URSI GASS 2020, Rome, Italy, 29 August - 5 September 2020

A Cross-Correlation based Spectral Kurtosis RFI Detector

Gelu M. Nita*(1) and Gregory Hellbourg(2)
(1) New Jersey Institute of Technology, Newark, NJ, 08536, USA
(2) California Institute of Technology, Pasadena, CA, USA

Abstract

Accurate flagging of Radio Frequency Interference (RFI) is necessary to recover instrumental efficiency and avoid false astronomical detections. Spectral Kurtosis ($\widehat{SK}$) is a popular operator in RFI flagging for radio astronomy due to its detection sensitivity to non-Gaussian emissions and its competitive computational cost. Most $\widehat{SK}$ detection pipelines are applied to single antennas or auto-correlations products. This paper investigates the application of the $\widehat{SK}$ to antennas cross-correlations, and demonstrates an improved detection performance compared to the auto-correlation-based approaches.

1 Introduction

Due to its conceptual simplicity, the feasibility of its real-time implementation, and its proven effectiveness in discriminating radio frequency interference (RFI) from astronomical data, the Generalized Spectral Kurtosis Estimator ($\widehat{SK}$, [1]) has become a component of the data processing pipelines of an increasing number of newly designed single-dish radio telescopes [2, 3] and interferometric arrays [4, 5].

Under the null-hypothesis (H_0), i.e. a RFI- free independent and identically distributed (iid) white Gaussian noise, the expectation of the $\widehat{SK}$ estimator is unity and the probabilities of false alarm (PFA) corresponding to a pair of user-defined RFI detection thresholds located below and above the unity expectation can be analytically computed [1] and straightforwardly employed to flag RFI contamination of the data streams produced by single antennas. However, in the case of an interferometric array, one may choose between several already proposed methods that are based on combining the antenna-based $\widehat{SK}$ estimators, or its proposed extension to cross-correlation between antennas.

To do so, we state the problem in §2, briefly discuss in §3 the mathematical formalism of an antenna-based $\widehat{SK}$ estimator and the way a pair of such $\widehat{SK}$ estimators may be combined to flag RFI-contaminated baseline visibilities, in §4 we introduce a cross-correlation-based $\widehat{SK}$ estimator, in §5 we present a performance analysis, and in §6 we state our conclusions.

2 The RFI detection problem

The rapid development of radio frequency commercial applications threatens radio astronomy as next generation radio telescopes featuring higher sensitivity and bandwidths come online. The accurate detection of corrupted time-frequency data is essential to recover telescope efficiency and avoid potential false astronomical detection due to RFI artifacts.

We consider here an array of N_a antennas. After amplification and digitization, the signal produced by each antenna is channelized and correlated with all other antennas. RFI-decorrelation occurs over large antenna-pair distances (i.e. baselines) and large signal bandwidth. Short baselines, sampling large scale structures in interferometric data [6], are therefore more vulnerable to RFI corruption.

The expected correlation between two antennas i and j, assuming one source of RFI affecting the array, is expressed as:

$$\sigma^2_{i,j} = \underbrace{\sum_{s}^{N_s} a_{s_{i,j}} \sigma^2_s + \sigma^2_{n_{i,j}}}_{\sigma^2_{H0}} + a_{r_{i,j}} \sigma^2_r(\Theta), \qquad (1)$$

where N_s astronomical sources with power σ^2_s are located in the field of view sampled by the interferometer, $a_{s_{i,j}} \in \mathbb{C}$ translates the spatial information (location and structure of the source) and potential gains variation between the antennas, $a_{r_{i,j}}$ is the same quantity for the RFI, $\sigma^2_r(\Theta)$ is the autocorrelation function of the RFI evaluated at time-lag Θ, and $\sigma^2_{n_{i,j}}$ is the system noise. The latter quantity includes the sample estimation noise due to the finite sample size over which $\sigma^2_{i,j}$ is estimated. The parameter Θ is an antenna-pair delay correction maximizing the instrument sensitivity towards the astronomical sources of interest.

RFI-decorrelation occurs over long baselines (or alternatively large signal coherence bandwidth). The RFI detection performance is therefore improved on pairs of individual antennas separated by short distances, where the cross-correlation noise is dominated by the sky and sample estimation noise, and the individual antenna noise is decorrelated.

The RFI detection problem is formulated as a binary hypothesis problem [7]:

$$\sigma^2_{i,j} = \begin{cases} \sigma^2_{H0} & (H0) \\ \sigma^2_{H0} + a_{r_{i,j}}\sigma^2_r(\Theta) & (H1) \end{cases} \quad (2)$$

where $(H0)$ is the RFI-free hypothesis, and $(H1)$ corresponds to the RFI-corrupted hypothesis.

The $\widehat{SK}$ detector has previously been suggested in the case $i = j$. This paper extends the $\widehat{SK}$ detection in the case $i \neq j$.

3 Spectral Kurtosis based flagging strategies for array telescopes

3.1 The Generalized Spectral Kurtosis Estimator

In the case $i = j$, the generalized $\widehat{SK}$ estimator is given by the expression [1]

$$\widehat{SK} = \frac{Md+1}{M-1}\left(\frac{MS_2}{S_1^2} - 1\right) \quad (3)$$

where $S_1 = \sum^M \left(\sum_{n=1}^N |x_i[n]|^2\right)$, $S_2 = \sum^M \left(\sum_{n=1}^N |x_i[n]|^2\right)^2$, with $x_i[n]$ being the complex voltage signal produced by antenna i at time sample n, and $d = N\delta$, where δ represents the shape factor of the Gamma distribution expected to be obeyed by the RFI-free auto-correlation samples. In the case of complex-valued data samples, $\delta = 1$.

3.2 Multi auto-correlation flagging strategies

Despite the reduced RFI detection sensitivity, flagging individual auto-correlations of an array radio telescope reduces the computational cost. Two main flagging strategies have been suggested for auto-correlation based antenna array flagging, which are described in this section.

3.2.1 OR operator flagging (OOF)

This strategy, proposed in [4], flags antenna cross-correlations depending on the **OR** operation between the flags evaluated on the individual antennas.

The expected PFA for this strategy follows $\text{PFA}_{\textbf{OR}} = (2 - \text{PFA}_{\text{auto}})\text{PFA}_{\text{auto}}$, where PFA_{auto} is the PFA of the $\widehat{SK}$ for the individual antennas auto-correlations. The single antenna PFA may be computed using Equations 61 or 62 of [8], which provide two alternative 4^{th}-order approximations of the true cumulative probability function (CDF) of the generalized $\widehat{SK}$ estimator, or Equation 20 of [1], which provides a 3^{rd}-order approximation that has been proven to be accurate enough for practical applications. The pair of non-symmetrical thresholds needed to ensure a predefined PFA below or above the unity expectation of the $\widehat{SK}$ estimator may be computed by using a numerical algorithm such as the one provided by [4].

3.2.2 Mean auto-correlation flagging (MACF)

The MACF strategy [5] consists in averaging the N_a individual, antenna-based $\widehat{SK}$ estimators, $\widehat{SK}_{N_a} = \frac{1}{N_a}\sum_i^{N_a}\widehat{SK}_i$ which has unity expectation and a variance N_a times smaller than that of the individual antenna-based estimators,

$$\sigma_{\langle\widehat{SK}\rangle} = \frac{1}{n}\sigma_{\widehat{SK}_i} \simeq 2\left(1+\frac{1}{d}\right)\frac{1}{n\times M} + n\times O\left[\frac{1}{(n\times M)^2}\right] \quad (4)$$

When compared with an antenna-based $\widehat{SK}$ detector tuned to have the same PFA, the multi-receiver $\widehat{SK}$ detector is expected to have a higher PD, due to its smaller variance for the same accumulation length, M.

The first order approximation of the $\sigma_{\langle\widehat{SK}\rangle}$ variance given by equation 4 also suggests a method to compute first order approximations of the expected PFA associated with a pair of arbitrarily defined RFI detection thresholds by making the substitution $M \to n \times M$ in the equations or numerical algorithms suggested in section 3.2.1.

4 Cross-Correlation Flagging (CCF)

As described earlier, cross-correlation products are more sensitive to RFI over short baselines. Flagging all individual array baselines is therefore not necessary, and the amount of processed baselines can depend on the computational capacity of the telescope back-end. Due to the signal coherence bandwidth, assuming a frequency channelization in δf bandwidths, baselines exceeding $b_{\max} = \frac{c}{\delta f}$ will have a reduced sensitivity to RFI due to its de-correlation.

The extension of the $\widehat{SK}$ detector to antenna cross-correlation follows the same expression as Equation 3, with

$$S_1 = \sum^M \left|\sum_{n=1}^N x_i[n]x_j^*[n]\right|_m$$
$$S_2 = \sum^M \left|\sum_{n=1}^N x_i[n]x_j^*[n]\right|_m^2$$

Under the arguable assumption that the inner sums entering S_1 and S_2 obey a PDF that can be satisfactorily approximated by a Gamma distribution having a shape factor d ultimately determined by possible dependence between the noise terms of each antenna, as well as by further quantization effects of the correlator [9, 10], the expectation of $\widehat{SK}$ estimator may be normalized to unity by estimating such instrument-specific shape factor through an empirical calibration process.

To do so, one may begin with evaluating the $\widehat{SK}$ estimator under the assumption $d = 1$, and then use the mean of a

statistical significant series of such $\widehat{SK}$ estimators involving RFI-free visibility squared amplitudes, $\mu = \left\langle \widehat{SK}(d \equiv 1) \right\rangle$, to estimate the unknown shape factor required to normalize to unity the expression given by Equation 3, which is

$$d = \frac{M - \mu + 1}{\mu M}. \tag{5}$$

Having the instrumental shape factor estimated by these empirical means, the PFA for any pair of user-defined RFI detection thresholds may be computed using the CDF of the generalized $\widehat{SK}$ estimator provided by [1].

5 Comparison and performance tests

5.1 Numerical Simulations

To evaluate and compare the performance of the various detection strategies presented in section 3.2, we conducted a Monte-Carlo simulation to empirically evaluate both PD and PFA. For each independent trial, a 2-antennas data set of voltage streams is distributed as:

$$x_i[n] = \begin{cases} x_{\text{noise}}[n] & (H0) \\ x_{\text{RFI}}[n] + x_{\text{noise}}[n] & (H1) \end{cases} \tag{6}$$

where $x_{\text{noise}}[n]$ is a zero-mean complex white iid Gaussian noise with unit variance, and $x_{\text{RFI}}[n] = a_i \sqrt{2 \times 10^{\xi/10}} \cdot e^{i2\pi(fn+\phi)}$ simulates a complex continuous wave at frequency $f \in [0,1]$ and phase $\phi \in [0, 2\pi]$. The factor a_i is a random complex phase term associated with the i^{th} antenna, and ξ is the Signal-to-Noise Ratio (SNR) in dB. Astronomical sources are here neglected.

For each independent trial, $N_{\text{sam}} = 10,000$ data samples are randomly generated under both $(H0)$ and $(H1)$. The correlator and $\widehat{SK}$ engine parameters chosen for this simulation are $M = 1000$ and $N = 10$. Sets of $4,096$ independent trials have been generated for each data point.

Both PD and PFA, evaluated under $(H1)$ and $(H0)$ respectively, are computed empirically over the Monte-Carlo simulation, and displayed on Figure 1 as Receiver Operational Characteristics (ROC) curves. Perfect separation between both $(H1)$ and $(H0)$ hypotheses is achieved at (PFA = 0, PD = 1) on these ROC curves. The closer to this point, the better the detector performs.

The ROC curves have been evaluated for the four approaches presented in this paper (including the $\widehat{SK}$ detector for single antenna data) for three different SNRs: $\xi = -10$ dB, $\xi = -8$ dB, $\xi = -5$ dB. The range $\xi \in [-10, -5]$ seems to be a performance inversion range for all detectors.

The CCF detection strategy outperforms all other detectors at $\xi > -8$ dB. The performance comparison between the

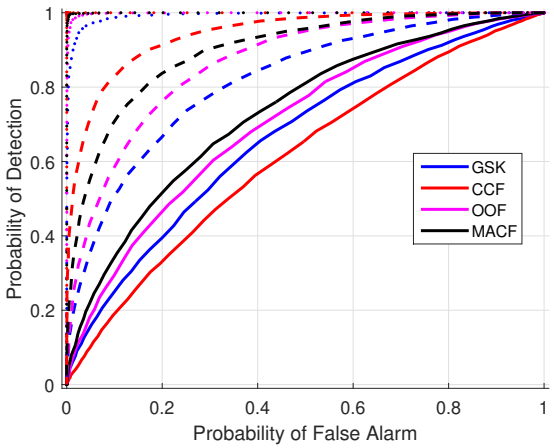

Figure 1. ROC curves for the single antenna $\widehat{SK}$ operator, and 2-antenna CCF, OOF, and MACF detection strategies, for SNR = -10 dB (plain curves), SNR = -8 dB (dashed curves), and SNR = -5 dB (dotted curves), evaluated empirically with Monte-Carlo simulation over 4096 trials, M = 100, N = 10.

three other detectors appear to evolve consistently with the SNR. At $\xi = -10$ dB, none of the detectors is able to perform accurate detection.

As a detection example, Figure 2 shows a comparison of the "above-theshold" events detected over 1,000 time samples with the OOF, MACF, and CCF methods on a channelized dataset made of 4,096 frequency channels, featuring a white Gaussian iid noise and a simulated Binary Phase-Shift Keying (BPSK) signal, centered at reduced frequency 0.

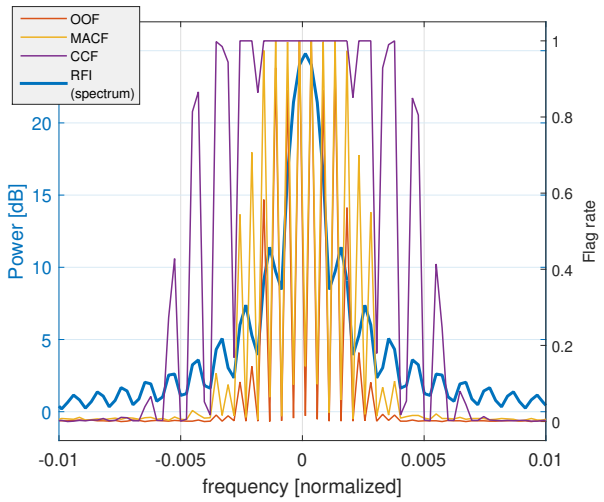

Figure 2. Comparison of detection sensitivities of the OOF, MACF, and CCF flagging strategies at detecting a BPSK signal with 2 antennas, N=10, M=100, SNR=-8dB PFA=0.013. Statistics evaluated over 1,000 independent realizations.

The figure highlights the detection sensitivities of the dif-

ferent methods, and shows the superiority of the CCF approach as it detects the BPSK signal at lower SNR channels than other methods. All methods plateau at PFA = 0.013 (arbitrarily set) at low SNR channels.

6 Conclusion

Interferometers have good RFI rejection properties due to RFI de-correlation over large baselines, but are also more sensitive to weak RFI in cross-correlation products over short baselines (shorter than the coherence wavelength of the channelization bandwidth). We propose therefore the use of the $\widehat{SK}$ detector over such data product to improve the detection performance of other $\widehat{SK}$ flagging strategies for array radio telescopes.

A Monte-Carlo simulation shows that the cross-correlation $\widehat{SK}$ detector outperforms all other detection strategies in a simple data model involving a white Gaussian system noise corrupted by a continuous wave RFI. Moreover, the PFA for the $\widehat{SK}$ detector can be analytically formulated, and accurate detection threshold can be derived under realistic RFI-free data modelling. Further improvement in detection performance can be expected by combining the flagging information of multiple array baselines, similarly to the auto-correlation approaches.

Further detection performance assessments remain to be conducted on more realistic scenarii, e.g. involving system uncalibration, complex information-bearing sources of RFI, non-negligible astronomical sources, non-iid or non-white system noise.

The low computational complexity of the $\widehat{SK}$ detector and its detection performance make it an interesting online RFI flagging solution for the next generation radio interferometers with large number of antennas.

References

[1] G. M. Nita, D. E. Gary, "The Generalized Spectral Kurtosis Estimator," Monthly Notices of the Royal Astronomical Society: Letters, 406, 1, July 2010, Pages L60–L64, https://doi.org/10.1111/j.1745-3933.2010.00882.x

[2] A. Liu, D. E. Gary, G. M. Nita, S. White, and G. Hurford, "A Subsystem Test Bed for the Frequency-Agile Solar Radiotelescope," Publications of The Astronomical Society of The Pacific, 119, 303-317, 2007. https://doi.org/10.1086/512825.

[3] D. E. Gary, Z. Liu, G. M. Nita, "A wideband spectrometer with RFI detection," Publications of the Astronomical Society of the Pacific 122 (891), May 2010, pp. 560

[4] G. M. Nita, J. Hickish, D. MacMahon, D. E. Gary, "EOVSA Implementation of a Spectral Kurtosis Correlator for Transient Detection and Classification," Journal of Astronomical Instrumentation, 5 (04), 1641009 (2016)

[5] J. Taylor, N. Denman, K. Bandura, P. Berger, K. Masui, A. Renard, I. Tretyakov and K. Vanderlinde,"Spectral Kurtosis-Based RFI Mitigation for CHIME," Journal of Astronomical Instrumentation-Vol. 08 (01), 1940004 (2019)

[6] Thompson, Anthony Richard, James M. Moran, and George Warner Swenson. Interferometry and synthesis in radio astronomy. New York et al.: Wiley, 1986.

[7] Van Trees, Harry L. Detection, estimation, and modulation theory, part I: detection, estimation, and linear modulation theory. John Wiley & Sons, 2004.

[8] G. M. Nita and D. E. Gary, "Statistics of the Spectral Kurtosis Estimator," Publications of the Astronomical Society of the Pacific 122 (891), May 2010, pp. 595–607

[9] G. M. Nita, D. E. Gary & G. Hellbourg, "Spectral Kurtosis statistics of quantized signals," 2016 Radio Frequency Interference (RFI), Socorro, NM, October 2016, pp. 75-80. doi: 10.1109/RFINT.2016.7833535

[10] G. M. Nita, A. Keimpema, & Z. Paragi, "Statistical Discrimination of RFI and Astronomical Transients in 2-bit Digitized Time Domain Signals," Journal of Astronomical Instrumentation, 8, 1940008 (2019)