PARAMETERS AND DOA ESTIMATION BASED ON WAVELET PACKETS SELECTION FOR TRANSIENT SIGNALS IN COLORED NOISE

Marwa CHENDEB [5] *, Guy PLANTIER* [5]*, Anthony SOURICE* [5] *and CODALEMA Group* [1,...,7]

[1] SUBATECH, Université de Nantes/École des Mines de Nantes/IN2P3-CNRS, 4 rue Kastler F44307, Nantes France.
[2] LESIA, USN de Nançay, Observatoire de Paris-Meudon/CNRS, Meudon France.
[3] LAL, Université Paris-Sud/CNRS, Orsay France. [4] LPSC, Université Joseph Fourier/INPG/CNRS, Grenoble France.
[5] Groupe Signal Image et Instrumentation, École Supérieure de l'Électronique de l'Ouest
4, rue merlet de la Boulaye 49000, Angers, France
phone: + (33) 2 41 86 67 67, fax:+ (33) 2 41 87 99 27, email: marwa.chendeb@eseo.fr
[6] LAOB Université de Besançon/CNRS, Besançon France. [7] LPCE Université d'Orléans/CNRS, Orléans France.

ABSTRACT

The aim of this paper is to estimate the directions of arrival of ultra-high-energy cosmic rays using an antenna array. Extensive air showers produced generate wideband transient signals which have to be detected and characterized by the use of wavelet packet (WP) transform. An adaptive energy detector is used as a criterion to choose the best basis of WP. After reconstruction of the selected WP, the Maximum Likelihood Estimator is applied to estimate the parameters and the directions of arrival. Results evidence the efficiency of the method for simulated and real signals to detect the pulses from the input signal to noise ratio of -20 dB. Moreover, the performances of the proposed method are studied via extensive Monte Carlo simulations and compared with the Cramer Rao bounds. For this purpose, the colored noise disrupting signals is proposed.

1. INTRODUCTION

For almost 70 years, physicists and astronomers have studied cosmic rays and gained a good knowledge of the flux as a function of energy up to 10^{19} eV. However, in this energy range, the problem of the origin and the nature of ultra-high-energy cosmic rays (UHECR) is unsolved and stands as one of the most challenging questions in astroparticle physics. In order to collect the elusive events above 10^{19} eV (which present an integrated flux of less than 1 event per km^2 per steradian and per year) giant detectors are presently being designed and built [1]. To overcome some of the difficulties already encountered after several studies [1, 2], a new experimental radio air shower detector CODALEMA (COsmic ray Detection Array with Logarithmic ElectroMagnetic Antennas) has been proposed and operated using a MIMO (Multiple Input Multiple Output) antenna array [1]. The principle of the measurement is to estimate the parameters of the Electrical Transient (ET) signals induced by UHECR such as the amplitude, the time scale and the time of arrival used to estimate the Directions Of Arrival (DOA) of the UHECR.

In practical, these transient signals which are typically of about 10 to 60 nanoseconds (ns) are observed in the frequency band of about 0 to 250 MHz and even 0 to 500 MHz. Therefore, these transient signals are corrupted mainly by many Radio Frequency (RF) Telecom signals. In this work, a new detection procedure of the ET signals corrupted by highly colored noise, with low signal to noise ratio (SNR), is firstly proposed. The parameters and directions of arrival, which are in the center of preoccupations for the astrophysicists, are secondly estimated. As the capability to detect the ET signals strongly depends on the choice of the frequency band [1], our approach consisted of choosing multiscale decomposition using Wavelet Packets (WP) [3]. WP decomposition leads to a redundant tree where only a few WP are relevant for a specific task like change detection. The Wavelet Packet Transform (WPT) is used for its capacity to chose an optimal basis for the detection problem and reduction noise by rejecting the noise frequency bands while keeping the bands of interest. The choice of the basis depends on criteria determined by analysis goals, such as compression, filtering (smoothing) [4], and detection [5]. Recently wavelet theory has been introduced to array signal processing in two ways : the wavelet denoising [6], and subbanding [7] methods. Parameters and DOA estimations are in many fields including radar, sonar [8], medical imaging [9] and communications. Almost all methods presented in the literature are essentially limited to narrowband or stationary data [10]. But in other cases (e.g. sonar or cosmic rays) [11], received signals are assumed to be wideband. In most of the methods [12], wideband signals are decomposed into many narrow band signals with Fourier transform. Another class of possible wideband methods are time-domain methods such as delay-and-sum (DS) beamforming [11].

In this study, the astrophysical signals are wideband and highly drowned in colored noise. For this application, a specific criterion is proposed in order to select the *best basis* for further event detection. The adaptive energy detector is used as an index of each WP to detect the pulses. The idea is to determine the maximum of the estimated adaptive energy detector on each WP and select the packets which have the higher values. After WP selection, parameters and DOA are estimated using a Non Linear Least Squares (NLLS) method [13] on the reconstructed chosen WP. To generate the test sequences for Monte Carlo simulations and to evaluate the Cramer Rao Lower Bounds (CRLB's), modeling of the noised signals is required. According to the spectral analysis of the real signals, a Rice decomposition seems to be the most adequate to model the real noise. Results show that the proposed procedure detects well the pulses even for low signal to noise ratio of -20 dB and also leads to good estimates, as evidenced by comparison with the CRLB's.

This paper is organized as follows. In Section 2, a signal and noise models are presented. In Section 3, the WP

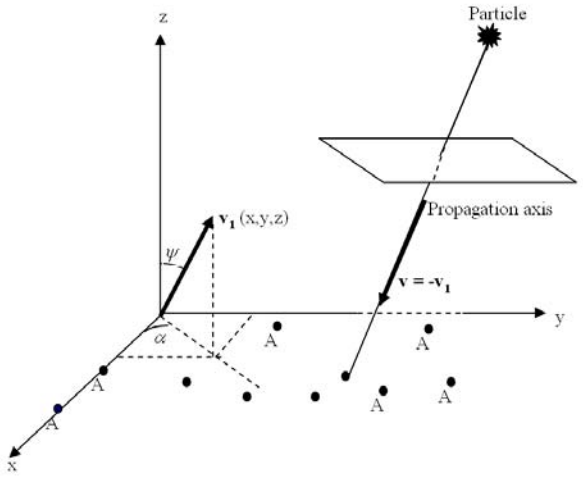

Figure 1: Experimental setup for particle detection and DOA estimation.

decomposition to select the *best basis* for event detection and estimation is described. NLLS estimation procedure and CRLB's are briefly cited. Results and comparisons with Monte carlo simulations are shown in section 4. Finally, section 5 concludes the paper.

2. PROBLEM FORMULATION

2.1 Signal model

Consider a MIMO array composed of M antennas receiving ET pulses $\mathbf{s}$ with broadband frequency characteristics. These pulses are emitted by sources with very high energy. The $M \times 1$ vector of antenna outputs can be modeled as

$$\mathbf{y}(k) = \mathbf{s}(k-\tau) + \mathbf{n}(k), \tag{1}$$

where $\mathbf{n}(k)$ is the $M \times 1$ colored noise vector, ($k = 0, \ldots, N-1$, N is the number of the available observed samples). $\mathbf{y}$, $\mathbf{s}$ and $\mathbf{n}$ are $M \times N$ matrices. The $M \times 1$ vector of timing differences τ corresponds to the propagation times of the wavefront for the different antennas, and also includes electronics and cable delays [1]. For one antenna m, the delay is written as

$$\tau_m = F_s \frac{\mathbf{v}^T \mathbf{p}_m}{c}, \tag{2}$$

where $\mathbf{p}_m = (x_m, y_m, z_m)(m = 0, \ldots, M-1)$ are the antennas locations. F_s is the sampling frequency, c is the propagation velocity and $\mathbf{v}$ is a unit vector which defines the direction of propagation and can be written as

$$\mathbf{v} = \begin{bmatrix} -\sin\psi\cos\alpha \\ -\sin\psi\sin\alpha \\ -\cos\psi \end{bmatrix}, \tag{3}$$

where α is measured counter clockwise from the x axis, and ψ is measured clockwise from the z axis as shown in figure 1. Introducing (3) in (2), τ_m becomes

$$\tau_m = -\frac{F_s}{c}\left(x_m \sin\psi\cos\alpha + y_m \sin\psi\sin\alpha + z_m\cos\psi\right). \tag{4}$$

The pulse waveform can be defined as [1]

$$s(k) = \varepsilon(k) A e^2 \left(\frac{k}{2\Delta}\right)^2 e^{\frac{-k}{\Delta}}, \tag{5}$$

where $\varepsilon(k) = 1$ for $k > 0$ and 0 for $k < 0$. This expression is derived from various research works in different fields of theoretical and experimental physics including Electromagnetism, Nuclear Science, wave propagation and antennas. Δ fixes the time scale of the positive-amplitude part (the maximum is reached for $k = 2\Delta$) and A is the maximum amplitude. The vector of parameters to be estimated is ($\Theta = [A, \Delta, \alpha, \psi]^T$). Therfore, the pulse waveform can be rewritten as

$$\mathbf{s}(k-\tau) = \mathbf{s}(k, \Theta). \tag{6}$$

The next section investigates the noise modeling and its autocorrelation function in order to calculate the Cramer Rao Lower Bounds.

2.2 Noise model

The colored noise is mainly due to the Radio Frequency telecom signals. To propose a model to such noise signals, a spectral and temporal study has been made on the real array output signals in analyzing all the correlations and estimated power spectral densities between antennas. Based on the experimental observations and a statistical analysis of RF telecom signals, the Rice decomposition is proposed to model the noise [14]:

$$\mathbf{n}(k) = \sum_{i=1}^{N_h} \left(\mathbf{a}_i(k) cos(2\pi f_i k) + \mathbf{b}_i(k) sin(2\pi f_i k)\right), \tag{7}$$

where N_h is the number of harmonic frequencies. $\mathbf{a}_i$ and $\mathbf{b}_i$ are obtained after a low pass filtering of two independent gaussian processes for each harmonic frequency. The autocorrelation function of $\mathbf{n}(k)$ is given by :

$$\varphi_n(k) = \varphi_{lp}(k) \sum_{i=1}^{N_h} A_i(k) cos(2\pi f_i k), \tag{8}$$

where the amplitudes A_i and frequencies f_i are extracted automatically from the careful spectral analysis of the real astrophysical signals. $\varphi_{lp}(k)$ is the autocorrelation sequence of the impulse response for the synthesis of low pass random processes $\mathbf{a}_i(k)$ and $\mathbf{b}_i(k)$. These methodologies allow us to characterize the noise and to generate the test sequences closest to the reality.

3. PROPOSED WAVELET PACKETS DETECTION AND ESTIMATION METHOD

3.1 Wavelet Packet Transform (WPT)

WPT is an extension of Discrete Wavelet Transform. Each detail coefficient vector is decomposed into two parts using the same approach as in approximation vector splitting [3]. We start with $h(n)$ and $g(n)$, the two impulsive responses of low-pass and high-pass analysis filters, corresponding to the scaling function and the wavelet function, respectively [3].

3.2 Cost function for detection purposes

To determine the *best basis*, a cost function must be chosen to represent the goal of the application. In our approach we propose to use an estimated adaptive energy detector of second order as a basis for the definition of the cost function. The total energy estimator (sum of suqres) or the maximum squared wavelet coefficients ($W_{j,i}^2$) are not adequate in this study because they do not take into account the local character of the transient signals existence. Estimated adaptive version of the energy is applied to each wavelet packet $W_{j,i}$ (j is the scale parameter and i is the sequence parameter). It can be written as

$$\widehat{m}_{j,i}^2(k) = \widehat{m}_{j,i}^2(k-1) + \mu(W_{j,i}(k)^2 - \widehat{m}_{j,i}^2(k-1)), \qquad (9)$$

where μ is the the adaptive step which controls the convergence of the estimator. The parameter μ is to be determined *a priori* but the performance curves show that there is a little influence on the quality of detection. In general, the value $\mu = 0.1$ is well appropriate for the transient signals detection. As a criterion for WP selection, we used the maximum ($D_{j,i}^{max}$) of the adaptive energy detector of equation (9) for each wavelet packet $W_{j,i}$:

$$D_{j,i}^{max} = max(\widehat{m}_{j,i}^2). \qquad (10)$$

3.3 Wavelet Packet selection

The goal of this part is to retain only the WP that are able to detect the electrical transient pulses. The main idea is that the maximum $D_{j,i}^{max}$ is greater than a threshold when a WP contains the transient signal. Two approaches can be imagined for WP selection : either defining a threshold on $D_{j,i}^{max}$, or selecting $D_{j,i}^{max}$ in descending order and limiting the number of selected WP to a predefined number. Whatever the approach, a node in the WP tree will be put to "1" if the corresponding WP has been selected, the others being put to "0". The previous step identified all nodes (i.e. WP) where significant activities were detected. As the tree is highly redundant, there is a need for a further step to reduce the number of selected nodes to only non-redundant ones. The current implementation of this second step is detailed in [5]. Let us define a Father Node (FN) as any connection between two branches whose ends are called Children Nodes (CN). A Father Node at level j is a Children Node for level $(j-1)$. A component is the association of a FN and its two CN.
Our algorithm selecting the *best basis* can be described as follows :

1. Take samples of signal model (1) which forms a $M \times N$ matrix.
2. From the binary tree of wavelet packets, we build a corresponding tree with the same structure, and where a given node takes the value "0" or "1" according to the value of the corresponding $D_{j,i}^{max}$ (figure 2.a).
3. Modify the node values according to the following rules:
   - If all nodes in a component have the value "1", put "0*" to the FN (all information is contained in CN).
   - If FN = "1" and both CN = "0" or "0*" in a component, put FN to 0 (aberrant case or information already taken into account).
   - Select all CN = "1" with FN = "0" in a component (information is only detectable in the CN).

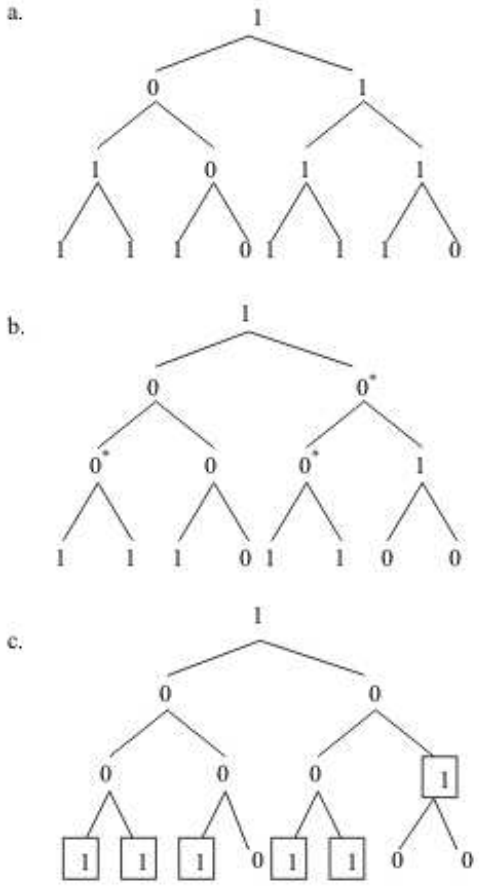

Figure 2: Steps for selection of the *best basis*.

   - If FN = "1" with CN1 = "1" and CN2 = "0", select FN (FN and CN1 display the same information).
   - If FN = "1" with CN1 = "1" and CN2 = "0*", select CN1 (the only information not taken into account yet is in CN1). All these cases are displayed in figure 2.b.
4. Select all nodes at "1" (figure 2.c).
5. Reconstruct the data matrix from the selected packets.
6. Apply Non Linear Least Squares (NLLS) estimation method to the new data model.

3.4 Non Linear Least Squares Estimator

After the detection and reconstruction steps, available signals can be written as

$$\mathbf{y}_r(k) = \mathbf{s}_r(k,\Theta) + \mathbf{n}_r(k). \qquad (11)$$

where $\mathbf{s}_r(k,\Theta)$ is the reconstructed pulse defined in equation (5). Therefore, Θ can be estimated with the NLLS estimator [13] :

$$\widehat{\Theta} = \arg\min_{\Theta} \left(\|\mathbf{y}_r(k) - \mathbf{s}(k,\Theta)\|^2 \right). \qquad (12)$$

3.5 Cramer Rao Lower Bounds (CRLB's)

To demonstrate the performance of this estimation method, the CRLB's were calculated in the case of colored noise. The

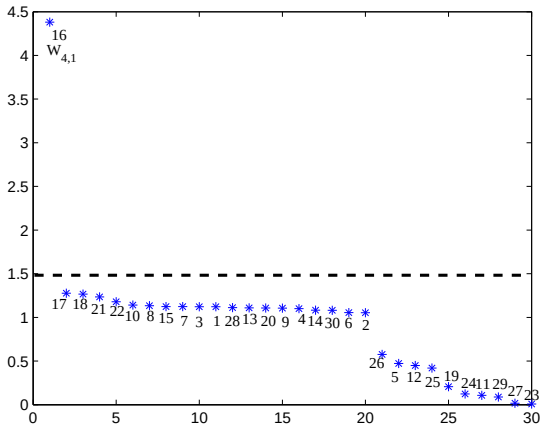

Figure 3: Average of $D_{j,i}^{max}$ from 500 simulated training signals on 30 packets. The number of each packet is associated to each value of $D_{j,i}^{max}$ (SNR = -15 dB). X-axis: arbitrary unit. Yaxis: $D_{j,i}^{max}$ value.

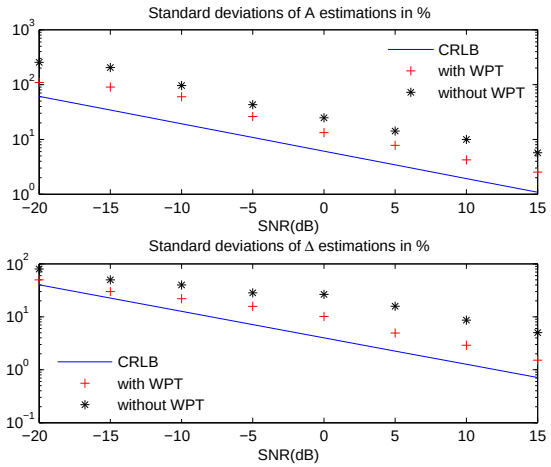

Figure 4: Relative standard deviations of A and Δ in % versus SNR.

CRLB's are used as a theoretical lower bound for any unbiased estimator [13]. Most of the derivations of the CRLB's for wideband source localization found in the literature are in terms of relative time-delay estimation error. This application is a deterministic case, a more general CRLB's is directly calculated from the signal model. In first approximation, the correlations between antennas are neglected. Thus the CRLB's which are are the inverse of Fisher matrix can be written as [13]:

$$[I(\Theta)]_{pq}^{-1} = \sum_{m=1}^{M} \left[\frac{\partial s_m(\Theta)}{\partial \Theta_p}\right]^T C_m^{-1}(\Theta) \left[\frac{\partial s_m(\Theta)}{\partial \Theta_q}\right], \qquad (13)$$

where C_m is the noise covariance matrix of antenna m calculated from equation (8).

4. RESULTS OF WP SELECTION FOR DETECTION AND ESTIMATION

4.1 Results on simulated signals

In this section, Monte Carlo simulations have been conducted with $M = 15$ antennas and $N = 2520$ samples. The algorithm described in section 3.3 is first applied to a training data set composed of 500 simulated signals. Each signal is obtained with the pulse $\mathbf{s}$ defined by equation (1) drowned in the colored noise modeled by equation (7). The UHECR DOA are supposed to be $\alpha = 70°$ and $\psi = 40°$. The sampling frequency is 1 Ghz, $\Delta = 8$ ns, $A = 1$. The wavelet Daubechies 10 is used, with a tree developed until level 4. This choice is done after the analysis of the performance curves which have been traced for different wavelets (Wavelets Symlet 1 to 10, Coiflet 1 to 5 and Daubechies 1 to 10). To decide which packets are able to highlight the presence of ET pulses, the average of $D_{j,i}^{max}$ is calculated for each WP (there are 30 packets). Figure 3 shows the $D_{j,i}^{max}$ values in descending order in the case of SNR = -15 dB. By examining figure 3, there is a clear threshold = 1.5. The adaptative choice of the threshold is important because the experimental conditions can change. In this study, thechoice of the threshold 1.5 gave good results. Value "1" is given to the packet $W_{4,1}$ (the packets on figure 3 are numbered by according the number 1 to the packet $W_{1,0}$ and the number 30 to the packet $W_{4,15}$). This threshold is used for all SNRs. The SNR influences widely the choice of the packets revealing the presence of the pulses. The same packet $W_{4,1}$ is selected for SNR = -20 dB to 5 dB. For SNR $\geq$ 10 dB, the same procedure is repeated. We found that the packets which have $D_{j,i}^{max} \geq 1.5$ are $W_{1,0}$, $W_{2,0}$, $W_{3,0}$, $W_{4,0}$ and $W_{4,1}$. After applying the algorithm of selection of *best basis*, only packets $W_{4,0}$ and $W_{4,1}$ are retained for the *best basis*. The NLLS estimation procedure is then applied to the reconstructed selected packets for 500 signals of a test set (built in the same way as the previous training set) to estimate Θ. To demonstrate the performances of the algorithm, the results are compared with CRLB's. Relative standard deviation (SD) of A and Δ estimations ($\sigma_{\widehat{A}}/A$ and $\sigma_{\widehat{\Delta}}/\Delta$ respectively) given in % are shown in the figure 4 according to SNR (defined as the square of the ratio of the pulse amplitude and the noise SD). Figure 5 shows the standard deviations (SD) of the DOA (angles α and ψ) estimations as functions of SNR. Figures 4 and 5 present the performances of combined WP and NLLS method versus classical NLLS procedure. The simulation results are closer to the theoretical CRLB's after applying the WP selection. As expected, the SD's decreases rapidly as the SNR increases.

4.2 Results on real signals

The antenna signals are recorded after RF signal amplification (1-200 MHz, gain 35 dB) by LeCroy digital oscilloscopes (8-bit ADC, 1 GHz sampling frequency, 2.5 μs recording time). The training set consists of a total of 100 runs and each run contains 15 antenna recordings. The signals are decomposed by WPT by using the wavelet Daubechies 10. For SNR = 10 dB, only packets $W_{1,0}$, $W_{2,0}$, $W_{3,0}$, $W_{4,0}$ and $W_{4,1}$ (bandwidths: [0–250], [0–125], [0-62.5], [0–31.25] and [31.25-62.5 MHz] respectively) are first selected after applying the first step of the selection algorithm. The WP are easily selected according to the $D_{j,i}^{max}$ by setting the threshold to 1.5 (the threshold is obtained in the same way that in the case of the simulated signals). Only packets

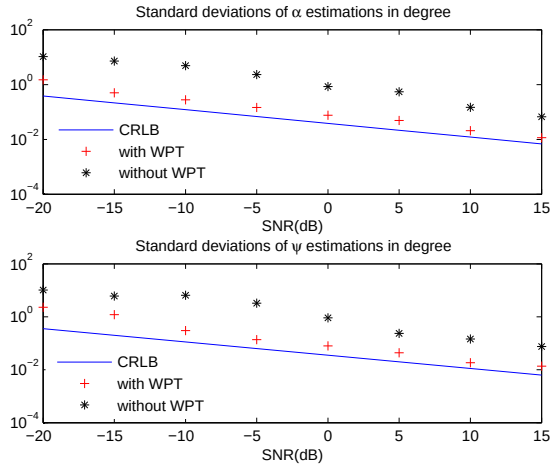

Figure 5: α and ψ standard deviations in degrees versus SNR.

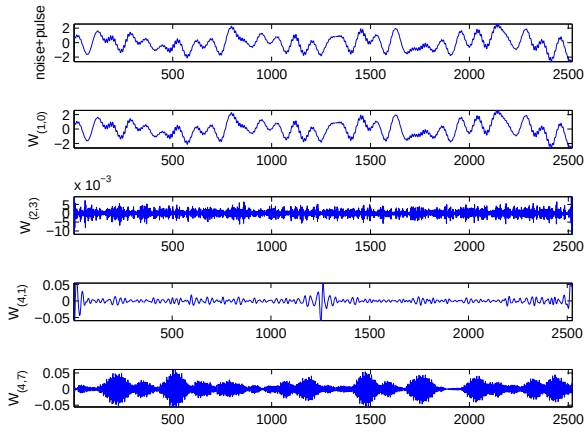

Figure 6: Detection of ET pulse on the selected WP ($W_{4,1}$) and on three no selected WP ($W_{1,0}$, $W_{2,3}$ and $W_{4,7}$) of an arbitrary real signal (SNR = -10 dB).

$W_{4,0}$ and $W_{4,1}$ are finally retained. For SNR = -20 dB to 5 dB, only the packet $W_{4,1}$ which its bandwidth corresponds to the pulses is selected. To test the performances of the algorithm, 100 other test runs are used. The same packets are chosen for each SNR. Figure 6 illustrates the detection of ET pulse on the selected packet $W_{4,1}$ for SNR = -10 dB. The pulse is not detected on the other packets (e.g packets $W_{1,0}$, $W_{2,3}$ and $W_{4,7}$). This shows the performance of the WP selection for detection the electrical transient pulses on real data.

5. CONCLUSION

In this paper, the application of WP selection algorithm has been investigated to increase the performance of detection and estimation for wideband electrical transient pulses in colored noise. The factors that affect the accuracy are the parameters of transient pulses, the signal to noise ratio, the sampling rate, and the geometry of the sensor array. All these parameters have been included in the calculation of the CRLB's. The NLLS method applied on the reconstructed selected WP improves the estimation variances which are closer to the CRLB's. The same best packets are selected in the case of simulated and real signals showing that the proposed models are in agreement with the real data.
The transient signals detection procedure is satisfactory and permits to identify almost all pulses received on the antennas of CODALEMA experiment. The results give more precision on the nature and the origin of the cosmic rays. They permit to locate with more exactitude the positions of the particles.

REFERENCES

[1] D. Ardouin, A. Belltoile, D. Charrier et al., "Radio-Detection Signature of High Energy Cosmic Rays by the CODALEMA Experiment," *Nuclear Instruments and Methods in Physics Research*, vol. 555, pp. 148–163, Aug. 2005.

[2] N. Hayashida, K. Honda, M. Honda et al., "Observation of a Very Energetic Cosmic Ray Well Beyond the Predicted 2.7 K Cutoff in the Primary Energy Spectrum," *Physics Review Letters*, vol. 73, pp. 3491–3494, 1994.

[3] S. Mallat, *A Wavelet Tour of Signal Processing*. Address: Academic Press, San Diego, CA, 1999.

[4] R.R. Coifman and M.V. Wickerhauser, "Entropy based algorithms for best basis selection," *IEEE Information Theory*, vol. 38, pp. 1241-1243, 1992.

[5] M. Chendeb, M. Khalil and J. Duchne. *Methodology of wavelet packet selection for event detection*. Signal Processing, vol. 86, no.12, pp.3826–3841, March 2006.

[6] R. Sathish and G. V. Anand, "Wavelet denoising for plane wave DOA estimation by MUSIC," *IEEE TENCON03*, vol. 1, pp. 104–108, 2003.

[7] B. Wang, Y. Wang and H. Chen, "Spatial wavelet transform preprocessing for direction-of-arrival estimation," *IEEE Antennas Propag. Societ. Internat. Sympos.*, vol., pp. 672–675, June 2002.

[8] F. Chevalier, *Principes de traitement des signaux Radar et Sonar*. Address: Masson, 1989.

[9] H. Mermoz, *Imagerie, corrélation et modèle*. Address: Ann Télécommunication, 2002.

[10] H.L. Van Trees, *Optimum Array Processing. Part VI of Detection, Estimation, and Modulation Theory*. Address: A John Wiley and Sons, 2002.

[11] P. Graffouliere, *Méthodes actives spatio-temporelles large bande, Techniques et performances, Application en SONAR*. Address: Thesis of Institut National Polytechnique de Grenoble, 1997.

[12] J. C. Chen, R. E. Hudson, and K. Yao, "Maximum-likelihood source localization and unknown sensor location estimation for wideband signals in the near field," *IEEE Trans. Signal Processing*, vol. 50, pp. 1843-1854, Aug. 2002.

[13] S. M. Kay, *Fundamentals of staistical signal processing: estimation theory*. Address: Prentice-Hall, 1993.

[14] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. Address: Prentice-Hall, 1991.