URSI GASS 2023, Sapporo, Japan, 19 – 26 August 2023

Spectral Kurtosis applied to tied-array beamformer for pulsar observations

V. van Tonder*(1)(2), L. Schwardt(1), A. Faustmann(2), J. Gilmore(2) and S. Büchner(1)
(1) South African Radio Astronomy Observatory, Cape Town, South Africa, www.sarao.ac.za
(2) Department of Electrical and Electronic Engineering, Stellenbosch University, Stellenbosch, South Africa, www.sun.ac.za

Abstract

Astronomical sources are characterised using second-order power statistics because they have Gaussian distributions. Spectral Kurtosis is a tool that can suppress Gaussian noise and can therefore be used as an interference mitigation technique. However, care should be taken as transient sources such as pulsars may contain non-Gaussian components. This paper provides an overview of the Spectral Kurtosis estimator and how one obtains the thresholds in order to apply the estimator as an interference mitigation technique. The technique is applied to Vela pulsar data observed in tied-array beamformer mode by the MeerKAT radio telescope. Flagged data portions are replaced with uncontaminated portions of the observation. The results are compared to the corresponding unmitigated data set.

1 Introduction

Radio frequency interference (RFI) is that which corrupts radio astronomical data and decreases the sensitivity of the instrument. It can be man-made communication, air traffic control, or any other electronic signal emitting in the observed band. It may also be caused by natural sources. The appropriate RFI mitigation technique depends on the location and type of telescope – whether it is a single dish or an interferometer. It also depends on the RFI and observation type [1].

Advances in radio telescope technology have increased the sensitivity of some telescopes to such an extent that terrestrial interference has become indistinguishable from the astronomical data. Therefore, one cannot rely solely on power statistics for RFI mitigation. Researchers have started looking into how statistics of RFI and astronomical sources differ in order to discriminate between them in one data set. One such technique is to make use of the cyclostationary criterion of RFI as demonstrated by [2]. It assumes that RFI statistics are periodic to identify interfering sources. The bispectrum, a third-order statistic, has also been used to potentially discriminate RFI from Gaussian noise as simulations have shown in [3].

An astronomical signal is a sum of many stochastic signals, which has a Gaussian distribution by virtue of the central limit theorem. On the contrary, RFI has non-Gaussian components within its distribution. Gaussian signals are completely characterised by their mean and co-variances. Higher-order statistics can be used to suppress the Gaussian components of a signal and allows one to identify the non-Gaussian components, and based on this one can mitigate the RFI. This is the basis of the Spectral Kurtosis (SK) estimator, which relies on the fact that the excess kurtosis of a Gaussian signal is zero. SK estimators have become increasingly popular in the last decade and have demonstrated its effectiveness as an online technique on systems as documented in [4–6]. Its response to various types of RFI has been characterised in [7]. The SK estimator has also been applied to pulsar data obtained from the Large European Array for Pulsars (LEAP) experiment [8]. However, transient astronomical sources, such as pulsars, can contain non-Gaussian components [9, 10]. The SK estimator has also been used to detect transient astronomical sources, whilst using only a 2-bit system [11].

MeerKAT is a 64-antenna interferometric array situated in the Karoo region of South Africa. The telescope is situated within a protected radio quiet zone, mitigating RFI at the source [12]. A multidimensional RFI characterisation in terms of time, frequency, baseline length, and telescope direction in both azimuth and elevation is applied to the processing pipeline of the telescope [13]. This paper looks into applying SK to data obtained from the tied-array beamformer pipeline. It acts as a pilot study, investigating how to apply this technique to pulsar observations, given that MeerKAT is an excellent pulsar instrument [14].

An overview of the beamformer data pipeline of MeerKAT, the instrument used to capture the data, is provided in Section 2. The SK estimator is also explained in this section. In Section 3 a discussion of the results are provided. Section 4 concludes this work and also briefly describes future research.

2 Methodology

2.1 Spectral Kurtosis

The following summary of SK is adapted from [15, 16]. SK is calculated as the spectral variability V_k^2, equal to

$$V_k^2 = \frac{\sigma_k^2}{\mu_k^2} \quad , \tag{1}$$

This paper's copyright is held by the author(s). It is published in these proceedings and included in any archive such as IEEE Xplore under the license granted by the "Agreement Granting URSI and IEICE Rights Related to Publication of Scholarly Work."

where k indicates the dependence on frequency and σ_k^2 is the variance and μ_k is the mean of the power spectral density (PSD). The SK can be estimated from the power, S_1, and power squared , S_2, per frequency channel as

$$\widehat{SK} = \frac{MNd+1}{M-1}\left(\frac{MS_2}{S_1^2}-1\right) \quad ,$$
$$S_1 = \sum_{i=1}^{M} \langle P \rangle_N \quad , \tag{2}$$
$$S_2 = \sum_{i=1}^{M} \langle P \rangle_N^2 \quad ,$$

where $\langle P \rangle_N$ is the averaged power spectrum. The factor M is the number of PSD samples in the SK window, and d is the scaling factor depending on the power signal distribution. The factor N allows the SK estimator to work with S_1 and S_2 quantities that have already been accumulated N times. Ideally this should be one, but it extends the SK estimator definition, to apply to spectrometers that were not specifically designed with SK in mind. Equation 2 describes the unbiased generalised SK, which is equal to one for a Gaussian signal.

The channelisation process is frequently computed using a polyphase filterbank (PFB). For a Gaussian signal, a PFB modifies the spectral variability to become

$$V_k^2 = 1 + |W_{2k}|^2 \quad , \tag{3}$$

where W_{2k} only has to be evaluated at even frequencies and is defined by:

$$W_{2k} = \frac{1}{\sum w_n^2} \sum_{n=0}^{N-1} w_n^2 e^{-4\pi i k n/N}. \tag{4}$$

The w_n are the weights of the chosen time domain windowing function and N the number of data points. The frequency and time domain indices are indicated by k and n respectively. The normalisation defined in Equation 3 is applied after the SK estimator and usually becomes negligible for frequencies other than $k = 0$ and $k = \frac{N}{2}$.

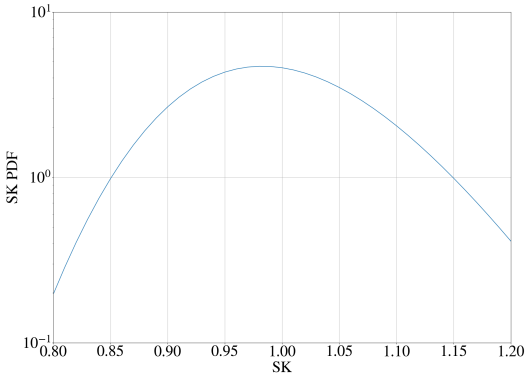

Figure 1. PDF of $\widehat{SK}$ for M=512.

We need to calculate the cumulative distribution function (CF) and complementary cumulative distribution function (CCF) in order to obtain the thresholds for RFI discrimination. For this we first need to determine the probability density function (PDF) and decide upon an acceptable probability of a false alarm (PFA). The PFA is usually based on the 3σ range of a normal distribution, which gives a PFA of 0.13499%. Depending on the choice of M, the PDF can be approximated by a Pearson distribution of type I, IV, or VI. The PDF and CF equations are given by Equations 56, 57 and 61 in [15]. These have been implemented to obtain Figures 1 and 2, which depict the PDF and CF for $M = 512$ respectively. Figure 2 also demonstrates how one can obtain the thresholds for the mitigation implementation. Note, to obtain an estimate of the upper and lower thresholds, one can use the standard deviation derived in [17] to obtain the $\pm 3\sigma$ values. The derived formula is given by,

$$\sigma_{V_k^2}^2 = \frac{4}{M}. \tag{5}$$

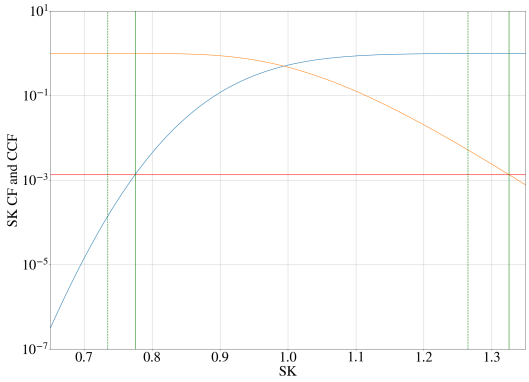

Figure 2. CF and CCF of $\widehat{SK}$ for M=512. The blue plot is the CF and the orange plot is the CCF. The horizontal red line is the PFA and is drawn at 0.13499%. The two dashed vertical green lines indicate the $\pm 3\sigma$ lines. The true lower and upper thresholds are at the intersection of the CF and CCF with the red PFA line. These are indicated in solid green vertical lines.

2.2 Data pipeline

The observations used 51 of the MeerKAT antennas configured in tied-array beamformer mode, operating over a frequency band of 856 to 1712 MHz. Incoming signals are downconverted to baseband by making use of anti-aliasing filters with a steep cutoff frequency and bandpass sampling. The signals are sampled at 1712 MHz in the second Nyquist zone. This causes spectral inversion which is corrected for on the field programmable gate array (FPGA) onboard the digitiser. The correction is obtained by multiplying the nth sample with $(-1)^n$ [18].

The baseband data is transported to the Karoo Array Processing Building (KAPB) where the next step is to channelise the data. Channelisation is achieved by using a PFB consisting of a 16-tap Hann window, followed by a 2048-point real FFT. In tied-array beamformer mode, the data from all the antennas are coherently summed together to

This paper's copyright is held by the author(s). It is published in these proceedings and included in any archive such as IEEE Xplore under the license granted by the "Agreement Granting URSI and IEICE Rights Related to Publication of Scholarly Work."

form beam(s). The phase correction is performed by a complex weight multiplication [18, 19]. For our observations only one beam was formed at boresight. Our data products are eight bits real and imaginary filterbank data for both horizontal and vertical polarisations.

3 Results and Discussion

The lower and upper thresholds for the various M values that were used are given in Table 1. The table also lists the time zapping resolutions for the different M values that were used during experimentation. A constant zapping frequency resolution of one channel is used. Zapping refers to flagged portions of the observation that are replaced by non-RFI contaminated data from the same observation. This portion was manually identified by inspection of the folded pulse profile in order to prevent copying portions of the pulsar into the mitigated area.

Table 1. M values, thresholds, and time zapping resolutions.

M	T_{low}	T_{upper}	τ ms
512	0.77	1.33	0.61
1024	0.83	1.22	1.22
2048	0.88	1.15	2.45
10240	0.94	1.06	12.25

Figure 3 depicts a comparison of different SK mitigation strategies on the Vela pulsar. In Figure 3a no RFI mitigation is applied and the folded pulse profile is shown. Figure 3b indicates portions of the profile that falls above the upper and below the lower thresholds, whereas 3c shows only the portions falling below the lower threshold. Both these experiments had $M = 512$. In Figure 3d, M is increased to 2048 and only the portions below the lower thresholds are indicated.

Recall that pulsars are transient sources that may contain non-Gaussian components. Transient signals with a low duty cycle usually exhibit excess kurtosis due to its bursty nature, resulting in large SK values. Comparing Figures 3b and 3c, one can see that the pulsar is flagged by the SK estimator when also flagging data above the upper threshold.

Furthermore, looking at the lower frequencies of 3c, one can see that the SK estimator is again triggered by the pulsar. If the number of samples, M, with which the SK is computed is a significant fraction of the pulsar pulse width, then the estimator will produce a negative SK, triggered by the lower threshold. Pulsar widths are wider at lower frequencies than at higher frequencies due to pulse broadening. Therefore, the SK estimator will cause false positives for smaller M values at lower frequencies. This is evident by comparing Figures 3c and 3d which use $M = 512$ and $M = 2048$ respectively.

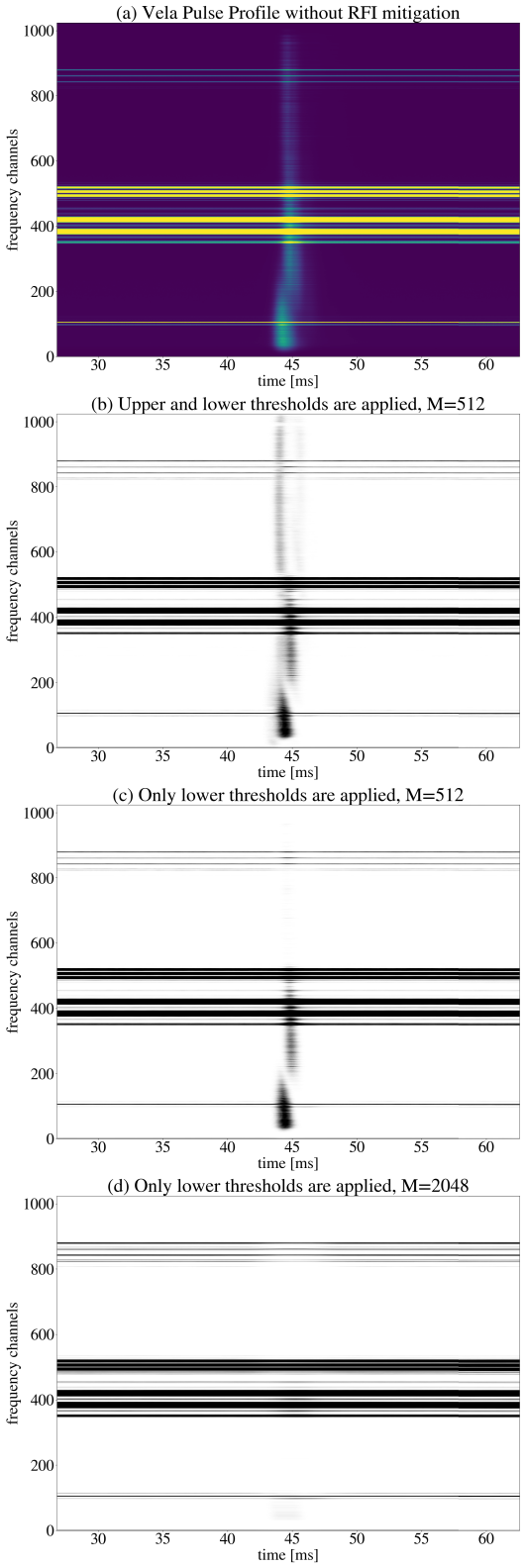

Figure 3. Comparison of different SK mitigation strategies. In (a) the folded Vela pulsar profile without RFI mitigation is shown, with power levels increasing from blue to yellow. The portions of the profile that are flagged using the upper and lower thresholds are shown in black in (b) with $M = 512$. In (c) only the portions flagged using the lower threshold are shown in black whereas in (d) M is increased to 2048.

This paper's copyright is held by the author(s). It is published in these proceedings and included in any archive such as IEEE Xplore under the license granted by the "Agreement Granting URSI and IEICE Rights Related to Publication of Scholarly Work."

4 Conclusion

In this paper, we have shown the feasibility of applying the SK RFI mitigation technique. The technique was applied to a Vela pulsar observation obtained via the tied-array beamformer mode of the MeerKAT radio telescope. It is noted that if a pulsar contains non-Gaussian components, using both the upper and lower thresholds would cause the pulsar itself to be flagged. Therefore, it is advised to only use the lower thresholds in such a case. Further, the SK window length, M, must be wider than the pulse width, otherwise the pulsar will also register as RFI due to an SK value falling below the lower threshold.

MeerKAT is an interferometric antenna array and forms visibilities via cross correlations of voltages from individual antennas. Future work includes obtaining raw voltages from a subset of the entire array and applying a cross correlator SK estimator to these baselines. The Generalised SK has a limitation in that it can not distinguish between RFI with a 50% duty cycle and Gaussian noise. A multi-scale SK has been formalised in order to address this limitation. This work can be extended in order to implement the multi-scale SK.

Acknowledgements

The MeerKAT telescope is operated by the South African Radio Astronomy Observatory, which is a facility of the National Research Foundation, an agency of the Department of Science and Innovation.

References

[1] P. A. Fridman and W. A. Baan, "RFI mitigation methods in radio astronomy," *Astronomy and Astrophysics*, vol. 378, pp. 327–344, 2001.

[2] R. Feliachi *et al.*, "RFI mitigation: cyclostationary criterion," in *Proceedings of Wide Field Astronomy & Technology for the Square Kilometre Array — PoS(SKADS 2009)*, vol. 132, p. 033, 2011.

[3] V. van Tonder, L. Schwardt, A. Faustmann, and J. Gilmore, "Bispectra of simulated GPS data for potential RFI mitigation," in *2022 3rd URSI Atlantic and Asia Pacific Radio Science Meeting (AT-AP-RASC)*, pp. 1–4, 2022.

[4] G. M. Nita *et al.*, "EOVSA Implementation of a Spectral Kurtosis Correlator for Transient Detection and Classification," *Journal of Astronomical Instrumentation*, vol. 5, no. 4, 2016.

[5] Y. Dou *et al.*, "The Korean Solar Radio Burst Locator (KSRBL)," *Publications of the Astronomical Society of the Pacific*, vol. 121, pp. 512–526, May 2009.

[6] J. Taylor *et al.*, "Spectral Kurtosis-Based RFI Mitigation for CHIME," *Journal of Astronomical Instrumentation*, vol. 08, no. 01, p. 1940004, 2019.

[7] E. Smith, R. S. Lynch, and D. J. Pisano, "Simulating Spectral Kurtosis Mitigation against Realistic RFI signals," *The Astronomical Journal*, vol. 164, Oct 2022.

[8] M. Purver *et al.*, "Removal and replacement of interference in tied-array radio pulsar observations using the spectral kurtosis estimator," *Monthly Notices of the Royal Astronomical Society*, vol. 510, pp. 1597–1611, Jan 2022.

[9] A. Faustmann *et al.*, "A Bispectral Analysis of the Radio Emissions of Pulsar J0437-4715," *The Journal of Astronomical Instrumentation*, vol. 09, no. 04, p. 2050018, 2020.

[10] A. Faustmann, L. Schwardt, V. van Tonder, J. Gilmore, and S. Buchner, "Investigating the high time-resolution statistics of pulsar radio signals using spectral self-noise," *Monthly Notices of the Royal Astronomical Society*. Forthcoming.

[11] G. M. Nita, A. Keimpema, and Z. Paragi, "Statistical discrimination of RFI and astronomical transients in 2-bit digitized time domain signals," *Journal of Astronomical Instrumentation*, vol. 8, no. 1, 2019.

[12] *The Astronomy Geographic Advantage Act 21 of 2007*. Republic of South Africa, April 2009.

[13] I. Sihlangu, N. Oozeer, and B. A. Bassett, "Multi-dimensional radio frequency interference framework for characterizing radio astronomy observatories," *Journal of Astronomical Telescopes, Instruments, and Systems*, vol. 8, pp. 1–13, Oct 2021.

[14] M. Bailes *et al.*, "The MeerKAT telescope as a pulsar facility: System verification and early science results from MeerTime," *Publications of the Astronomical Society of Australia*, vol. 37, 2020.

[15] G. M. Nita and D. E. Gary, "Statistics of the Spectral Kurtosis Estimator," *Publications of the Astronomical Society of the Pacific*, vol. 122, pp. 595–607, May 2010.

[16] G. M. Nita and D. E. Gary, "The generalized spectral kurtosis estimator," *Monthly Notices of the Royal Astronomical Society*, vol. 406, no. 1, pp. 60–64, 2010.

[17] G. Nita *et al.*, "Radio Frequency Interference Excision Using Spectral-Domain Statistics," *Astronomical Society of the Pacific*, vol. 119, pp. 805–827, July 2007.

[18] A. van der Byl *et al.*, "MeerKAT correlator-beamformer: a real-time processing back-end for astronomical observations," *Journal of Astronomical Telescopes, Instruments, and Systems*, vol. 8, July 2021.

[19] A. Faustmann, J. Gilmore, V. van Tonder, and M. Serylak, "Higher-Order Spectral Analysis of Radio Pulsar Bursts using MeerKAT," in *2021 15th European Conference on Antennas and Propagation (EuCAP)*, pp. 1–5, 2021.

This paper's copyright is held by the author(s). It is published in these proceedings and included in any archive such as IEEE Xplore under the license granted by the "Agreement Granting URSI and IEICE Rights Related to Publication of Scholarly Work."