95% upper limit comparison between the cross-spectrum and the spectrum average with 5 radio-telescopes

A. Baudiquez, E. Rubiola, F. Vernotte
FEMTO-ST, Time and Frequency Department
Observatory THETA, UBFC
Besançon, France
antoine.baudiquez@femto-st.fr

E. Lantz
FEMTO-ST, Département d'Optique P.M. Duffieux
UBFC
Besançon, France

***Abstract*—This paper compares the 95% upper limit on the signal level between 2 estimators: the cross-spectrum and the spectrum average. This is generalized in the case where we have more than 2 radio-telescopes and an example is given in the case of 5 radio-telescopes.**

***Index Terms*—Bayesian statistics, confidence interval, cross-spectrum, spectrum average, QR decomposition, characteristic function, probability density function.**

I. Introduction

This paper aims to compare the efficiencies of the cross-spectrum (c-s) estimator and the spectrum average (s.a) estimator generalized in the case where more than 2 radio-telescopes (RTs) measure a signal. For that a comparison of the 95% Bayesian upper limit is computed. Consequently the estimator which is the more stringent will be the more efficient. To assess this confidence interval the probability density function (pdf) of both estimators knowing the parameters σ_S^2, the sought signal level, and σ_N^2, the noise level of each RT (assumed to be the same), is required. This is generalized in the case where 5 RTs, e.g. Effelsberg (Ge), Cagliari (It), Nançay (Fr), Jodrell Bank (UK) et Westerbork (NL) presently included in the Large European Array for pulsars (LEAP), are involved. This signal can be a red noise originated from gravitational waves on the line of sight from the millisecond pulsars observations. Our previous paper [1] shows that the variance-gamma distribution is the exact solution of the c-s pdf. However this is no longer the case for more than 2 RTs that is why a description included the Fourier transform of the characteristic function is proposed as a solution. Then the inverse problem is computed to assess the confidence interval on the signal level σ_S^2, and 2 sets of measurement are given in order to compare both estimators. Nonetheless, whereas both estimators give nearly the same 95% upper limit what about when we increase the number of degree of freedom, i.e. the number of RTs ?

II. Statement of the estimators

Let us consider at a given Fourier frequency the spectra of the i^{th} RT as

$$X_i = X_S + X_{N_i} \tag{1}$$

where X_S is a normal complex rv of variance σ_S^2 i.e. the signal level and X_{N_i} are uncorrelated normal complex rv of variance σ_N^2 i.e. the white noise level. Then the s.a and c-s estimators are respectively

$$\begin{cases} S_{sa} = \mathcal{R}\left[N\frac{\sum_i^n X_i}{\sigma_{N_i}^2}\right]^2 + \mathcal{I}\left[N\frac{\sum_i^n X_i}{\sigma_{N_i}^2}\right]^2 - N \\ S_{cs} = \langle X_i \cdot \tilde{X}_j\rangle_m \qquad \text{with} \quad i \neq j \end{cases} \tag{2}$$

where $\langle\cdot\rangle$ stands for the m average over the different combinations of RTs with $m = \binom{n}{2}$ and $\tilde{\cdot}$ stands for the complex conjugate of the quantity which is below. Moreover $\mathcal{R}[\cdot], \mathcal{I}[\cdot]$ stands respectively for the real and the imaginary parts, appearing through the Fourier transform operation, of the quantities within the brackets whereas n is the number of RTs. The N factor is the noise variance ponderation normalization corresponding to

$$N = \left(\sum_i^n \sigma_{N_i}^{-2}\right)^{-1}. \tag{3}$$

These estimators are **unbiased**, as shown in [2], since their expectation is the sought signal level σ_S^2.

III. Probability Density Function

A. Spectrum Average

The s.a estimator leads to the following χ^2 distribution with 2 degrees of freedom resulting from the real and imaginary part of the spectrum,

$$p(S_{sa}|\sigma_S^2) = \frac{e^{-\frac{S_{sa}+N}{2\sigma^2}}}{2\sigma^2} \tag{4}$$

where,

$$\sigma^2 = N + \sigma_S^2. \tag{5}$$

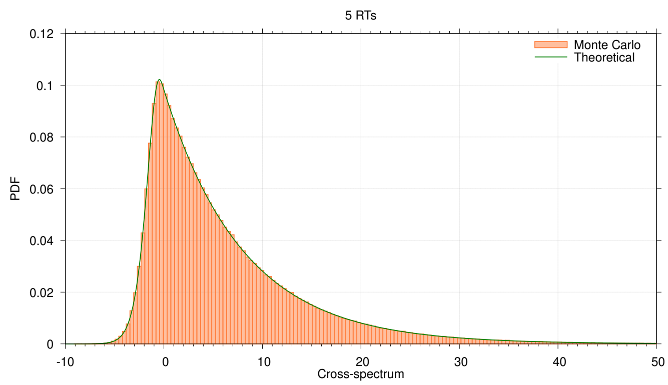

Fig. 1: Comparison of the empirical and theoretical PDF of the cs for 5 RTs where the variances are $\sigma_S^2 = 3$ and $\sigma_N^2 = 5$.

B. Cross-spectrum

The c-s estimator leads to the variance-gamma (VΓ) distribution for 2 RTs as established in [1] but for more than 2 RTs it is no longer the case. Having no exact solution known nowadays, we give an approximation of it. First we perform a QR decomposition by using the Householder transformation in an orthogonalization process. Second we compute the eigenvalues l_i of the resulting components and obtain a linear combination of χ^2 distribution as followed,

$$S_{cs} = \sum_i^n l_i \chi_k^2 \tag{6}$$

where k is the number of degree of freedom of each eigenvalue. It can be shown that the white noises induce negative eignvalues where the signal implies a positive one. Then we define the characteristic function of χ_k^2 as

$$\phi_{cs_i}(t) = (1 - 2jl_i t)^{-k/2} \tag{7}$$

where j is the imaginary unit complex number and we apply a variable change of $-t$ for the negative eignvalues. The χ^2 distribution being independent, the characteristic function of the c-s becomes $\phi_{cs}(t) = \prod_i^n \phi_{cs_i}(t)$. The probability density function of the c-s is finally define as

$$p(S_{cs}|\sigma_S^2) = \frac{1}{2\pi} \int_{\mathbb{R}} \exp^{-jtS_{cs}} \phi_{cs}(t) dt. \tag{8}$$

Figure 1 shows that theoretical probability density function (PDF) fits very well the histogram obtained by 10^7 Monte Carlo simulations for 5 RTs. The variance of each white noise is the same $\sigma_N^2 = 5$ whereas the signal level is $\sigma_S^2 = 3$.

IV. Comparison of the 95% upper limit

We have set the direct problem, i.e. the statistics of the s.a or c-s knowing the signal level and noise level (which is assumed to be known). Now we have to deduce the inverse problem from the direct problem, i.e. the statistics of the signal level knowing the s.a or c-s estimate and this is the Bayes theorem which enables us to establish this link. A full descrition of the direct and inverse problem will be detailed in the paper.

TABLE I: Measurement set for the outputs of each RT (5 in total) where $\sigma_S^2 = 3$ and $\sigma_N^2 = 5$.

	measurement set 1		measurement set 2	
	Real part	Imaginary part	Real part	Imaginary part
X_1	-3.8947	-1.7994	-0.1494	8.9456
X_2	-5.0950	-3.9125	-0.5275	4.4659
X_3	-25133	-5.5431	0.2176	5.7742
X_4	0.6433	-1.9566	1.6044	3.2146
X_5	-0.2294	-2.5738	-0.5284	0.3563

Let us give an example of such a process. We set the number of RTs to 5 and the variances of the signal and noise are respectively $\sigma_S^2 = 3$, $\sigma_N^2 = 5$. Then we produce 2 sets of random measurement with these parameters, shown in Table I. We obtain respectively for the first and second measurement set the c-s values $S_{cs_1} = 13.2256$ and $S_{cs_1} = 18.5636$. It leads for the first set to the 95% upper limit on the signal σ_S^2 following value, 58.8 for the s.a and 12.3 for the c-s. Whereas the second set gives us 66.6 for the s.a and 14.7 for the c-s. These results show that the c-s estimator give a more stringent upper limit and is then more efficient than the s.a estimator.

V. Conclusion

The efficiency of both estimators, the spectrum average versus the cross-spectrum, is highlighted through the comparison of the 95% Bayesian upper limit. We found an advantage not the least for the cross-spectrum estimator especially that increases with higher degree of freedom which means with higher number of radio-telescopes. This estimator has also a particular interest since it is not biased whereas the spectrum average estimator is. However both estimators required a complete knowledge of the noise introduced by the measurement instruments.

The generalization to more than 2 radio-telescopes implies a numerical integration of the characteristic function and has no longer an exact density solution to the cross-spectrum probability density function. However whereas it was wiser to compute both estimator for 2 radio-telescopes since one can be stringent than the other depending on the measurement, it is clearly better to take into account the cross-spectrum for higher degrees of freedom.

Acknowledgement

This work was partially funded by the ANR Programmes d'Investissement d'Avenir (PIA) Oscillateur IMP (Project 11-EQPX-0033) and FIRST-TF (Project 10-LABX-0048).

References

[1] A. Baudiquez, E. Lantz, E. Rubiola, and F. Vernotte. Cross-spectrum measurement statistics: uncertainties and detection limit. *IEEE Trans. Ultras. Ferroelec. Freq. Contr.*, pages 1–1, june 2020.

[2] E. Rubiola and F. Vernotte. The cross-spectrum experimental method. arXiv:1003.0113v1, March 2010.