

A NOVEL RFI DETECTION METHOD FOR MICROWAVE RADIOMETERS USING MULTI-LAG CORRELATORS

R. Díez-García, A. Camps

Universitat Politècnica de Catalunya - BarcelonaTech and IEEC/UPC,
CommSensLab - ‘María de Maeztu’ unit,
Campus Nord, Edifici D4, 08034 Barcelona, Spain.
E-mail: raul.diez@upc.edu, camps@tsc.upc.edu

ABSTRACT

Radio Frequency Interference (RFI) is an increasing problem in Microwave Radiometry, particularly for Earth Observation, because the antennas are pointing towards the Earth. RFI has been observed at L-band in ESA’s SMOS (Soil Moisture and Ocean Salinity) Earth Explorer mission, as well as in NASA’s SMAP (Soil Moisture Active and Passive) and Aquarius missions, as well as in AMSR-E and WindSat missions at 10.7 GHz and 18.7 GHz [1]. Therefore, dedicated on-board systems are nowadays a must to detect and remove contaminated measurements, improving radiometric accuracy, and increasing the spatial coverage. In this work, a novel detection technique, specially tailored for Synthetic Aperture Interferometric Radiometers (SAIR), is proposed and its performance analysed. It is based on the change of the shape of the cross-correlation function at lags different from zero under the presence of RFI. The performance of the proposed technique is compared to other common RFI detection algorithms, and its limitations and advantages are discussed. Post-correlation detection performance is found to out-perform other commonly used algorithms such as Kurtosis. In addition, it presents some convenient properties for its practical application in correlation and real aperture radiometers.

Index Terms— Microwave, Radiometer, Interferometer, RFI, Detection, Mitigation, Correlation, Quantization

1. INTRODUCTION

Microwave radiometers are seriously affected by Radio Frequency Interference (RFI). The high radiometric resolution required is prone to be interfered by non-natural signals, usually of much larger power than the radiometric emissions. These interferences conceal the underlying natural signal, corrupting, or even preventing, the retrieval of geophysical variables. Passive remote sensing bands are protected by ITU-R recommendations, and considerable effort is devoted to enforce proper spectrum usage [2]. Nevertheless, RFI remains a considerable threat to the accuracy and spatial coverage of passive instruments. RFI are prevalent in large areas

of the Earth, making the retrieval of geophysical parameters difficult or impossible over wide regions. Its impact has been noticed in several Earth Observation applications [3]. Some affected missions are: at L-band (1.4 GHz) ESA’s SMOS (Soil Moisture and Ocean Salinity) Earth Explorer mission [4, 5, 6], as well as NASA’s SMAP (Soil Moisture Active and Passive) and Aquarius missions, and at Ku-band (10.7 GHz and 18.7 GHz) in AMSR-E and WindSat missions [1]. While there is an ongoing effort to track and identify RFI sources in order to switch them off [6], it is unlikely that the problem is going to be solved in a near future.

The adoption of on-board detection and mitigation techniques is then required to reduce the impact of RFI, increasing the radiometric accuracy and data reliability, and improving the spatial coverage over areas prevalently obscured by interference. A wide range of RFI detection techniques has been developed over the last years. Some examples include parametric techniques (where the RFI type is known a priori), statistical detection methods (where the statistics of the received signal are estimated and compared to Gaussian ones, e.g. [7, 8]), polarimetric methods (where the cross-polarization components may indicate the presence of RFI, [9]), and time and/or frequency analysis (where the Time/Frequency properties of the signal are studied to infer the presence of RFI, [10]). In the present work, a new family of techniques is proposed, based on the changes induced by the RFI to the autocorrelation shape of the signal.

In the next sections, the practicalities and limits of using the correlation shape as an RFI detection method are discussed in detail. Thanks to the widespread access to fast correlators, this novel RFI-detection technique is applicable to all kinds of radiometers, but it is specially useful in radiometer architectures where the computation of the correlations is central, such as in SAIRs [11] and some types of spectrometers or polarimetric radiometers [12]. It should be noted that, for the sake of simplicity, the analysis presented here is for the autocorrelation, but the same concepts can be applied to the cross-correlation between pairs of antennas looking at the same radiometric source as well, or even to the cross-correlation be-

tween the outputs of the ports of a single antenna (e.g. in polarimetric radiometers).

2. EFFECT OF RFI OVER THE CORRELATION SHAPE OF A SIGNAL

Without loss of generality, assuming a sharp band-pass filters whose frequency response can be approximated by a rectangular function of bandwidth B_w , the auto-correlation of the Low-Pass Equivalent (LPE) signal of an uncorrelated white noise source is given by:

$$R_{n_i}(t) = T_A \cdot \text{sinc}(B_w \cdot t), \quad (1)$$

where T_A is the so-called noise temperature, with $P_n = T_A \cdot B_w \cdot k_B$, k_B being the Boltzmann constant, B_w the system bandwidth, and P_n the noise power. If the radiometric signal is contaminated by RFI, however, this expression is no longer valid. In this latter case, $s_i = n_i + r_i$, where n_i is the radiometric signal and r_i the RFI. The auto-correlation can then be expressed as a function of the thermal noise and RFI components:

$$R_{s_i}(t) = R_{n_i}(t) + R_{r_i}(t) + R_{n_i, r_i}(t) + R_{r_i, n_i}(t), \quad (2)$$

where R_{r_i} denotes the autocorrelation of the RFI, and R_{n_i, r_i} and R_{r_i, n_i} the cross-correlation between n_i and r_i . Since the RFI and radiometric noise are uncorrelated, $R_{r_i, n_i} = R_{n_i, r_i} = 0$ and R_{s_i} is then just the sum of the auto-correlation of the separate components:

$$R_{s_i}(t) = R_{n_i}(t) + R_{r_i}(t). \quad (3)$$

By substituting eq. 1 in eq. 3,

$$\hat{R}_s(t) = T_A \cdot \text{sinc}(B_w \cdot t) + R_r(t), \quad (4)$$

where R_r is the autocorrelation of the RFI signal. The noise autocorrelation shape will be therefore modified by the autocorrelation shape of the RFI. This fact can be used as an indicator of RFI presence.

The shape of the fringe-wash function is directly dependent on the frequency contents of the RFI. Spectrally narrow signals will present distinctive autocorrelation signatures, and are therefore more prone to detection. For example, CW signals present distinctive cosinus-like behaviour, easy to discriminate from the 'sinc-shaped' baseline at long correlation lags. On the contrary, spectrally wide signals, as per example some types of PRN sequences, have nearly sinc-shaped autocorrelations. In these cases, the resulting shape of the

combined signal will be nearly sinc-shaped as well, and the only factor to discriminate the RFI will be the signal power, assuming it is known beforehand. Moreover, it is well known that the autocorrelation of a generic sine-wave follows the following expression:

$$R_{r, r}(\tau) = P_r \cdot \cos(2\pi f \tau), \quad (5)$$

where P_r is the power of the sine-wave and f its frequency. The dependence of the autocorrelation shape on the central frequency of the RFI signal (and not only on its bandwidth) is readily observable. The dependence on the RFI type and Time-Frequency properties is not exclusive of this new family of detection methods, and indeed it is a common limitation in most techniques (see for example [10, 13, 14, 15]).

Note that, in a real system, filters cannot be considered ideally sharp. In this case, and assuming that $h(t)$ is the filter's impulse response, the autocorrelation of an RFI-free noise signal is:

$$R_{n_i}(t) = T_A \cdot (h * h)(t), \quad (6)$$

where $*$ is the convolution operator. In this work, all filters have been assumed ideal for simplicity, but the same principles apply if the filters' impulse response is known.

3. METHODOLOGY

The implementation of the RFI detection method is described hereafter. Indeed, multiple implementations may exist for the same basic principle, the fact that a non-sinc shaped correlation is indicative of RFI presence. Without trying to be exhaustive, some possibilities are presented here.

3.1. Zero-crossing Correlation Ratio (ZCR)

If sampling at the Nyquist frequency is considered, the autocorrelation function (eq. 1) is sampled exactly at the zero-crossings. Therefore, in RFI-free conditions, the autocorrelation of signal s is just:

$$R_{ss}(\tau_i) = \begin{cases} \sigma_s^2 & \tau_i = 0 \\ 0 & \text{otherwise,} \end{cases} \quad (7)$$

where σ_s^2 is the variance of the signal s . It should be noted, however, that in practical terms, the exact value of R_{ss} is unknown, so an estimator is used to approximate it. Being $\overline{R_{ss}}$ the unbiased estimator of R_{ss} :

$$\overline{R_{ss}}(m) = \begin{cases} \frac{1}{N-m} \sum_{n=0}^{N-m-1} s(n+m) \cdot s^*(m) & m \geq 0 \\ \overline{R_{ss}}(-m) & m < 0. \end{cases} \quad (8)$$

Its variance follows [16, eq. 14.1.27]:

$$\sigma_{\overline{R_{ss}}}^2(m) \approx \frac{N}{(N-|m|)^2} \sum R_{ss}^2(n) + R_{ss}(n+m)R_{ss}(n-m), \quad (9)$$

that, under the assumptions on sampling of eq. 7, simplifies to:

$$\sigma_{\overline{R_{ss}}}^2(m) \approx \frac{N}{(N-|m|)^2} \cdot R_{ss}^2(0). \quad (10)$$

As it has been seen, RFI presence may induce non-zero correlation values for $\tau_i \neq 0$. This may be used as a metric to infer RFI presence. Let's define the Zero-Crossing Correlation Ratio of R_{ss} as:

$$ZC = \frac{R_{ss}(\tau_1)}{R_{ss}(0)}. \quad (11)$$

That is, the ratio of $R_{ss}(\tau_1)$, the the correlation at the first zero crossing of the correlation (τ_1), with its maximum at 0 (i.e. the variance σ_s^2). By design, $ZC = 0$ in RFI-free conditions, and hence, if $ZC \neq 0$ indicates RFI presence. However, both the numerator and denominator of equation 11 are only known through estimation. Therefore, it is necessary to define \overline{ZC} as an estimator of ZC :

$$\overline{ZC} = \frac{\overline{R_{ss}(\tau_1)}}{\overline{R_{ss}(0)}}. \quad (12)$$

Being an estimator, \overline{ZC} will be a random variable with its own statistics. Even in the RFI-free case, there is a non-negligible probability that the decision metrics depart from 0 enough to trigger the detection of RFI. It is necessary then to define a confidence interval, for witch an acceptable probability of False Alarm (P_{fa}) must be set. A high P_{fa} implies a conservative approach with a large number of false alarms, but it also implies a prompt reaction of the method when an RFI appears. On the contrary, a low P_{fa} implies a lower number of false alarms, but it also means more laxity on RFI detection. The value of P_{fa} to be chosen depends on the specific application, but it enables to define the confidence intervals and decision thresholds. To do so, the specific statistics of the decision variable must be known or estimated.

\overline{ZC} is the ratio of two normal random variables. Let's denote a generic random variable ratio as $Z = \frac{X}{Y}$. Such a variable has no moments in the general case, and its probability density has an analytic, but very complicated, expression [17, eq. 1 and 2]. Such expression can be found simplified at [18, eq. 1], where the following parametrization is used:

$$\beta = \frac{\mu_X}{\mu_Y}, \quad \delta_y = \frac{\sigma_Y}{\mu_Y}, \quad \rho = \frac{\sigma_Y}{\sigma_X}, \quad (13)$$

being $\mu_{X,Y}$ and $\sigma_{X,Y}$ the mean and standard deviation of X, Y . Under some conditions on the parameters of eq. 13, the distribution Z may be considered Gaussian with a high degree of accuracy. Several bounds for the parameters have been established empirically and, as a rule of thumb, the approximation can be done if $\delta_y < 0.1$ [18]. In such a case,

$$Z \approx \mathcal{N}(\beta, \delta_y^2(\rho^{-2} + \beta^2)). \quad (14)$$

Taking into account eq. 10, for \overline{ZC} ,

$$\begin{aligned} \mu_X &= 0, & \sigma_X &= \frac{N}{(N-m)^2} \sigma_s^2, \\ \mu_Y &= \sigma_s, & \sigma_Y &= \frac{\sigma_s^2}{N}. \end{aligned} \quad (15)$$

Therefore, by substitution into equation 13,

$$\begin{aligned} \beta &= 0, \\ \delta_y &= \sqrt{\frac{2}{N}} \ll 0.1, \\ \rho &= \sqrt{2} \frac{N-m}{N}, \end{aligned} \quad (16)$$

and, with $m = 1$,

$$\overline{ZC} \approx \mathcal{N}\left(0, \frac{N}{(N-1)^2}\right). \quad (17)$$

will have Gaussian statistics with a mean value $\mu_{\overline{ZC}} = 0$ and a variance:

$$\sigma_{\overline{ZC}}^2 \approx \frac{N}{(N-1)^2}. \quad (18)$$

In order to confirm the assumptions on the Gaussianity of \overline{ZC} , its exact distribution PDF given by [17, eq. 1 and 2] has been compared to the approximation given by [18]. As it can be seen in Fig. 1, for reasonable values of μ_X, μ_Y, σ_X and σ_Y ,

the agreement is excellent, and it enables to use the approximation in practical terms¹.

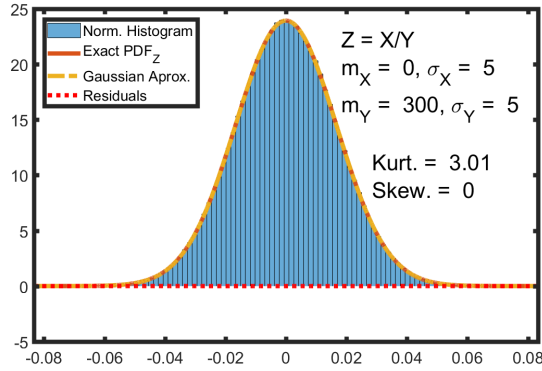


Fig. 1: Comparison between the exact distribution of the ratio [17] and its Gaussian approximation [18].

Given that the statistics can be approximated by a Gaussian, bilateral decision thresholds may be computed by integrating the tails of its PDF. Considering the Cumulative Distribution Function (CDF) of a generic Gaussian of mean μ and standard deviation σ :

$$CDF(x) = \frac{1}{2} \left(1 + \operatorname{erf} \left(\frac{x - \mu}{\sqrt{2}\sigma} \right) \right). \quad (19)$$

The lower threshold, th_1 , can be obtained by equating eq. 19 with the intended P_{fa} (with the $1/2$ factor to account for the bilaterality of the distribution).

$$CDF(th_1) = \frac{1}{2} \left(1 + \operatorname{erf} \left(\frac{th_1 - \mu}{\sqrt{2}\sigma} \right) \right) = \frac{P_{fa}}{2}. \quad (20)$$

Solving for th_1 yields:

$$th_1 = \mu - \sigma\sqrt{2} \operatorname{erf}^{-1}(1 - P_{fa}), \quad (21)$$

where the property $\operatorname{erf}(-x) = -\operatorname{erf}(x)$ has been used. Given the symmetry of the Gaussian distribution function, the upper threshold th_2 is readily obtained by:

$$th_2 = \mu + \sigma\sqrt{2} \operatorname{erf}^{-1}(1 - P_{fa}). \quad (22)$$

By substituting eq. 18 into 21 and 22:

¹Note that, since the filters response is known, it is possible to compute the position and depth of the notches. Also, since the exact expression is available [17, eq. 1 and 2], it is possible to compute the thresholds numerically without the Gaussian approximation if deemed necessary.

$$th_{1,2} = \pm\sqrt{2} \operatorname{erf}^{-1}(1 - P_{fa}) \cdot \frac{N}{(N-1)^2}. \quad (23)$$

Performance of the method can be assessed by computing the detection probability (P_{dec}) for a certain RFI power and type. To do so, $th_{1,2}$ are applied to the measured \overline{ZC} . If $\overline{ZC} < th_1$ or $\overline{ZC} > th_2$, RFI is detected, and the sequence is flagged as such and discarded. In this work, P_{dec} has been approximated numerically by conducting $N_{MC} = 5000$ Monte Carlo simulations:

$$P_{dec} = \frac{N_{rfi}}{N_{MC}}, \quad (24)$$

being N_{rfi} the number of sequences flagged as RFI-affected. If the thresholds have been properly computed, $P_{dec} \approx P_{fa}$ in the absence of RFI. Performance results of the described method will be given and discussed in section 5. It will be shown that, without RFI, $P_{dec} \approx P_{fa}$, demonstrating that all assumptions discussed above on the Gaussianity of \overline{ZC} hold. This is a direct consequence of how the bilateral thresholds are derived.

It should be noted that the proposed method is only valid if the sampling conditions of eq. 7 are fulfilled. In a more general case (for example, if oversampling is considered, or if the filters are not perfectly ideal), the correlation may not be sampled exactly at the zero-crossings. Nonetheless, the same general idea may be applied at the cost of system complexity by, for example, interpolating the correlation values at the zero-crossing.

3.2. Pearson Correlation Coefficient Detection (PCD)

As it will be shown later, Zero-Crossing Ratio provides good results with a fairly simple implementation. It relies, however, only in one sample point (the first zero-crossing of the correlation), making it prone to blind spots. In order to overcome some of the drawbacks of the ZCR, a metric based on the computation of the Pearson correlation coefficient is proposed as well. The Pearson coefficient ρ of two ensembles X, Y can be computed as:

$$\rho_{x,y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}, \quad (25)$$

where $\mu_{X,Y}$ and $\sigma_{X,Y}$ are the mean and standard deviation of X, Y , respectively. The Pearson coefficient, which satisfies $-1 < \rho_{X,Y} < 1$, can be interpreted as metrics of the similarity of the shape of two signals. As our interest is to determine if a signal is contaminated by RFI, X will be the measured correlation shape R_{ss} . The reference to compare with, Y , will be the theoretical correlation shape without RFI, that is,

equation 1. With this definition, a ρ close to 1 indicates non-contamination, and a value close to 0 is indicative of severe contamination. Note that the Pearson coefficient is independent of the power of the signals considered, so an estimation of the radiometric power level is not required.

It is important to remark that, in practical terms, only the sample correlation coefficient, $\overline{\rho_{X,Y}}$, can be computed, which is an estimation of the real Pearson coefficient. As the sample correlation coefficient is derived from a limited number of samples, it is a random variable itself, with its own statistics. From eq. 25, and substituting the sample mean and standard deviation, $\overline{\mu_x}, \overline{\mu_y}$ and $\overline{\sigma_x}, \overline{\sigma_y}$, the sample Pearson coefficient is:

$$\overline{\rho_{X,Y}} = \frac{\sum_{i=0}^N (x_i - \overline{\mu_X})(y_i - \overline{\mu_Y})}{\overline{\sigma_X} \overline{\sigma_Y}}, \quad (26)$$

where $X = [x_1 \dots x_N]$ is the estimation of the correlation given by eq. 8, and $Y = [y_1 \dots y_N]$ is a deterministic sequence given by appropriately sampling equation 1. It should be noted that from equation 10, the variance of R_{ss} grows for large delays, as this is a property of the unbiased estimator of the autocorrelation. Therefore, in order to keep the correlation coefficients close to 1, it is better to restrict the computation to the central $2m + 1$ samples of the correlation:

$$\overline{\rho_{X,Y}} = \frac{\sum_{N-m}^{N+m} (x_i - \overline{\mu_X})(y_i - \overline{\mu_Y})}{\overline{\sigma_X} \overline{\sigma_Y}}. \quad (27)$$

Several possibilities for m have been evaluated in this work. The statistics of $\overline{\rho}$ are in general non-normal, and the exact distribution will depend also on the statistics of X and Y . If they are jointly normal an exact expression for the PDF exists, but it is not mathematically tractable [19]. Thankfully, Fisher transformation can be used to approximately normalize its statistics and make possible to compute a confidence threshold [20]. Fisher transformation is defined as:

$$z = \tanh^{-1}(x). \quad (28)$$

Indeed,

$$z = \tanh^{-1}(\overline{\rho_{x,y}}) \quad (29)$$

is approximately normally distributed. Assuming Gaussianity of the variable under consideration, bilateral decision thresholds are given by eq. 21 and 22 in order to guarantee a given P_{fa} :

$$th_{1,2} = \mu_z \pm \sqrt{2} \operatorname{erf}^{-1}(1 - P_{fa}) \cdot \sigma_z, \quad (30)$$

where μ_z and σ_z are the mean and standard deviation of z in the absence of RFI. Under some requirements on the statistics of X, Y , σ_z and μ_z have closed forms. For the case of interest, however, no closed expression exists, as the distribution of the sequence X is not Gaussian. Therefore, to conduct this exercise, both parameters have been estimated numerically by using an RFI-free sequence X , mimicking, for example, an internal calibration load. Note that, in order to apply $th_{1,2}$ to the sample Pearson coefficient $\overline{\rho}$, they must be anti-transformed first by using:

$$x = \tanh(z). \quad (31)$$

Similarly to the Zero-Crossing Ratio, the performance is assessed by computing the probability of detection P_{dec} . Results are shown and discussed at section 5 of this work, demonstrating as well that the assumptions taken in the computation of the thresholds are good enough ($P_{dec} \approx P_{fa}$ in the absence of RFI).

As a reference to compare with, two additional common metrics have been computed: a trivial power detection method and a signal Kurtosis statistical test.

3.3. Total Power Detection

This trivial detection method relies on the assumption that the RFI-free signal power is known beforehand. Therefore, the measured power and the RFI-free power can be compared. Measured power is derived by using the sample variance of the LPE signal, being $\mathbf{s} = (s_1 \dots s_N)$ the signal vector with $\overline{\mu_s}$ the sample mean:

$$\overline{\sigma_s^2} = \frac{1}{N-1} \sum_{i=0}^N (s_i - \overline{\mu_s})^2. \quad (32)$$

The sample variance follows a scaled chi-squared distribution of $N - 1$ degrees of freedom, $\frac{\sigma_s^2}{N-1} \chi_{N-1}^2$ [21, eq. 2.11]. For simplicity, if N is large enough, this distribution can be considered normal for all practical purposes as $\mathcal{N}\left(\sigma_s^2, \frac{2\sigma_s^4}{N-1}\right)$ [22, p. 118], being σ_s^2 the true variance of \mathbf{s} . Provided that $\sigma_s^2 = \sigma_n^2$ for the RFI-free case, bilateral decision thresholds can be computed in order to guarantee a given P_{fa} (eqs. 21 and 22):

$$th_{1,2} = \sigma_n^2 \pm 2 \operatorname{erf}^{-1}(1 - P_{fa}) \frac{\sigma_n^2}{\sqrt{N-1}}. \quad (33)$$

The above can only be defined if, as mentioned, the radiometric signal power σ_n^2 is known beforehand, which is of course not applicable in a general case, as this is usually the parameter obscured by the RFI. This simple metric is, however, a helpful benchmark to compare with in laboratory conditions. Total Power Detection can also be used as a practical method

if the radiometric power is estimated [23], either by taking previous measurements as a reference or by comparison with a range of possible geophysical variation. Performance is then directly linked to the accuracy of the estimation of the baseline noise level. This is the fundamental drawback of using this technique in practical conditions. Note that ZCR, PCD or Kurtosis, below, do not require any prior estimation on the signal properties.

3.4. Signal Kurtosis

Signal Kurtosis is, on the other hand, a practical detection method widely used and tested in the literature [24]. Like similarly-minded statistical tests (see, for example, Anderson-Darling [8]), it aims to test the statistics of the signal for a hint of non-normality. Signal Kurtosis relies on the computation of the scaled fourth moment of the signal distribution (statistical kurtosis), which provides a metric of the probability of finding outliers in the data. In other words, provides a metric on the shape of the PDF, indicating whether the tails of the probability distribution are fatter (leptokurtic) or thinner (platykurtic) when compared to the standard Gaussian distribution. A purely normal distribution will have a Kurtosis value of 3. Any divergence from this value indicates deviation from normality and thus RFI contamination. Kurtosis is defined as:

$$K = \frac{m_4}{m_2^2}, \quad (34)$$

with m_4 and m_2 being the fourth and second moments of the distribution under test, respectively. In practice, Kurtosis is computed by using the sample kurtosis estimator, \overline{K} , given by:

$$\overline{K} = \frac{\frac{1}{N} \sum_{i=0}^N (s_i - \overline{\mu})^4}{\left(\frac{1}{N} \sum_{i=0}^N (s_i - \overline{\mu})^2 \right)^2}. \quad (35)$$

While the probability distribution of the Kurtosis estimator has no closed form, except for the smallest of the samples, it reaches almost normality when N is sufficiently large ($N > 50000$). In practical contexts N is not so big, so the exact distribution has to be numerically estimated and accounted for [15]. The moments for this distribution are well known, and its variance follows [25]:

$$\sigma_K^2 = \frac{24N(N-1)^2}{(N-3)(N-2)(N+3)(N+5)} \approx \frac{24}{N}. \quad (36)$$

Regardless of the distribution's non-normality, in the context of this work is enough to define the decision thresholds as if the distribution was normal, that is, $\mathcal{N}\left(3, \sigma_K^2\right)$. With this

assumption, symmetric decision thresholds can be computed as (eqs. 21 and 22):

$$th_{1,2} = 3 \pm \sqrt{2} \operatorname{erf}^{-1}(1 - P_{fa}) \cdot \sigma_{\overline{K}}. \quad (37)$$

As it will be seen, for an $N = 1024$, the P_{fa} obtained using $th_{1,2}$ closely matches the intended value, with a deviation error of around 1%. As the main objective of this paper is not evaluating the performance of Kurtosis (a topic extensively studied in the literature) but to use it as a reference, this approximation is considered good enough.

4. MATERIALS AND METHODS

In order to evaluate the performance of the methods proposed, a simulated processing chain has been implemented, mimicking as much as possible the receiving elements of a correlation radiometer.

The radiometric signal has been simulated as a digital complex random Gaussian process of 300 K of power with N independent samples. A generic complex waveform of varying type and power is added into it, simulating contamination by RFI. The combined signal is then passed through a digital anti-aliasing filter of $B_w = 20$ MHz, and sampled at the Nyquist frequency $F_s = 2B_w$. This simulates the complex Low-Pass Equivalent (LPE) centered at $f_c = 0$ of the acquired band-pass signal.

In this study, 6 different types of RFI have been considered:

1. A continuous wave (CW): A single tone signal (sinusoidal), simulating a narrowband modulation, of configurable frequency.
2. A burst of pulses of 10% duty cycle: a train of Gaussian pulses with a pulse repetition period (PR) of $N/256$ samples, a pulse width of $PR/10$ samples, and configurable central frequency. This is representative of DME-like signals [26].
3. A burst of pulses of 50% duty cycle: a train of rectangular pulses with a pulse repetition period (PR) of $N/128$ samples, a pulse width of $PR/2$ samples, and configurable central frequency. Kurtosis detection exhibits a blind spot for this duty cycle, and is therefore interesting to study how correlation-based RFI detection behave in the same scenario.
4. A narrowband chirp signal: A chirp signal sweeping linearly with an arbitrary bandwidth of $B_w/2$ and a $PR = N/16$ samples. Chirp signatures are representative of RADAR signals and jammers.
5. A wideband chirp signal: A chirp signal sweeping linearly with an arbitrary bandwidth of B_w and a $PR = N/16$ samples.

6. A generic wideband signal modulation: Simulated using a pseudo-random noise code (PRN) of $PR = N/2$, with its bandwidth overlapping the entire noise bandwidth. Note that, even if overlapping it, its spectrum is not flat (see Fig. 2.e), hence being still prone to detection.

Their spectrograms are presented in Fig. 2, and their normalized autocorrelation is depicted in Fig. 3, along with the normalized autocorrelation of Gaussian noise for reference.

The real and imaginary components of the LPE signal mimic the I and Q components of the generic digital receivers. Each component is then treated independently at separate processing chains. In order to take into account the influence of digitalization, a configurable quantizer has been implemented as well. Quantization is performed following several uniform quantization schemes: 1 bit, 2 bits, 3 bits, 4 bits, and 8 bits, in addition to the unquantized option. All quantizers are mid-riser, where 0 is a decision threshold. Small, near-0 values are quantized as the first positive or negative level, depending on their sign. Quantization is applied to each I, Q components separately.

Complex correlation between antennas m, n is computed as:

$$R_{m,n} = R_{m_I,n_I} + R_{m_Q,n_Q} + (R_{m_I,n_Q} - R_{m_Q,n_I}) \cdot j, \quad (38)$$

where $m_{I,Q}$ are the IQ components of signal from antenna m , and $n_{I,Q}$ the IQ components of signal from antenna n . If quantization has been applied, then the denormalization function F has to be computed following [27] and applied to each of the correlation components, thus obtaining the 'unquantized' correlation. The resulting correlation sequence is tested for RFI with the detection methods described above, ZCR and PCD. With regards to the reference metrics (Total Power and Kurtosis), they are obtained directly from the signal: signal power is obtained by combining the variance of both I, Q signals, and average signal kurtosis is computed as the mean of the kurtosis of each I, Q chains, a solution often implemented in practice [28]:

$$\bar{K} = \frac{\overline{K_I} + \overline{K_Q}}{2}. \quad (39)$$

5. RESULTS

In this section, the performance achieved by the proposed detection methods will be analyzed and compared with the reference metrics. In the first place, the performance achieved will be shown, for all RFI types, and a fixed central frequency. Later, the dependence on central frequency, quantization type and sample size will be discussed in detail.

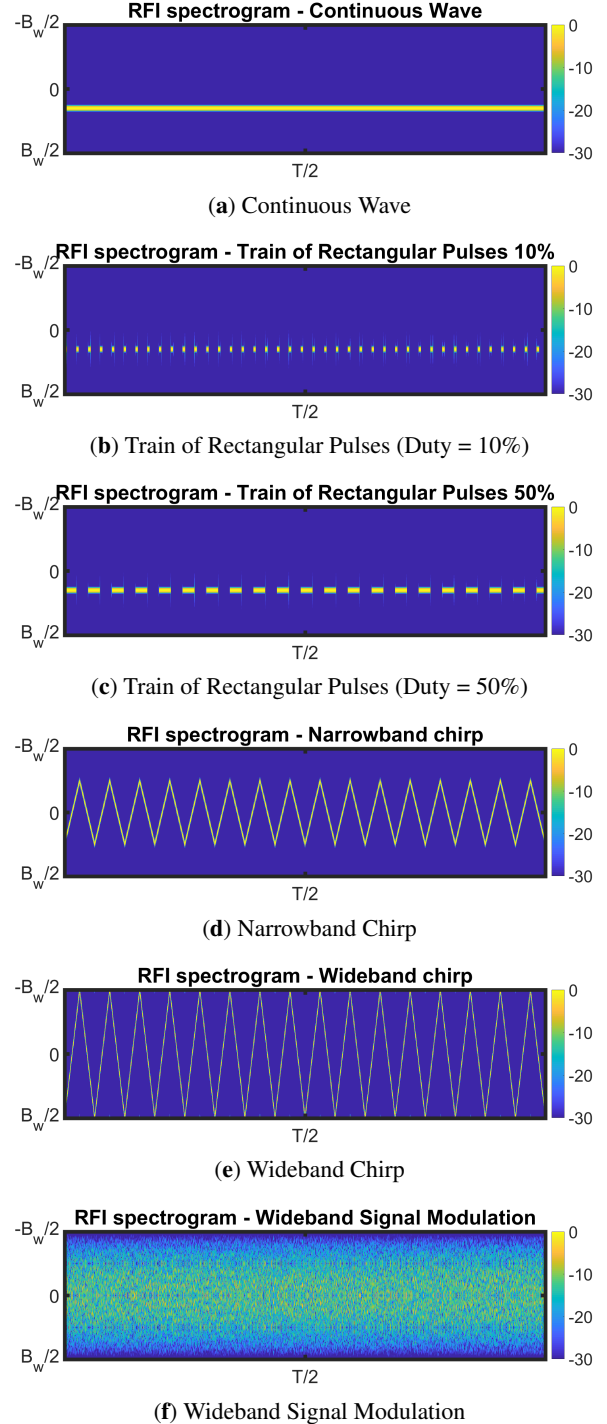


Fig. 2: Spectrograms of the different types of RFI considered in this study.

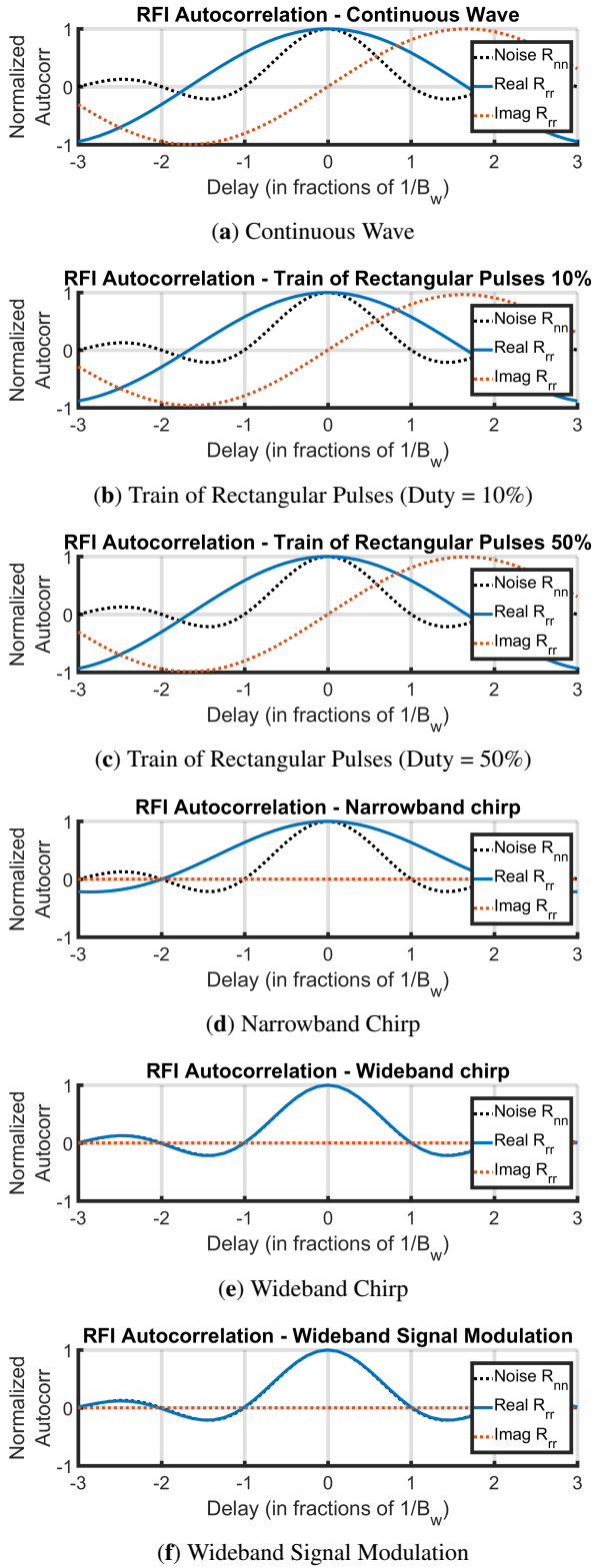
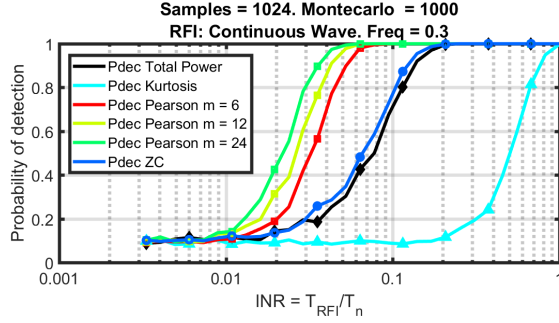


Fig. 3: Autocorrelation of the different types of RFI considered in this study, being $R_{n,n}$ the autocorrelation of the noise, and $R_{r,r}$ the autocorrelation of the RFI.

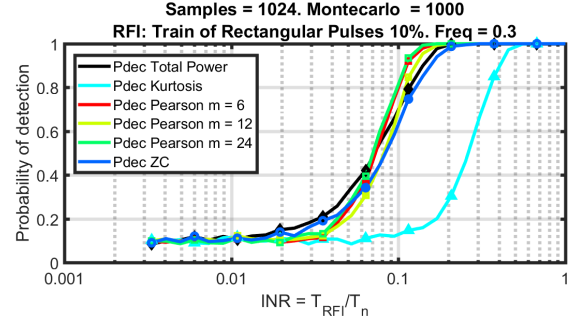
5.1. Detection performance of Correlation-based RFI detection

As mentioned, the main figure of merit to evaluate the methods' performance is the Probability of Detection P_{dec} . In Fig. 4, P_{dec} has been depicted in function of the Interference to Noise Ratio ($INR = TB_{RFI}/TB_{Noise}$, where TB_{RFI} and TB_{Noise} are the Brightness Temperature for the RFI and Noise, respectively) for the 6 types of RFI considered in this study. As a first approximation, RFI central frequency has been chosen to be $f_c = 0.3 \cdot B_w$, and $N = 1024$. Given that $F_s = 2B_w = 40$ MHz, this corresponds to a sequence length of $T = 25\mu s$. The results obtained demonstrate that the proposed methods work, being able to detect all types of RFI. Performance is discussed qualitatively for each of the methods below:

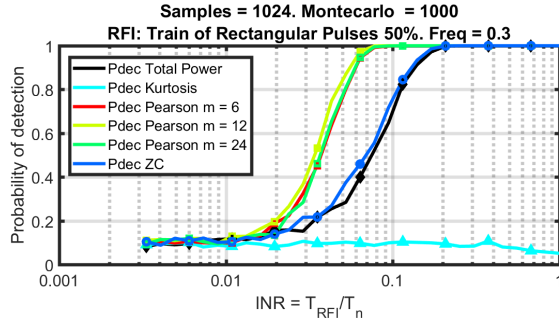
1. Zero-Crossing Ratio: ZCR performs comparatively to Total Power Detection for narrowband RFI. In these cases, ZCR is able to detect RFI for $INR < 0.2$, outperforming Signal Kurtosis. For spectrally wide RFI, however, ZCR performs worse. For Wideband Chirp, for example, ZCR is not able to guarantee detection for $INR < 1$. ZCR has a marked dependence of the frequency contents of the RFI. For some selected cases, performance is notable, even outperforming more complex methods like the Pearson method. For other cases, however, performance is much poorer. This is related to how the autocorrelation of the RFI impacts the zero-crossings of the noise autocorrelation. In Fig. 3 it can be seen how, for some RFIs, autocorrelation shapes are very similar to the expected noise autocorrelation at the first zero-crossing, a fact that impairs RFI detection by using it. Nevertheless, and taking into account the simplicity of the implementation, ZCR is a strong candidate for RFI detection where system complexity is an issue. It should be remarked how, even if the performance obtained is similar than using TPD, the latter requires a good estimation of the baseline noise level, while ZCR does not.
2. Pearson Coefficient Detection: Pearson outperforms Kurtosis for all types of RFI, and even Total Power Detection in selected cases. For RFI with well-concentrated spectra, such as the Pulse Train and CW, Pearson is able to detect RFI of the order of $INR = 0.05$. For wider spectra RFI, Pearson performs worse, but achieving comparable results to Signal Kurtosis. There is a notable dependence on the number of samples considered to compute Pearson. Indeed, for spectrally narrow RFI, these differences are minor, with larger samples implementation giving marginally better results. For spectrally-wide RFI, however, lower sample computation exhibits lower performance or even non-detection for $INR < 1$ (see, as an example, Fig. 4.e). The origin



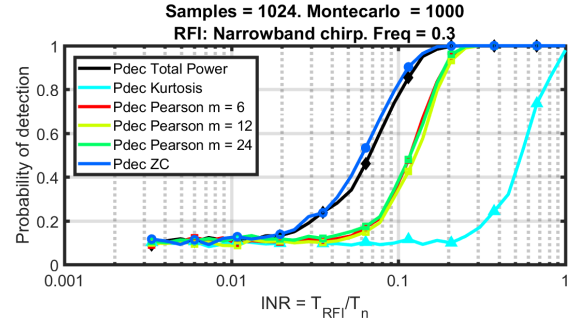
(a) Continuous Wave



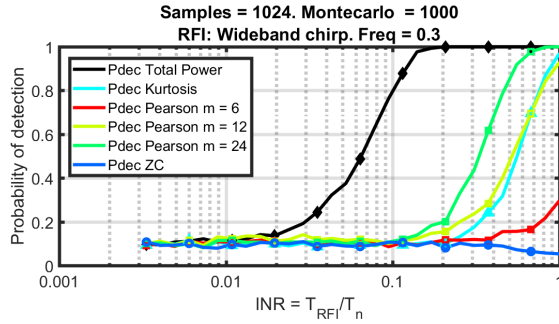
(b) Train of Rectangular Pulses (Duty = 10%)



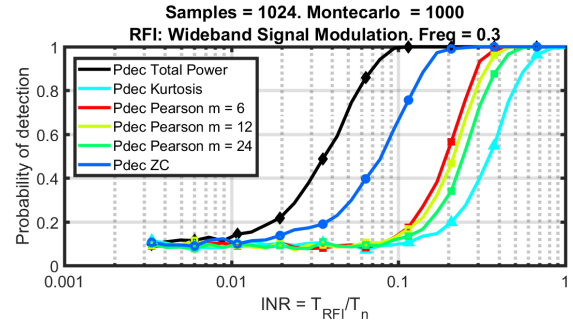
(c) Train of Rectangular Pulses (Duty = 50%)



(d) Narrowband Chirp



(e) Wideband Chirp



(f) Wideband Signal Modulation

Fig. 4: P_{dec} in function of the INR for several RFI typologies

for this poor performance is explained because, in this case, only the lower delays of the correlation are sampled. As the main lobe of the correlation is less prone to distortion, this makes detection harder. As it can be appreciated in Fig. 4), if m is chosen to be $m > 12$, detection is achieved with reasonable performance for all the cases studied, providing a good trade-off between performance and system complexity.

It should be stressed how, for an $INR = 0$, all detection methods give a $P_{dec} \approx P_{fa}$. This confirms the approximations done for some thresholds computation, as explained in detail in section 3.

In order to quantitatively compare the performance of the methods, it is interesting to define the Minimum Detectable RFI. For a given RFI type and P_{fa} , this can be defined arbi-

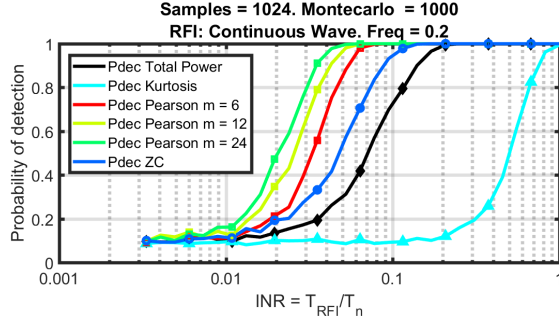
trarily as the minimum required INR to detect an RFI with a probability equal to $1 - P_{fa}$:

$$INR_{min} = INR(1 - P_{fa}). \quad (40)$$

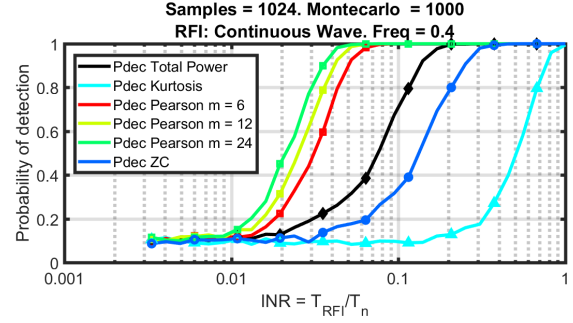
In table 1, INR_{min} is given for each type of RFI, under the same conditions than Fig. 4.

5.2. Effects of RFI frequency on Correlation-based RFI detection

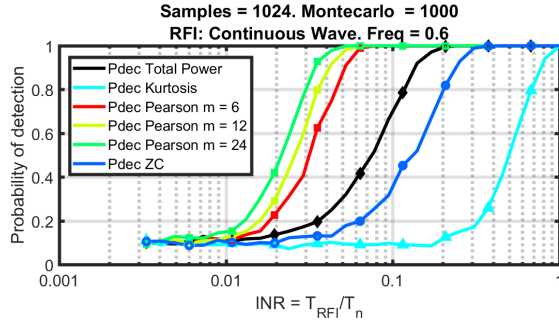
As discussed in section 1, the autocorrelation shape of the RFI is directly depending on its frequency contents. Therefore, any technique that relies on checking the autocorrelation may have a dependence on it. In this section, the effect of choosing a different central frequency for the RFI on the performance



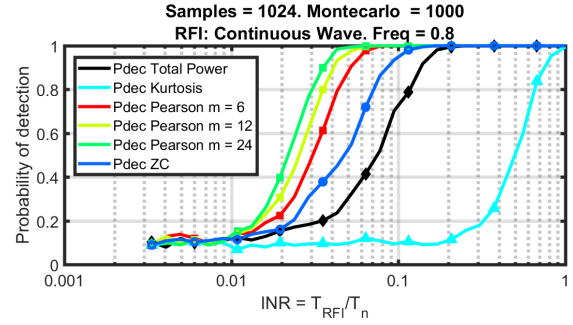
(a) Continuous Wave with $f_c = 0.1$



(b) Continuous Wave with $f_c = 0.2$



(c) Continuous Wave with $f_c = 0.3$



(d) Continuous Wave with $f_c = 0.4$

Fig. 5: P_{dec} in function of the INR for CW of different frequencies

of the proposed techniques is studied. In Fig. 5, P_{dec} as a function of the INR has been plotted for an CW RFI with four different central frequencies: 0.2, 0.4, 0.6 and 0.8.

As it can be seen, ZCR is considerably affected by frequency. This is related to the fact that only one non-central sample is taken into consideration. As the CW the central frequency of the RFI determines the zero-crossings of its autocorrelation, depending on where they are located the sensitiveness of the detection varies. The limiting case is for 0.5: the first zero crossing of the RFI autocorrelation is located exactly at $\tau = \frac{1}{B_w}$, constituting a blind spot for detection. This latter case is exemplified in Figs. 6 and 7. Pearson Correlation Detection, on the other hand, does not exhibit significant variation across frequencies. This is due to the fact that a much larger number of samples is taken into consideration, effectively removing

the frequency dependence for the CW.

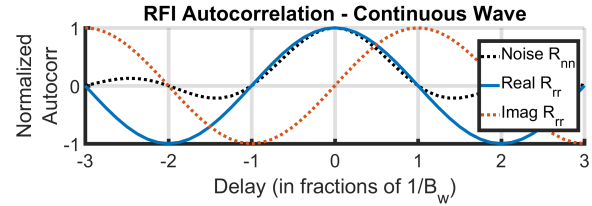


Fig. 6: Autocorrelation of an CW RFI with $f_c = 0.5$

5.3. Quantization effects on Correlation-based RFI detection

It is interesting to consider the effects of digitalization in the correlation shape. Real-time computation of the cross-

Table 1: INR_{min} to guarantee detection for all types of RFI considered

	Total Power	Kurtosis	PCD (m = 6)	PCD (m = 12)	PCD (m = 24)	ZCR
Continuous Wave	0.13	0.77	0.05	0.04	0.03	0.12
10% DC Pulse Train	0.13	0.40	0.11	0.13	0.11	0.15
50% DC Pulse Train	0.14	N/D	0.06	0.05	0.06	0.13
Narrowband Chirp	0.13	0.85	0.19	0.20	0.19	0.11
Wideband Chirp	0.12	0.89	N/D	0.93	0.54	N/D
Wideband Modulation	0.07	0.58	0.29	0.33	0.39	0.15

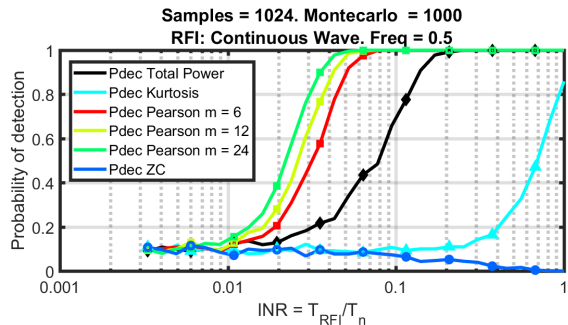


Fig. 7: Blind spot of the ZCR technique, for an RFI with $f_c = 0.5$

correlation is a costly procedure when considering finely quantized signals. Consequently, it is usually implemented efficiently using coarse-quantization schemes (e.g., 1-bit quantization) prior to the correlation. Thanks to some properties of the radiometric signals, the original, non-quantized correlation of Gaussian signals can be recovered from the correlation of quantized signals [29]. If the quantized signal is Gaussian, a denormalization function can be computed regardless of the specific quantization scheme used [27]. Quantization, and specially but not limited to the coarser schemes, has been demonstrated to be a limiting factor for several RFI mitigation families. For example, quantization impairs considerably the performance of Time/Frequency techniques [30], and statistical methods [15]. By using the correlation shape, however, these limitations can be overcome thanks to the denormalization function. In this section, the impact of quantization is studied numerically to demonstrate the applicability of the proposed techniques to SAIR. In order to apply the proposed methods to the quantized case, some additional considerations must be taken into account with regards the computation of the thresholds, as some assumptions do not longer hold. In particular, the thresholds for ZCR have been recalculated by estimating numerically σ_{ZC} .

Fig. 8 shows P_{dec} as a function of the INR for 1-bit quantization. As a reference, the same metrics in the unquantized case has been included. As it can be readily observed, the use of the denormalization function prior of RFI detection allows to obtain almost equivalent results for Pearson Coefficient Detection, albeit marginally degraded. This is confirmed for the most coarse quantization, using just 1-bit to quantize the IQ components. Identically, ZCR exhibits similar results, with a slightly degraded performance with respect the unquantized case. Nevertheless, in both cases the observed performances are still comparable with Total Power Detection, making them viable candidates as detection techniques if quantization is required. Given the considerable advantages that 1-bit quantizers have in the framework of correlation radiometers, this is an important result towards the practical implementation of RFI detection for SAIRs.

5.4. Sample size impact for Correlation-based RFI detection

The above study has been conducted, without loss of generality, for a sample size of $N = 1024$. This demonstrates the applicability of the methods for reduced integration times (e.g., with $F_s = 40$ MHz, $T_{int} = 25\mu s$). It is interesting however to evaluate which is the impact of the sample size in the method's performance. As described in detail in section 3, decision thresholds are a function of the standard deviation of the estimator used. As the standard deviation directly depends on the number of independent samples considered, sample size impacts performance in a direct way. In general, larger sample sizes will allow better performances (i.e. detection for lower power RFI) and vice-versa.

As described, larger samples sizes allow to detect smaller RFI powers. This, however, leads to flag entire ensembles of data as contaminated by RFI. Lower sample sizes, albeit providing lower raw performance, may allow to use them in combination with time binning, that is, splitting N samples in M bins, and apply the correlation detection method individually to each bin. This kind of implementation provides the capability to remove only those bins that are contaminated, and therefore, allows the use of the rest of the signal. There is a compromise between raw performance and data availability that has to be tailored to the specific application and RFI environment where the instrument has to operate.

In order to illustrate the performance dependence on the sample size, the INR_{min} (eq. 40) has been computed for each detection method for different number of sample sizes between two extreme scenarios: on one hand, an $N = 32$ samples, and on the other, more than 1 million samples (i.e. 2^{20} samples). This latter scenario is representative of an instrument with a $T_{int} = 26$ ms assuming $F_s = 40$ MHz. Results are shown in Fig. 9 for a CW RFI with $f_c = 0.3$.

From Fig. 9, it is apparent that for $N > 10^3$, INR_{min} decreases proportionally as $N^{-1/2}$. This demonstrates that correlation-based RFI detection methods are capable of detecting arbitrary low power RFIs at the cost of increasing the number of samples considered. In all cases, correlation-based methods are capable of detecting lower power RFI than Total Power or Kurtosis detection, evidencing that the better performance shown in previous sections is observed regardless the number of samples considered.

6. CONCLUSIONS

In this work, a new family of RFI detection methods is proposed. They are based on the computation of the autocorrelation of the received signal, and the comparison with the theoretical autocorrelation of the radiometric noise, which should be an approximately sinc-shaped function in RFI-free conditions. Two implementations based on this idea have been proposed.

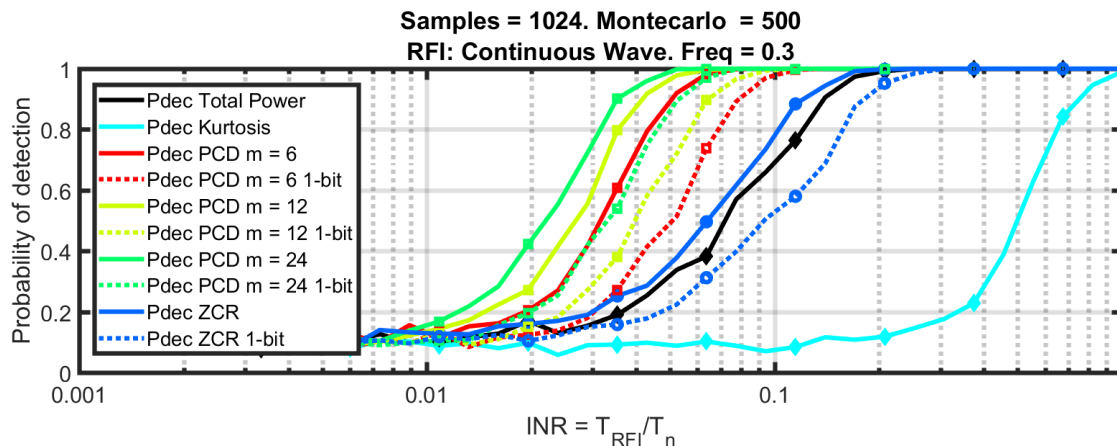


Fig. 8: Detection performance for unquantized and 1-bit quantized signals

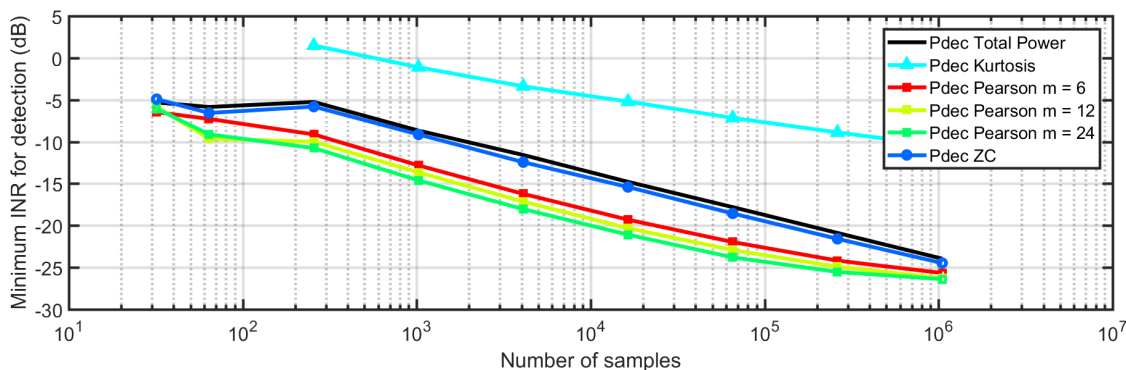


Fig. 9: Minimum INR required for detection in function of the sample size for an CW RFI with $f_c = 0.3$

posed in this work: the Zero-Crossing Ratio (ZCR), based on measuring the correlation at its zero-crossings, and the Pearson Coefficient Detection (PCD), based on the goodness of fit of the autocorrelation with the ‘sinc-shaped’ baseline. Their performance has been tested and evaluated by simulation for several types of RFI types, demonstrating that they outperform the reference metrics for most of the considered cases. Correlation methods are well suited for detection of frequency concentrated signals. For example, for a CW RFI of 0.3 frequency, and considering 1024 samples, PCD is able to detect RFI for $INR \approx 0.03$, and ZCR for $INR > 0.12$ outperforming Total Power Detection ($INR > 0.13$) or Kurtosis methods (detection for $INR > 0.77$, see Fig. 4.a). The dependence of the performance with frequency and sample size has been studied, confirming the applicability of the proposed methods in practical conditions. In addition, the proposed methods do not require any assumption on the RFI type or properties, nor the baseline noise level to operate. This makes them good candidates with respect other techniques with comparable performance levels, such as Total Power detection, which requires a good estimation of the noise level. The proposed novel detector family is suitable for a wide range of radiome-

ter topologies, such as Real Aperture Radiometers, Polarimetric Radiometers, etc. In addition, it has been demonstrated that correlation detection methods are applicable to quantized signals as well, opening the door for its application to digital systems, in particular to 1-bit quantized Synthetic Aperture Radiometers.

7. REFERENCES

- [1] S. Misra and P. de Matthaëis, “Passive Remote Sensing and Radio Frequency Interference (RFI): An Overview of Spectrum Allocations and RFI Management Algorithms,” *IEEE Geoscience and Remote Sensing Magazine*, vol. 2, no. 2, pp. 68–73, 2014.
- [2] National Research Council, *Spectrum Management for Science in the 21st Century*, The National Academies Press, 2010.
- [3] D. M. Le Vine and M. Haken, “RFI at L-band in synthetic aperture radiometers,” in *Proceedings of the 2003 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2003)*, 2003, vol. 3, pp. 1742–1744.

- [4] E. Daganzo-Eusebio, R. Oliva, Y. H. Kerr, S. Nieto, P. Richaume, and S. M. Mecklenburg, "SMOS radiometer in the 1400–1427-MHz passive band: Impact of the RFI environment and approach to its mitigation and cancellation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 10, pp. 4999–5007, 2013.
- [5] R. Oliva, E. Daganzo, Y. H. Kerr, S. Mecklenburg, S. Nieto, P. Richaume, and C. Gruhier, "SMOS radio frequency interference scenario: Status and actions taken to improve the RFI environment in the 1400–1427-MHz passive band," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 5, pp. 1427–1439, 2012.
- [6] R. Oliva, E. Daganzo, P. Richaume, Y. Kerr, F. Cabot, Y. Soldo, E. Anterrieu, N. Reul, A. Gutierrez, J. Barbosa, et al., "Status of Radio Frequency Interference (RFI) in the 1400–1427 mhz passive band based on six years of smos mission," *Remote sensing of environment*, vol. 180, pp. 64–75, 2016.
- [7] S. Misra, P. N. Mohammed, B. Guner, C. S. Ruf, J. R. Piepmeier, and J. T. Johnson, "Microwave radiometer radio-frequency interference detection algorithms: A comparative study," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 11, pp. 3742–3754, 2009.
- [8] T. W. Anderson and D. A. Darling, "A test of goodness of fit," *Journal of the American statistical association*, vol. 49, no. 268, pp. 765–769, 1954.
- [9] A. Camps, J. Gourrion, J. M. Tarongi, M. Vall Llossera, A. Gutierrez, J. Barbosa, and R. Castro, "Radio-frequency interference detection and mitigation algorithms for synthetic aperture radiometers," *Algorithms*, vol. 4, no. 3, pp. 155–182, 2011.
- [10] J. Querol, R. Onrubia, A. Alonso-Arroyo, D. Pascual, H. Park, and A. Camps, "Performance Assessment of Time-Frequency RFI Mitigation Techniques in Microwave Radiometry," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 10, pp. 1–11, 2017.
- [11] K. D. McMullan, M. A. Brown, M. Martin-Neira, W. Rits, S. Ekholm, J. Marti, and J. Lemanczyk, "SMOS: The payload," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 46, no. 3, pp. 594–605, March 2008.
- [12] P. Gaiser, K. St Germain, E. Twarog, G. Poe, W. Purdy, D. Richardson, W. Grossman, W. Jones, D. Spencer, G. Golba, et al., "The WindSat spaceborne polarimetric microwave radiometer: Sensor description and early orbit performance," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 11, pp. 2347–2361, 2004.
- [13] J. Querol, A. Perez, and A. Camps, "A Review of RFI Mitigation Techniques in Microwave Radiometry," *Remote Sensing*, vol. 11, no. 24, pp. 3042, 2019.
- [14] G. Forte, J. M. Tarongí Bauza, V. dePau, M. Vall, A. Camps, et al., "Experimental study on the performance of RFI detection algorithms in microwave radiometry: Toward an optimum combined test," *IEEE transactions on geoscience and remote sensing*, vol. 51, no. 10, pp. 4936–4944, 2013.
- [15] R. D. De Roo, S. Misra, and C. S. Ruf, "Sensitivity of the kurtosis statistic as a detector of pulsed sinusoidal RFI," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 7, pp. 1938–1946, 2007.
- [16] J. G. Proakis, *Digital signal processing: principles algorithms and applications*, Pearson Education India, 2001.
- [17] D. V. Hinkley, "On the ratio of two correlated normal random variables," *Biometrika*, vol. 56, no. 3, pp. 635–639, 1969.
- [18] E. Díaz-Francés and F. J. Rubio, "On the existence of a normal approximation to the distribution of the ratio of two independent normal random variables," *Statistical Papers*, vol. 54, no. 2, pp. 309–323, 2013.
- [19] H. Hotelling, "New light on the correlation coefficient and its transforms," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 15, no. 2, pp. 193–232, 1953.
- [20] R. A. Fisher, "Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population," *Biometrika*, vol. 10, no. 4, pp. 507–521, 1915.
- [21] K. Knight, *Mathematical Statistics*, Chapman & Hall, 2000.
- [22] G. E. Box, W. H. Hunter, S. Hunter, et al., *Statistics for experimenters*, vol. 664, John Wiley and sons New York, 1978.
- [23] B. Guner, J. T. Johnson, and N. Niamsuwan, "Time and frequency blanking for radio-frequency interference mitigation in microwave radiometry," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 11, pp. 3672–3679, 2007.
- [24] S. Misra, S. S. Kristensen, S. S. Sobjaerg, and N. Skou, "CoSMOS: Performance of kurtosis algorithm for radio frequency interference detection and mitigation," in *2007 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2007, pp. 2714–2717.
- [25] R. A. Fisher, "The moments of the distribution for normal samples of measures of departure from normality," *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, vol. 130, no. 812, pp. 16–28, 1930.
- [26] G. X. Gao, "DME/TACAN interference and its mitigation in L5/E5 bands," in *ION Institute of Navigation Global Navigation Satellite Systems Conference*, 2007.
- [27] X. Bosch-Lluis, I. Ramos-Perez, A. Camps, N. Rodriguez-Alvarez, E. Valencia, and H. Park, "A general analysis of the impact of digitization in microwave correlation radiometers," *Sensors*, vol. 11, no. 6, pp. 6066–6087, 2011.

- [28] A. J. Schoenwald, D. C. Bradley, P. N. Mohammed, J. R. Piepmeier, and M. Wong, "Performance analysis of a hardware implemented complex signal kurtosis radio-frequency interference detector," in *2016 14th Specialist Meeting on Microwave Radiometry and Remote Sensing of the Environment (MicroRad)*. IEEE, 2016, pp. 71–75.
- [29] J. Hagen and D. Farley, "Digital-correlation techniques in radio science," *Radio science*, vol. 8, no. 8-9, pp. 775–784, 1973.
- [30] R. Díez-García and A. Camps, "Impact of signal quantization on the performance of RFI mitigation algorithms," *Remote Sensing*, vol. 11, no. 17, pp. 2023, 2019.