NORMALITY TESTS ANALYSIS OF RADIOMETRIC SIGNALS FOR RADIO FREQUENCY INTERFERENCE DETECTION

J. M. Tarongi, A. Camps

Remote Sensing Lab, Departament del Teoria del Senyal i Comunicacions,
Universitat Politecnica de Catalunya, Barcelona, Spain & IEEC CRAE/UPC
Jordi Girona 1-4, Barcelona, P.O. Box: 08034, Barcelona, Spain
jose_miguel_tarongi@tsc.upc.edu, camps@tsc.upc.edu

Abstract: *Radio-frequency interference (RFI) present in microwave radiometry measurements leads to erroneous radiometric results. RFI addition to the radiometric signal modifies the detected power and the estimated antenna temperature from which the geophysical parameters will be retrieved. In recent years, techniques to detect the presence of RFI in radiometric measurements have been developed. They include time- and/or frequency domain analyses, or statistical analysis of the received signal which, in the absence of RFI, must be a zero-mean Gaussian process. The statistical analysis of the received signal can be performed by the normality study of the signal, as in statistical literature several normality tests are available. The motivation of this paper is the study of a set of normality tests applied to the received signal as the radiometric signal presents a Gaussian nature; observing the behavior of the different normality tests. A description of the normality tests and the RFI detection results are presented.*

Keywords - Microwave *Radiometry, Radio Frequency Interference, Detection*

I. INTRODUCTION

The performance of microwave radiometers can be seriously degraded by the presence of radio-frequency interference (RFI). Undetected RFI present in the radiometric measurement leads to an erroneous estimated power. Furthermore, RFI can be present even in the calibration, producing a systematic error in the whole data set. Nowadays, several RFI detection and mitigation methods have been developed, which are: time, frequency or statistical analyses of the received signal. Time analysis consists of the detection of power peaks in the received signal that are larger than the variance of the measured parameter. These values are discarded and only the RFI-free data are left [1]. Frequency analysis is based on the study of the received signal spectrum, and consists of discarding sub-bands with power higher than the variance of the measured parameter [1]. Finally, statistical analysis is based on the fact that the radiometric signal (thermal noise) follows the probability distribution of a zero-mean random Gaussian variable; therefore, several statistical parameters such as the moments, and the probability density function are perfectly known. The best known statistical analysis in microwave radiometry is the Kurtosis [2], although, other algorithms [3-5] have also been studied.

In this research, the suitability of several normality tests for RFI detection in microwave radiometry has been analyzed. The normality tests used in this study are: Saphiro-Wilk (SW), Anderson-Darling (AD), Lilliefors (L), Cramer-von Mises (CM) in addition to the Kurtosis (K) statistical parameter. The ultimate objective is to compare these normality tests to obtain the best test to detect RFI. Monte Carlo simulations have been used to compare tests performance.

A brief description of these tests is given in Section II. Section III analyzes the validity of these tests in relation to its probability of false alarm. Section IV shows the results obtained in the different analyses of a thermal noise signal contaminated with RFI. Finally Section V summarizes the conclusions obtained in this work.

II. NORMALITY TESTS

The main idea of using normality tests to detect RFI in microwave radiometry is the fact that the thermal noise, follows a zero-mean Gaussian distribution, and in general, man-made RFI are not Gaussian.

Normality tests used in this study have been widely used in statistical literature:

- Kurtosis test: Kurtosis is a statistical parameter related to the shape of the probability density function (PDF) of a random variable. Assuming a random process X, the Kurtosis (K) follows:

$$K = \frac{\mu_4}{\sigma^4}, \tag{1}$$

where μ_4 is the fourth moment about the mean, and σ is the standard deviation. In case that the random process is a zero-mean Gaussian process, the value of K tends to 3 as the number of samples increases. The Kurtosis test consists of comparing the estimated Kurtosis value of the received signal with tabulated values of the cumulative distribution function (CDF) of the Kurtosis of a Gaussian random variable of n samples. Kurtosis has been used in microwave radiometry RFI detection, although it exhibits some problems in detecting determined signals [2, 4].

- Lilliefors (L) test: L test is based on the empirical distribution function (EDF), which is defined as:

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \le x), \quad (2)$$

where $I(\cdot)$ is the indicator of the event, X_i is the i^{th} element of the sample to be tested. EDF function is a step function that increases by $1/n$ at the value of each ordered data point. The L test is a slight modification of the Kolmogorov-Smirnov (KS) test [6]. L test consist of:

$$\mathrm{L} = \max_{1\le i\le N}\left|F(X_i) - \hat{F}_N(X_i)\right|, \quad (3)$$

where $\hat{F}_N(X_i)$ is the value of the i^{th} element of the EDF of the sample X, and $F(X_i)$ is the value of the i^{th} element of the normal CDF with mean and variance equal to:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad (4)$$

$$\sigma^2{}_X = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2, \quad (5)$$

A requisite of L test, and the following tests, is that their samples must be ordered; as all are EDF based tests.

L confidence values are obtained from 2^{15} Monte-Carlo simulations [6] and are tabulated as the previous Kurtosis values. L test tends to be more sensitive near the center of the distribution than at the tails.

- Anderson-Darling test (AD): AD test is a modification of the L test that gives more weight to the tails than the L test. This test consists of:

$$\mathrm{AD}^{*2} = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left(\ln\Phi(Y_i) + \ln\left(1-\Phi(Y_{n+1-i})\right)\right), \quad (6)$$

with:

$$Y_i = \frac{X_i - \overline{X}}{\sigma_X}, \quad (7)$$

$\Phi(\cdot)$ represents the standard normal cumulative distribution function (CDF) operator. As it is described in, [7], AD^{*2} must be adjusted for the sample size:

$$\mathrm{AD}^2 = \mathrm{AD}^{*2}\left(1 + \frac{0.75}{n} + \frac{2.25}{n^2}\right). \quad (8)$$

- Shapiro-Wilk test (SW): SW test is defined as:

$$\mathrm{SW} = \frac{\left(\sum_{i=1}^{n} a_i X_i\right)^2}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}, \quad (9)$$

The main part of the SW test is the a_i, $i = 1...n$ vector of coefficients, that can be analytically calculated. A drawback of this test is the limitation of the sample size to a maximum of 2000 values [8]. To avoid this issue, SW test can be transformed to present a normal distribution in the case of normality of the tested signal [8]. Hence, longer sample lengths can be test by dividing it in several shorter length sets of samples, calculating the SW test on each set, and averaging the results as they are normally distributed [5].

- Cramer-von Mises test (CM): CM test is similar to the Ad test and L test, and consist of:

$$\mathrm{CM} = \frac{1}{12n} + \sum_{i=1}^{n}\left(\frac{2i-1}{2n} - F(X_i)\right)^2, \quad (10)$$

III. NORMALITY TEST VALIDATION

In order to understand the concepts defined in this section, two error types must be introduced first:

Probability of false alarm (P_{fa}): This error is produced when, in the absence of RFI sources, the algorithm “detects” the presence of RFI in a determined sample, leading to the elimination of correct data, reducing the total integration time.

Probability of detection (P_{det}): This error is produced when an undetected RFI is present in a sample, leading to an erroneous measurement which is assumed to be correct.

Objective is to obtain a low P_{fa} and a high P_{det}, but both types of errors have a strong correlation. This way, if it is desired to minimize the P_{fa}, the RFI detection threshold must be set to a relative high value, leading to a low P_{det}. On the other hand, if it is desired to maximize the P_{det}, the RFI detection threshold must be set to a relative low power value, leading to a high P_{fa}. A good way to evaluate the compromise between P_{det} and P_{fa}, is the calculation of Receiver Operating Characteristic (ROC) curves [4, 5]. In addition, ROC plots of the interference-to-noise ratio (INR) in function of the sample size are shown.

Validation of the normality tests is performed in order to minimize the errors in the threshold calculation for a determined pair of values P_{det} and P_{fa}. If the normality test introduces an error due to its nature, changes in the values of P_{det} and P_{fa} must be acceptable. The method followed to check the normality tests is the calculation of the ROC curve by means of 2^{15} Monte Carlo simulations of a Gaussian signal in the absence of RFI, for every test. In this study a test is validated when the error between the ROC curve of the test and the ideal case ($P_{det} = P_{fa}$) is less than the 5% (e.g. for $P_{fa} = 0.1 \rightarrow 0.095 < P_{det} < 0.105$). Sample size and quantization level are the most decisive parameters in the normality test validation.

The quantization has been modeled for a dynamic range of $\pm 8\sigma$, where σ^2 is the RFI free noise power. This way, digital input signal has less chance to be saturated in case of high power RFI contributions and the probability of clipping is almost zero.

As seen in fig. 1 validity of the normality tests is in function of the quantization depth (the deeper the better) and the sample size (the lower the better). Kurtosis test does not appear as the number of quantification bits needed lower than 6 bits for any sample size.

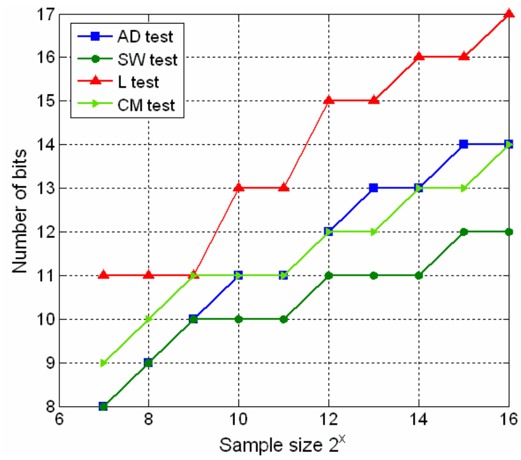

Fig. 1. Minimum number of bits to neglect the quantization error in function of the sample size.

IV. NORMALITY TEST PERFORMANCE

The main objective of this work is a comparison of different normality tests in the detection of RFI in microwave radiometry signals. To make a fair comparison between the different normality tests, two RFI contributions over a thermal noise signal will be tested. The selected RFI signals are:

- Pulsed sinusoidal signal: this signal has been extensively studied in microwave radiometry [2-5]. Signal model of this RFI is described as:

$$PS[n] = A\cos(2\pi f_0 T_s n + \varphi_0) H[n], \qquad (11)$$

where $PS[n]$ is the sampled pulsed sinusoidal signal, A, f_0 and φ_0 are the amplitude, the frequency and the initial phase of the RFI respectively, T_s is the sampling period and $H[n]$ is a train of pulses function described as:

$$H[n] = \begin{cases} 1 & \mathrm{mod}_M(n) \le M \bullet DC \\ 0 & \text{otherwise} \end{cases}, \qquad (12)$$

where N is the sample length, M is the pulse length and DC is the duty cycle of every pulse.

- Pseudo-Random Noise (PRN) signal: a PRN signal is a signal similar to noise that satisfies one or more standard tests for statistical randomness. This signal consists of a deterministic sequence of pulses that repeats itself after a long period, leading to a spread spectrum behavior of the signal. In this work, the firsts 10230 output bits of a maximum length sequence generator of 14 stages (13) are used as the PRN interfering signal.

$$PRN = X^{14} + X^8 + X^7 + X^4 + X^3 + X^2 + 1, \qquad (13)$$

Figures 2-4 represent the INR value required to obtain a ROC curve with a P_{fa} = 0.1 for P_{det} = 0.9. Results obtained in all figures by Monte-Carlo sets of 2^{15} simulations.

IV.1 Pulsed Sinusoidal Signal

Figure 2 shows the performance of the different normality tests for sample sizes of 1024 and 16384 respectively. The performance is measured in terms of the required INR to obtain a ROC curve with a P_{Det} = 0.9 for a P_{Fa} = 0.1. In order to get reliable results, performance has been calculated with 2^{15} Monte-Carlo simulations. Values of f_0 and φ_0 have been selected to be different and random for every simulation.

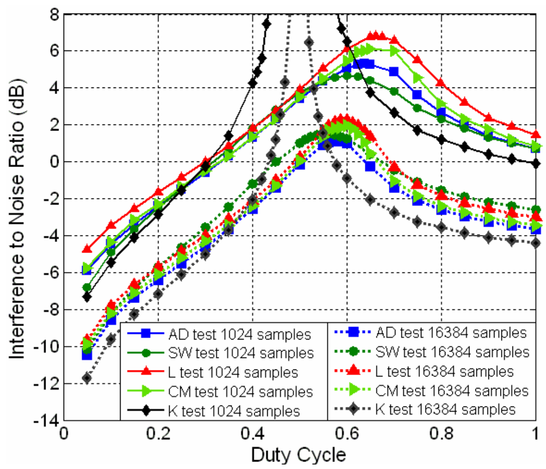

Fig. 2. Normality test performance in the detection of a pulsed sinusoidal interference of 1024 (solid line) and 16384 (dotted line) samples as a function of signal's duty cycle.

EDF normality tests perform similarly in the detection of a pulsed sinusoidal interfering of duty cycle around 0.5 with improving performance as the sample size increases. SW test has the better performance for a sample size of 1024 while AD test does for the sample size of 16384. Kurtosis test outperform for duty cycle values different from 0.5, while for 0.5 value has a blind spot [2-5].

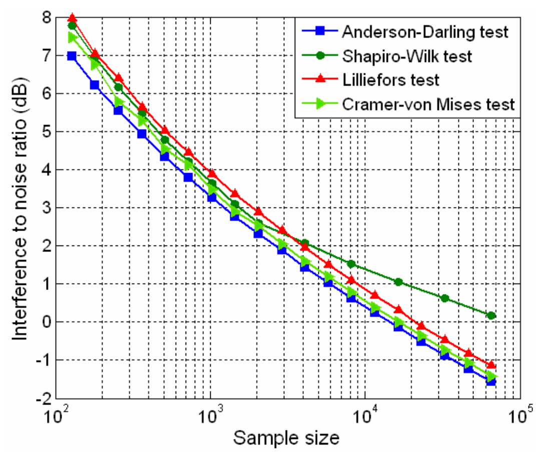

Fig. 3. Normality test performance in the detection of a 0.5 duty cycle pulsed sinusoidal interference as a function of the signal's sample size.

In fig. 3 the performance of this four normality tests in the detection of a pulsed sinusoidal signal of exactly 0.5 of duty cycle is compared as a function of the sample size. In this case, AD, L and CM tests follow almost the same trend while SW test has a different trend for sample length of 4096 and above due to averaging of the shorter sample length sets.

IV.2 PRN Signal

Kurtosis test achieves the best performance in the detection of this kind of interfering signal performing better than all EDF based test in the detection of PRN (fig. 4). Best EDF test performance is obtained with the AD test followed by the CM test. SW test performance fails for high number of samples due to averaging.

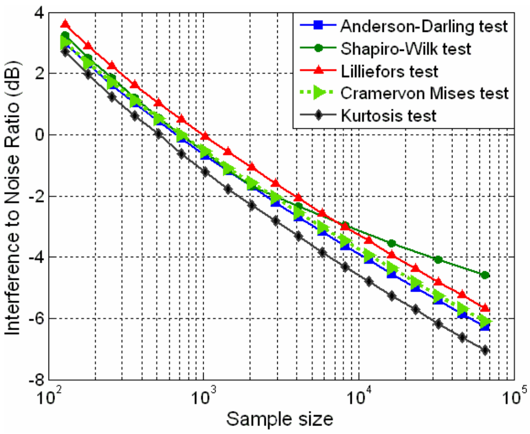

Fig. 4. Normality test performance in the detection of a PRN interference as a function of the signal's sample size.

V. CONCLUSIONS

As it has been shown, Kurtosis is the best RFI detection algorithm for both interfering signals, although it is known that it has a blind spot for sinusoidal interfering signals of 0.5 duty cycle.

The four EDF based normality tests have a similar performance in the detection of the presented interfering signals. For a low number of samples AD and SW tests work better that the CM and L test, but as number of samples increases, SW test performance gets worse in front of CM and L tests, as SW test must be averaged above 4096 samples length to obtain a correct performance. AD test has a better performance than L test as gives more weight to the tails, and P_{det} is located in the tails (0.9).

In summary, the Kurtosis is the best RFI detection algorithm for PRN and sinusoidal interfering signals, although it has a blind spot for sinusoidal signals of 0.5 of duty cycle. A complementary normality test that covers this blind spot can be the AD test that has a very good performance for all the studied sample sizes. The performance of the detection tests improves with the number of signal samples and depends on the duty cycle of the pulsed RFI.

VI. ACKNOWLEDGEMENTS

This work, conducted as part of the award "Passive Advanced Unit (PAU): A Hybrid L-band Radiometer, GNSSReflectometer and IR-Radiometer for Passive Remote Sensing of the Ocean" made under the European Heads of Research Councils and European Science Foundation EURYI (European Young Investigator) Awards scheme in 2004, was supported by funds from the Participating Organisations of EURYI and the EC Sixth Framework Programme, and to the Spanish National Research and EU FEDER Project AYA. 2008-05906-C02-01-ESP.

REFERENCES

[1] B. Guner, J. T. Johnson, and N. Niamswaun, "Time and frequency blanking for radio frequency interference mitigation in microwave radiometry". *IEEE Trans. Geosci. Rem. Sens.*, Vol. 45(11), pp. 3672-3679, Nov. 2007.

[2] C. S. Ruf, C. S., S. Gross, and S. Misra, "RFI Detection and Mitigation for Microwave Radiometry with an Agile Digital Detector", *IEEE Trans. Geosci. Rem. Sens.*, Vol. 44, No. 3, pp. 694-706, Mar. 2006.

[3] R. D. De Roo, and S. Misra, "Effectiveness of the Sixth Moment to Eliminate a Kurtosis Blind Spot in the Detection of Interference in a Radiometer". *IEEE Geosci. Rem. Sens. Symp.*, Vol. 2, pp. 331-334, Jul. 2008.

[4] J. T. Johnson, and L. C. Potter, "Performance Study of Algorithms for Detecting Pulsed Sinusoidal Interference in Microwave Radiometry". *IEEE Trans. Geosci. Rem. Sens.*, Vol. 47(2), pp. 628-636, Feb. 2009.

[5] B. Guner, M. T. Frankford, and J. T. Johnson, "A study of the Shapiro-Wilk test for the detection of Pulsed Sinusoidal Radio Frequency Interference". *IEEE Trans. Geosci. Rem. Sens.*, Vol. 47(6), pp. 1745-1751, Jan. 2009.

[6] H. W. Lilliefors, "On the Kolmogorov-Smirnov test for normality with mean and variance unknown", *J. Amer. Stat. Assn.*, Vol. 64, pp. 387-389, Jun. 1967.

[7] R. B. D'Agostino, and Stephens, "Godness-of-Fit Techniques" 1st Ed.; Marcel Dekker: New York, United States. 1986; pp. 122-133.

[8] J. P. Royston, "An extension of Shapiro and Wilk's W test for normality to large samples". *Appl. Statist.*, Vol. 31, pp. 115-124. 1982.