Available online at www.sciencedirect.com
SCIENCE DIRECT®
Planetary and Space Science 52 (2004) 1479–1491
Planetary and Space Science
www.elsevier.com/locate/pss

Search of exoplanetary radio signals in the presence of strong interference: enhancing sensitivity by data accumulation

V.B. Ryabov[a,*], P. Zarka[b], B.P. Ryabov[c]

[a]Compelx Systems Department, Future University Hakodate, 116-2 Kamedanakano-cho, Hakodate-shi, Hokkaido 041-8655, Japan
[b]Laboratoire d'Etudes Spatiales et d'Instrumentation en Astrophysique (LESIA), Paris Observatory, 92195 Meudon, France
[c]Institute of Radio Astronomy, 4 Krasnoznamennaya St., 61002 Kharkov, Ukraine

Accepted 3 September 2004

Abstract

We develop a statistical approach aimed at the detection of weak sporadic pulses on the noise background. The results are applied to modeling the observational time series where pulsed radio emissions have to be recognized against the sky background fluctuations. The proposed methodology demonstrates the efficiency of using the statistics of peak values (integrated tail of probability distribution function of the intensity) for the purpose of signal detection. It is established that the highest sensitivity is reached with this method at low values of the filling factor (duty cycle) of the pulsed signals. If in addition the pulses have sufficiently high intensity, the discussed approach performs better than simple integration over the observational time. Then we discuss the possibility of detecting radio pulses from exoplanetary magnetospheres, especially from known "hot Jupiters" found by radial velocity measurements in the visible and we report our results from extensive observations of several candidate exoplanets with the world largest decameter telescope UTR-2. Although no detection of pulsed emission from exoplanets has been found to date, the analysis demonstrates the feasibility of detection with more stable receivers and longer observational time.

Keywords: Exoplanets; Radio-bursts; Jupiter; Signal detection

1. Introduction

The auroral regions of magnetized planets in the Solar system are known to be the sources of intense radio emissions, sometimes comparable in brightness temperature to those generated by the Sun during flare events. Of all the known planets, Jupiter is perhaps the most familiar and extensively studied source of powerful radiation that reaches highest intensity, in particular, during the so-called S-burst storms generated in the decameter (DAM) wave band. "S" stands for "short" to account for millisecond timescale variability of such events and to distinguish them from other types of Jovian DAM emissions (L-bursts) whose characteristic timescales are of the order of seconds to minutes. Due to the sharp beaming properties of the source of S-radiation, the bursts can be detected with ground-based instruments only about 10% of the time when Jupiter, its satellite Io, and the Earth are in specific geometrical configurations constrained by the location and radiation pattern of the emitting area in the magnetosphere of Jupiter.

From the point of view of a distant observer studying the Sun and its surroundings, Jupiter would be the brightest object in the DAM wave band during a typical S-burst storm except, maybe, at rare occasions when a solar type III emission is generated at the same time. However, even when such a coincidence happens, there are ways to distinguish planetary radio emissions from those of the parent star through either modulation by

*Corresponding author. Tel.: +81 138 34 6129; fax: +81 138 34 6301.
E-mail address: riabov@fun.ac.jp (V.B. Ryabov).

0032-0633/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.pss.2004.09.019

the planet's rotation or polarization measurements, as it had been discussed in our earlier work (Zarka et al. (1997) to be referred hereinafter as Paper I) and also in Bastian et al. (2000). This line of reasoning put forward the conjecture that planetary systems around nearby stars (so-called exoplanets) can be detected in principle with the largest radio telescopes like, e.g., UTR-2 in Ukraine, operating in the DAM wave band between 8 and 32 MHz, or the very large array (VLA) in New Mexico, at higher frequencies, from 74 MHz to 50 GHz. Up to now, several attempts aimed at the detection of radio emissions from exoplanets have been reported, but all revealed unsuccessful (Paper I; Bastian et al., 2000). Despite the absence of positive results achieved so far, the search for radio emissions from magnetized extra-solar planets has still a high priority in decameter radio astronomy, especially in view of the new perspectives that LOFAR will open. Without doubt, more extensive observations and refined data analysis with already existing instruments are also worth trying (Farrell et al., 1999; Zarka et al., 2001).

On another hand, the absence of detection from earlier observations indicates the necessity of reassessing and possibly revising the existing hardware involved in these observational campaigns as well as the data processing algorithms used for the search of exoplanetary signals on the cosmic noise background. From a general point of view, the key issue to be addressed is the problem of sensitivity or in other words, the feasibility of weak radio bursts detection with ground-based instruments. Paper I used the following simple formula for estimating the maximal distance $d_{\max}$ (in parsecs) to an exoplanet whose radio emission can be detected from Earth

$$d_{\max} = 5 \times 10^{-6} \left(\frac{I}{N\sigma} \right)^{1/2}, \quad (1)$$

where I is the intensity of the planetary radio bursts normalized to a distance of 1 A U (for example, in the case of Jupiter $I \equiv I_J \approx 10^{-16}\,\mathrm{W\,m^{-2}\,Hz^{-1}}$ is the peak intensity of S-bursts from a 1 A U range), σ is the level (standard deviation) of the sky background fluctuations

$$\sigma = \frac{2kT_s}{A_e\sqrt{\Delta f \Delta t}}, \quad (2)$$

N is the number of standard deviations defining the confidence level of the detection, $k = 1.38 \times 10^{-23}\,\mathrm{JK^{-1}}$ is the Boltzmann constant, T_s the sky background temperature, A_e the effective area of the radiotelescope used, Δf the spectral bandwidth of observations, and Δt, their integration time. Under realistic assumptions on the values of the parameters at the largest existing decameter radiotelescope, UTR-2 ($T_s = 30\,000$ K, $A_e = 50\,000\,\mathrm{m^2}$, $\Delta f = 6$ MHz, $\Delta t = 0.3$ s, $N = 2.6$), and taking into account the instrumental constraints imposed by the stability of the acousto-optical spectrometer (AOS) used as the receiver (Rohlfs and Wilson, 2000), Zarka et al. concluded that Jupiter S-bursts would not be detectable at distances greater than 0.25 parsec. Although it is argued later in the same paper that much higher flux density levels can be expected from exoplanets located closer to their host stars (so-called "hot Jupiters"—see also Farrell et al., 1999; Zarka et al., 2001), thus giving additional support to the idea of detecting radio bursts from exoplanets with ground-based radiotelescopes, serious questions related to the feasibility of such a study in general remain open. Note, for example, that in formulas (1), and (2) it is assumed that the flux of intensity I is received permanently, whereas it is well known that planetary radio emissions are very sporadic so that time averaging reduces the average flux density by orders of magnitude. Recent results on the intensity distribution of individual S-bursts (Queinnec and Zarka, 2001) indicate that the peak flux level $I_J \approx 10^{-16}\,\mathrm{W\,m^{-2}\,Hz^{-1}}$ is rarely reached, whereas most of the time pulse intensity is in the range 10^{-17}–$10^{-18}\,\mathrm{W\,m^{-2}\,Hz^{-1}}$. Time averaging over an interval of duration about 3 s decreases the average flux densities down to 10^{-19}–$10^{-20}\,\mathrm{W\,m^{-2}\,Hz^{-1}}$ due to low (20–30%) filling factor of the time–frequency plane and to the abundance of low intensity bursts. These remarks suggest a reduction by one order of magnitude of the maximum distance of detectability of the radio emission, thus making the prospects of ground-based detection even more unfavorable.

However, as follows from the discussion presented below, the minimum detectable flux density can be much decreased if we take advantage of the total time of observation. Intuitively, it is clear that longer observation should increase the probability of detection, i.e. the "effective sensitivity", but in order to precise this idea, we need to give it a statistical framework that includes a detailed examination of all the constituents of formula (1) as well as of the algorithms developed for signal detection at the final stages of data processing.

Apart from purely instrumental limitations difficult to overtake like, e.g., the effective area of the antenna array or the maximum observation bandwidth, there is one parameter—the integration time—that apparently could be assigned much larger values than those used in the Paper I, i.e. less than 1 s. Indeed the increase of Δt by, say, a factor $\sim 10^4$ (up to 1 h integration time) would allow to detect Jupiter at a distance of about 2–3 parsec with the UTR-2 telescope. There are, however, two main factors discussed in the Paper I that restrict the integration time and, hence, the sensitivity of the observations: those are man-made radio frequency interference and stability limit of the receiver (AOS). The high intensity interference signals from local communication systems and short-wave broadcasting stations are prominent in the DAM band, making the

problem of detecting weak cosmic radio signals superimposed to them an extremely difficult task. In addition, the AOS gain shows fluctuations at the ~1% level on timescales of tens of seconds, susceptible to cause spurious signal detection if larger integration times are used for achieving higher sensitivity. The detailed account of various types of interference made in the Paper I resulted in an efficient computer algorithm for filtering out the signals of terrestrial origin and the receiver's gain fluctuations from the post-detector spectrograms. However, the proposed algorithm is crucially dependent on high-pass filtering procedures that impose a severe restriction on the effective value of the integration time.

In the present paper, we reconsider some of the estimates made in Paper I, through a more careful analysis of the parameters used in formula (1), together with the statistical procedures applied at the final stages of the post-detector signal analysis. In particular, we focus on the very definition of the notion of "signal detection", as well as those of "signal-to-noise ratio (SNR)" and "integration time". As a result, we develop a statistical approach that allows lowering the final detection threshold by one order of magnitude. The goal is reached by taking the total observation time into consideration, which is equivalent to increasing effective integration time in formula (1). In practical terms, this means that we can accumulate and then combine the results from several observational sequences or campaigns in order to reach higher sensitivity. Finally, we apply the proposed procedure to the data collected at UTR-2 with an AOS receiver between 1999 and 2002 in the frame of a radio-exoplanet search program. The observation targets, listed in Table 1, included 19 exoplanetary systems (candidate exoplanets or brown dwarfs) known from optical measurements. Although no positive detection can be reported yet, our results provide upper limits for the flux densities of the objects studied.

2. Statistical analysis. Peak detection algorithm

As already mentioned in the Introduction, such factors as the presence of interference signals within the observation band and instability of the AOS receiver require to implement data cleaning procedures which involve, in particular, filtering out low-frequency fluctuations from the analyzed time series. In terms of probability distributions discussed in the appendix, this results in an absence of separation between the (maxima of the) density curves or, equivalently, in the impossibility of using larger integration times for increasing sensitivity. We demonstrate below that this restriction can be substantially weakened if we use alternative ways to quantify the difference between the two statistical hypotheses. If mean values can no longer be used as discriminating statistics, then other statistical indicators must be found. The next natural choice seems to be the r.m.s. deviation which is independent of the time

Table 1
List of observations

Parent star	Spectral type	Distance from the Sun (pc)	Number of acquisition files (~5 min each)	Maximal sensitivity reached (at 300 ms integration time) [mJy]
Normal stars				
Ups Andromeda (HD 9826)	F7V/F8V	13.5	57	240
55 Cnc (HD 75732)	G8V	12.5	113	130
Rho CrB (HD 143761)	G0Va/G2V	17.4	93	160
Tau Bootes (HD 120136)	F6IV	15.6	92	160
70 Vir (HD 117176)	G4V	18.1	88	165
HD114762	F9V	28	71	190
Hd130322	K0V	30	4	1560
Hd187123	G5	50	14	1300
Hd38529	G4	42.4	30	1000
Hd52265	G0V	28	12	1300
Bd103166	K0V	<200	44	400
eps Eri	K2V	3.3	76	190
Lalande 21185 (HD 117176)	M2	2–2.5	134	100
Brown dwarfs				
HD 283750	K2	16.5	102	130
Hd110833	K3V	17	21	1250
Hd112758	K0V	16.5	11	1300
Hd140913	G0V	48	4	1560
Hd89707	G1V	25	49	400
Hd98230	F8.5 V	7.7	32	1000

average and, hence, could be used as a discriminating statistics. However, the temporal instability of the receiver gain causes fluctuations of this quantity thus bringing additional spurious signals that cannot be removed by the high-pass filtering. In order to make the analysis independent of the non-stationary value of σ, we normalize the series of data $A(t)$ by σ and obtain the time series $z(t)$ of unit dispersion and zero mean value

$$z(t) = \frac{A(t) - A_{\text{out}}}{\sigma}. \tag{3}$$

It is worth noting here that although this normalization procedure looks rather trivial, the combined effect of filtering and normalization—as it will be shown below—may have a crucial effect on the resulting signal detection procedure, because it may suppress the signal at high (or even moderate) values of the filling factor η (defined in the appendix).

For finding a suitable statistical characteristic that can be used for the purpose of signal detection, we assume by analogy with Jupiter S-bursts that the signal searched for is pulsed with a typical timescale τ ($\tau \leqslant 300\,ms$). It seems thus reasonable to search the data for pulses of high amplitude ($z > a$ in normalized units) and characteristic time scales close to those of S-bursts. The occurrence rate v_{a} (integrated tail of the probability density function) of high values in the recorded time series appears as a good candidate to serve as a discriminating statistics. Counting the number of peaks higher than a given level (typically 3σ) is a method widely used in (radio) astronomy for detecting bursts on the noise background.

In statistical terms, the problem of signal detection via analysis of occurrence rates is similar to the one discussed in the appendix for mean values distributions, i.e. one has to analyze the separation between two Gaussian density curves and find out when the normalized distance between the maxima is sufficient for providing the predetermined level of confidence. But it should be noted here that the value of a used as a threshold in the peak counting procedure needs not to be high, as can be expected from a simple reasoning based on common sense. As we demonstrate below, the optimal value of a depends on such signal characteristics as the filling factor η and the peak SNR N, so that a range of values of a should be tested for optimizing the detection procedure.

Let us perform the statistical analysis of cumulated distribution tails above a threshold a. If the signal is absent in the time series, then the distribution f_z of the random variable z is normal (Gaussian with zero mean and unit dispersion)

$$f_z(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2)$$

and the statistical properties of the distribution of occurrence rate v_a can be easily estimated. The probability density function of v_a is well approximated by a Gaussian with mean value $d_{\text{n}a}$ and dispersion $\sigma^2_{\text{n}a}$ given by

$$d_{\text{n}a} = \int_a^\infty f_z \, \mathrm{d}z, \quad \sigma^2_{\text{n}a} = \frac{d_{\text{n}a}(1 - d_{\text{n}a})}{M}. \tag{4}$$

Somewhat more complicated are the corresponding estimates for the case of signal and noise mixture in the analyzed data. Here the distribution function for the occurrence rate v_a is also Gaussian and defined by the shifted mean value $d_{\text{s}a}$ and dispersion $\sigma^2_{\text{s}a}$

$$d_{\text{s}a} = (1-\eta)\int_{a\sigma_{\text{s}}+N\eta}^\infty f_z \, \mathrm{d}z + \eta \int_{a\sigma_{\text{s}}-N(1-\eta)}^\infty f_z \, \mathrm{d}z; \quad \sigma^2_{\text{s}a} = \frac{d_{\text{s}a}(1 - d_{\text{s}a})}{M}. \tag{5}$$

The confidence level of detection writes almost identically to Eq. (A.3),

$$\gamma = \frac{\Delta d}{\sigma_{\text{n}a} + \sigma_{\text{s}a}} \tag{6}$$

where

$$\Delta d = d_{\text{s}a} - d_{\text{n}a} = \eta \int_{a\xi - N(1-\eta)}^{a\xi+N\eta} f_z \, \mathrm{d}z - \int_a^{a\xi+N\eta} f_z \, \mathrm{d}z \tag{7}$$

and since in this case we deal again with two Gaussian distributions, the 95% confidence level corresponds to the same value of $\gamma = 1.6$.

There are, however, differences compared to the approach based on mean values (see the appendix), that turn out to be important for the question of signal detection. The one more parameter used in the calculation of cumulated occurrence rates, the value of threshold a, makes the analysis of the function $\gamma(N, \eta, a, M)$ non-trivial. In Fig. 1 we show several dependencies of γ on a computed numerically through formulas (6) and (7) at different fixed values of η and N. The remarkable feature of the set of curves in Fig. 1 is the presence of a maximum located at a certain value of the parameter a depending on the filling factor η. For small values of η the maximum separation is achieved at about $a \approx 3$, whereas for a larger fraction of time occupied by the signal the maximum moves toward smaller values of a. This means that setting a threshold in the analysis of peak values implies automatically restricting the range of detectable η-values that correspond to high confidence levels. Another important property of the analysis based on the discriminating statistics v_a is the possibility that the distance between the maxima of the probability distribution functions, $d_{\text{s}a} - d_{\text{n}a}$, may become small or even negative for some values of the parameter a. For example, whatever is the SNR value of N in the example shown in Fig. 1a at $\eta = 0.5$, if the threshold is chosen at the usual value of 3σ, the cumulated occurrence rate value for the signal + noise

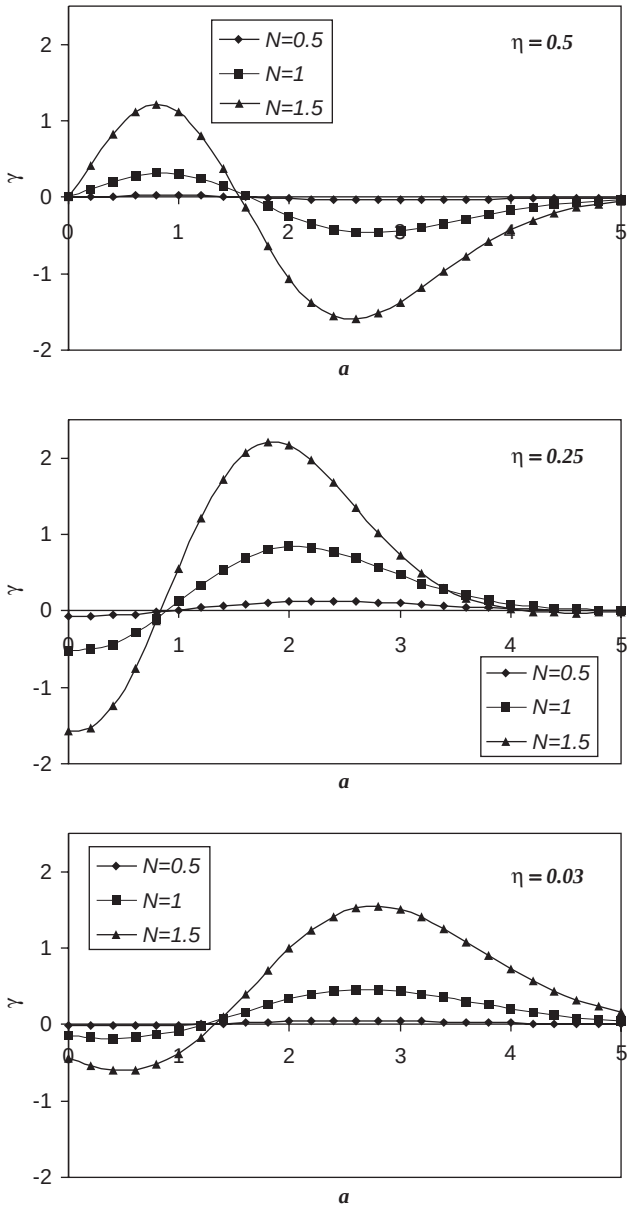

Fig. 1. Normalized separation between the maxima in Gaussian distributions corresponding to the cases of noise only and signal+noise mixture at $M = 16\,384$.

mixture is much below that for pure noise, seemingly indicating the absence of signal in the analyzed data. On another hand, the selection of $a = 0.8$ would result in this case in signal detection at sufficiently large N. This means that due to filtering and normalization the signal can be suppressed and, if the filling factor is high, the discriminating statistics v_a cannot be exploited. In practice, for using efficiently a signal detection algorithm based on the discriminating statistics v_a, it is desirable to have a priori information about the signal, especially on the parameter η. Otherwise the choice of too large or too small value of the threshold a results either in low detection efficiency or even complete suppression of the useful signal ($\gamma < 0$). If such information on η is not available, then it becomes necessary to explore a range of a-values looking for abnormally high magnitudes of v_a. This is what we have done in our analysis of exoplanetary search data as discussed below.

As it has been already mentioned, the statistical analysis of peak values in longer data series is equivalent to increasing the integration time of an ideal receiver considered in Appendix A. For translating this statement in quantitative terms, it is convenient to introduce an "equivalent integration time" that would be necessary for an ideal receiver to reach the same confidence level at the same values of the other parameters. This goal is reached by equating formulas (A.3) and (6) and calculating the ratio of the corresponding values of M. In Fig. 2a we plot the resulting curves for the case of $N = 2.6$, where M_1 stands for the length of the time series required for reaching the 95% confidence level with an ideal receiver (with the mean value used as a discriminating statistics), and M_2 represents the equivalent number when a peak detection algorithm is used for the same purpose (discriminating statistics v_a). Interestingly the dependencies displayed in Fig. 2a reveal a "selection property", i.e. any predefined level of a implies a corresponding interval of η-values over which the peak detection procedure is most effective. For example, high value of the threshold $a = 3$ favors the search of high amplitude peaks with low filling factor $\eta < 0.2$, while moderate values about $a = 1.5$ correspond to search for weaker signal with filling factor in the range $\eta \in [0.1; 0.4]$. This result is better illustrated in Fig. 2b where we plot areas of the plane $a-\eta$ allowing for detectability of a signal with peak intensity $N = 1$ for different numbers of points in the data series. As expected, the area shrinks with decreasing number of data points, and the highest sensitivity is reached at about $a = 2.3$ and $\eta = 0.15$. When the peak intensity N decreases below $N = 1$, the required length of the observation time series increases drastically as shown in Fig. 2c. This can be easily understood by noting that $N = 1$ corresponds to a signal intensity at the $1 \times \sigma$ level, i.e. equal to r.m.s. deviation of the sky background noise. Therefore, in order to detect signals with peak amplitude below this level, huge data sets have to be analyzed.

Another important property of the algorithm based on the analysis of peak values is that for very small filling factors ($\eta \leqslant 0.05$), its performance is even better than that of an "ideal" receiver (as shown in Fig. 2a) if, of course, the SNR N is high enough. This fact is not surprising: it is a direct consequence of the dilution effect of averaging when using mean value as the discriminating statistics with an "ideal" receiver. The final conclusions that can be derived from the above

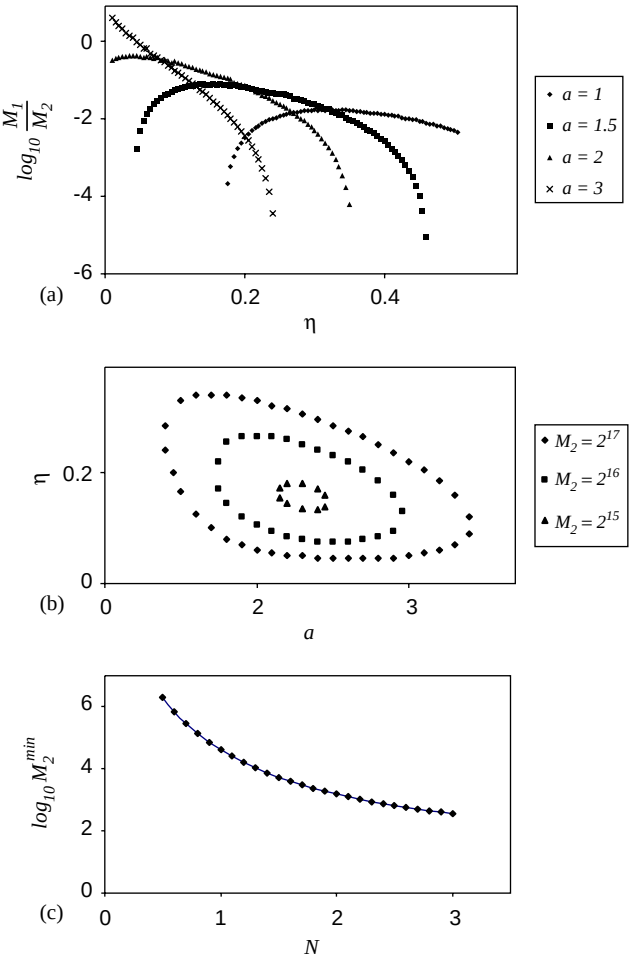

Fig. 2. Analysis of the confidence function $\gamma(N, \eta, a, M)$ at the level $\gamma = 1.6$: (a) comparative efficiency of the discriminating statistics v_a vs. mean value analysis (ideal receiver) at $N = 2.6$. $M_{1,2}$ are the numbers of data points necessary for reaching 95% confidence level of signal detection with ideal receiver and peak detection algorithm, respectively. Note the high performance of the algorithm based on the analysis of v_a-statistics at low values of the filling factor $\eta \leqslant 0.05$ and $a = 3$; (b) detectability areas of the peak analysis algorithm at $N = 1$; (c) minimum number of points in the time series $M_2^{\min}$ necessary for the signal detection with the peak detection algorithms vs. signal to noise ratio N.

calculations are: (i) signal detection based on the statistical analysis of peaks is efficient only at low levels of the filling factor η (typically $\eta \leqslant 0.1$), and (ii) a threshold about $a = 2.5$ is an optimal level for the discriminating statistics v_a to be used in this detection algorithm.

3. Exoplanet data processing

3.1. Instrumentation and data

The statistical approach developed in the previous section has been applied to the data obtained at the UTR-2 radio telescope between November 1999 and February 2002 for 19 candidate exoplanets and brown dwarfs detected earlier by optical radial velocities measurements, as well as a few pulsars to serve as tests for the detection capability. The observational time has been scheduled during the season with lowest man-made interference (November to April, nighttime). We have used the multi-beam capability of UTR-2 to perform simultaneous "ON-source" and "OFF-source" measurements, implementing a special interface between the telescope and the receiver, as described in the Paper I. The common band observed from the target and from a direction 1° away of it is 10 MHz wide. It is divided in 333 frequency channels separated by 30 kHz. Integration time is 17.4 or 33.8 ms/spectrum. One observation sequence consists of the acquisition of about 17 000 consecutive spectra (~30 Mbytes volume). Data were acquired with a PC driving the AOS receiver, then stored on high capacity disks for post-processing on a workstation. Observations for a given radio source have been performed preferably around the meridian transit, corresponding to maximum elevation. This ensures better immunity to interference propagating by reflection under the ionosphere and a less distorted radiotelescope beam thus reducing confusion. The list of targets and the total observation times is given in Table 1.

3.2. Data preprocessing: interference elimination, low pass filtering

Any data file typically contains many spurious signals of various origins (radio stations, lightning, satellite signals, etc.) that must be removed prior to search for exoplanetary signals. These interferences have been eliminated through the off-line processing procedure described in Paper I. The spectral response of the system UTR-2 + receiver is first corrected (flattened). Then, interference signals are identified and eliminated if they can be attributed to one of the following classes of time–frequency patterns in the dynamic spectra:

- broadband pulse-like signals caused by antenna switching, terrestrial lightning, or other reasons detected simultaneously in the data corresponding to the ON- and OFF-source beams;
- constant-frequency stationary narrow-band signals (in most cases short-wave broadcasting stations);
- high-intensity narrow-band pulses with typical timescale of a few tens Δt_s (i.e. 0.7–1 s).

An example of interference elimination for a typical data files is shown in Fig. 3. It is clear that interference signals have been removed (masked) fairly well in the bottom panel, making possible the next steps of signal detection.

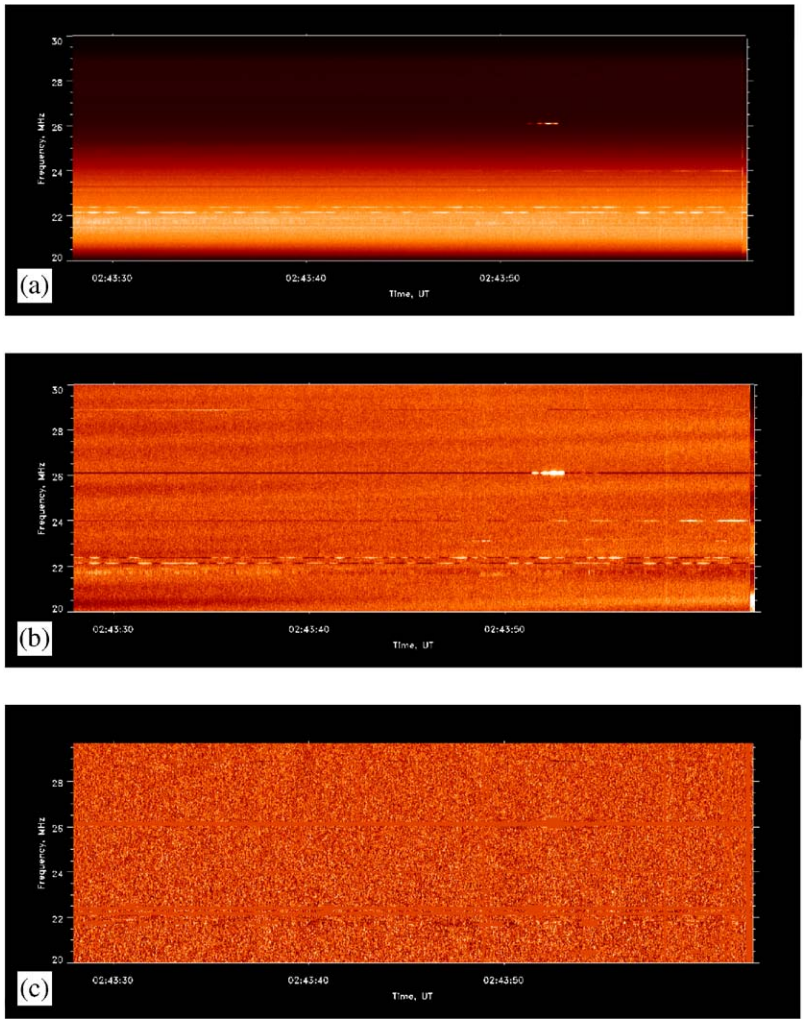

Fig. 3. An example of interference elimination and high-pass filtering algorithm applied to one of the data files for the candidate HD187123 on March 16th, 2000: (a) the raw dynamic spectrum; (b) the same spectrum after high-pass filtering and dispersion normalization; (c) the clean spectrum after identification and elimination of interferences.

3.3. Analogy with pulsar signals. Dedispersing

After interference have been removed, dynamic spectra can be converted to a single time series through frequency integration before performing the peak detection analysis described above. Prior to that, we must correct the effects of propagation of radio signal through the interstellar medium, and especially the dispersion of broadband short pulses due to frequency-dependent group velocity. This well-known effect that results in apparent negative frequency drift of the pulses on the time-frequency plane is most familiar in pulsar signal analysis (see, e.g., Beskin et al., 1993). Defining the dispersion measure (DM) as

$$\mathrm{DM} = \int n_\mathrm{e}\,\mathrm{d}l = \langle n_\mathrm{e}\rangle D[\mathrm{pc\,cm^{-3}}],$$

where n_e is the electron density along the propagation path and D the distance to the radio source, the time delay in pulse arrival time at two frequencies f_1, f_2 (in MHz) is (Manchester and Taylor, 1977)

$$\Delta t = 10^7\left(\frac{1}{f_1^2}-\frac{1}{f_2^2}\right)\frac{\mathrm{DM}}{2.41}\ \mathrm{ms}. \tag{8}$$

Compensation of the pulse distortion (dedispersing) can be thus performed by introducing time delays calculated in accordance with formula (8) to the output of each frequency channel of the receiver. An example of application of this procedure to pulsar signals detected at UTR-2 with the AOS is shown in Fig. 4.

Distances of known pulsars are generally much larger than those of exoplanet candidates, therefore expected dispersion measures are considerably smaller for the latter. But dedispersing is still expected to improve substantially the SNR and, hence, increase the probability of detection of possible exoplanetary radio signals. Note that dedispersing may even be partly

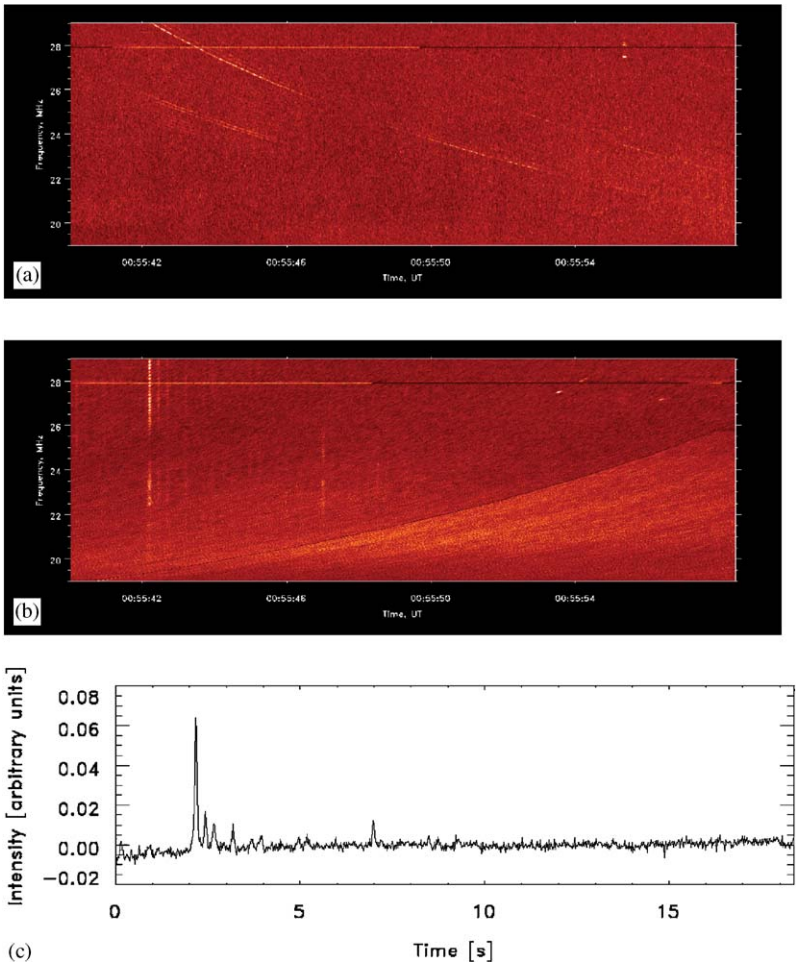

Fig. 4. The result of dedispersing a data file of December 29th, 2001, for the pulsar PSR0950+08. Dispersion measure is DM = 2.97 pc cm^{-3}: (a) raw dynamic spectrum after high-pass filtering. The pulses are drifting from high to low frequencies; (b) dedispersed dynamic spectrum (no frequency drift); (c) time series after frequency integration of the dynamic spectrum shown in (b).

correct for intrinsic frequency drifts of exoplanetary bursts.

In many other respects (typical periods of tens to hundreds of milliseconds, high intensity variations at slow timescales, 5–30% filling factor) pulsar signals resemble those expected from exoplanets. We have thus also observed several known pulsars with the same procedure as exoplanet candidates, and corresponding data have been processed in the same way, providing an additional possibility to test and validate the hardware, observation procedures, and signal processing software developed for pulse detection purposes.

3.4. Time series analysis. Detection diagrams

The price to pay for an increased SNR due to dedispersing is the presence of an additional parameter (DM) in the detection procedure. Since its value is generally poorly known, we applied the conventional procedure used in searches for new pulsars. It consists in testing a range of DM values for the presence of a clear peak in the corresponding power spectra of dedispersed time series (Nice, 1995, 1999). Technically, one looks for peaks above a predefined threshold value in a two-dimensional array of Fourier amplitudes in DM–frequency coordinates. Actually, such a search is equivalent to the statistical analysis discussed above, except that the amplitudes of Fourier harmonics are used as a frequency-dependent random variable instead of time variable amplitudes. The final result is conveniently displayed on a two-dimensional color diagram, where color denotes the intensity of corresponding Fourier components in the coordinate frame of DM versus frequency.

In the case of exoplanetary signals Fourier analysis is inefficient since emissions are not expected to be periodic. However other statistical characteristics as, e.g., the occurrence rates of the peak values v_a can be used in a similar way. We implemented this idea by working in the time domain instead of the frequency domain, and analyzed the statistical distribution of amplitudes depending on the value of DM. The final

detection diagram can be depicted in coordinates of DM versus acquisition sequence number, where occurrence rate for a given file is represented by a color or gray scale intensity variation. In principle, the same data processing can be applied to pulsar searches as well but of course with substantially reduced sensitivity as compared to search in the Fourier space. Due to many similarities with expected exoplanetary pulses, pulsar data appear as well adapted for testing the detection algorithm. In Fig. 5 we show an example of detection diagrams calculated for two low-frequency radio pulsars, PSR0950+08 (Fig. 5a) and PSR1919+21 (Fig. 5b). In the former case the SNR is high, as can be concluded from the presence of a bright strip on the detection diagram around the known value DM = 2.95. Note however the large flux variations (due to interstellar scintillation) that result in the apparent absence of the signal, e.g., for acquisition sequences No. 80–90. The flux level of the second pulsar PSR1919+21 is much lower than that of PSR0950+08, which results in a negative result of our peak detection algorithm for most of the available observation sequences. A few bright patches around the (correct) value of DM = 12.4 nevertheless reveal the presence of the expected signal, for example in sequence No. 14.

The efficiency of our peak detection algorithm is clearly demonstrated by our ability to detect pulsar signals, even in cases of low SNR ratio. All our "exoplanet search" observations have been processed accordingly and tested for several combinations of the control parameters defining the signal detection procedure and for DM varying within the interval [0; 2]. No positive detection has been obtained for any of our 19 targets. An example of detection diagram for the target HD283750 (brown dwarf) is shown in Fig. 6. The diagram has been calculated at the threshold value of $a = 2.5$, which corresponds to a pulsed emission with filling factor $\eta \leqslant 0.2$ and the highest sensitivity reached at somewhat smaller than 2.5 value of η (see, Fig. 2b).

If we take into account the overall beaming of Jupiter S-bursts (bursts are detected 10% of the time by a fixed observer) together with the filling factor of the pulses in the time–frequency plane during a typical S-burst storm (~25%, Queinnec and Zarka, 2001), one obtains a long-term average of the filling factor of about 0.025, i.e. close to that for which peak detection algorithm is more efficient than mere integration. Reaching the area of high efficiency requires changing the value of the threshold a for maximizing the confidence level of detection. However, varying the parameter a over the broad interval $a \in [2; 3]$ resulted in qualitatively similar results, i.e. no positive detection.

4. Discussion and conclusions

The reasons for non-detection of exoplanetary radio emission in our data may be similar to those discussed by Bastian et al. (2000), and we refer the interested reader to this detailed discussion. Let us simply mention a few important factors that may restrict our detection capability before summarizing the peculiarities of the present study.

1. Radio emissions from our targets may exist in frequency ranges different from the one observed with UTR-2 (generally between 18 and 32 MHz). For cyclotron emission this range is defined by the

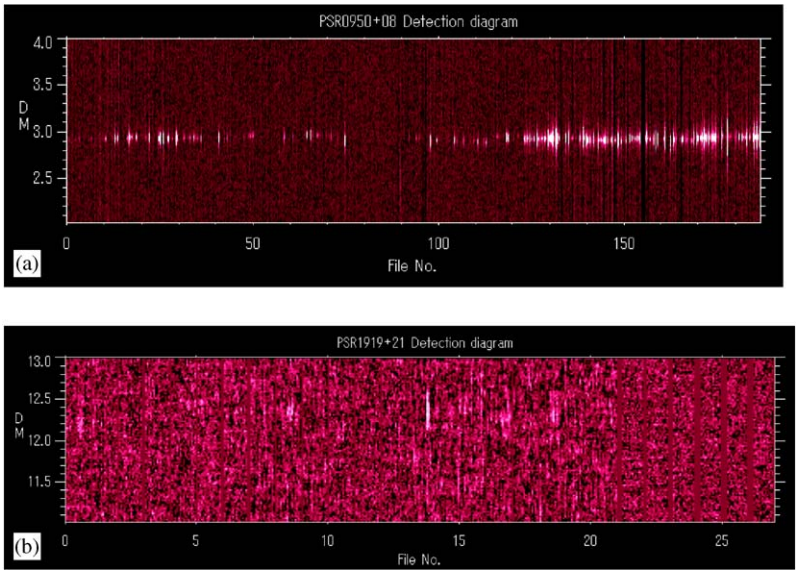

Fig. 5. Detection diagrams for pulsar signals calculated by means of peak detection algorithm (v_a is used as discriminating statistics): (a) high SNR pulsar PSR0950+08 with DM = 2.95; (b) low SNR pulsar PSR1919+21, DM = 12.4.

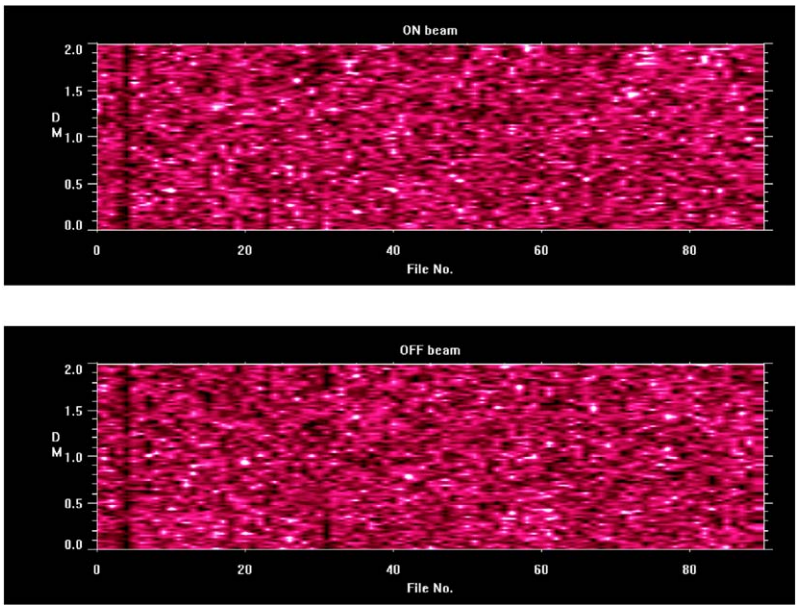

Fig. 6. Detection diagram for the brown dwarf HD283750. The data from both ON and OFF beams of the radio telescope are shown. The value of v_a is displayed by the color intensity in the coordinates of (dispersion measure)—(file number). The apparent homogeneity in the random distribution of color intensity indicates the absence of signal at any value of DM.

amplitude of the magnetic field at the planetary surface, and tidal locking of hot Jupiters may result in weak magnetic fields.
2. The power of generated bursts (if they exist) may be still below the detectability level of UTR-2.
3. The observed exoplanets may not generate at all radio bursts similar to those of the Jupiter–Io system.
4. Emissions from exoplanet candidates were not observable at the time of observation either because the emission process has an intermittent character or due to sharp beaming of the radiation pattern.

All those factors can combine to contribute to the negative result of our data analysis. There are, however, several additional reasons that may turn out to be crucial. The first one is the necessity to apply a refined and computationally extensive analysis of interference signals present in the data in order to recognize and eliminate them. This part of our processing may have two significant impacts on the final result. On the one hand, in spite of the detailed study and classification described in Paper I, there is no guarantee that all interference signals have been filtered out. Indeed, some of the final detection diagrams shown in Section 3 are still polluted with interference that can be recognized and removed by additional use of the correlations of the diagrams corresponding to the ON and OFF observation beams. These residual interferences may influence the statistical analysis at the final stages of calculating the discriminating statistics by masking the possibly much weaker signals from exoplanets. On the other hand, too severe interference elimination may result in eliminating also pulses of exoplanetary origin, especially if they are relatively less intense.

This situation with two competing requirements of both improving interference elimination analysis and simplifying the signal processing algorithm in order not to eliminate useful signal through iterative process is probably typical for any problem of signal classification where detection is based on predefined thresholds, filter cut-off frequencies, etc. Since all the types of interference signals are not known a priori, any algorithm requires optimization by trials and errors for fine-tuning the parameters controlling the detection procedure.

From a general point of view, a statistical analysis aimed at the detection of signal on the noise background should be based on assumptions concerning the characteristics of the signal that can be detected and separated from those of the pure noise. In our analysis, we assume that the emissions searched for are pulses, possibly frequency drifting, sparsely distributed over the time–frequency plane. The statistical procedure we have developed in the present paper for detection of this type of signal is thus based on the specific expected characteristics of the emission and should therefore have better performance than simple time averaging. However, if the signals we look for have very different characteristics (e.g., high filling factor, or not containing any high intensity outbursts), the proposed algorithm will reveal inefficient. Our detection capability is thus restricted by the following properties of the signal implemented in the detection procedure.

1. Pulses occupy a small part of the time–frequency plane, i.e. the filling factor η is below 0.1.
2. Peak intensity of the signal is much higher than the average level. This requirement is related to the parameter N in the detection algorithm, and consists in the demand of $N \gg 1$, where the value of $N \sim 1$ roughly corresponds to the level of sky background fluctuations.
3. The frequency drift of pulses, either due to interstellar dispersion or to intrinsic generation conditions within the radiation source, is consistent with a dispersion measure within the interval $\mathrm{DM} \in [0; 2]$.

If all of the above conditions are met, one can obtain an estimate of the equivalent sensitivity reached by our observations of exoplanet candidates, depending on the total observation time for each of them. These estimates for the observed exoplanet candidates are shown in the last (to the right) column of the table. Note, that prior to running the peak detection algorithm, the data series has been integrated over the time interval of 300 ms that is equivalent to reaching the sensitivity of ~1.6 Jy as can be estimated with formula (2). This establishes the upper bound on the sensitivity shown in the table. Enhancement in sensitivity due to the analysis of peak values distribution results in smaller magnitudes of the flux density. However, due to nonlinear character of the dependence of sensitivity on the total length of the time series, the substantial improvement in sensitivity can be reached for the number of acquisition files greater than ~30 (total observation time > 150 min).

Acknowledgements

This work has been supported by INTAS (Project No. 99–1183).

Appendix A. Signal detection with an ideal receiver

A typical scheme used for signal detection on the noise background (see, e.g., Bendat and Piersol, 1986) includes the development of a system of two alternative statistical hypotheses—one for the signal + noise mixture, and the other for pure noise. The problem thus consists in applying an algorithm of hypothesis testing in order to discard one of them, therefore making the decision about the presence or absence of signal in the analyzed data. As for the hypothesis-testing algorithm, it is usually developed together with the hypotheses themselves, based on a set of criteria defined by physical considerations or other a priori information about the signal, the noise, and their statistical properties. Technically, the straightforward approach just described implies calculating a representative number (called discriminating statistics, d_s) and comparing it with the predefined threshold. If the discriminating statistics lies above the threshold, one concludes that useful signal is present.

In the problem of detecting signals from exoplanetary magnetospheres with the combination of the UTR-2 radio telescope and an AOS receiver, the analyzed data can be considered as the output of the idealized receiver, i.e. square detector + low-pass filter (with time constant defined by the integration time), with white Gaussian noise of intensity (dispersion) D_{inp} at the input. At the output of such receiver a user deals with a time series distributed according to a Gaussian law with mean value A_{out} and dispersion D_{out}, determined by the average power of the input signal D_{inp}, the transfer coefficient β of the antenna + receiver system, the total frequency band Δf and integration time Δt (Burke and Graham-Smith, 2002; Davenport and Root, 1958)

$$A_{\mathrm{out}} = \beta \Delta f D_{\mathrm{inp}}, \quad D_{\mathrm{out}} = \frac{\kappa}{\sqrt{\Delta t}} A_{\mathrm{out}}, \tag{A.1}$$

where κ is a positive constant defined by the frequency transfer function of the receiving system. The SNR N at the output of receiver can be then introduced (see also the Eq. (1)) as the shift in the A_{out} value due to the presence of signal divided by the amplitude of sky background fluctuations, i.e. $\sigma = \sqrt{D_{\mathrm{out}}}$. It should be also noted, that if the integration time is sufficiently large, the ratio $\kappa/\sqrt{\Delta t}$ can be considered small (for the experimental setup discussed in Section 3 it typically constitutes about 0.001), that results in practical independence of the D_{out} value from the power of the useful signal within a reasonable range of SNR levels, e.g., $N \leqslant 10$. This conclusion enables us to put $D_{\mathrm{out}} \approx \sigma^2$ for both cases of noise and signal + noise at the receiver output, in all the calculations of Section 2 and further discussion of the experimental data processing.

The above-mentioned hypotheses to be used at the final stage of the signal detection correspond to the two cases of (i) sky background noise only or (ii) the combination of the same noise with an exoplanetary signal superimposed. An example of computer generated time series without (a) and with (b,c) useful signal is given in Fig. 7 with SNR $N = 2.6$. In Fig. 7 the bursts (signal) occupy approximately 25% of the observational time, which is consistent with the typical value of S-burst "filling factor" in the time–frequency plane $\eta = 0.26 \pm 0.1$ (Queinnec and Zarka, 2001). Throughout this paper we make a distinction between two parameters, N and η, defining the principal characteristics of the analyzed time series, i.e. maximal flux density (N) and average flux calculated over the whole data set ($N\eta$).

It should also be noted that a high magnitude of SNR is not enough to have a statistically significant detection. The problem of signal detection requires also the definition of the confidence level which is conventionally

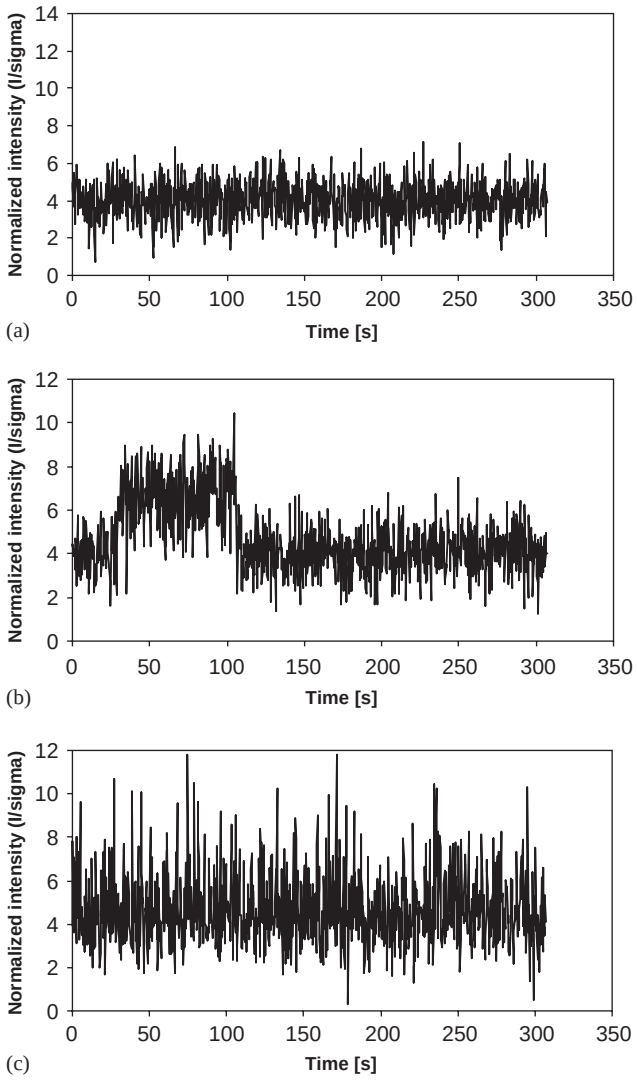

Fig. 7. Computer simulated time series at the output of the receiver system after integration with the time constant of $\Delta t = 300$ ms: (a) sky background fluctuations (noise) only; (b) sky background + long burst signal (characteristic time scale ~75 s) with signal/noise ratio $N = 2.6$ and filling factor $\eta = 0.25$; (c) same as (b) but with the short burst signal (characteristic time scale ~300 ms).

introduced in statistical terms by the probabilities of false detection (P_{I}) if the signal is not present in the analyzed data, and of non-detection (P_{II}) if the signal is present (hypotheses testing errors of type I and II according to Bendat and Piersol, 1986). In order to simplify the analysis, it is convenient to use a single parameter α that controls the confidence level. For example, it can be defined as the (half of the) area under the tails of the probability density functions $p_{\mathrm{I}}(x) = \mathrm{d}P(x)/\mathrm{d}x$ and $p_{\mathrm{II}}(x) = \mathrm{d}P_{\mathrm{II}}(x)/\mathrm{d}x$ to the right and left of their intersection point $x = x_c$, i.e. $p_{\mathrm{I}}(x_c) = P_{\mathrm{II}}(x_c)$. Then, $\alpha = (P_{\mathrm{I}}(x_c) = P_{\mathrm{II}}(x_c))/2$.

A standard way of quantifying the difference between the two time series with and without signal (e.g., Fig. 7(a) relative to Fig. 7(b) or (c)) used in radio astronomy (Burke and Graham-Smith, 2002) would be to calculate their mean values as discriminating statistics, which corresponds to the estimation of the average power of the received signal or, equivalently, to the result of an increase of the integration time up to the total length of the time series. The signal detection procedure would thus consist in comparing the calculated average $A_{\mathrm{out}} = d_{\mathrm{s}}$ with that expected in the case of pure noise at the input $A_{\mathrm{out}} = d_{\mathrm{n}}$. The probability density function for A_{out} is in both cases well approximated by Gaussian distributions with dispersions σ_{s}^2 and σ_{n}^2, respectively, centered on different mean values d_{s} (if the signal is present) and d_{n} (if it is absent). The dispersion σ_{n}^2 is completely defined by the total length of the time series M (observation time) and the intensity of sky background fluctuations σ, whereas σ_{s}^2 depends also on other parameters like the SNR N and filling factor η.

$$\sigma_{\mathrm{n}} = \sigma/\sqrt{M}, \quad \sigma_{\mathrm{s}} = \sigma\sqrt{(1+\eta(1-\eta))/M} \equiv \xi\sigma_{\mathrm{n}},$$
$$\text{where } \xi = \sqrt{1+\eta(1-\eta)N^2} \tag{A.2}$$

The confidence level γ can then be defined via the shift between the mean values normalized to the sum of r.m.s. deviations

$$\gamma = \frac{d_{\mathrm{s}} - d_{\mathrm{n}}}{\sigma_{\mathrm{n}} + \sigma_{\mathrm{s}}} \equiv \frac{N\eta\sqrt{M}}{1+\xi} \tag{A.3}$$

and the parameter α defining the confidence probability $P_\alpha = 1 - \alpha$ can introduced as half of the overlap area under the corresponding probability density curves (see Fig. 8).

$$2\alpha = \frac{1}{\sigma_{\mathrm{n}}\sqrt{2\pi}}\int_{x_c}^{\infty} \exp\left[-(x-d_{\mathrm{n}})^2/2\sigma_{\mathrm{n}}^2\right]\mathrm{d}x + \frac{1}{\sigma_{\mathrm{s}}\sqrt{2\pi}}\int_{-\infty}^{x_c} \exp\left[-(x-d_{\mathrm{s}})^2/2\sigma_{\mathrm{s}}^2\right]\mathrm{d}x, \tag{A.4}$$

where

$$x_c = d_{\mathrm{n}} + \frac{\xi\sqrt{M\eta^2N^2 + 2(\xi^2-1)\ln(\xi)} - \eta N\sqrt{M}}{\xi^2 - 1}\frac{\sigma}{\sqrt{M}}.$$

Under such a definition, the probability P_{I} consists in that of detecting an abnormally high value of the $A_{\mathrm{out}} > x_c$ in the case of observing sky background noise only, whereas the value of P_{II} defines the probability of detecting an abnormally low ($A_{\mathrm{out}} < x_c$) value of the average amplitude when the signal + noise mixture is observed.

The direct calculation of the dependence of $\alpha(\gamma)$ for various values of N and η shows that it is quasi-invariant with respect to these parameters, i.e. there is one-to-one correspondence between the confidence probability α and the normalized distance γ between the maxima of the two Gaussian curves. Therefore, any of those

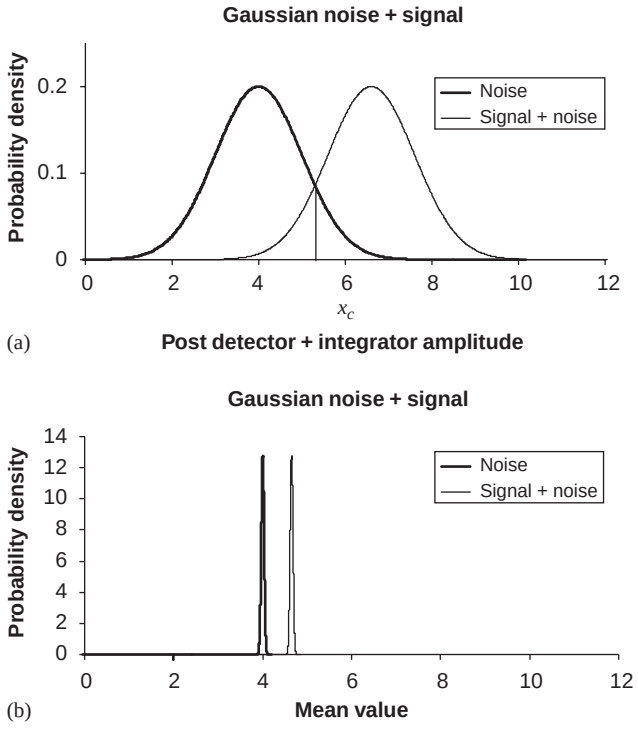

Fig. 8. Probability distributions (Gaussian) for the noise and signal + noise corresponding to Fig. 7: (a) the value of filling factor $\eta = 1$, integration time $\Delta t = 300\,\text{ms}$ ($M = 1$); (b) filling factor $\eta = 0.25$, integration time $\Delta t = 300\,\text{ms}$ ($M = 1000$).

parameters (α or γ) can be used for specifying the confidence level of detection. Without loss of generality we can assume, e.g., that the value of $\gamma = 1.6$ is sufficient for ensuring detection at the confidence level of 95% (i.e., $\alpha = 0.05$). Note that the mean value d_s of the signal + noise mixture (if $M > 1$) depends on the filling factor η ($\eta = 0.25$ is assumed in Fig. 8), therefore the separation between the maxima in Fig. 8b is reduced by a factor $1/\eta$ compared to Fig. 8a.

From formula (A.3) it is easy to estimate the amount of data (length of the time series) necessary for reaching the desired confidence level γ at a given value of the peak intensity N:

$$M = \left(\frac{\gamma(1+\xi)}{N\eta}\right)^2. \tag{A.5}$$

For the time series of Fig. 7, a separation of $\gamma = 1.6$ (95% confidence level) is achieved at about $M = 35$, which corresponds to an integration time of $35\Delta t \approx 10\,\text{s}$. Actually, formula (A.5) establishes a fundamental limit on signal detection with an ideal (integrating) receiver and data ideally cleaned from interference. For a given time series of M data points with integration time Δt, it also provides the maximum sensitivity (minimum detectable peak intensity) expressed in units normalized to the sky background fluctuations

$$N = \frac{\gamma(1+\xi)}{\eta\sqrt{M}}. \tag{A.6}$$

The physical meaning of formula (A.6) is quite trivial. It just reestablishes the well-known fact that sensitivity is inversely proportional to the square root of the total integration time $\tau = M\Delta t$.

References

Bastian, T.S., Dulk, G.A., Leblanc, Y., 2000. A search for radio emission from extrasolar planets. Astrophys. J. 545, 1058–1063.

Bendat, J.S., Piersol, A.G., 1986. Random data. Analysis and Measurement Procedures. Wiley, New York.

Beskin, V.S., Gurevich, A.V., Istomin, Ya.N., 1993. Physics of the Pulsar Magnetosphere. Cambridge University Press, Cambridge.

Burke, B.F., Graham-Smith, F., 2002. An Introduction to Radio Astronomy. Cambridge University Press, Cambridge.

Davenport, W.B., Root, W.L., 1958. An Introduction to the Theory of Random Signals and Noise. McGraw-Hill, New York.

Farrell, W.M., Desch, M.D., Zarka, P., 1999. On the possibility of coherent cyclotron emission from extrasolar planets. J. Geophys. Res. 104, 14025–14032.

Manchester, R.N., Taylor, J.H., 1977. Pulsars. W. H. Freeman, San Francisco.

Nice, D.J., 1995. Radio pulsars: an observational perspective. In: Alpar, M.A., Kiziloglu, U., van Pradijs, J. (Eds.), The Lives of the Neutron Stars. Kluwer, Dordrecht, p. 225.

Nice, D.J., 1999. Radio pulses along the galactic plane. Astrophys. J. 513, 927–932.

Queinnec, J., Zarka, P., 2001. Flux, power, energy and polarization of Jovian S-bursts. Planet. Space Sci. 49, 365–376.

Rohlfs, K., Wilson, T.L., 2000. Tools of Radio Astronomy. Springer, Berlin.

Zarka, P., et al., 1997. Ground-based high sensitivity radio astronomy at decameter wavelengths. In: Rucker, H.O., Bauer, S.J., Lecacheux, A. (Eds.), Planetary Radio Emissions IV. Austrian Academic Science Press, Vienna, pp. 101–127.

Zarka, P., Treumann, R.A., Ryabov, B.P., Ryabov, V.B., 2001. Magnetically-driven planetary radio emissions and application to extrasolar planets. Astrophys. Space Sci. 277, 293–300.