

RESEARCH ARTICLE

OPEN  ACCESS

A method for the automated detection of solar radio bursts in dynamic spectra

Houssam Salmane^{1,2,*}, Rodolphe Weber^{1,3}, Karim Abed-Meraim¹, Karl-Ludwig Klein², and Xavier Bonnin²

¹ PRISME, University of Orléans, EA4229, 8 rue Leonard de Vinci, 45072 Orléans, France

² LESIA-UMR 8109, Observatory de Paris, PSL Res. Univ., CNRS, Sorbonne Univ., Univ. Paris-Diderot, 5 place Jules Janssen, 92190 Meudon, France

³ Station de Radioastronomie de Nançay, Observatoire de Paris, PSL Res. Univ., CNRS, Univ. Orléans, OSUC, 18330 Nançay, France

Received 9 October 2017 / Accepted 26 July 2018

Abstract – The variability of the solar corona, including flares and coronal mass ejections, affects the space environment of the Earth (heating and ionization of the atmosphere, magnetic field disturbances, and bombardment by high-energy particles). Electromagnetic emissions are the first signatures of a solar eruptive event which by modifying the electron density in the ionosphere may affect airborne technology and radio communications systems. In this paper, we present a new method to detect automatically radio bursts using data from the Nançay Decametre Array (NDA) in the band 10 MHz–80 MHz. This method starts with eliminating unwanted signals (Radio-Frequency Interference, RFI and Calibration signals) by analyzing the dynamic spectrum of the signal recorded in time. Then, a gradient median filter is applied to smooth and to reduce the variability of the signal. After denoising the signal, an automated solar radio burst detection system is applied. This system is based on a sequential procedure with adaptive constant-false-alarm rate (CFAR like detector) aimed to extract the spectra of major solar bursts. To this end, a semi-automatic software package is also developed to create a data base of all possible events (type II, III, IV or other) that could be detected and used for our performance assessment.

Keywords: solar radio bursts / automatic detection / dynamic spectra / events type II, III and IV

1 Introduction

Eruptive activity in the solar corona can lead to severe disturbances of the space environment of the Earth. This activity produces enhanced intensities of ionising photons at EUV and X-ray energies, high-energy particles and ejected plasma structures. Among the potential technological impacts are damage to technological systems, disturbances of radio wave propagation in the ionosphere, satellite malfunctioning (Bothmer & Daglis, 2007), and enhanced radiation exposure to astronauts (Fry, 2012), sometimes even aboard aircraft (Kataoka et al., 2011).

The prediction of major solar disturbances and their impact on Earth is one element to mitigate space weather hazards, provided alerts can be emitted on time with a low false alarm rate (Zhang et al., 2014). The reliable prediction that no major disturbance is expected in the near future is also a relevant piece of information. It is presently not possible to predict

the space weather impact of a solar event before the first signatures of the solar event can be observed. Radio emission is among the earliest observed signatures of an eruptive event in the corona.

A dynamic spectrum shows the evolution of the emission of the whole Sun in the frequency-time plane. The work presented in this paper is an attempt to develop an automated procedure to recognize eruptive solar activity in the dynamic spectra of solar radio emission at metric and decametric waves. In this context, we introduce a new method that can be applied to automatically detect radio bursts. We illustrate the procedure using the solar radio spectral data provided by the Nançay Decametre Array (NDA) (Lecacheux, 2000), which operates at frequencies between 10 and 80 MHz (wavelengths 4.5 m–30 m). The proposed algorithm proceeds in two steps: the first consists in filtering the received spectral data to remove artifacts and to emphasise solar radio bursts. In this step, three kinds of noise and unwanted signals are subtracted from the original signal: Calibration signals, Radio-Frequency Interference RFI (due mostly to communications systems) and galactic noise.

*Corresponding author: houssam.salmane@univ-orleans.fr

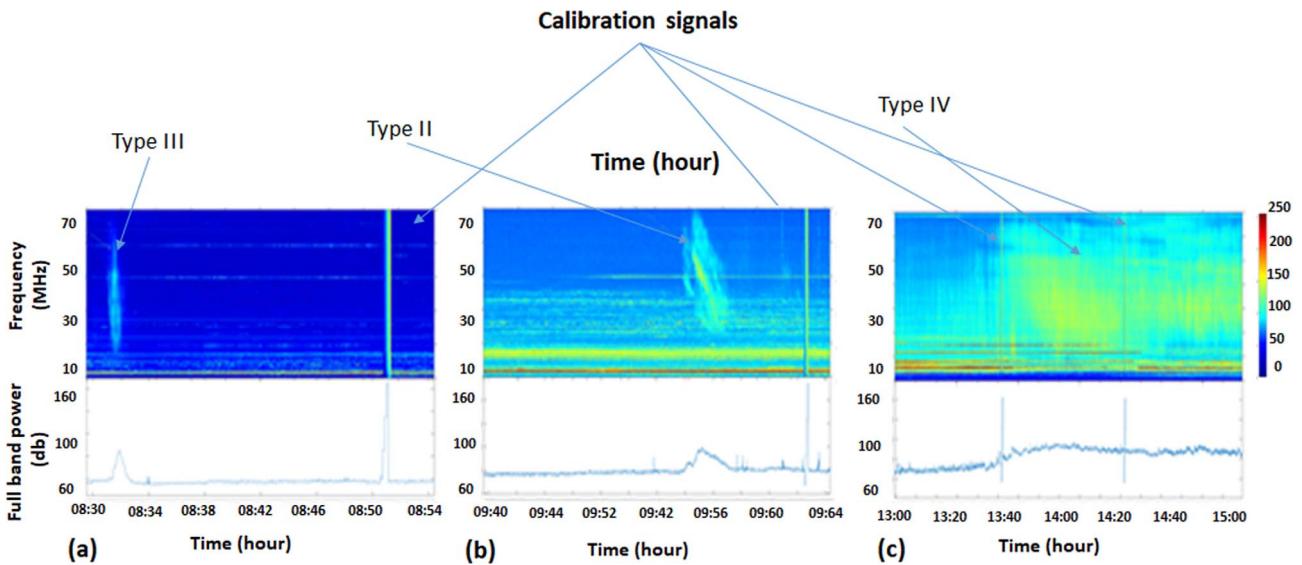


Fig. 1. Type II, III and IV bursts registered by the Nançay Decametre Array.

We propose an adaptive noise removal method taking into account the nature (origin) and noise variation (non-stationarity) of the signal with time. The second step of the algorithm is to detect solar radio bursts in the dynamic spectra in order to enable an automated identification that can be used either in real-time applications or in observations of a given type of events from a large data base. In this paper, we show that a constant false alarm rate (CFAR)-like technique (Cai et al., 2011; Bandiera et al., 2009) with optimized parameters can achieve a high detection performance. This work has been developed within the ORME (Observations Radioastronomiques pour la Météorologie de l'Espace) project. It represents the first phase to achieve a complete and efficient detection and classification system. Indeed, the solution proposed in this paper helps selecting the regions of interest where a radio burst occurs. The second phase would be the classification system which would focus on the detected events for their appropriate cataloging. This is an ongoing work that will be presented in future publications.

The paper is organized as follows. In Section 2, a brief overview on solar radio bursts and on previous attempts at an automated detection is given. In Section 3 we present the rationale of this work and we summarize our new contribution. The interference and noise removal procedure for Time-Frequency dynamic spectra is described in Section 4. Section 5 summarizes the proposed solar radio burst detection technique. Section 6 presents the obtained results and evaluates the performance of the proposed method. The software and database used in this study are described in this section, too. Section 7 is for concluding remarks.

2 A brief overview of solar radio bursts with relevance to space weather

Radio bursts at decimetre and longer wavelengths have characteristic spectra in the frequency-time plane, which reveal the nature of the solar disturbance. Figure 1 illustrates different

types of solar radio burst spectra between 80 and 10 MHz observed by the Nançay Decametre Array, located at the Nançay radio astronomy station in central France. A detailed discussion of such spectra can be found, e.g., in Pick & Vilmer (2008) and Nindos et al. (2008).

The rather short burst in Figure 1a is called a type III burst. It is generated by electron beams propagating outwards through the corona along open magnetic field lines. The electron beams excite the plasma waves in the solar corona, which create radio emission at the electron plasma frequency or its first harmonic. The emission frequency is hence proportional to the square root of the ambient electron density. As the electron beams travel outward through the corona, towards regions of decreasing electron density, they emit radio waves at decreasing frequency. The characteristic feature of type III bursts that can be exploited for an automated recognition is their rather short duration and rapid drift from high to low frequencies.

A second type of radio burst is seen in Figure 1b. It consists of one or several narrow lanes that drift towards low frequencies, but much more slowly than the type III bursts. This type of emission is called a type II burst. It is emitted by electrons that are accelerated by an outward propagating shock wave. The shock wave also propagates into regions of decreasing electron density, but more slowly than non thermal electron beams. As a consequence, the frequency drift is slower, too. Sometimes these bursts of type II can be associated with fine structures known as herringbone bursts. These structures are direct signatures of particle acceleration which occurs at coronal shocks (Cairns & Robinson, 1987).

The third burst associated with space weather effects is known as type IV burst. The emission is produced by energetic electrons trapped in magnetic structures that are part of a Coronal Mass Ejection (CME). This is a broadband emission that lasts several minutes or even longer and has a broad instantaneous bandwidth. An example is shown in Figure 1c. The type IV bursts may also have a drift towards lower frequencies, but this is not always the case. Type IV bursts are defined by their long duration and broad overall bandwidth.

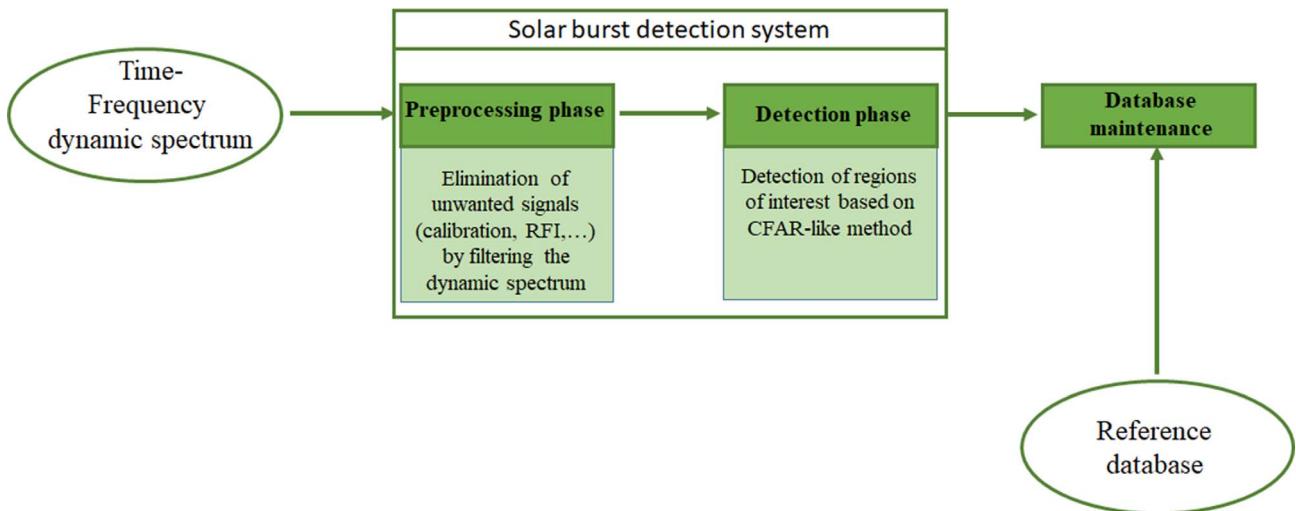


Fig. 2. Architecture of the proposed radio burst detection system.

All of these particular radio bursts provide important information for space weather because they show particle propagation to the interplanetary space or because of their association with CMEs (Munro et al., 1979). There have been a number of attempts to identify solar radio bursts in an automated way. These efforts pertained mostly to type III bursts, and to a lesser extent to type II bursts. Lobzin et al. (2009, 2010) and Bonnin et al. (2011) propose to use the Radon transform (Deans, 2007) for the detection of the solar bursts. This operation transforms drifting features into vertical straight lines in dynamic spectra. Time-Frequency binary spectra are used in this algorithm to characterize and hence detect the desired burst. More specifically, this method is based on the following successive steps: noise filtering, image binarization, Radon transform and thresholding. Jones & Richards (2014) also use binary spectra to implement a mathematical morphology technique to extract features that represent solar radio bursts. Carley et al. (2015) develop a method based on the Hough transform (Mukhopadhyay & Chaudhuri, 2015) to identify herringbone bursts (Kong et al., 2012) associated with type II solar radio emissions. Another interesting method is proposed by Gunes & Erdol (2006). They have developed an efficient Hidden Markov Model (HMM) (Tumilaar et al., 2015) to track fine structures in spectrograms. However, these algorithms are used to identify straight lines in time-frequency dynamic spectra, whereas even transformed spectra of type III and type II bursts nearly always show some curvature, especially at their high-frequency and low-frequency borders. As a result, these methods fail to precisely detect the start and stop times of the bursts, and they are not well adapted to detect large structures like in type IV solar radio bursts.

3 Concept of the radio burst detection system

The detection and classification of the radio bursts discussed in the previous section is a tool to alert about the arrival of solar disturbances in the space environment of the Earth.

In order to detect these particular bursts, basic spectrum properties such as frequency drift rate, bandwidth, duration, brightness and signal variation must be determined with a high degree of accuracy. However the presence of some unwanted signals, like Radio Frequency Interference (RFI) signals, calibration signals and some other unknown noise signals, make the bursts detection problem more difficult.

In this work we develop a method to automatically detect solar burst events (in particular events of type II, III and IV) from a noisy time frequency dynamic spectrum¹. After eliminating unwanted signals by analyzing the dynamic spectrum of the signal recorded in time, we propose to use an automated radio burst detection based on the Constant False Alarm Rate (CFAR) method cited above. Our contributions can be summarized as follows (see Fig. 2):

- Dedicated denoising and interference mitigation techniques for the enhancement of the signals of interest and consequently the improvement of the detection performance.
- Development of a new detection method based on a CFAR-like technique.
- Creation of a labeled database using the recorded NDA data for performance assessments and comparisons of different detection methods.
- A performance analysis of the proposed method on real-world data. As a byproduct, we have developed an ergonomic software interface for event detection and performance evaluation, which can be used in particular to further enrich our database for future investigations.

4 Noise removal technique

The main contribution of this section is the introduction of a denoising algorithm to reduce the amplitude of RFI,

¹ Note that the spectrum phase information is not provided by solar radio spectrographs and so we have access only to the time varying spectrum amplitude information.

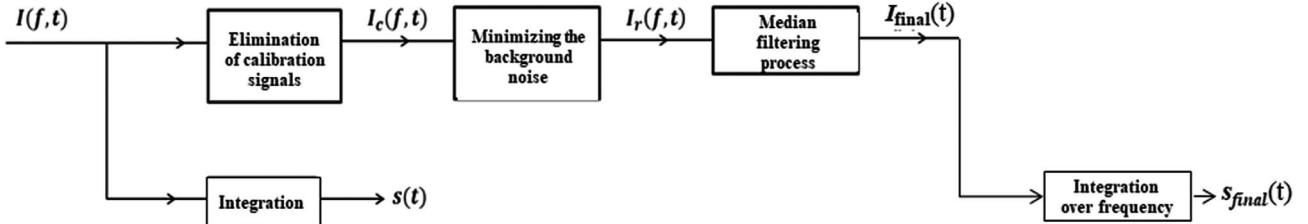


Fig. 3. Proposed denoising and interference mitigation procedure.

calibration and background noise signals that affect the time-frequency (T-F) dynamic spectrum at metric-to-decametric wavelengths.

To remove noise from the dynamic spectrum, we propose in this paper a solution which takes into account the nature of the corrupting signals. This approach is based on three main steps (see Fig. 3): The first step serves to remove the unwanted calibration signals provided by NDA (the resulting T-F signal after this first step is denoted $I_c(f, t)$). The second step aims to mitigate the background and RFI noise in the T-F dynamic spectrum with minimal distortion of the signals produced by solar bursts. The output of this step is denoted $I_r(f, t)$. In the third step, we use a specific median filter in order to remove artifacts and discontinuities due to the denoising process. This step also enhances the contrast of the dynamic spectrum resulting in $I_{\text{final}}(f, t)$ and the denoised signal energy $s_{\text{final}}(t)$. Note that the original signal $s(t)$ and the denoised signal $s_{\text{final}}(t)$ (in which a CFAR algorithm is applied to detect solar burst events) are obtained by integrating the T-F dynamic spectra $I(f, t)$ and $I_{\text{final}}(f, t)$ with respect to the frequency axis (Fig. 3). The CFAR algorithm which is to detect the solar bursts is applied to the signal $s_{\text{final}}(t)$.

4.1 Elimination of the calibration signals

The Nançay Decametre Array must be calibrated every 1 h in order to maintain the accuracy of flux density measurements as seen in Figure 4. It is an injection of a signal that is identical at each frequency and varies stepwise in flux density in the course of time.

To remove the calibration signal completely without seriously affecting the desired signal, we have developed a dedicated algorithm which exploits the known time profile of the signal to localize the patterns that correspond to the calibration signals on the dynamic spectrum and to remove them (see Appendix A).

Figure 5 shows an example of removing the calibration signal without much affecting the desired signal of the solar burst.

4.2 Background noise mitigation

Applying a filter to suppress the fluctuations of the background noise obtained on solar radio spectrographs allows us to enhance the features of solar radio events. Since both the background of the dynamic spectrum and solar events display a large variety of morphological structures, a background subtraction by setting a simple threshold does generally not lead to an satisfying result: If we set the threshold value very high,

we will remove some of the solar radio events. If we set a low threshold value, the background noise level will be reduced but not enough to separate solar event features from unrelated background. In the method proposed by Zarka et al. (2004), the authors represent the background noise by a Gaussian distribution computed at each frequency f_i during a long time interval. The problem of this method is that it reduces the flux density level of both technique without significantly affecting the features of solar radio bursts. Given a T-F dynamic spectrum $I_c(f, t)$ (obtained after elimination of calibration signals), we proceed by comparing the Gaussian variability of the signals received during a long and a short period. As shown in Figure 6, the statistical fluctuations of the background noise level at a given frequency, evaluated over the long period, are Gaussian-distributed around the average value μ_r of the flux density $I(f, t)$ at that frequency. This is due to the Gaussian nature of the solar signal and to the instruments (Nita et al., 2007).

The rationale of our approach is based on the following observations:

- If we take into consideration a long observation period for a given frequency f_i not dominated by any terrestrial emitter (using an interval of 7 h of duration), the produced events related to solar bursts are rare and relatively sparse. Therefore the mean value of the signal over the long period represent a rough estimate of the averaged noise power μ_r that follows a Gaussian distribution G_r (see the histogram of the distribution in Figure 6).

At a given time-frequency point (f_i, t) in the absence of a radio burst (see yellow intervals in Fig. 6) the local mean value represents a noise power estimate of the dynamic spectrum that follows the Gaussian noise distribution G_r . In that case, we would use the local mean value for the background noise mitigation. At a given time-frequency point (f_i, t) in the presence of a radio burst event (see red interval in Fig. 6) the local mean value would be much higher than the long period mean value μ_r (which represents roughly the noise term value), and so we use this as an indicator to distinguish between “noise” TF points and “desired signal” TF points. In that case, the denoising would rely on the long-term mean value of the signal.

- Now, if we consider a long observation period for a frequency dominated by a terrestrial emitter, the averaged signal power would represent an estimate of the RFI signal at this particular frequency (see Fig. 7). The measure of the spectrum $I_c(f, t)$ of the RFI spectrum is characterized by a high average flux density level over a long period.

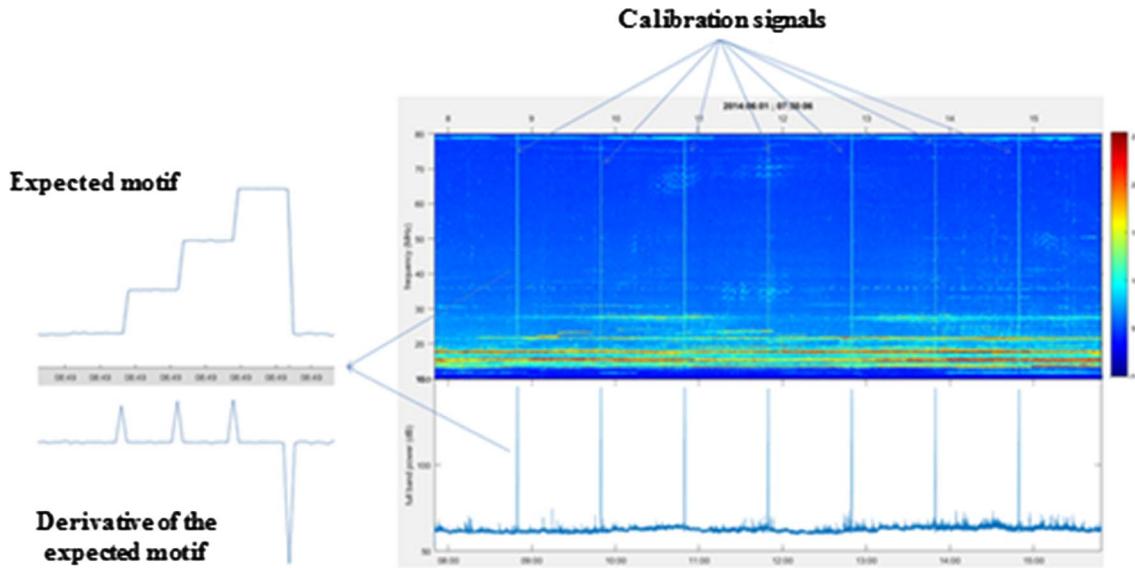


Fig. 4. Time and T-F signatures of calibration signals.

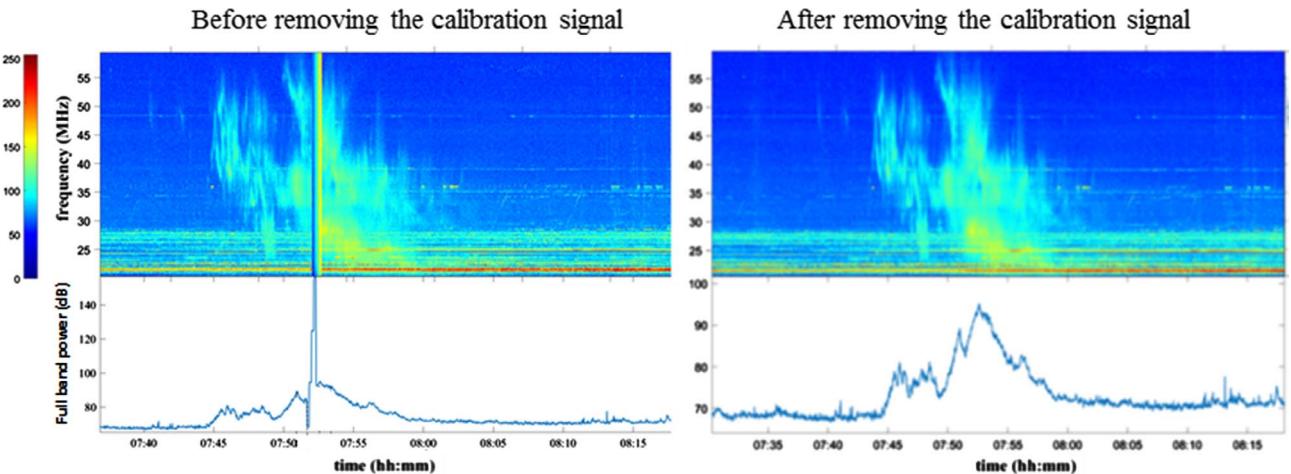


Fig. 5. Example of removing the calibration signal by the method described in Section 4.1. Top panels: dynamic spectra. Bottom panels: time histories integrated over the entire frequency band.

For this reason, the mean value of the RFI dynamic spectrum during a short period and the mean value of the observed signal during a long period are close to each other.

Based on this work, the background noise of the T-F dynamic spectrum $I_c(f, t)$ (obtained after elimination of calibration signals) is reduced using the following equation:

$$I_r(f, t) = \begin{cases} I_c(f, t) - \mu_b(f, t) & \text{if } \mu_b(f, t) \leq (\mu_r(f, t) + k\sigma_r(f, t)) \\ I_c(f, t) - \mu_r(f, t) & \text{if } \mu_b(f, t) > (\mu_r(f, t) + k\sigma_r(f, t)) \end{cases} \quad (1)$$

where $I_r(f, t)$ is the spectrum obtained after removing the background noise. We consider that the background noise of the dynamic spectrum $I_c(f, t)$ follows a Gaussian distribution

$N(\mu_r(f, t), \sigma_r(f, t))$. $\mu_r(f, t)$ and $\sigma_r(f, t)$ are evaluated using a long period of observation. The variability of any point on the dynamic spectrum $I_c(f, t)$ is represented by the mean value of the spectrum $\mu_b(f, t)$ during a short period of $T = 60$ s (Note: the value of 60 s is only used to represent the variability of the spectrum in a short time period). The domain of coverage of the normal random variable $N(\mu, \sigma)$ is $[\mu - k\sigma; \mu + k\sigma]$ where k is a real constant chosen in the interval [1 3] (Because in a Gaussian distribution, 99.7% of the data are within 3 standard deviations of the mean (Czaplicki, 2014)).

$$\mu_b(f, t) = \frac{1}{T} \sum_{t-T}^t (I_c(f, t)) \quad T = 60 \text{ s} \quad (2)$$

$$\mu_r(f, t) = \frac{1}{T} \sum_{t-T}^t (I_c(f, t)) \quad T = 7 \times 3600 \text{ s} \quad (3)$$

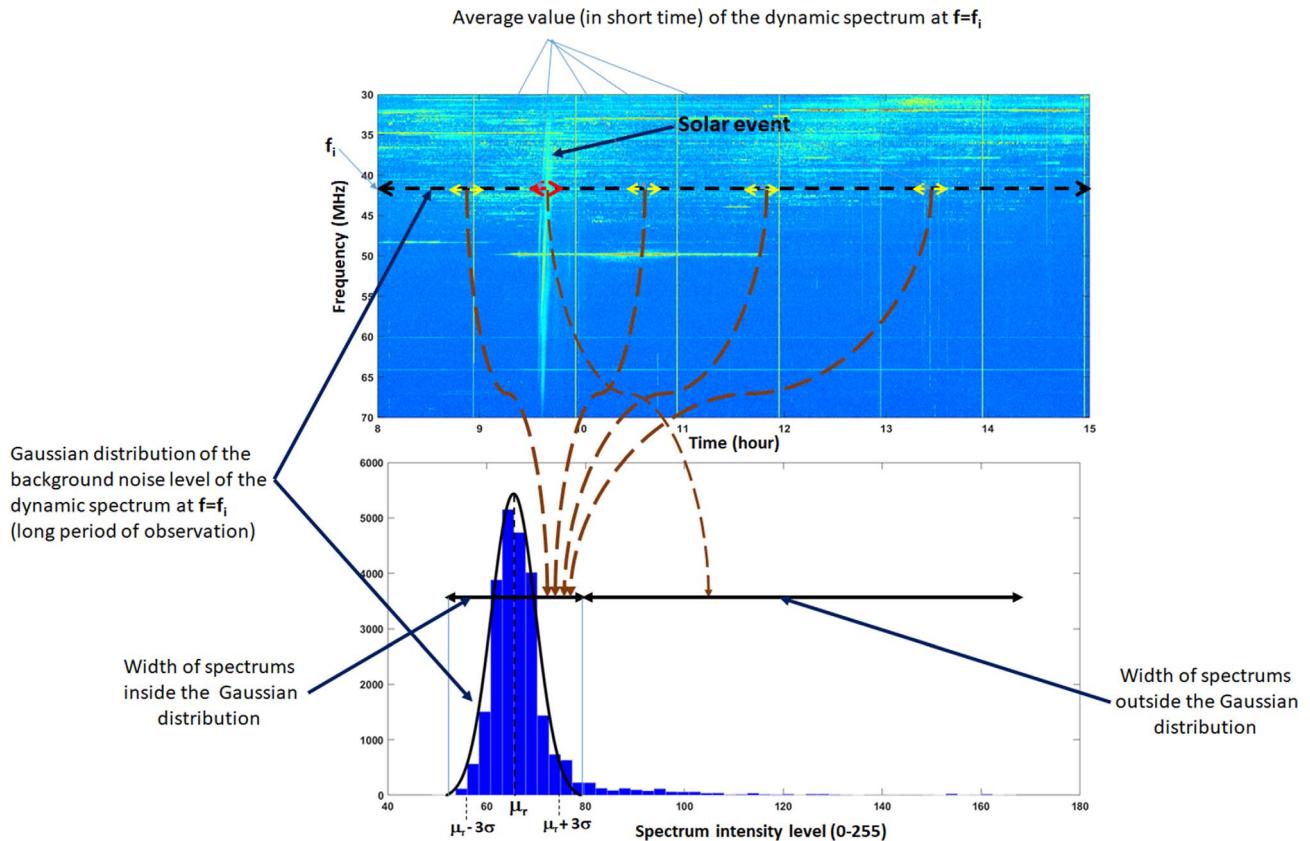


Fig. 6. Representation of the Gaussian distribution of the Histogram of flux densities of T-F dynamic spectrum for a long observation period and at $f = f_i$.

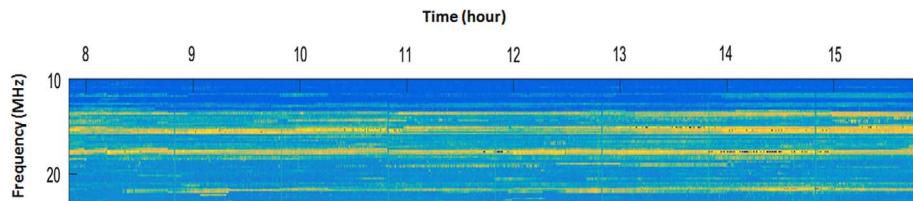


Fig. 7. Example of the RFI spectrum (yellow part) on a Time-Frequency dynamic spectrum.

$$\sigma_r(f, t) = \sqrt{\frac{1}{T} \sum_{t-T}^t (I_c(f, t) - \mu_r(f, t))^2} \quad T = 7 \times 3600 \text{ s} \quad (4)$$

4.3 Median filtering

To complete this process, we use a median filter that takes into account the direction of the flux density gradient in the time-frequency plane to emphasize the solar burst spectra and to remove parasitic signals, artifacts, and remaining RFI outliers on the T-F dynamic spectrum (Storath & Weinmann, 2017). The idea of using this median filter based on the direction of the spectral gradient is to smooth the spectrum of the

solar events (vertical filtering for events of type III, horizontal filtering for events of type IV or inclined filtering for events of type II) in order to remove outliers values (an outlier is the value that is distant from the other values in the window of a median filter). In Appendix B we show the procedure that we use to calculate the median-filtered dynamic spectrum $I_d(f, t)$.

As a conclusion, to eliminate the noise from the original spectrum, we proceed in our proposed method to subtract different mean values from the spectrum $I_c(f, t)$. Then once the noise is reduced, we add to the filtered spectrum $I_d(f, t)$ and the mean value of the spectrum $I_c(f, t)$ over a png time interval in order to rescale the spectrum and properly visualize the signal after reducing noise and RFI background signals (see Fig. 8). The final spectrum $I_{final}(f, t)$ obtained

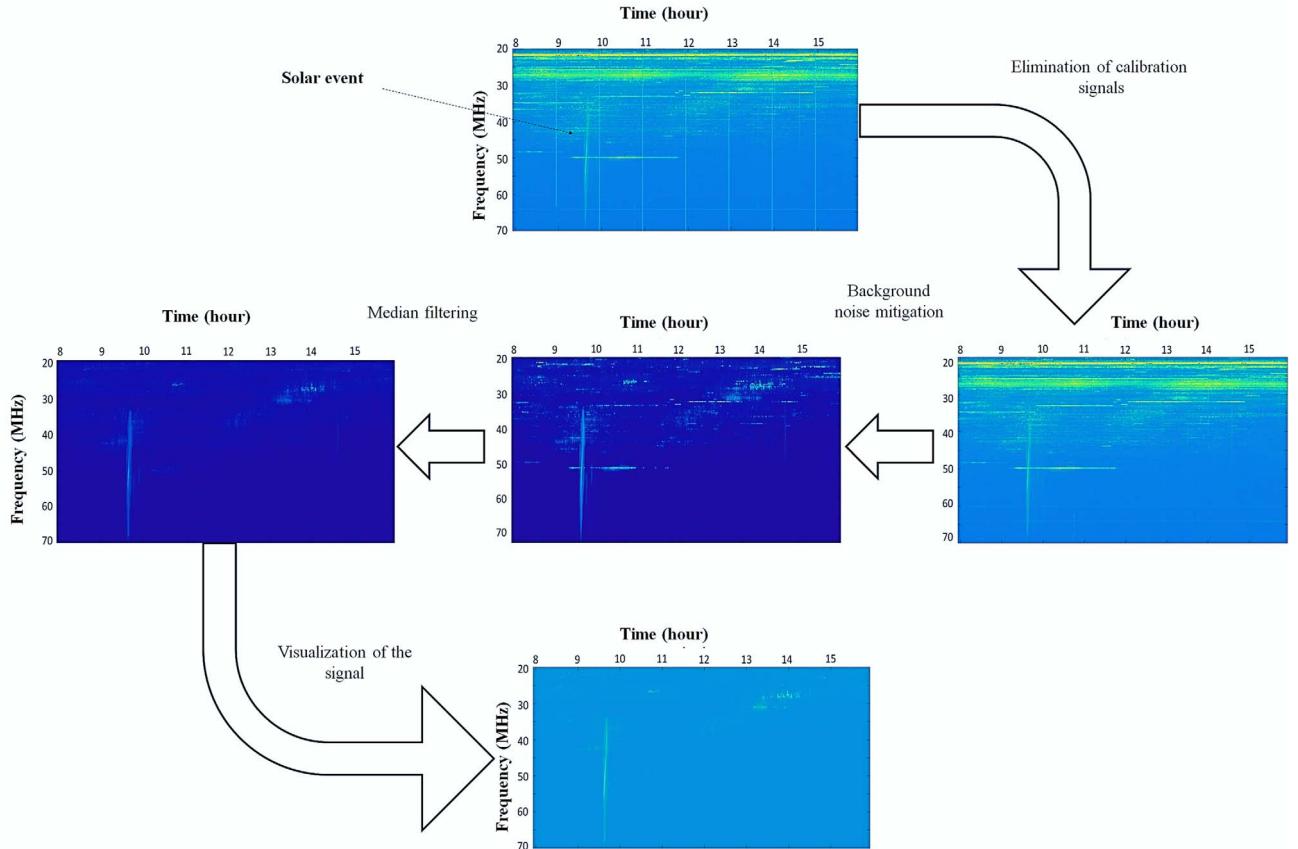


Fig. 8. Example showing the dynamic spectrum obtained at each step of the calibration and background noise removal process.

by the proposed noise removal process is expressed as follows:

$$I_{\text{final}}(f, t) = I_d(f, t) + \mu_g(f, t), \quad (5)$$

$$\mu_g(f, t) = \frac{1}{T \times F} \sum_{f=10 \text{ MHz}}^{80 \text{ MHz}} \sum_{t=1}^{t-T} I_c(f, t), \quad (6)$$

where $\mu_g(f, t)$ is the mean value of the dynamic spectrum data $I_c(f, t)$ for a long period $T = 7 \times 3600 \text{ s}$ (7 h) and for a frequency range 10–80 MHz. This choice indicates that the overall average of the real signal received in the dynamic spectrum reaches at instant t and for the frequency f the value μ_g .

Figure 9 shows the dynamic spectrum as it appears after the treatment described so far (top panel), together with the original observations (third panel from top). For comparison, the second panel shows the results of the method of Zarka et al. (2004) (based on the assumption of the Gaussian distribution of the signal). The three bottom panels represent for each method the signal to noise ratio (SNR) obtained for five solar events that appear in the dynamic spectra in the upper panels. This example illustrates the fact that the process of combining the natural Gaussian distribution of the signal and its variability in time is an effective way to remove both the RFI and galactic noise signals. This strategy is also effective to focus on solar events (higher SNR ratio), as we can see in

Figure 10, where the signal to noise ratios (SNR) of all the events detected on 2 April 2004 (17 events) are significantly increased after applying the proposed noise removal process.

5 Event detection

5.1 Detection of the signal of interest

Once the T-F dynamic spectra have been cleaned, the next step is to identify all solar radio events in these spectra. Figure 11 shows the methodology used to detect the signal of interest that could represent a solar event. The proposed method is a decision rule between two main hypotheses:

The first hypothesis H_0 represents the non-event case where we have only the background noise on the dynamic spectrum or the baseline reference of the signal $s_{\text{final}}(t)$. The second hypothesis H_1 represents the sum of the background noise and the signal of interest (corresponding to an event).

Given a denoised T-F dynamic spectrum $I_{\text{final}}(f, t)$, the denoised signal $s_{\text{final}}(t)$, obtained by integrating the T-F dynamic spectra $I_{\text{final}}(f, t)$ with respect to the frequency axis is expressed as follows:

$$s_{\text{final}}(t) = \sum_{f=10 \text{ MHz}}^{80 \text{ MHz}} (I_{\text{final}}(f, t)), \quad (7)$$

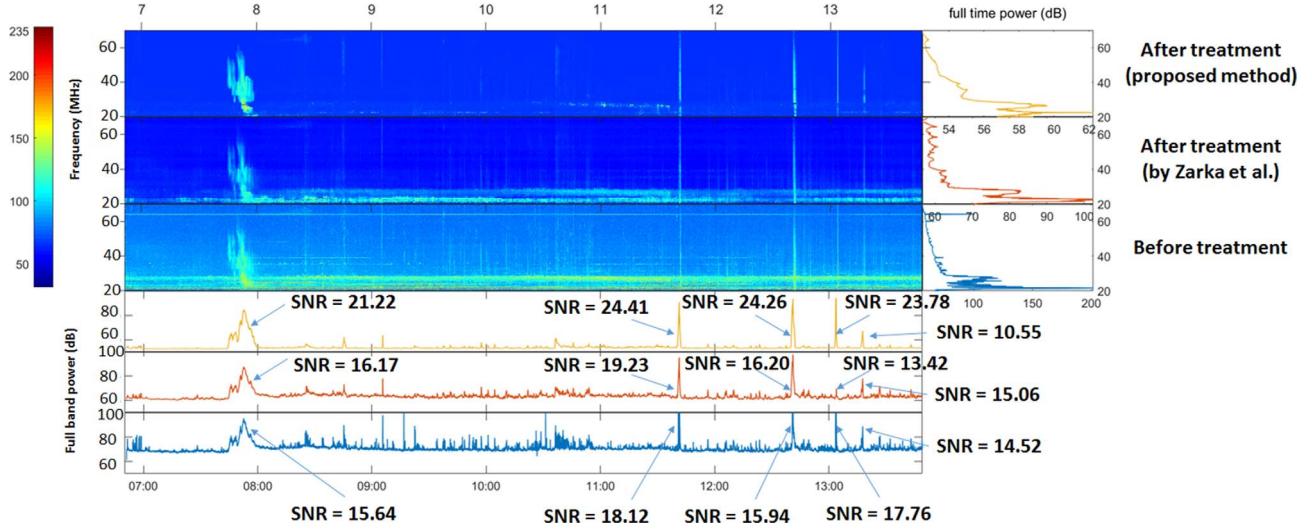


Fig. 9. Example of the dynamical spectrum of a signal by applying Zarka et al. (2004) method and the proposed background noise removal process.

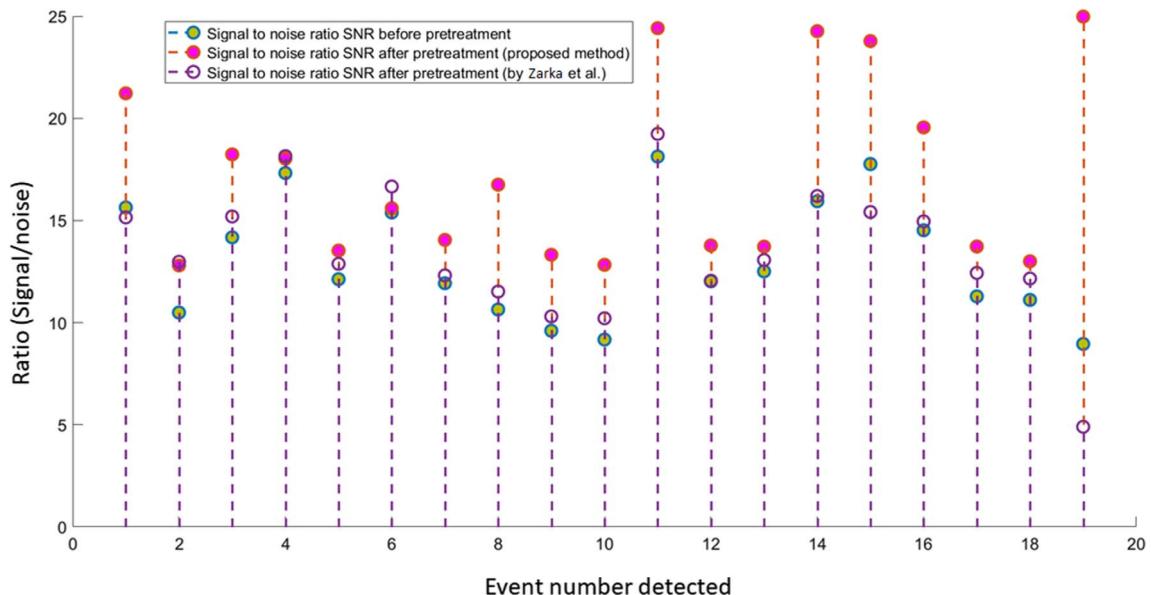


Fig. 10. SNR before and after pretreatment of the data recorded on 2 April 2004.

Many studies, e.g. Horvatic et al. (2011), McCauley et al. (2008) show that data detrending techniques can reduce the baseline shifts or amplitude fluctuations in a signal. As illustrated in Figure 12, due to its non-stationary nature, the signal shows a baseline shift and therefore it is complicated to represent the true amplitude of the signal. In order to remove baseline shifts, we apply a large median filter to the signal $s_{\text{final}}(t)$ and we subtract the result represented by $\Delta lb = \text{median}(s_{\text{final}}(t_i), s_{\text{final}}(t_{i-1}), \dots, s_{\text{final}}(t_{i-\text{win}}))$ from the original signal $s_{\text{final}}(t)$. “win” is the window size of the median filter fixed experimentally to a value of the order of 3600 seconds.

$$s_f(t) = (s_{\text{final}}(t) - \Delta lb)^+, \quad (8)$$

where $s_f(t)$ is the detrended signal and $(x)^+ = \max(x, 0)$. After data detrending, a signal of interest can be detected by thresholding $s_f(t)$ using an adaptive threshold “Th” given by:

$$\text{Th} = K \times Pn_{H_0}, \quad (9)$$

where Pn_{H_0} corresponds to the estimated power of the background noise level evaluated over a time window $T = 7$ h. As presented in the literature of CFAR techniques (Jen, 2011; Barkat, 2005), K is a threshold factor instantly adjusted in order to maintain a constant False Alarm Rate (FAR).

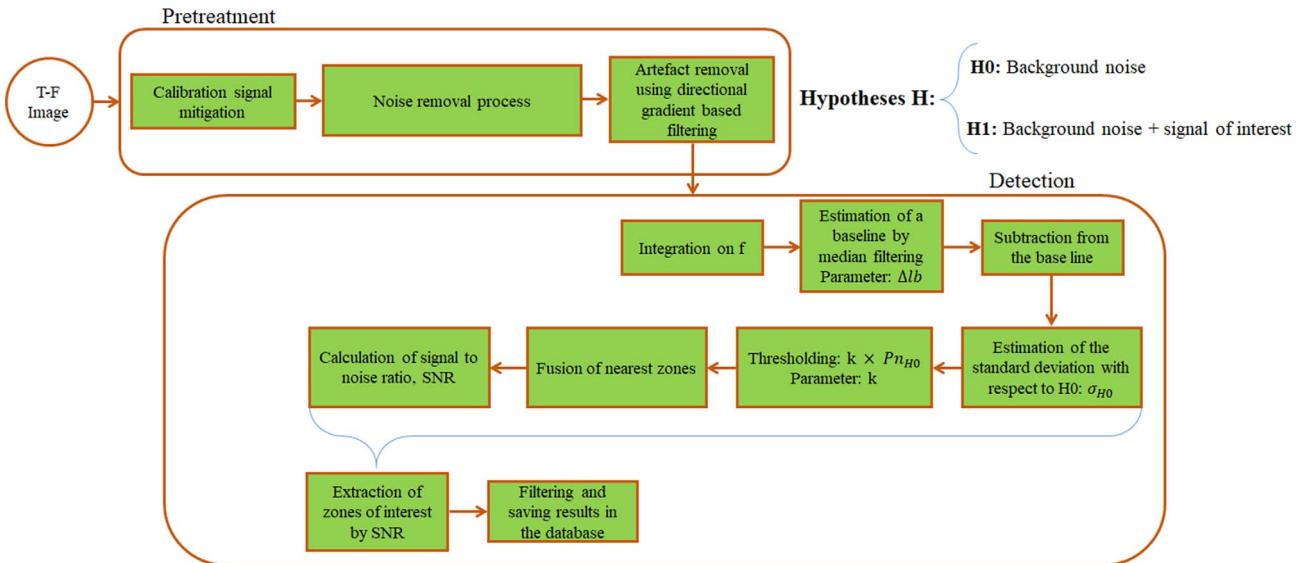


Fig. 11. Synopsis of the proposed detection system.

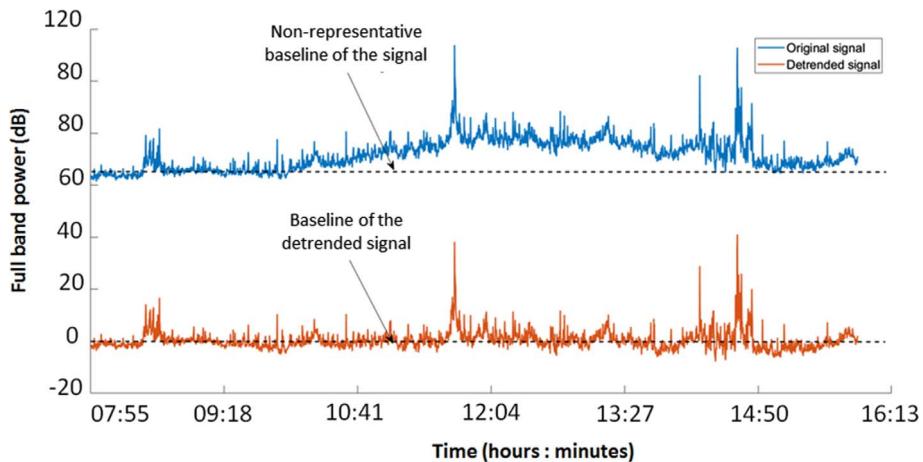


Fig. 12. Solar signal before and after detrending.

From the work of Richards (2005), the threshold factor K can be written as a function of the size of the reference window N (see Fig. 13) and the desired FAR as follows:

$$K = N(\text{FAR}^{-\frac{1}{N}} - 1), \quad (10)$$

Note that the reference window can be chosen left-sided for online implementation purpose. Also, due to mismatch between the assumed i.i.d. Gaussian model and the effective data model, one might not use the theoretical threshold value in equation (10), but rather a value corresponding to the target FAR and based on a learning process from the data.

5.1.1 CFAR-like approach

The role of this method is to compare the signal to a threshold “ Th ”. As shown in Figure 13, the detection of a signal of

interest occurs when the Signal Under Test (SUT) exceeds a threshold. The threshold level (Th) is calculated from the estimated noise level from the N -samples reference window (Eq. (9)). Some guard signals that are immediately adjacent to the SUT are ignored from training signals. The threshold (Th) is also function of the desired FAR. From equations (9) and (10), we can observe that a higher FAR will lead to a lower threshold level. Then a large number of false alarms will mask the detected solar events. Conversely, a lower FAR will lead to a higher threshold level. Then the number of false alarms will be decreased, but only major solar events will be detected, and a lot of low energy solar events will be missed (see illustration in Fig. 14). One of the most widely used methods to address the false-alarm problem is the cell averaging CFAR detector. In this method, the threshold level is calculated by averaging the level of the noise of the reference window according to:

$$\text{Th} = K \times P_{n,H_0}, \quad (11)$$

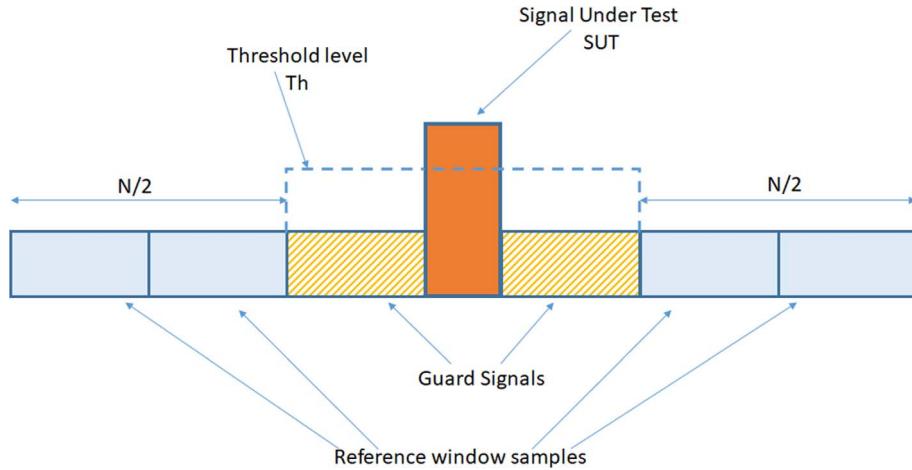


Fig. 13. Principle of CFAR detector.

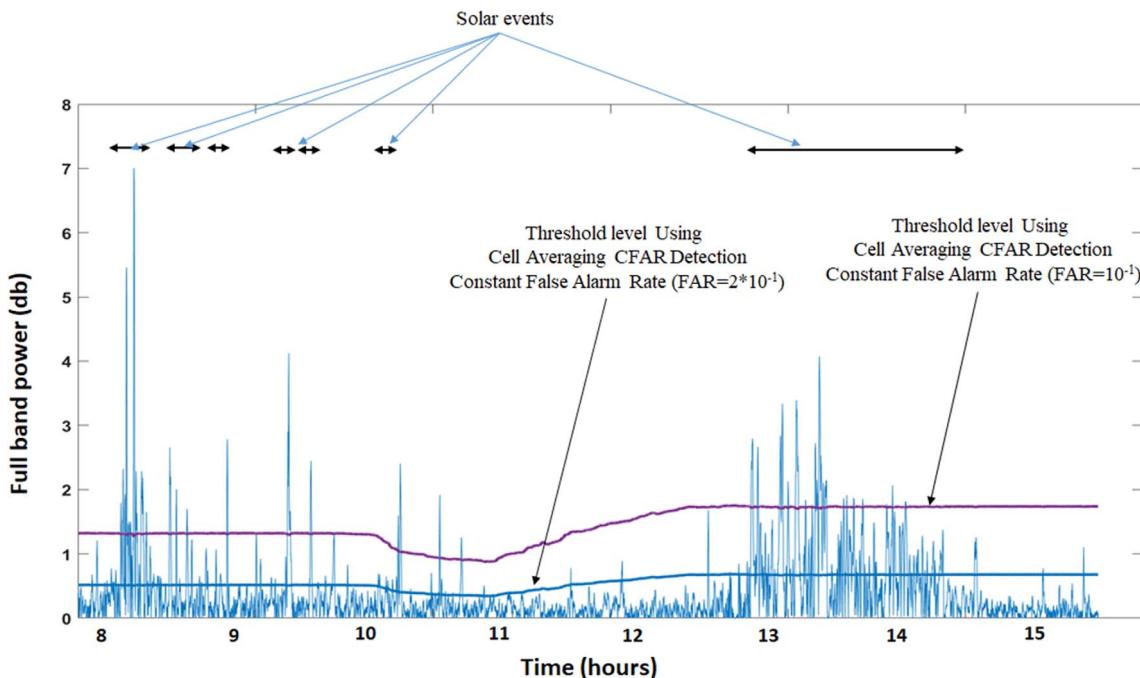


Fig. 14. Comparing the threshold curves obtained by the cell averaging CFAR detector (with two different desired false alarm rate).

$$Pn_{H0} = \frac{1}{N} \sum_{t_1}^{t_N} s_f(t_i). \quad (12)$$

Note that we exclude from the reference window all the past points already detected as “event” signal as well as the SUT surrounding samples (referred to as guard signals) to prevent the energy spillover phenomenon.

On the other hand, the scaling factor K is chosen according to the target FAR and is computed either theoretically (e.g., Eq. (11)) or “experimentally” using training data.

In our case, to take into account the so far ignored time correlation of the signal (i.e. the desired energy signal is not i.i.d.) and the fact that, in practice, a “high energy” event lasts for a

certain duration (e.g. type III or type IV solar events) we use two values of the scaling factor K corresponding to two target FAR values. More precisely, if a “high energy” event is detected, then we switch to a lower value of K (corresponding to a larger value of the FAR) in order to better detect all “small events” surrounding the considered “high energy” event.

5.2 Extraction of the regions of interest

In this section, we proceed to extract all of these regions of solar radio emission based on the signals of interest. The signals of interest (red circles in Fig. 15) are the signals above the threshold level (Th) described in Section 5.1. A region of interest known as a group of signals of interest (events 1–13 in

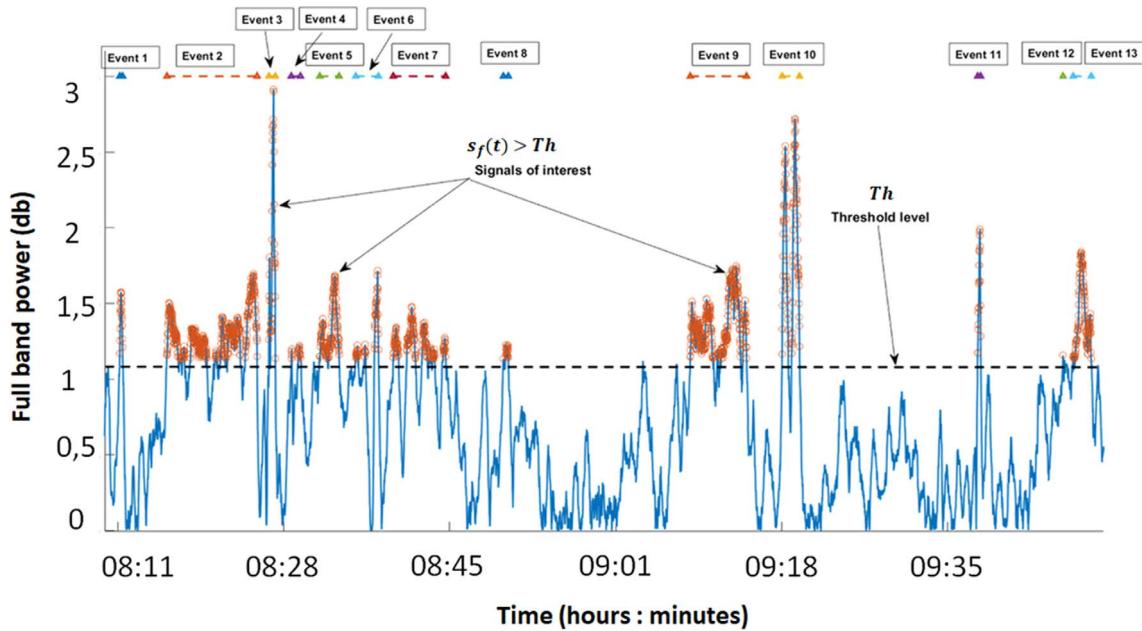


Fig. 15. Illustration of some regions of interest (solar events) extracted by the system (Note that the signal $s_f(t)$ is obtained by integrating the T-F dynamic spectra with respect to the frequency axis).

Fig. 15) is defined as the one corresponding to a possible solar event spectrum of types II, III or IV, which are characterized by specific features described in [Section 2](#).

[Figure 15](#) shows an illustration of the process that we use to find these regions. First, if the distance between two detected signals of interest (or two groups of signals) is less than the minimum distance d_{\min} , calculated by the system, both signals (or groups of signals) are grouped together and are considered to belong to the same event. Second, each group of detected signals of interest ($s_f(t) > Th$) represents a new event if and only if it is far from another group of signals by a minimum distance d_{\min} ².

6 Performance assessment

6.1 Software interface

We present in this section results obtained with the automatic detection of solar radio bursts and compare them with solar events detected manually by an expert. A solar radio interface is then developed for experts to create a data base of all possible events (type II, III, IV or other) or regions of interest that could be used for our performance assessment.

[Figure 16](#) presents an example of the solar bursts labeled manually by experts, using our platform, and bursts detected automatically with the proposed method. We can see on the picture that the largest events (especially type IV bursts) are automatically detected. More details (type III bursts) could also be detected by lowering the threshold of detection (see Eq. (9)).

² Because the shortest solar events are characterized by a duration of 3–10 s ([Boiko et al., 2011](#)), we suppose that two signals (or groups) having a distance less than $d_{\min} = 10$ s represent the same event (signal).

6.2 Data validation

To validate our work, we used archived data (stored in files) from the Nançay Decametre Array. The data are provided in the form of time-frequency dynamic spectra. Each file contains two Time-Frequency (T-F) spectra (400 (frequency) \times 28 700 (time)) that represent left-hand and right-hand circular polarization. In this study, we use low resolution spectral data archive of the left-hand T-F spectra (Frequency resolution: 175 kHz, temporal resolution 1 s).

The daily spectra of June 2012 and June 2014³ were analyzed in detail by experts using our software interface to label solar radio events in T-F spectra (major and minor events) and to build up the ground truth (see [Fig. 16](#)). These labeled data sets are used next to evaluate the performance of our proposed detection system. [Figure 17](#) and [Table 1](#) present the distribution of the data set as a function of event duration (width of events) and SNR. The blue symbols in [Figure 17](#) represent all regions of interest that could correspond to solar events (especially type III bursts that are characterized by short durations and sometimes low signal to noise ratio, as shown in the figure). The big number of detected regions of interest is due to type III bursts. In general a group of type III events which on occasion may be part of a longer-lasting type III storm, is identified as a single solar event, but in the present case we have separately detected all regions of interest that correspond to type III bursts.

6.3 Evaluation criteria

Since the goal of our work is to develop an automated detection method, we defined an evaluation protocol based on

³ We have selected these two months for our analysis because they comprise many solar events of type III and storms and some events of type II and IV.

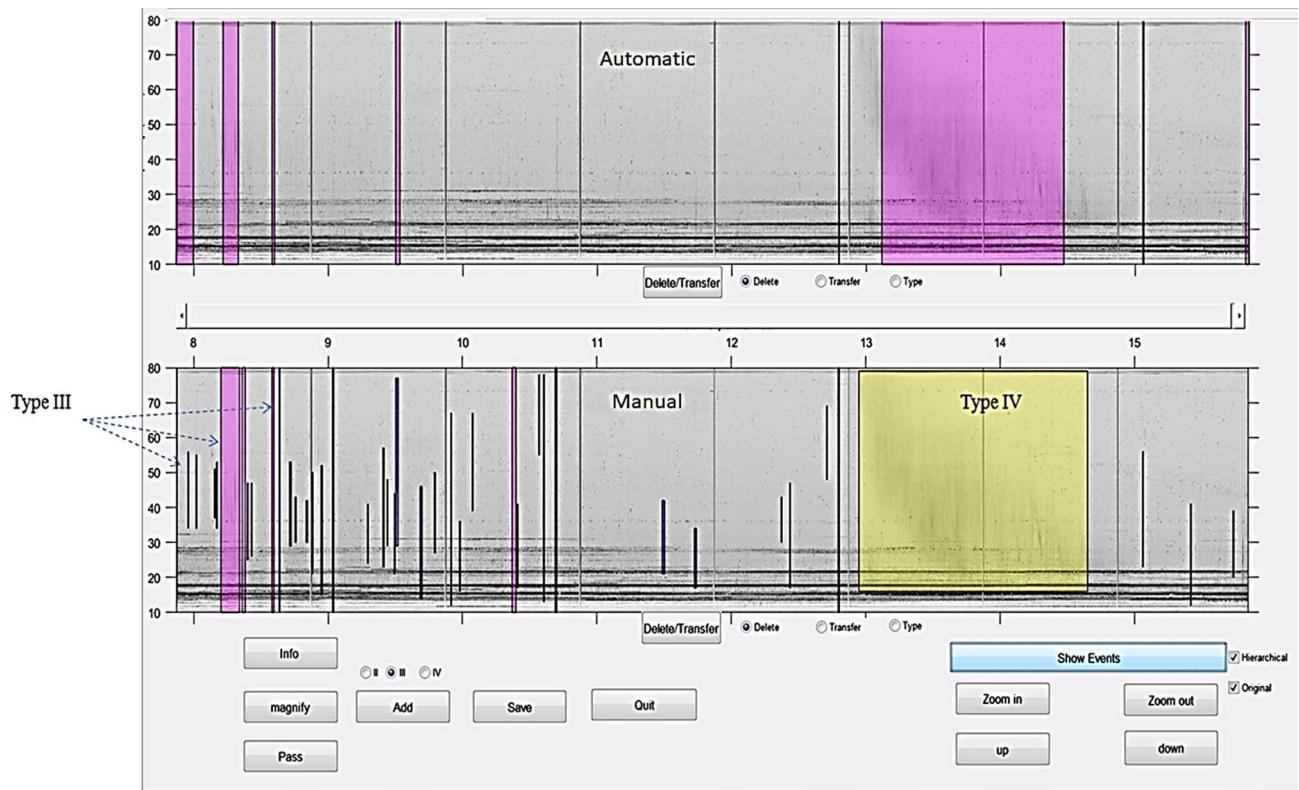


Fig. 16. Events detected by the proposed method and manually for the day of 15 June 2014. The spectrum on the top shows the patterns detected automatically, and the spectrum on the bottom the patterns detected manually.

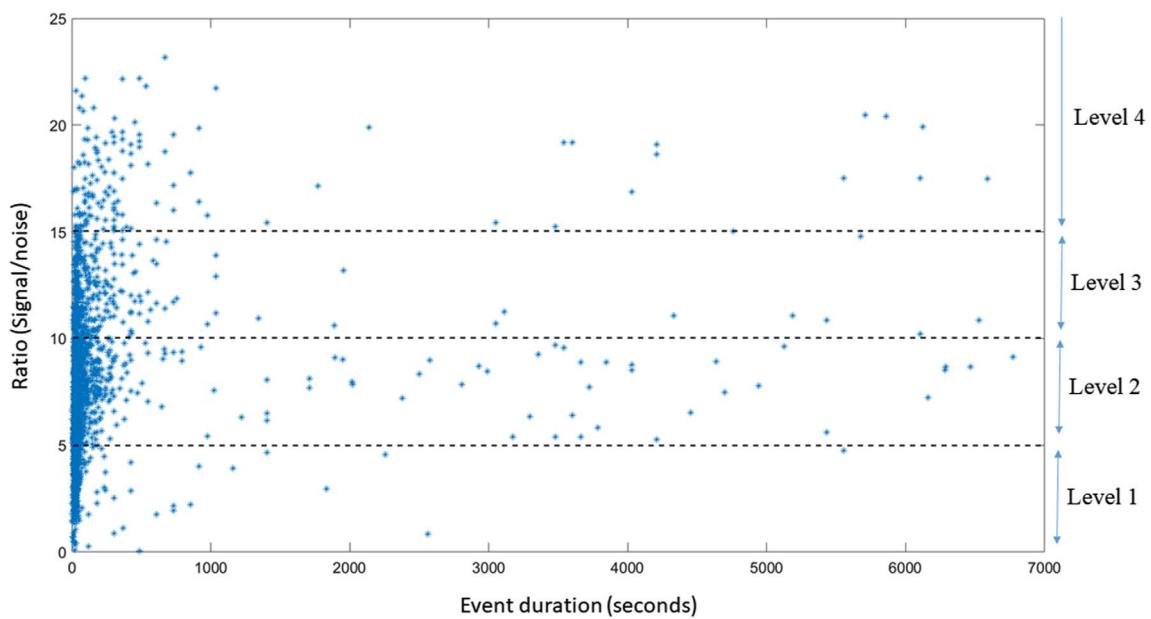
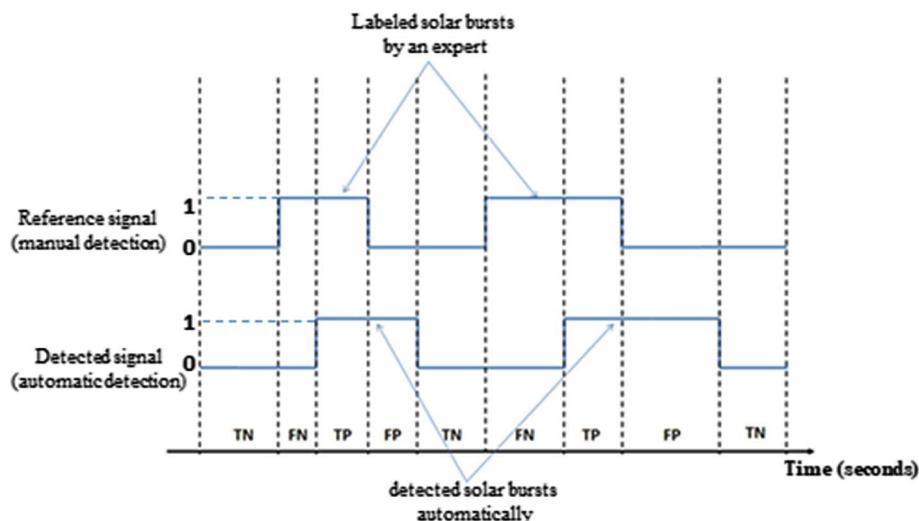


Fig. 17. Presentation of the data-set detected manually to label solar events of June 2012 and June 2014.

Table 1. Table of distribution of the data-set of solar events detected manually from June 2012 and June 2014.

Type of events	Number of events	Mean duration (s)	Signal to noise ratio (SNR)			
			Level 1 (SNR ∈ [0,5]) (%)	Level 2 (SNR ∈ [5,10]) (%)	Level 3 (SNR ∈ [10,15]) (%)	Level 4 (SNR > 15) (%)
III	2182	222	20.6	55	18.3	6.1
IV	7	5131	6	47	29.4	17.6
II	6	2821	0	66.7	9	24.3

**Fig. 18.** Evaluation protocol used to evaluate the events detected automatically by the proposed method (1 if a solar burst event is detected and 0 if no).

finding the common period between events detected by the proposed method and events detected manually (reference signals; see Fig. 18).

The detection process is then evaluated by computing the sensitivity, the specificity, the precision and the accuracy (Loong, 2003) as follows:

TP: True Positive represents the time period (in seconds) of the received signal where the solar burst exists in the reference signal and in the detected signal (see Fig. 15).

TN: True Negative represents the time period of the received signal where the solar burst exists neither in the reference signal nor in the detected signal.

FP: False Positive represents the time period of the received signal where the solar burst does not exist in the reference signal but exists in the detected signal.

FN: False Negative represents the time period of the received signal where the solar burst exists in the reference signal, but not in the detected signal.

Sensitivity: $Se = 100 \times \frac{TP}{(TP+FN)}$ measures the capacity of the system to detect the events labeled by an expert (reference signal).

Specificity: $Sp = 100 \times \frac{TN}{(TN+FP)}$ measures the capacity of the system to not detect an event unlabeled by an expert.

Precision: $P = 100 \times \frac{TP}{(TP+FP)}$ measures the capacity of the system to detect exactly (same time period) the events labeled by an expert.

Accuracy: $R = 100 \times \frac{(TP+TN)}{(TP+FP+TN+FN)}$ measures the probability that the system can detect the same situations (bursts or not) obtained in the reference signal.

In fact, the concept of TP, TN, FP, FN is quite standard and often used in detection and classification theory (Fawcett, 2006). In our work, we use this method of evaluation to compare the duration between the labeled events by an expert (representing ground truth) and the duration of the events detected by the Automated Radio Burst Identification System (ARBIS) (Lobzin et al., 2009) or by our own method. Figure 19 represents the detection of a solar event and the evaluation protocol used in our simulations. From this figure, we show that the correct detection of a solar event (called TP) is represented by the period of time identified as a solar event both manually by an expert and automatically by using our method or ARBIS. The false detection of solar event (called FP) is represented by the period of time not identified by an expert as a real solar event but only detected automatically by our method as an event. By the same manner we can define on Figure 19 the period of times that represent the evaluation criteria of False Negative FN and True Negative TN.

Looking at the definition of these evaluation criteria, we can note that high values of sensitivity and precision rate can be achieved if and only if the start and the end of the events detected by the system are the same as those labeled by an

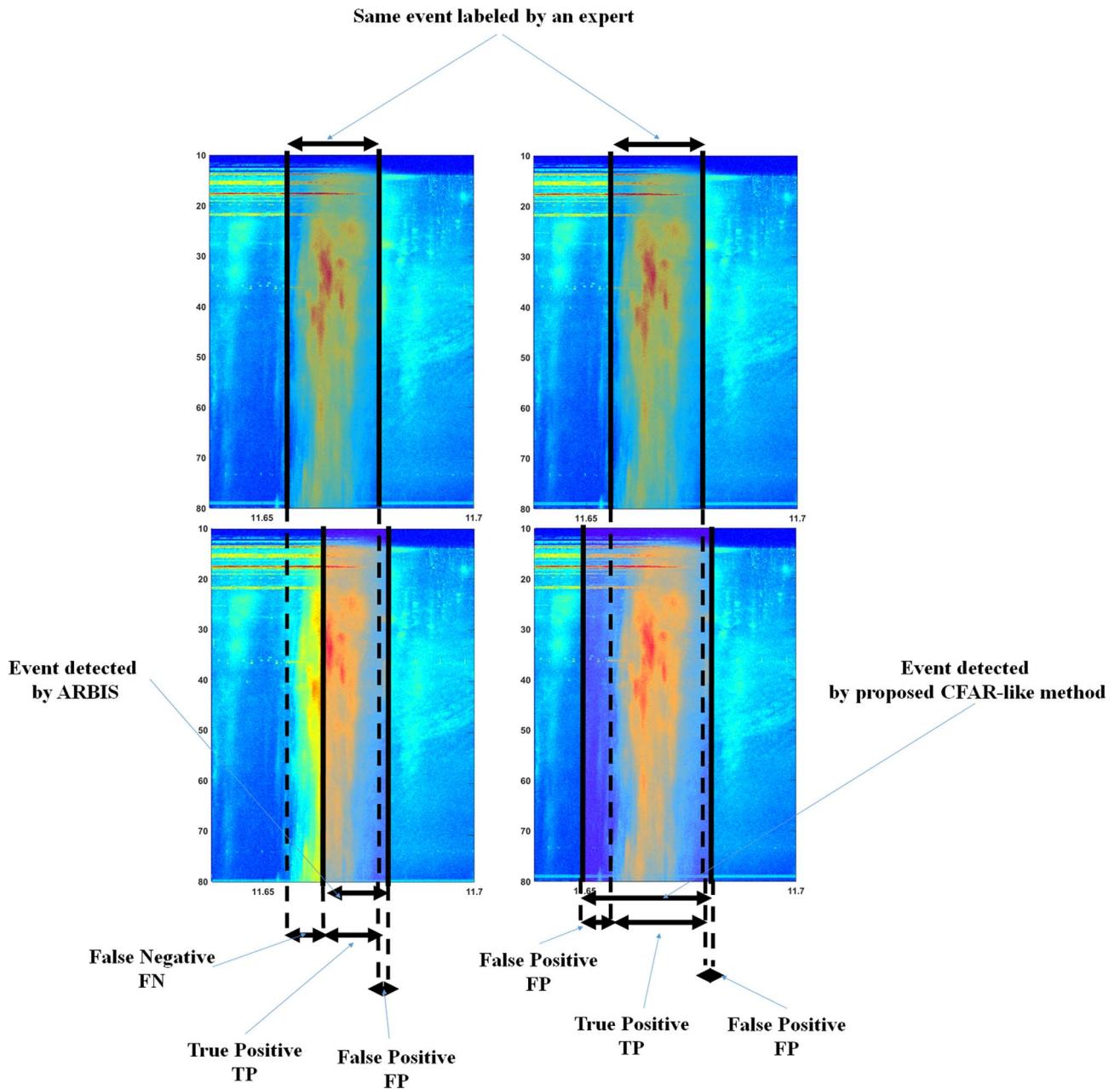


Fig. 19. Protocol used to evaluate the duration of the events detected by the method proposed in the present work and the ARBIS system in comparison with the duration of the event (start and end of the event) labeled by an expert.

expert. This is almost impossible to have, because most events are detected in short time intervals (few seconds). For this reason, we are more interested to the measurements obtained in specificity and accuracy in order to evaluate the global performance of the system.

6.4. Automatic detection by the ARBIS method

ARBIS is an automated method for recognizing solar radio bursts (especially type III bursts) based on radio spectrograph data. The method is considered here for comparison purposes. Its central idea is to use the Radon transform for a more objective burst detection in dynamic

spectra. The algorithm of this method is summarized as follows:

1. The first stage of the algorithm is a moving average filtering of the dynamic spectrum $I(f, t)$ with respect to time. Filtering of dynamic spectra is usually required to reject undesired signals and to emphasize the regions of interest (solar bursts).
2. In the second stage, the filtered spectrum $I_{\text{filtered}}(f, t)$ is converted to a binary spectrum $B(f, t)$. The binary spectrum is chosen to be equal to 1, i.e., $B(f, t) = 1$, if the spectrum value $I_{\text{filtered}}(f, t)$ corresponds to a local maximum with respect to time, i.e., $I_{\text{filtered}}(f, t - 1) \leq I_{\text{filtered}}$

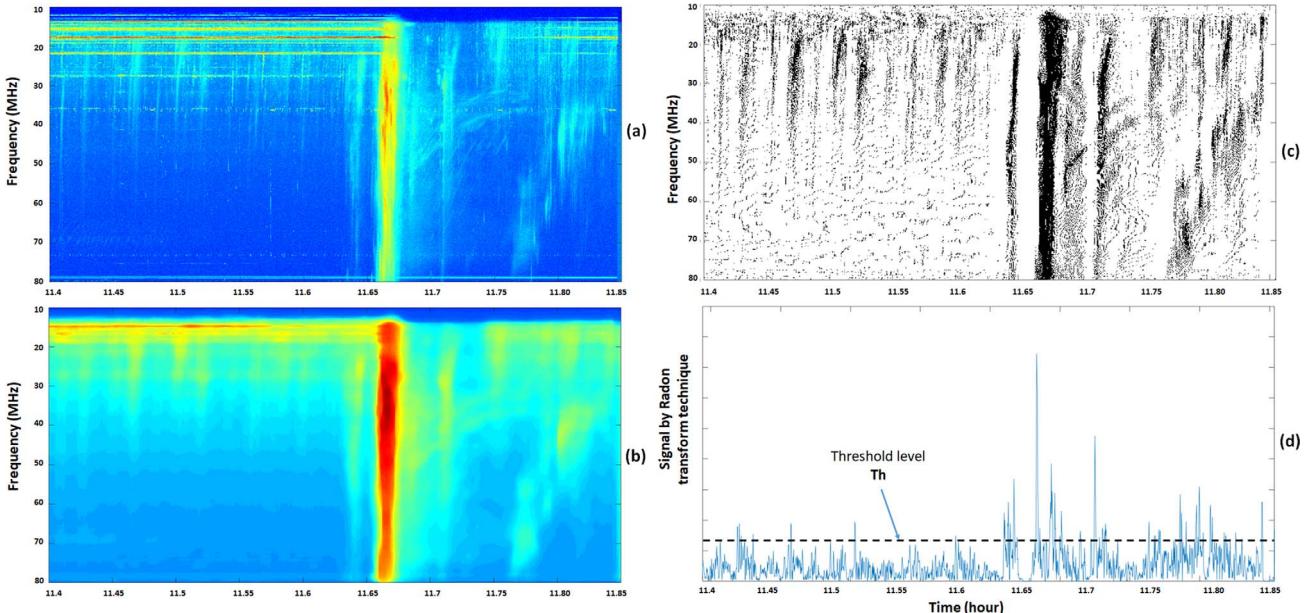


Fig. 20. Visualization of the different stages of the ARBIS system.

Table 2. Performance results of the detection of type III bursts obtained by the proposed solar burst detection method and the ARBIS system.

	Original spectrum removal process (%)	Background removal process (%)	ARBIS detection method (%)
Sensitivity = $\frac{TP}{TP+FN}$	97.6	40.3	30.5
Specificity = $\frac{TN}{FP+TN}$	23.8	89.2	75.8
Precision = $\frac{TP}{TP+FP}$	31.9	70.6	56.7
Accuracy = $\frac{TP+TN}{TP+FP+FN+TN}$	43.2	81.2	70.1

$(f, t) \geq I_{\text{filtered}}(f, t + 1)$ otherwise $B(f, t) = 0$. This step aims to find whether there is a significant enhancement in the spectrum.

- Then a technique based on a Radon transform of the binary spectrum $B(f, t)$ is used to find the features corresponding to solar bursts (in particular type III bursts) (Lobzin et al., 2009). The signal obtained by applying this technique is compared with a threshold to decide whether a signal of interest is observed or not. When a signal is detected, the optimal threshold value is determined empirically by assuming that the acceptable probability of finding a false solar burst in one spectrum is 10^{-5} .
- Finally, the detected signals of interest are combined into a group if they are less than 1 min distant (this choice by the authors is defined based on the resolution of the radio spectrograph and the average duration of solar events).

Figure 20 presents the different steps of ARBIS system. It shows the original and filtered dynamic spectrum with a solar burst in the middle (Figs. 20a and b), the corresponding binary

spectrum (Fig. 20c), and the signal obtained after applying the Radon transform technique (Fig. 20d).

6.5 Performance evaluation

Table 2 shows the results (sensitivity, specificity, precision and accuracy) obtained for the months of June 2014 and June 2012 using the ARBIS system (Lobzin et al., 2009) and the proposed event detection method. Because the ARBIS system is used to detect solar bursts of type III, only such events are evaluated. As we can see, the method proposed in the present work detects a huge number of noisy events along with the solar bursts. This is why the system is missing the real duration of events, which appears in the results as a combination of high sensitivity with weak precision and accuracy. After denoising the data, the comparison proves that our method has the ability to detect at least 80% of the events labeled manually by an expert (accuracy = 81.2%, specificity = 89.2%). On the other hand, the sensitivity of the system is decreased without the proposed preprocessing method (sensitivity = accuracy = 40.3%). This is due to the inability of the system to detect the solar events with the durations labeled by experts, i.e. to detect the correct start and end times of the solar bursts.

7 Conclusions

The effective detection of solar radio bursts is a key problem to forecast space weather events related to solar eruptive activity. It is also a tool for searches of specific types of radio bursts in large data sets. In this work, an automated method for the detection of solar radio bursts in dynamic spectra is proposed. This method starts by eliminating unwanted signals (Radio-Frequency Interference, RFI, Calibration ...). Then a specific filter is applied to denoise and improve the quality of the dynamical spectra to be interpreted. When the preprocessing phase is finished, an automated solar radio burst

detection system based on an adaptive constant false alarm rate (CFAR-like) is applied. To evaluate the performance of our method, a semi-automatic software package has been developed to create a data set of all possible events (type II, III or IV) that could be recognized. Finally, our proposed system performs better than the Automated Radio Solar Burst Identification System (ARBIS), which has been developed for type III burst detection objective (accuracy level: proposed system 81% vs. ARBIS 70%). In addition to this gain, the proposed method allows the detection of any high or moderate energy event that can be of interest for radio astronomers.

Acknowledgements. This work is developed within the framework of the *Agence Nationale pour la Recherche* (ANR/ASTRID,DGA) project *Outils radioastronomiques pour la météorologie de l'espace* (ORME, contact No. ANR-14-ASTR-0027). The editor thanks two anonymous referees for their assistance in evaluating this paper.

References

- Bandiera F, Orlando D, Ricci G. 2009. CFAR detection strategies for distributed targets under comic constraints. *IEEE Trans Signal Process* **57(9)**: 3305–3316.
- Barkat M. 2005. *Signal detection and estimation*, 2nd edn. Artech House.
- Boiko A, Mel'Nik V, Konovalenko A, Rucker H, Abranin E, Dorovskyy V, Lecacheux A. 2011. Frequency drift rates of powerful decameter Type III bursts. *Adv Astron Space Phys* **1**: 57–60.
- Bonnin X, Aboudarham J, Fuller N, Renie C, Perez-Suarez D, Gallagher P, Higgins P, Krista L, Csillaghy A, Bentley R. 2011. Automated detection and tracking of solar and heliospheric features in the frame of the European project HELIO. *SF2A-2011: Proc. Annu. Meeting French Soc Astron Astrophys* **373**: 377.
- Bothmer V, Daglis IA. 2007. *Space weather: physics and effects*. Springer Science & Business Media.
- Cai L, Ma X, Xu Q, Li B, Ren S. 2011. Performance Analysis of Some New CFAR Detectors under Clutter. *J Comput* **6(6)**: 1278–1285.
- Cairns I, Robinson R. 1987. Herringbone bursts associated with type II solar radio emission. *Sol Phys* **111(2)**: 365–383.
- Carley EP, Reid H, Vilmer N, Gallagher PT. 2015. Low frequency radio observations of bi-directional electron beams in the solar corona. *A&A* **581**: A100.
- Czaplicki JM. 2014. *Statistics for Mining Engineering*. CRC Press.
- Deans SR. 2007. *The Radon transform and some of its applications*. Dover Books on Mathematics Series. Dover Publications, ISBN 9780486462417. URL <https://books.google.fr/books?id=xSCc0KGiu0C>.
- Fawcett T. 2006. An introduction to ROC analysis. *Pattern Recogn Lett* **27(8)**: 861–874.
- Fry EK. 2012. The risks and impacts of space weather: Policy recommendations and initiatives. *Space Policy* **28(3)**: 180–184.
- Gunes T, Erdol N. 2006. HMM based spectral frequency line tracking: Improvements and new results. In: *Acoustics, Speech and Signal Processing, 2006, ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, **2**, IEEE.
- Horvatic D, Stanley HE, Podobnik B. 2011. Detrended cross-correlation analysis for non-stationary time series with periodic trends. *EPL* **94(1)**: 18007.
- Jen JJ. 2011. A study of CFAR implementation cost and performance tradeoffs in heterogeneous environments, *Ph.D. Thesis*, California State Polytechnic University, Pomona.
- Jones J, Richards GP. 2014. Automated recognition of type III solar radio bursts using mathematical morphology. In: *Advanced Maui Optical and Space Surveillance Technologies Conference*.
- Kataoka R, Sato T, Hiroshi Y. 2011. Predicting radiation dose on aircraft from solar energetic particles. *Space Weather* **9(8)**: 1–2.
- Kong X-L, Chen Y, Li G, Feng S-W, Song H-Q, Guo F, Jiao F-R. 2012. A broken solar type II radio burst induced by a coronal shock propagating across the streamer boundary. *ApJ* **750(2)**: 158.
- Lecacheux A. 2000. The Nançay decameter array: A useful step towards giant, new generation radio telescopes for long wavelength radio astronomy. *Radio Astronomy at Long Wavelengths* **119**: 321–328.
- Lobzin VV, Cairns IH, Robinson PA, Steward G, Patterson G. 2009. Automatic recognition of type III solar radio bursts: automated radio burst identification system method and first observations. *Space Weather* **7(4)**.
- Lobzin VV, Cairns IH, Robinson PA, Steward G, Patterson G. 2010. Automatic recognition of coronal type II radio bursts: the automated radio burst identification system method and first observations. *ApJ* **710(1)**: L58.
- Loong T-W. 2003. Understanding sensitivity and specificity with the right side of the brain. *BMJ* **327(7417)**: 716–719.
- McCauley JL, Bassler KE, Gunaratne GH. 2008. Martingales, detrending data, and the efficient market hypothesis. *Phys A* **387(1)**: 202–216.
- Mukhopadhyay P, Chaudhuri BB. 2015. A survey of Hough transform. *Pattern Recognition* **48(3)**: 993–1010.
- Munro R, Gosling J, Hildner E, MacQueen R, Poland A, Ross C. 1979. The association of coronal mass ejection transients with other forms of solar activity. *Sol Phys* **61(1)**: 201–215.
- Nindos A, Aurass H, Klein K-L, Trottet G. 2008. Radio emission of flares and coronal mass ejections. *Sol Phys* **253(1–2)**: 3.
- Nita GM, Gary DE, Liu Z, Hurford GJ, White SM. 2007. Radio frequency interference excision using spectral-domain statistics. *PASP* **119(857)**: 805.
- Pick M, Vilmer N. 2008. Sixty-five years of solar radioastronomy: flares, coronal mass ejections and Sun–Earth connection. *A&ARv* **16(1–2)**: 1–153.
- Richards MA. 2005. *Fundamentals of radar signal processing*. Tata McGraw-Hill Education.
- Storath M, Weinmann A. 2017. Fast median filtering for phase or orientation data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Tumilaar K, Langi Y, Rindengen A. 2015. Hidden Markov Model. *Cartesian* **4(1)**: 86–94.
- Zarka P, Cecconi B, Kurth W. 2004. Jupiter's low-frequency radio spectrum from Cassini/Radio and Plasma Wave Science (RPWS) absolute flux density measurements. *J Geophys Res: Space Phys* **109 (A9)**: 1–12.
- Zhang Y, Du A, Du D, Sun W. 2014. Evaluation of a Revised Interplanetary Shock Prediction Model: 1D CESE-HD-2 Solar-Wind Model. *Sol Phys* **289(8)**: 3159–3173.

Cite this article as: Salmane H, Weber R, Abed-Meraim K, Klein K & Bonnin X. 2018. A method for the automated detection of solar radio bursts in dynamic spectra. *J. Space Weather Space Clim.* **8**, A43.

Appendix A

Methodology to eliminate the calibration signals

The core of the process to eliminate the calibration signals (see Fig. A.1) consists of finding the start time t_s and the end time $t_e = t_s + 40$ (40 s is the known duration of the calibration signal) of the desired pattern as illustrated in Figure A.2. To detect these calibration pulses, we proceed every 1 h to find the signal of the highest amplitude.

We consider that the astrophysical information between the instants t_s and t_e is completely lost due to the calibration process. For this reason, we chose to replace the canceled signal by a smoothed version of the local T-F spectrum, in order to avoid discontinuities and artefacts. For that, we take into account in our calculation the instant $t_b = t_s - \Delta T$ just before the calibration process and the instant $t_a = t_e + \Delta T$ just after the calibration process. $\Delta T = 15$ s is fixed experimentally to

represent a short duration around the calibration pulse. Let mean_b be the mean value of the flux density $I(f, t)$ calculated just before the calibration, between the instants t_b and t_s and mean_a the mean value of the flux density $I(f, t)$ calculated just after the calibration, between the instants t_e and t_a . For each frequency at time $t_i \in [t_s, t_e]$ we replace the data $I(f, t_i)$ ⁴ by a linear interpolation between the mean values mean_b and mean_a .

Appendix B

Algorithm of median filtering

Given the dynamic spectrum $I_r(f, t)$ obtained after background noise removal process, let (f_i, t_i) be the coordinates of the i-th pixel of the spectrum $I_r(f, t)$.

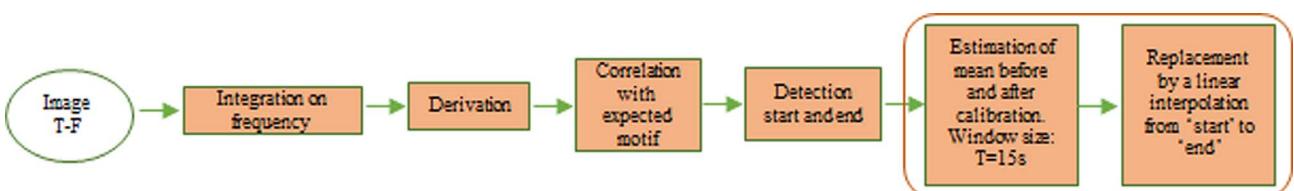


Fig. A.1. Synopsis of eliminating calibration signals.

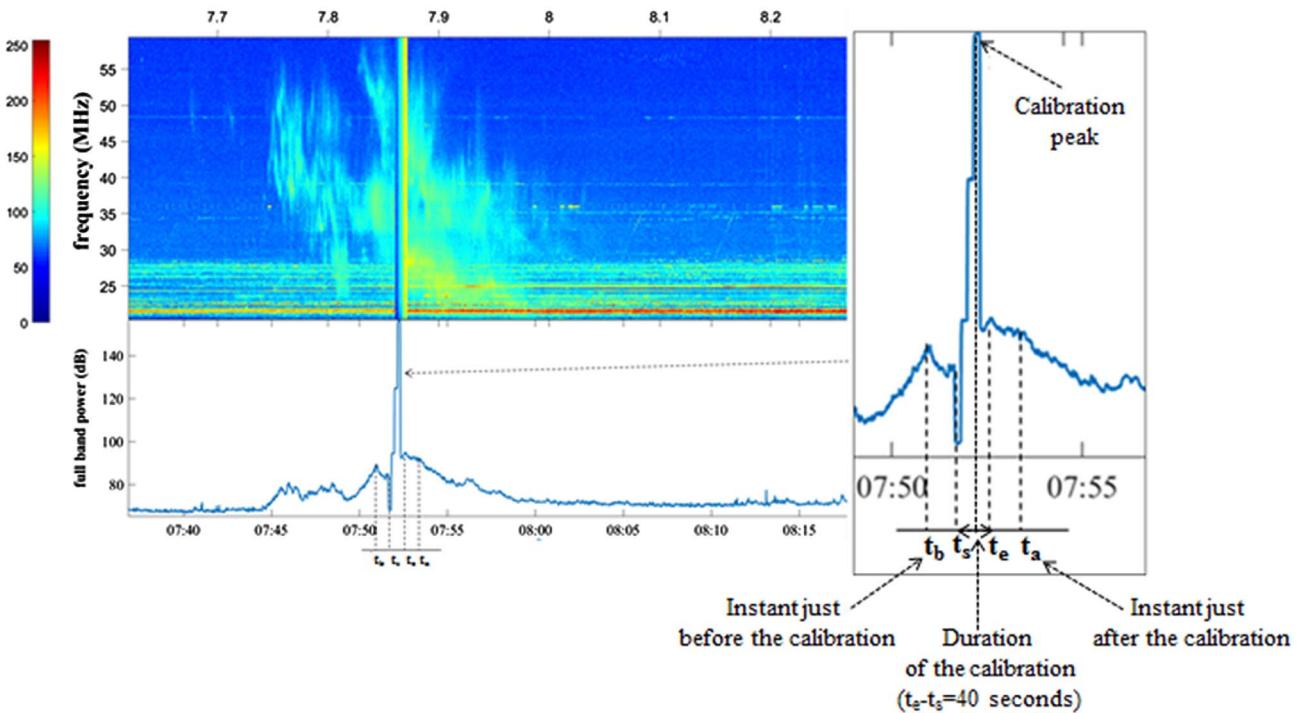


Fig. A.2. Pattern of a calibration signal.

⁴ The spectrum is given with a frequency sampling interval of $\Delta f = 0.175$ MHz in the range [10–80 MHz] and a time sampling interval of $\Delta t = 1$ s.

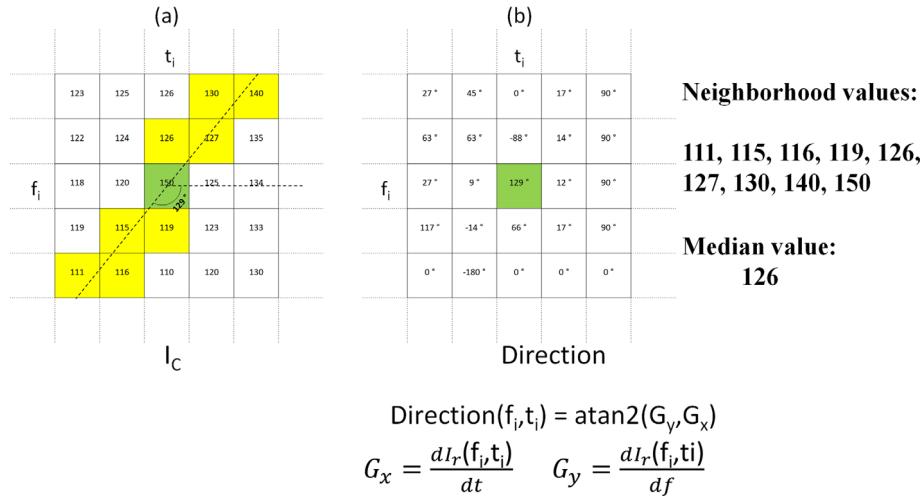


Fig. B.1. Calculating the median value of a pixel neighborhood $I_r(x_i, y_i)$. As can be seen, the pixel value of 150 (outlier value) is replaced with the median value: 126. Squares neighborhood around the calculated $\text{Direction}(f_i, t_i)$ (129°), are used here.

We start by calculating the spectrum gradients in the x -direction $G_x = \frac{dI_r(f_i, t_i)}{dt}$ and in the y -direction $G_y = \frac{dI_r(f_i, t_i)}{df}$.

The pixels with the largest gradient values (called edge pixels) are those corresponding to the normal direction of the gradient computed as $\text{Direction}(f_i, t_i) = \text{atan2}(G_y, G_x)$ (see Fig. B.1b).

To construct the median filter, only pixels $(i_0, i_1 \dots, i_k)$ that are located no farther than a maximum Euclidean distance D from the pixel (f_i, t_i) and that are situated around the

direction of edge pixels (called $\text{Direction}(f_i, t_i)$) are selected (see Fig. B.1a). The Euclidean distance D is fixed experimentally to the value 15^5 . The filtered dynamic spectrum pixel $I_d(f_i, t_i)$ is then replaced by the median value of the spectral flux densities $I_r(f_{i_0}, t_{i_0}), I_r(f_{i_1}, t_{i_1}) \dots, I_r(f_{i_k}, t_{i_k})$:

$$I_d(f_i, t_i) = \text{median}(I_r(f_{i_0}, t_{i_0}), I_r(f_{i_1}, t_{i_1}) \dots, I_r(f_{i_k}, t_{i_k})), \quad (\text{B.1})$$

⁵ The value 15 represents a short Euclidean distance. It only depends on the resolution of the dynamic spectrum of the instrument.