An Algorithm for Mitigating Transient RFI in Pulsar Observation

Yining Song1,2, Zhiyong Liu1,3, Na Wang1,3, Jun Li4, and Rai Yuen1

1 Xinjiang Astronomical Observatory, Chinese Academy of Sciences, Urumqi 830011, People’s Republic of China; liuzhy@xao.ac.cn
2 University of Chinese Academy of Sciences, Beijing 100094, People’s Republic of China
3 Key Laboratory of Radio Astronomy, Chinese Academy of Sciences, Nanjing 210008, People’s Republic of China; na.wang@xao.ac.cn
4 Xichang University, Sichuan 615000, People’s Republic of China

Received 2021 July 16; revised 2021 October 14; accepted 2021 October 21; published 2021 November 24

Abstract

We propose an algorithm, referred to as the pulsar phase and standard deviation (PPSD), to mitigate transient radio frequency interference (RFI) in pulsar observations. PPSD uses the model for pulsar time of arrival to identify pulsar phase and extract the pulse profile to protect the original pulsar profile. PPSD sets a threshold based on the statistics empirical rule to label the transient RFI in the off-pulse data until all unlabelled off-pulse data obeys the white Gaussian noise (WGN) distribution. The transient RFI data is then substituted with WGN. Finally, we use PPSD to process the pulsar observation data obtained from the NanShan 25 m Radio Telescope. Our results show that PPSD can effectively mitigate the transient RFI and improve the signal-to-noise ratio of the pulsar observations.

Unified Astronomy Thesaurus concepts: Pulsars (1306); Astronomical techniques (1684)

1. Introduction

Pulsars are rapidly rotating neutron stars with extremely high density and magnetization. They are natural laboratories for studying the state of matter under extreme physical conditions. Because of the extremely stable rotation period in most pulsars, the pulse time of arrival (TOA) can be used as a pulsar clock for gravitational wave detection, deep space navigation, and applications in other fields (Will 2014; Wu et al. 2018). The emission of pulsars is mainly detected at radio frequencies, which means that the observation data is inevitably subjected to radio frequency interference (RFI) originating from communication and electronic equipments. To some extent, RFI can be mitigated or eliminated by shielding and other technical methods (Sureshkumar et al. 2016; Liu et al. 2019; Li et al. 2020), as well as through the RFI management regulations (Spoelstra 2002; Liu & Liu 2019). However, the impact of RFI still remains in the observed data, and subsequent processing is required for further mitigation in the post-correlation (Baan 2010). RFI can be divided into stable RFI and transient RFI. Typical stable RFI includes alternating current at 50 Hz and satellite signals, whereas transient RFI typically involves sparks from plug ignitions and switching in power supplies. Transient RFI is difficult to identify and mitigate, yet it can easily affect pulsar observations (Czech et al. 2018).

There are many RFI mitigation methods that have good applications in radio astronomy (Fridman & Baan 2001; Baan 2010; An et al. 2017). However, emission from pulsars has unique profile characteristics and regular periodicity; and hence ordinary RFI mitigation methods may remove those characteristics. Therefore, some algorithms were specially designed to mitigate the RFI in pulsar observations. Existing RFI mitigation algorithms typically fall into three categories (Akeret et al. 2017). The first category involves extension from the basis of the linear method, such as the adaptive filter system used at the Parkes 64 m radio telescope for mitigating RFI in pulsar observations (Kesteven et al. 2005). Kocz et al. (2012) designed automated RFI detection in multipixel feeds to reduce false candidates. Fourier-domain RFI excision was proposed by Maan et al. (2020), which increases the effective sensitivity of astrophysical sources. The second category uses threshold-based algorithms, such as (Spitler et al. 2012) excised narrowband RFI by applying a modulation index threshold. Baan (2019) described a threshold method based on kurtosis, which can mitigate time-variable interference in real-time. Zeng et al. (2021) proposed several novel strategies based on the SumThreshold algorithm for detecting different types of RFI. The third category employs machine learning, such as (Bethapudi & Desai 2018) separate pulsar signals from noise using supervised machine-learning algorithms. Wang et al. (2020) improved pseudoinverse learning to mitigate RFI, while retaining channels that are contaminated by RFI. Rodriguez Arzaga & Lynch (2021) combined machine-learning algorithm and traditional RFI excision techniques to identify and replace the RFI before the data are further integrated.

Although many algorithms have been developed for pulsar RFI mitigation, only a few of them have been widely used in practical applications. At present, PSRCHIVE provides zapping function to mitigate transient RFI (van Straten et al. 2012), but it does not work effectively for some pulsar observations. Consequently, most astronomers deal with transient RFI manually. With the increase in the size of observation data and the demand for real-time and online data processing, it has become an urgent problem to identify and mitigate the transient RFI in pulsar observations through algorithms.

We propose a pulsar phase and standard deviation algorithm (PPSD) based on the pulsar TOA model combined with the statistics empirical rule to mitigate the transient RFI in pulsar observations. This paper introduces the PPSD algorithm and data processing results in detail. Section 2 introduces the method for mitigating transient RFI in pulsar observations and the specific steps for implementation. Section 3 introduces the data characteristics, and the results of transient RFI mitigation using the PPSD algorithm. We conclude the paper in Section 4.
sub-band using
\[ \Delta t_{i}^{DM} = D \times \left( \frac{1}{v_{i}^{2}} - \frac{1}{v_{0}^{2}} \right) \times DM. \] (3)

Here, DM is the dispersion measure; \( \Delta t_{i}^{DM} \) is the dispersion delay time between \( v_{i} \) and \( v_{0} \); \( D \) is the dispersion constant; \( v_{i} \) is the \( i \)th sub-band observation frequency; and \( v_{0} \) is its corresponding reference frequency (Lorimer & Kramer 2012). Hence, the specific position of the phase at different frequencies can be obtained using Equation (3).

Since the pulse phase varies across the frequency channels, after obtaining the precise phase at each frequency channel, the pulse width needs to be solved separately to ensure accurate extraction of the pulse data. In our calculation, we select the standard profile phase at the center of the observing station. Then, a baseline data is extracted across a width that is two times greater than the width measured at 10% of the peak intensity (\( W_{p0} \)) of the integrated profile (Lyne & Graham-Smith 2012). For a stable observing system, the baseline data can be considered as linearly correlated with the system temperature. Hence, linear fitting is used to subtract the background data. Next, the positions on the left and on the right outside of the baseline data where negative values first appear are determined and marked as the left starting position and right ending position of the pulse phase, respectively. The pulse profile is then extracted based on these two positions.

2.2. Identify and Mitigate RFI

After extracting the pulsar profile, the off-pulse data are white background noise and possible RFI. We use the method of threshold setting and loop iterations to identify and label the RFI in the data.

First, we calculate the mean value, \( \mu \), and the standard deviation, \( \sigma \), of the off-pulse data. The threshold range is then set as \( (\mu - k\sigma, \mu + k\sigma) \), where \( k \in \mathbb{R}^+ \) is the coefficient of the threshold range.

In practice, we find that the standard 3\( \sigma \) rule does not work well in our case. Accordingly, we reduce the coefficient \( k \) for the threshold to adapt to the RFI situation. By analyzing the observation data, we arrive at the reasonable range for \( k \in [1.8, 3] \) (Lyon et al. 2016).

In addition, we label the data above the threshold as RFI. In order to determine whether all RFIs have been recognized, PPSD recalculates \( \mu \) and \( \sigma \), with the new values represented by \( \mu' \) and \( \sigma' \), of the unlabeled data and obtains a new threshold. If all unlabeled data are below the new threshold, the RFI labeling is complete. Otherwise, the new threshold is used to identify and label the remaining RFI until all unlabeled data are within the latest threshold.

Next, the latest \( \mu' \) and \( \sigma' \) are used to generate random data points that obey Gaussian white noise (WGN) \( N(\mu', \sigma') \) to substitute for the RFI. After that, all off-pulse data obeys the WGN distribution. The on-pulse data is then put back to the original position in the data, thereby completing the process for mitigating the transient RFI.

The above processing is repeated for other frequency channels until all frequencies have been completed for mitigation. Finally, the processing results after RFI mitigation are output and saved. The specific algorithm workflow diagram is shown in Figure 2.
In this section we focus on the effect of PPSD on mitigating
transient RFI in pulsar observations.

3.1. Data Source

With the rapid development and application of electronic and
communication technology, radio astronomical observatories
generally face the challenge of RFI. The NanShan 25 m Radio
Telescope (NSRT) is located far away from the city at the north
of the Tianshan Mountains in China. In the initial stage of the
construction, the radio astronomy observation environment at
the site was away from man-made RFI (Liu et al. 2019). Since
then, the RFI at the NSRT site has gradually worsened in
partial time and in specific observing directions, and the
existing mitigation method for transient RFI has become
unsatisfactory for scientific purposes.

For testing of PPSD, we selected some pulsar timing
observations from NSRT at L band that contained obvious
transient RFI. The receiver backend for the data collection was
Digital Filter Bank 3 (DFB3), which was developed by CSIRO,
Australia. The kernel data parameters included the frequency
range, which covered the range of 1300–1812 MHz, and the
sub-integration time of 30 s. The operating system for the
experimental environment was Ubuntu 20.4, and the program-
ming language was python3.7.

3.2. Experimental Results

We evaluate the effectiveness of the algorithm based on the
signal-to-noise ratio (S/N) of the pulsar signals. The S/N is
expressed as
\[
S/N = \frac{S_p}{N_{\text{rms}}} \tag{4}
\]
where \(S_p\) is the peak value of the original signal, \(N_{\text{rms}}\) refers
to the rms of the noise excluding the signal: \(\sum_{i=1}^{n} N_i^2 / n\)
(Gallager 2008; Lorimer & Kramer 2012).

We obtained a set of folded data of PSR J0332+5434 as a
test example to identify and mitigate the transient RFI. We
compare the results before and after the data processing, which
are shown in Figure 3. The plots in the upper row are obtained
from the original data, whereas the plots in the lower row are
results after mitigating the transient RFI. Several obvious
transient RFIs in the original data, including their shapes and
intensities, are indicated by the red rectangles in the upper row.
The pulsar profile is shown in the green rectangle in each plot,
and the real pulsar profile after de-dispersion is displayed in the
two plots in the right column.
The original data, without mitigation, is shown in the middle
plot in the upper row, which contains many strong RFIs,
including transient RFI, resulting in a low S/N. However, the
de-dispersion smooths the transient RFI to a certain extent as
shown in the right plot in the upper row. Nevertheless, the
average pulse profile has improved and clearly visible, and the
S/N has also increased. However, there are still obvious
transient RFIs remaining in the data. The PPSD uses TOA
model to calculate the pulse phase and the width of the pulse
window in every frequency subchannels, which are used to
mitigate RFI by threshold (see Section 2 for the specific details
on the process). From the lower row, the transient RFIs are
obviously mitigated, when compare with that in the upper row,
and without changing the pulsar profile. As shown in the two
plots in the middle column, the S/N has increased from
2.7–26.1, which is more than eight times. From the two plots in
the right column, the S/N has increased by nearly nine times
from 6.9–67.2.

In the above example, the specific coefficient for the
threshold is 2. The performance of RFI mitigation is related
to the threshold coefficient. To further mitigate the spikes in the
off-pulse baseline in the lower row in Figure 3, a smaller
threshold coefficient is needed. When RFI coincides with the
pulse window in the sub-band, PPSD has certain limitations.
The probability of RFI being present in the pulse window for a
sub-band may increase under the following conditions: (i) long
transient RFI duration, (ii) large pulse window, and (iii) the
source is a millisecond pulsar, etc. However, these conditions
are not common.
In general, the off-pulse transient RFI is effectively mitigated after using PPSD. The threshold coefficient can change dynamically according to the number, distribution, and strength of the transient RFI. We find that the best range for the threshold coefficient is between $[1.8, 3]$ (Lyon et al. 2016).

3.3. Comparison with PAZI

PSRCHIVE provides the PAZ and PAZI sub-routines to mitigate RFI in observation data. PAZ automatically zaps the RFI in frequency based on median smoothed and profile lawn mowing methods. However, it is limited to transient RFI. On the other hand, PAZI presents non-de-dispersion plot, which is helpful for manuallymitigating the RFI. The user manually zaps the RFI data, and PAZI will re-assign the data. For comparison, we use PAZI to mitigate transient RFI in the same data set of PSR J0332+5434 as described in Section 3.2, and the results are shown in Figure 4. As indicated by the red dashed rectangles in Figure 4, the values used to replace the mitigated transient RFI are less than the background noise, and
the transient RFI appears over zapped. Although this may not always cause obvious impact on the final result, the efficiency of manual zapping is limited. More importantly, PAZI relies heavily on user’s experience, meaning that different people performing the RFI zapping may lead to different results. In addition, PAZI mitigates all data indiscriminately, even if the data is inside the pulse window.

By comparison, PPSD algorithm remedies the shortcomings of PSRCHIVE for mitigation of transient RFI, and provides the ability to automatic identification and mitigation of the transient RFI in pulsar observation.

3.4. Result Analysis

We applied this method to 10 other pulsars and achieved an S/N improvement ranging from 40%–800%. The stronger the RFI is in an observation, the more significant the S/N improvement is after PPSD is used. To compare the PPSD results, the average profile from three typical observations are shown in Figure 5. In the upper panel, the data is not de-dispersed, which shows the real shapes of the transient RFI. On the contrary, the shapes of the transient RFI are affected by the de-dispersion as shown in the middle panel. In the lower panel, the plots show the pulse profile after transient RFI mitigation. The overall results show that PPSD can effectively mitigate transient RFI and enhance the S/N of pulsar observations.

In order to test the feature of PPSD algorithm on recovering the original pulsar signal, we also verified the validity of the data before and after the transient RFI mitigation on each frequency channel. We found that the original data of the pulse profile was recovered effectively.

4. Conclusions

We have developed an algorithm for mitigation of off-pulse transient RFI in pulsar observations based on pulsar phase and standard deviation. The algorithm, referred to as the PPSD, combines TOA model, de-dispersion, and other factors to accurately locate the pulsar phase and protect the original pulsar signals. During data processing, PPSD calculates the mean and standard deviation for each frequency and sets the threshold accordingly to achieve effective mitigation for various transient RFIs. In general, the algorithm significantly mitigates the transient RFI and improves the S/N of the contaminated pulsar observations, thus providing guarantee for subsequent scientific analysis. We apply our PPSD algorithm to the data obtained...
from NSRT to show the effectiveness of the algorithm. As for the next step, we will create a pipeline, which incorporates PPSD, to achieve real-time transient RFI mitigation.

This work was supported by the National Key Research and Development Program of China under grant No. 2018YFA0404603. The research work is also partly supported by the Operation, Maintenance and Upgrading Fund for Astronomical Telescopes and Facility Instruments, budgeted from the Ministry of Finance of China (MOF) and administrated by the Chinese Academy of Sciences (CAS).

**ORCID iDs**

Yining Song @ https://orcid.org/0000-0002-1274-4766  
Na Wang @ https://orcid.org/0000-0002-9786-8548

**References**

Akeret, J., Chang, C., Lucchi, A., & Refregier, A. 2017, A&C, 18, 35  
An, T., Chen, X., Mohan, P., & Lao, B. Q. 2017, AcASn, 58, 43  
Baan, W. 2010, in PoS (RFI2010), 107, 1  
Baan, W. A. 2019, JAL, 8, 1940010  
Bethapudi, S., & Desai, S. 2018, A&C, 23, 15  
Czech, D., Mishra, A., & Inggs, M. 2018, RaSc, 53, 656  
Edwards, R. T., Hobbs, G. B., & Manchester, R. N. 2006, MNRAS, 372, 1549  
Fridman, P. A., & Baan, W. A. 2001, A&A, 378, 327  
Galagher, R. G. 2008, Principles of Digital Communication (Cambridge: Cambridge University Press)  
Hobbs, G. B., Edwards, R. T., & Manchester, R. N. 2006, MNRAS, 369, 655  
Kesteven, M., Hobbs, G., Clement, R., et al. 2005, RaSc, 40, RS5806  
Kocz, J., Bailes, M., Barnes, D., Burke-Spolaor, S., & Levin, L. 2012, MNRAS, 420, 271  
Li, J., Wang, N., Liu, Z.-Y., et al. 2020, AR&T, 17, 2  
Liu, Q., Liu, Y., Chen, M. Z., et al. 2019, AcASn, 60, 54  
Liu, Y., & Liu, Q. 2019, Sci. China: Phys., Mech., Astron., 49, 099514  
Lorimer, D., & Kramer, M. 2012, Handbook of Pulsar Astronomy (Cambridge: Cambridge University Press)  
Lyne, A., & Graham-Smith, F. 2012, Pulsar Astronomy (Cambridge: Cambridge University Press)  
Lyon, R. J., Stappers, B. W., Cooper, S., Brooke, J. M., & Knowles, J. D. 2016, MNRAS, 459, 1104  
Maan, Y., van Leeuwen, J., & Vohl, D. 2020, arXiv:2012.11630  
Rodriguez Arzaga, E., & Lynch, R. 2021, AAS Meeting 53  
Spitler, L. G., Cordes, J. M., Chatterjee, S., & Stone, J. 2012, ApJ, 748, 73  
Spoelstra, T. 2002, CRAFT Handbook for Frequency Management (European Science Foundation)  
Sureshkumar, S., Raybole, P. A., & Rai, S. 2016, in 2016 Radio Frequency Interference (RFI), 123-5  
van Straten, W., Demorest, P., & Oslowski, S. 2012, AR&T, 9, 237  
Wang, H.-F., Yuan, M., Yin, Q., et al. 2020, RAA, 20, 114  
Will, C. M. 2014, LRR, 17, 4  
Wu, X. J., Qiao, G. J., & Xu, R. X. 2018, MAICHONGXING WULI (in Chinese) (Beijing: Peking University Press)  
Zeng, Q., Chen, X., Li, X., et al. 2021, MNRAS, 500, 2969