DATA PREPROCESSING FOR DECAMETRE WAVELENGTH EXOPLANET DETECTION: AN EXAMPLE OF CYCLOSTATIONARY RFI DETECTOR

R.Weber[1], P.Zarka[2], V. B. Ryabov[3], R.Feliachi[1], J.M. Grießmeier[2], L. Denis[4], R. V. Kozhyn[5], V. V. Vinogradov[5], P. Ravier[1]

[1]LESI – Polytech'Orléans, Université d'Orléans ,Orléans, France, *rodolphe.weber, rym.feliachi, philippe.ravier@univ-orleans.fr*

[2]LESIA, Observatoire de Paris, CNRS, UPMC, Université Paris Diderot, France *philippe.zarka, jean-mathias.griessmeier@obspm.fr*

[3]Complex Systems Dep., Future University Hakodate, Hokkaido, Japan *riabov@fun.ac.jp*

[4]Station de radioastronomie de Nançay, Observatoire de Paris, CNRS/INSU Nançay, France *laurent.denis@obs-nancay.fr*

[5]Institute of Radio Astronomy, National Academy of Sciences of Ukraine, Microwave Electronic Dep., Kharkiv, Ukraine *vin@radar.kharkov.com, krv@ira.kharkov.ua*

ABSTRACT

The relatively high contrast between planetary and solar low-frequency radio emissions suggest that the low-frequency radio range may be well adapted to the direct detection of exoplanets. Detection is based on temporal and spectral power integration to remove statistical fluctuations of the sky background noise. However, strong limitations, in particular due to radio frequency interference (RFI), may impair such detection. Indeed, if not detected, a RFI may be interpreted as an exoplanet signal or may conceal it. After a short summary of the detection principle in radio astronomy, we describe the algorithms proposed to remove RFI. First method, designed for rather strong RFI, is based on power threshold detection. The second one is based on the cyclostationary properties of amplitude modulation RFI. The main advantages of the latter are the detection of 1) weak RFI, 2) without confusion with potential signal. The statistics of this RFI detector are derived as a function of the integration time and the interference to noise ratio.

1. INTRODUCTION

The main limitation to the direct detection of exoplanets in the infra-red or visible range is the very high contrast between star and planet luminosity, together with their small angular separation as measured from stellar distance.

However, in the low-frequency radio range (10 MHz to 40 MHz), the relatively high contrast between planetary and solar radio emissions suggests that direct exoplanet detection could be performed. In particular, it has been shown [1] that a specific class of "Jupiter-like" planets, called hot Jupiters, may produce very intense radio emissions due to electromagnetic interactions with their nearby star. The expected flux in the vicinity of these exoplanets can be 10^3 to 10^6 times higher than that of our Jupiter. Thus, in spite of the stellar distance, detection of such low frequency emissions from ground-based radio telescopes seems possible.

In order to prove it, we have thus started an observational project which associates a very large decametre array in Ukraine and a modern digital spectrometer and waveform receiver (see Figure 2 and Table 1). The technical part of this project is described in the present paper.

However, this detection is not straightforward. Indeed, decametre radio observations suffer strong limitations due to severe sky background noise, ionospheric perturbations, and natural or man-made radio frequency interferences (RFI). The scope of this article is to describe the strategies chosen to overcome these limitations. In particular, we will focus on the digital signal pre-processing techniques used to clean the data before any exoplanet detection attempt.

In section 2, the principle and limitation of detection in low frequency radio astronomy are briefly described. The main characteristics of the radio telescope and the receiver used are also given. In section 3, the RFI mitigation strategy is presented. The case of low level RFIs is discussed in section 4 where a specific detector based on cyclostationarity is proposed. Its statistical properties are defined and its performances are analysed.

2. PRINCIPLE AND LIMITATIONS OF THE DETECTION PROCESS

From a radio astronomer point of view, an ideal world would be a centred Gaussian one. Thus, the ideal measured signal would be written as:

$$s(n) = \underbrace{N(0, p_{sys})}_{\text{system noise}} + \underbrace{N(0, p_{sky})}_{\text{skybackground noise}} + \underbrace{N(0, p_{source})}_{\text{source of interest if any}} \qquad (1)$$

where *N(0,p)* are centred Gaussian noise with variance *p*.

With this model, only second order statistics are needed and raw data provided by radio telescopes can be limited to time-frequency (t-f) representations of the measured power:

$$S_{\tau,B}(t,f) = \iint_{B,\tau} |s(n)|^2 \, dn \qquad (2)$$

where B is the frequency bandwidth with central frequency, f, and τ is the time interval centred at the instant, t.

From Equ.(1) and Equ.(2), the $S_{\tau,B}(t,f)$ statistics are easily derived. Its standard deviation, σ, is called the measurement sensitivity. In radio astronomy and, in particular, for deca-

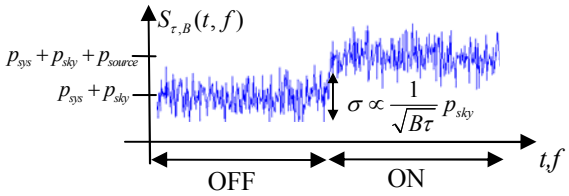

Figure 1 – Detection and sensitivity. During the OFF (respectively ON) procedure, the source is supposed to be out of (resp. in) the radio telescope beam. A source can be detected when the sensitivity, σ, is less than its mean power p_{source}. The sensitivity can be adjusted through the integrated bandwidth B, the integration time τ and the sky background noise power, p_{sys}. In our case, the expected exoplanet flux is between 5.10^{-2} *Jy* ($1Jy=10^{-23}$ $W/m^2/Hz$) and 5 *Jy*, to be compared with the sky background flux which is around 10^7 *Jy* [1].

Table 1 – Overview of the digital receiver designed for the project

Digital Receiver for exoplanet detection	
Sampling	2 channels @66MHz, 16 bits, baseband
Embedded processing	direct recording of waveform dynamic power spectrum (4096 channels) dynamic cross-spectrum (4096 channels)
recording	Waveform snapshots up to 1Gb Up to 24 MHz spectrum anywhere within the band of 0-33 MHz, spectral resolution 4kHz, current minimum temporal resolution 2 ms

metre wavelength we can consider that $p_{source} << p_{sys} << p_{sky}$. Thus, the sensitivity verifies the following relation:

$$\sigma \propto \frac{1}{\sqrt{B\tau}} p_{sky} \qquad (3)$$

As shown in Figure 1, assuming that $p_{sys} + p_{sky}$ is obtained by a calibration procedure, a source can be detected when the sensitivity σ is less than p_{source}. To improve this sensitivity, several options are possible:

- Using a large effective area telescope, *Ae*. Indeed, p_{sky} is inversely proportional to *Ae*. Figure 2 gives some characteristics of the radio telescope used for our observations. Its *Ae* will reduce p_{sky} by 50 dB.
- Frequency smoothing by enlarging the bandwidth B. A specific broadband receiver has been designed for the project. Its characteristics are given in Table 1. If the full bandwidth can be used, the sensitivity will gain 36 dB more.
- Temporal smoothing by adjusting the integration time τ.

In practice, other limiting factors may degrade the ultimate sensitivity:

- ionospheric propagation effects (strong scintillations for frequencies below f<30 MHz). Mitigation of these effects requires a specific ON-OFF procedure using the radio telescope multi-beam capability (description is beyond the scope of this article).
- natural RFI. They are generally broadband and impulsive (e.g. lightning).

Figure 2 – Overview of the UTR-2 radio telescope, located near Kharkov (Ukraine). It is a T-shape phased array with 2040 dipoles. Its effective area is around ~140000 m2. Its frequency range of operation is 7 MHz to 32 MHz.

- manmade RFI. They are most often narrowband amplitude modulation (AM) signals.

Sensitivity optimisation in the presence of RFIs is based on their detection and blanking (*i.e.* t-f blocks, $S_{\tau,B}(t,f)$, must be discarded when polluted) before any further t-f power integration. If not detected, a RFI may be interpreted as an exoplanetory signal or may conceal it.
So, the t-f resolution becomes one of the key parameters of RFI mitigation. In [2], a study on this question is proposed. Practically, there is a compromise in terms of t-f resolution between the hardware complexity and the RFI t-f characteristics. For the project, a 4 kHz spectral resolution and a 2 ms temporal resolution have been selected. Table 1 summarizes the other characteristics of the receiver.
The second key parameter for efficient RFI mitigation is the performance of RFI detection algorithms. This topic is a growing subject of research in radio astronomy [3,4], but not only [5]. In the following, our RFI detection strategy will be described.

3. RFI DETECTION STRATEGY

By including the RFI, the received signal becomes:

$$s(n) = b(n) + rfi(n) \qquad (4)$$

Where *b(n)* is the clean Gaussian signal defined by Eq. 1 and *rfi(n)* is the RFI, only defined for the moment by its mean power p_{rfi}. Let us define the interference to noise ratio:

$$inr = \frac{p_{rfi}}{p}$$

where $p = p_{sky} + p_{sys}(+p_{source})$ which are defined by Eq. 1.

Given that the RFI are impulsive or narrow band, they will pollute horizontal or vertical lines of $S_{\tau,B}(t,f)$ blocks on the t-f power plane (an example is given on Figure 3 for strong RFI).

A first algorithm has been developed to remove the polluted (t-f) blocks. The principle is based on a robust threshold

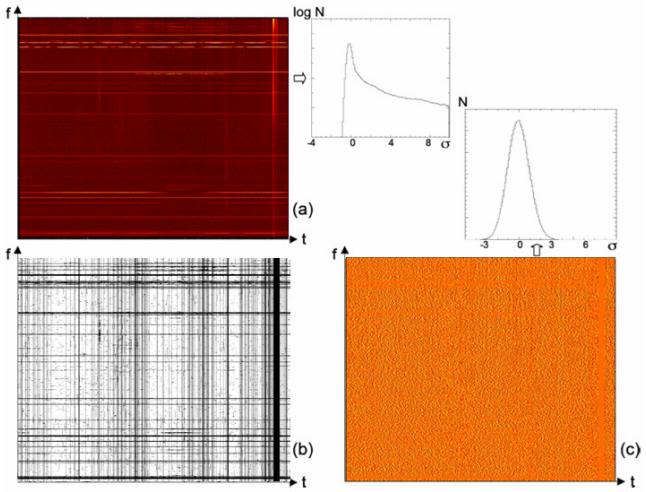

Figure 3 – (a) Dynamic spectrum of the galactic background covers a time interval of 100 s (time resolution 20 ms) and the frequency band 14–28 MHz (spectral resolution 7 kHz). The normalized distribution plotted on the right has a high–intensity tail corresponding to RFI. (b) is the map resulting from an automated offline recognition of spurious t-f blocks (in black). (c) = (a)×(b) is the original dynamic spectrum with spurious blocks masked out. Statistical background fluctuations are clearly visible due to the stretched dynamic range of the image. The corresponding distribution of intensities is plotted above it, and is nearly perfectly Gaussian, indicating that all RFI have been effectively suppressed down to the 3σ level.

detector. This kind of method is generally limited by the precision of the reference estimation [5,6]. The complete procedure is described in [7,8] and Figure 3 illustrates some of the steps. In this way, RFI above 3σ can be removed but it is not sufficient to reach the sensitivity which is necessary for exoplanet detection. For lower *inr* RFI detection, the two kind of RFI will be processed differently.

The weak impulsive case will be processed during the exoplanet detection phase. Indeed, to enhance the exoplanet detection, we must correct the effects of propagation of radio signals through the interstellar medium, and especially the dispersion of broadband short pulses. Compensation of the pulse distortion (dedispersing) can be performed by introducing precalculated time delays to the output of each receiver frequency channel. An example of application of this procedure to pulsar signals detected at UTR-2 is shown in Figure 4. Inversely, dedispersing will spread impulsive RFI power over time. Unfortunately, it will useless for narrowband AM RFIs. The next section will describe a narrowband AM detector based on the cyclostationary properties of such RFI.

4. CYCLOSTATIONARY DETECTOR

A signal is cyclostationary at order 2 if its correlation function is a periodic function of time [9]. From a spectral point of view, the following criterion can be derived:

$$C^{\alpha}(\tau) = \lim_{N\to\infty} \frac{1}{N+1} \sum_{n=-N/2}^{N/2} s(n+\frac{\tau}{2}) s(n-\frac{\tau}{2}) e^{-i2\pi\alpha n} \quad (5)$$

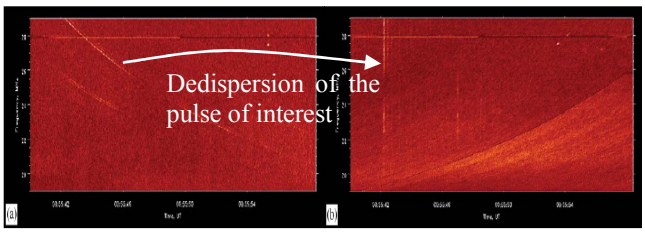

Figure 4 – (a) raw dynamic spectrum after the first step RFI detection algorithm described in Figure 6. The pulses (observed from a pulsar) are drifting from high to low frequencies; (b) dedispersed dynamic spectrum. The pulse is now vertical and can be easily detected by integrating all the frequencies at a given time.

where α is the frequency related to the periodicity (α is also called the cyclic frequency). If *s(n)* is stationary, $C^{\alpha}(\tau)$ will be null. Otherwise, if *s(n)* is cyclostationary with cyclic frequency α, then $C^{\alpha}(\tau)$ will be non zero. In other words, the quadratic signal $s(n+\frac{\tau}{2})s(n-\frac{\tau}{2})$ will present a spectral line at frequency α.

Most of the signals generated by telecommunication systems present such cyclostationarity due to their modulation. In particular the following AM modulation RFI is cyclostationary:

$$rfi(n) = \sqrt{p_{rfi}}\sqrt{2}a(n)\cos(2\pi f_o n + \varphi) \quad (6)$$

where *a(n)* is stationary signal with unit mean power. This AM signal has a cyclic frequency equal to $2f_o$ and Eq. 5 can be rewritten as [9]:

$$C^{2f_0}(\tau) = R_a(\tau)\frac{p_{rfi}}{2}e^{\mp i2\varphi}$$

where $R_a(\tau)$ denotes the autocorrelation function of *a(n)*. If we limit the previous equation to $\tau=0$, we obtain the following criterion:

$$C = \left| \lim_{N\to\infty} \frac{1}{N+1} \sum_{n=-N/2}^{N/2} s(n)^2 e^{-i2\pi 2 f_0 n} \right| = \begin{cases} 0 \text{ if no AM modulation at } f_0 \\ p_{rfi}/2 \text{ if AM modulation at } f_0 \end{cases}$$

These two cases are called respectively H0 and H1 hypothesis and they are perfectly separated. Unfortunately, this performance is asymptotic. In practice, the number, *N*, of sample is limited. Thus, C is no more deterministic and its statistics must be studied to derive the new contrast between these two hypotheses. The next 2 sections will state the statistics for the following functions:

$$C_N^{\alpha} = \left| \frac{1}{N} \sum_{n=0}^{N-1} s(n)^2 e^{-i2\pi\alpha n} \right| \text{ with } \alpha=0 \text{ and } 2f_0.$$

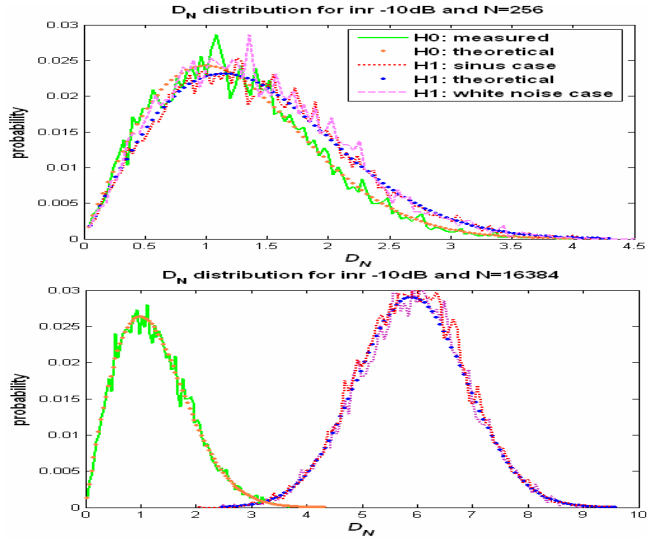

Figure 5 – D_N distribution comparison. For the theoretical distributions, Equ. 9 (respectively Equ. 10) is used for H0 (resp. H1) hypothesis. The measured histograms are based on 10000 trials of each configuration. Two extreme scenarios has be chosen: a pure sine wave (*a(n)* fully correlated) or a modulated white noise (no correlation between the *a(n)*). For low inr, the theory fits the simulations.

The case α=0 will be used to define a normalized criterion in section 4.3. Besides, the following assumptions will be used in the rest of the paper:

(A1) N is large enough to apply the central limit theorem.

(A2) The *inr* is low (*i.e inr <1*).

4.1 H0 hypothesis statistics

Under this hypothesis, *s(n)* is a *i.i.d* Gaussian noise with variance p. By using the assumption (A1), the statistics of C_N^{α} are easily derived:

$$C_N^0 : N(p, \frac{2p^2}{N})$$

$$C_N^{2f_0} : R(0, \frac{p^2}{N}) \text{ and } \begin{cases} E(C_N^{2f_0})_{H0} = p\sqrt{\frac{\pi}{2N}} \\ \text{var}(C_N^{2f_0})_{H0} = \frac{p^2}{N}(2 - \frac{\pi}{2}) \end{cases}$$

where $R(.,.)$ is a Raleigh-Rice distribution defined in the appendix. The first remark is that the mean of C_N^{2fo} is no more equal to zero.

4.2 H1 hypothesis

Under this hypothesis, *s(n)* is given by Equ. 4 and Equ. 6. The first following result is derived:

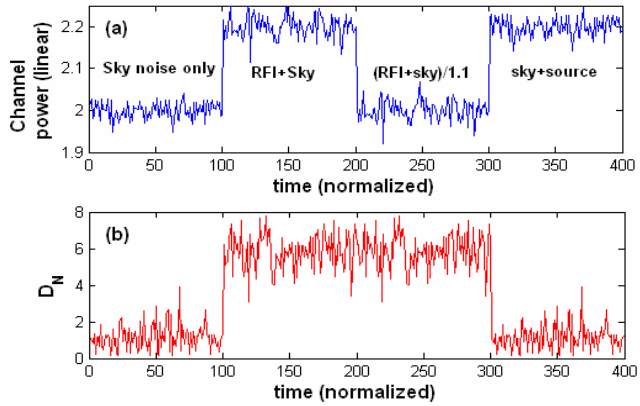

Figure 6 – Power detector vs. cyclostationary detector. One of the advantages of the cyclostationary detector is its insensitivity to power fluctuations. (a) In the example above, the *inr* is maintained at -10dB but the total power may fluctuate with time. It appears that the classical power detector may misinterpret at least 2 of the 3 last cases: 1) the (RFI+sky)/1.1 can not be detected, 2) if it rejects the (RFI+sky) case, it will also reject the (sky+source) case, 3) if it keeps the (sky+source) case, it will also keeps the (RFI+sky) case. (b) On the contrary, the cyclostationary detector D_N detects the RFI properly.

$$\begin{aligned} E(s^2) &= p + p_{rfi} \\ \text{var}(s^2) &\approx 2p^2 + 4p.p_{rfi} \end{aligned} \tag{7}$$

For the latter equation, we use the assumption (A2) to remove any contribution due the correlation between the *a(n)* samples. Then, with the assumption (A1) and the approximation given in Equ. 7, the statistics of C_N^{α} becomes:

$$C_N^0 : N(p + p_{rfi}, \frac{2p^2 + 4p.p_{rfi}}{N})$$

$$C_N^{2f_0} : R(\frac{p_{rfi}}{2}, \frac{2p^2 + p.p_{rfi}}{2N})$$

From these statistics, a normalized detector criterion is defined in the next section.

4.3 Normalized detection criterion

We define a normalized version of our criterion as:

$$D_N = \frac{\sqrt{N} C_N^{2f_0}}{C_N^0} \tag{8}$$

By using the following approximation, $C_N^0 \approx E(C_N^0)_{H_0 \text{ or } H_1}$, the theoretical D_N statistics becomes :

H0: *R(0,1)* under the H0 hypothesis (9)

$$\text{H1: } R(\frac{\sqrt{N} inr}{2(1+inr)}, \frac{1+2inr}{(1+inr)^2}) \tag{10}$$

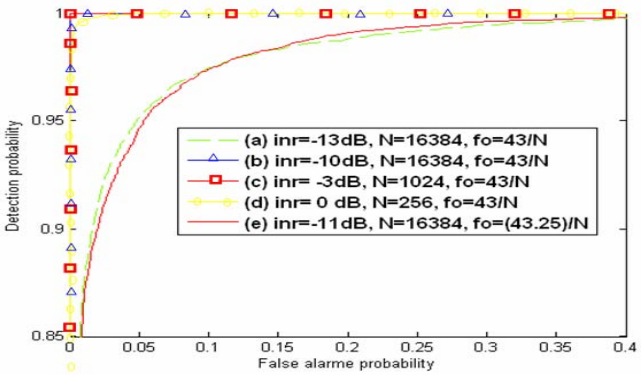

Figure 7 – Probability of detection vs. probability of false alarm for the detector presented in section 4.4. This simulation is based on 10000 runs of each scenario. In cases (a) to (d), f_o is chosen so that $2f_o$ is multiple of 1/N (*i.e.* the $2f_0$ spectral line is in the middle of the N bin FFT channel, no attenuation). In case (e), the $2f_0$ spectral line is between two N bin FFT channels, which is the worst case. The channel attenuation at nyquist frequency for a rectangular window is ~2 dB. Thus, (e) should be similar to (a).

In Figure 5, Monte-Carlo simulations are presented. For low *inr*, theoretical statistics from Equ. 9 and Equ. 10 are verified. One of interest of such detector is its insensitivity to power fluctuations (see Figure 6). From Equ. 9, the threshold can be easily derived and is equal to $\sqrt{-2\ln(pfa)}$ where *pfa* is the false alarm probability.

4.4 Final detection algorithm

From the previous analysis, we propose an operational detector which can be applied on complex (*i.e.* not real) channel with unknown carrier AM modulation RFIs:

1) Channelization of the signal coming from the radio telescope. The signal in each channel is supposed to be complex. This process is done in real time by the digital receiver (see Table 1). M is the number of channels.
2) $s_r(n)$ and $s_i(n)$ are respectively the real and imaginary part of the signal from one of the M channels. We compute the Fourier transform, $FFT_N^{r,i}(f)$, over N samples on both $s_r(n)^2$ and $s_i(n)^2$.
3) According to a given threshold ξ, we will consider that a RFI is present on this channel if:

$$\exists k > 0 \quad / \quad \frac{\sqrt{N}\left|FFT_N^r(k)\right|}{FFT_N^r(0)} + \frac{\sqrt{N}\left|FFT_N^i(k)\right|}{FFT_N^i(0)} \geq \xi$$

Figure 7 shows performance measurements of such detector. To be optimally detected, the RFI carrier frequency should be multiple of 1/2N. If not, the performance in term of *inr* will decrease. For example, with a rectangular windowing the maximum degradation will be around 2 dB.

5. CONCLUSIONS

In the context of exoplanet detection, we have described the algorithms proposed to remove RFI from a time-frequency power plane. In particular, we have defined an original method based on the cyclostationary properties of the RFI. This cyclostationary detector is useful for low level RFI. In this case, it can outperform classical power detectors. A blind detector is also proposed. Its operational implementation (on the receiver used for the exoplanet detection) is still under study.

6. APPENDIX

The Raleigh-Rice distribution, $R(\mu, \lambda^2)$, is defined by:

$$R(r/\mu,\lambda^2) = \frac{r}{\lambda^2} e^{-\frac{r^2+\mu^2}{2\lambda^2}} I_0\left(\frac{\mu r}{\lambda^2}\right) \quad r \geq 0$$

where $I_0(.)$ is the modified Bessel function of order 0.

Its mean is given by :

$$E(r) = \lambda\sqrt{\frac{\pi}{2}} e^{-\frac{\mu^2}{4\lambda^2}} \left[\left(1+\frac{\mu^2}{2\lambda^2}\right) I_0\left(\frac{\mu^2}{4\lambda^2}\right) + \frac{\mu^2}{2\lambda^2} I_1\left(\frac{\mu^2}{4\lambda^2}\right)\right]$$

where $I_1(.)$ is the modified Bessel function of order 1.

And its variance is given by:

$$\mathrm{var}(r) = \mu^2 + 2\sigma^2 - E(r)^2$$

ACKNOWLEDGMENTS

This project is supported by the french national research agency (ANR) under contract number NT05-1_42530, RADIO-EXOPLA.

REFERENCES

[1] P. Zarka, "Plasma interactions of exoplanets with their parent star and associated radio emissions", *Planetary and Space Science*, doi:10.1016/j.pss.2006.05.045, 2006

[2] V. Clerc, R. Weber, L. Denis, C. Rosolen, "high performance receiver for rfi mitigation in radio astronomy: application at decameter wavelengths" in *Proc. EUSIPCO 2002,* Toulouse, France, 2002

[3] P.A. Fridman and W.A. Baan, "RFI Mitigation Methods in Radio Astronomy", *Astronomy & Astrophysics*, vol.378, pp. 327-344, 2001.

[4] P.A. Fridman, "RFI Excision using a higher order statistics analysis of the power spectrum", *Astronomy & Astrophysics*,vol. 368, pp. 369.376, 2001.

[5] M. Ghozzi, M.Dohler, F.Marx, J. Palicot, "Cognitive radio: methods for the detection of free bands", *C.R. Physique* 7,pp. 794-804, 2006.

[6] R.Weber, C.Viou, A.Coffre, L.Denis, P. Zarka and A.Lecacheux, "DSP Enabled Radio Astronomy: Towards IIIzw35 Reconquest", *JASP*, Vol. 16, pp. 2686-2693, 2005.

[7] P. Zarka, and *al.* 'Ground-based high sensitivity radio astronomy at decameter wavelengths'. In: Rucker, H.O., Bauer, S.J., Lecacheux, A. (Eds.), *Planetary Radio*, 1997

[8] Ryabov, V. B., P.Zarka, and B. P. Ryabov, "Search of exoplanetary radio signals in the presence of strong interference : Enhancing sensitivity by data accumulation", *Planet. Space Science*., 52, 1479-1491, 2004.

[9] W. A. Gardner, "Statistical Spectral Analysis - A Non-probabilistic Theory". New Jersey: *Prentice Hall*, 1988.