than LSs, it is quite practical to merge them into the current pulsar timing packages.

7 CONCLUSIONS

In this paper, we introduce the robust fitting into the current pulsar timing process to cope with non-Gaussian noise. As a main approach to robust fitting, an M-estimator can resist the impact of non-Gaussian noise by modifying the loss function of the parameter fitting problem to be bounded. The SW test is employed to test whether the uncertainty in the observed ToAs is drawn from a non-Gaussian distribution. Two types of non-Gaussian distributions are simulated. M-estimators achieve RMS error smaller than LS at the cost of slightly more expensive computation. In addition, the exact expression of loss function of M-estimator has little impact on the final fitting result. M-estimators are also applied to real timing data of PSR J1713+0747. The recorded ToAs are found to be from a non-Gaussian distribution as the p value of SW test applied to the observed ToAs is less than 0.05. In this case, M-estimators could have fitting result more accurate than LS and are more close to that of VLBI.

ACKNOWLEDGEMENTS

YW is grateful to the China Scholarship Council, which supports his overseas study in the University of Manchester.

Table 5. Fitting result of PSR J1713+0747.

Model parameter	Measured quantities		
	LS	Huber	M-estimator Bisquare
Right ascension, α (h:min:s)	17:13:49.533 1799(5)	17:13:49.533 1775(3)	17:13:49.533 1773(3)
Declination, δ (deg:min:s)	+07:47:37.492 69(16)	+07:47:37.492 59(9)	+07:47:37.492 57(9)
Pulse frequency, ν (s ⁻¹)	218.811 840 417 1585(4)	218.811 840 417 158 12(17)	218.811 840 417 158 07(17)
First derivative of pulse frequency, $\dot{\nu}$ (s ⁻²)	-4.083 64(6) × 10 ⁻¹⁶	-4.083 71(4) × 10 ⁻¹⁶	-4.083 72(4) × 10 ⁻¹⁶
Dispersion measure, DM (cm ⁻³ pc)	15.9919(5)	15.9921(3)	15.9921(3)
First derivative of dispersion measure, \dot{DM} (cm ⁻³ pc yr ⁻¹)	3.9(24) × 10 ⁻⁵	-0.065(13) × 10 ⁻⁵	-0.75(13) × 10 ⁻⁵
DM2 (cm ⁻³ pc yr ⁻²)	2.8(7) × 10 ⁻⁵	1.7(4) × 10 ⁻⁵	1.6(4) × 10 ⁻⁵
Proper motion in right ascension, $\mu_\alpha \cos \delta$ (mas yr ⁻¹)	4.897(3)	4.9086(14)	4.9104(14)
Proper motion in declination, μ_δ (mas yr ⁻¹)	-3.944(6)	-3.925(4)	-3.921(4)
Parallax, π (mas)	1.05(4)	0.93(3)	0.90(3)
Orbital period, P_b (d)	67.825 130 952(6)	67.825 130 963(3)	67.825 130 963(3)
Epoch of periastron, T_0 (MJD)	48 741.9732(5)	48 741.9734(3)	48 741.9732(3)
Projected semimajor axis of orbit, x (lt-s)	32.342 4199(3)	32.342 419 74(12)	32.342 419 83(11)
Longitude of periastron, ω_0 (°)	176.196(3)	176.1976(12)	176.1966(12)
Orbital eccentricity, e	7.494 43(11) × 10 ⁻⁵	7.494 28(6) × 10 ⁻⁵	7.494 25(6) × 10 ⁻⁵
Companion mass, M_c (M_\odot)	0.241(12)	0.260(7)	0.255(7)
Longitude of ascending node, Ω (°)	66(6)	77(4)	76(3)
Orbital inclination angle, i (°)	76.4(6)	75.5(3)	75.7(3)
	Set quantities		
Epoch of frequency determination (MJD)	55 000		
Epoch of position determination (MJD)	55 000		
Epoch of dispersion measure determination (MJD)	55 000		
Sine of inclination angle, $\sin i$	0.981 561		

Table 6. Parameters provided by VLBI, LS and M-estimator.

	VLBI	LS	M-estimator	
			Huber	Bisquare
$\Delta\alpha$ (s)	0.5306(1)	0.530 7217(15)	0.530 7138(9)	0.530 7124(8)
$\Delta\delta$ (")	0.519(2)	0.522 11(5)	0.521 88(3)	0.521 83(3)
μ_α (mas yr ⁻¹)	4.75 ⁺¹⁷ ₋₇	4.897(3)	4.9079(15)	4.9102(14)
μ_δ (mas yr ⁻¹)	-3.67(16)	-3.944(6)	-3.926(4)	-3.921(4)
π (mas)	0.95(6)	1.05(4)	0.93(3)	0.89(3)

REFERENCES

- Bishop C., 2006, *Pattern Recognition and Machine Learning*. Springer, New York
- Chatterjee S. et al., 2009, *ApJ*, 698, 250
- Coles W., Hobbs G., Champion D., Manchester R., Verbiest J. P. W., 2011, *MNRAS*, 418, 561
- Desvignes C. et al., 2016, *MNRAS*, 458, 3341
- Folkner W., Border J., 2015, *Highlights of Astron.*, 16, 219
- Hampel F., Ronchetti E., Rousseeuw P., Stahel W., 1986, *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York
- Hobbs G., Edwards R., Manchester R., 2006, *MNRAS*, 369, 655
- Hobbs G. et al., 2009, *MNRAS*, 394, 1945
- Holland P., Welsch R., 1977, *Commun. Stat. - Theory Methods*, 6, 813
- Huber P., 1964, *Ann. Stat.*, 35, 73

- Huber P., 1972, *Ann. Math. Stat.*, 43, 1041
- Huber P., 1973, *Ann. Stat.*, 1, 799
- Huber P., 1981, *Robust Statistics*. Wiley, New York
- Janet F., Hobbs G., Lee K., Manchester R., 2005, *ApJ*, 625, L123
- Lentati L., Alexander P., Hobson M., Feroz F., van Haasteren R., Lee K. J., Shannon R., 2014a, *MNRAS*, 437, 3004
- Lentati L., Hobson M., Alexander P., 2014b, *MNRAS*, 444, 3863
- Llanos C., Sanchez M., Maronna R., 2015, *Ind. Eng. Chem. Res.*, 54, 5096
- Matsakis D., Taylor J., Eubanks T., 1997, *A&A*, 326, 924
- Razali N., Wah Y. J., 2011, *Stat. Modelling Anal.*, 2, 21
- Shapiro S., Wilk M., 1965, *Biometrika*, 52, 591
- Street J., Carroll R., Ruppert D., 1988, *Am. Stat.*, 42, 152
- Taylor J., Weisberg J., 1989, *ApJ*, 345, 434

APPENDIX A: USAGE OF ROBUST FITTING IN TEMPO2

Robust fitting is merged into the latest version of TEMPO2. It can be employed by adding ‘ROBUST option’ to the pulsar parameter file. There are three robust options, in which ‘H’ refers to Huber function, ‘B’ refers to Bisquare function and ‘W’ refers to Welsch function.

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