

Article

Diffusion Maximum Correntropy Criterion Based Robust Spectrum Sensing in Non-Gaussian Noise Environments

Xiguang Xu ^{1,*} , Hua Qu ^{1,2}, Jihong Zhao ^{1,2,3}, Feiyu Yan ¹ and Weihua Wang ¹

¹ School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China; qh@mail.xjtu.edu.cn (H.Q.); zhaojihong@mail.xjtu.edu.cn (J.Z.); yanfeiyu@stu.xjtu.edu.cn (F.Y.); wangweihua@stu.xjtu.edu.cn (W.W.)

² Suzhou Caiyun Network Technologies Co., Ltd., Suzou 215123, China

³ School of Telecommunication and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710061, China

* Correspondence: xuxiguang@stu.xjtu.edu.cn; Tel.: +86-182-9140-0073

Received: 21 March 2018; Accepted: 30 March 2018; Published: 3 April 2018



Abstract: Spectrum sensing is the most important task in cognitive radio (CR). In this paper, a new robust distributed spectrum sensing approach, called diffusion maximum correntropy criterion (DMCC)-based robust spectrum sensing, is proposed for CR in the presence of non-Gaussian noise or impulsive noise. The proposed distributed scheme, which does not need any central processing unit, is characterized by an adaptive diffusion model. The maximum correntropy criterion, which is insensitive to impulsive interference, is introduced to deal with the effect of non-Gaussian noise. Simulation results show that the DMCC-based spectrum sensing algorithm has an excellent robust property with respect to non-Gaussian noise. It is also observed that the new method displays a considerably better detection performance than its predecessor (i.e., diffusion least mean square (DLMS)) in impulsive noise. Moreover, the mean and variance convergence analysis of the proposed algorithm are also carried out.

Keywords: robust spectrum sensing; maximum correntropy criterion (MCC); diffusion scheme; non-Gaussian noise; cognitive radio networks

1. Introduction

Cognitive radio (CR), which can exploit the vacant spectrum dynamically, has been considered as a promising technology for response spectrum scarcity [1–3]. The most important component of CR is spectrum sensing. This is because cognitive users (CUs) should know their own operating environments to make sure that the primary users (PUs) are not interfered with by CUs. A large number of schemes has been proposed for this target, including energy detection [4], matched filtering detection [5], cyclostationary detection [6] and cooperative spectrum sensing [7,8].

Cooperative spectrum sensing is proposed to tackle the problem of shadow fading and hidden primary receivers [9]. Among the cooperative spectrum sensing methods, centralized cooperative sensing [9,10] lets each CU send information to a central processing unit, and then, the decisions about the presence or absence of the PU are made by a certain fusion algorithm. In [11], fuzzy data fusion Kalman filter-based cooperative spectrum sensing has been proposed to make a global sensing decision. The fuzzy data fusion Kalman filter, which has been applied to reduce failure risk in an integrated vehicle health maintenance system (IVHMS) [12], is an effective and reliable method to improve sensing performance in CR. Because of the central processing unit, high communication overhead and enormous computing power are needed [13]. To solve these problems, decentralized cooperative

sensing methods have been proposed, such as consensus algorithms [14,15], belief propagation [16] and diffusion methods [17–19].

Some recent literature [17–19] has proposed distributed diffusion methods (diffusion least mean square (DLMS)), which allow each CU to collect data from its neighbors and make detection decisions based on these data without any central processing unit. As it is a distributed and adaptive diffusion scheme, DLMS has an ability to enhance network failure resistance. However, this diffusion solution is only suitable for Gaussian noise. If the network is disturbed by non-Gaussian noise or so-called impulsive noise, the detection performance of this diffusion solution may deteriorate seriously.

Typical non-Gaussian noise distribution [20–23] has heavy tails, which may be generated by multiple impulsive interference. For example, the impulsive nature of the noise in industrial, scientific and medical (ISM) bands leads to impulsive noise. Impulsive noise is also caused by microwave ovens or devices with electromechanical switches. In addition, impulsive interference is also caused by various components on a computer platform. What is more, igniting car engines, power lines and heavy current switches in urban environments are the typical man-made impulsive noise sources. Impulsive interference in real-world environments degrades the performance of spectrum sensing.

In order to solve the non-Gaussian noise problem in CR, various robust spectrum sensing approaches have been published [24–27] in recent years. In [24], a p -th order moment-based spectrum sensing has been proposed to counteract impulsive noise. An accurate kernelized energy detector is displayed in [25]. Besides, Ref. [26] presents a soft-limited polarity-coincidence-array spectrum sensing to detect the primary signal distorted by non-Gaussian noise. However, these approaches are single-user detections, which have a limited capability to detect the signal in complex noise environments. In [27], a multi-user detection algorithm called Rao test-based cooperative spectrum sensing has been proposed for robust detection. Although this detection scheme performs well in centralized cooperation, it needs a central processing unit. The disadvantage of the central unit is obvious, that is once the central unit fails, it can easily lead to paralysis of the entire network. In order to solve the shortcoming of centralized cooperation and improve robustness for CR, a robust distributed cooperation spectrum sensing approach is needed. As far as we know, there are no previous studies in the literature dealing with the robust distributed spectrum sensing in non-Gaussian noise environments.

In this paper, a distributed robust spectrum sensing (without any central processing unit), called diffusion maximum correntropy criterion (DMCC)-based robust spectrum sensing, is proposed in non-Gaussian noise environments. The correntropy, as a nonlinear similarity measure in information theoretic learning (ITL), has been successfully used in non-Gaussian noise for its robust and cost-efficient function [28–30]. The new distributed robust spectrum sensing is motivated by the desirable features of correntropy and the diffusion model schemes [17,18]. The main contributions of this paper are three-fold: A DMCC-based robust spectrum sensing scheme is presented to solve the distributed power estimation with non-Gaussian noise; a version of DMCC-based robust spectrum sensing, namely adaptation to combination DMCC (ATC DMCC) algorithm, is derived, which can solve the non-Gaussian noise problem in spectrum sensing; the mean and mean square performance of the new power estimation algorithm have been analyzed. In addition, the simulation results show that the performance of the proposed method is excellent under impulsive noise environments.

This paper is organized as follows: in Section 2, we describe the signal model and non-Gaussian noise model; a brief review of MCC is given in Section 3; in Section 4, we derive the distributed DMCC-based power estimation algorithm and develop the DMCC-based robust spectrum sensing algorithm; in Section 5, we present the mean analysis, the variance convergence analysis and the detection performance analysis; in Section 6, the performance of the proposed sensing scheme is evaluated and compared with existing sensing algorithms. Finally, the conclusion is given in Section 7.

2. Signal Model and Non-Gaussian Noise Model

2.1. Signal Model

In this paper, every CU is interested in performing spectrum sensing in a distributed manner without any central processing unit, where CU gets the available information from its neighbors. It is assumed that the information transfer between neighbor CUs is lossless. Generally, the spectrum sensing problem in a distributed system can be described as a two-hypotheses decision as:

$$\begin{cases} H_0 : x_k(n) = v_k(n) \\ H_1 : x_k(n) = a_k s(n) + v_k(n), (n = 1, 2, \dots) \end{cases} \quad (1)$$

where $k = 1, 2, \dots, K$ is the CU number and $n = 1, 2, \dots, N$ is the sample index. $x_k(n)$ is the received signal of the CU k ; $s(n)$ is the signal emitted by the PU; $v_k(n)$ can be regarded as the non-Gaussian noise, which will be described in detail below. $s(n)$ and $v_k(n)$ both are independently and identically distributed (i.i.d). At the same time, they are statistically independent of each other. a_k represents the channel gain of the CU k . As the channel is assumed to be a slowly fading channel in this paper, the channel gain can be considered constant and obtained by: $a_k \sim N(0, 1)$. The two-hypotheses H_0 and H_1 denote the absence and presence of the PU signal, respectively.

In energy detection theory [4], an energy detector is an energy measurer that gauges the energy of the received signal. Based on these energy data, it decides whether the received waveform contains the PU signal. According to (1), we obtain the energy model at CU k :

$$\begin{cases} H_0 : E[|x_k(n)|^2] = E[|v_k(n)|^2] \\ H_1 : E[|x_k(n)|^2] = E[|a_k|^2 |s(n)|^2] + E[|v_k(n)|^2] \end{cases} \quad (2)$$

where $E[|x_k(n)|^2]$ denotes the average power of the received data samples, $E[|v_k(n)|^2] = \sigma_v^2$ is the average power of noise and the power of the PU signal is $E[|s(n)|^2] = S$. In this paper, the received power estimated by CU k is represented by $P_k(n)$.

2.2. Non-Gaussian Noise Model

The non-Gaussian noise, which is also called impulsive noise, is modeled as a Gaussian mixture in this paper. The Gaussian mixture model has been widely applied in wireless communications. We can get the impulsive noise by [31]:

$$v_k(n) = g_{1,k}(n) + d_k(n) g_{2,k}(n) \quad (3)$$

where $g_{1,k}(n)$ and $g_{2,k}(n)$, the two zero mean Gaussian noises with variances σ_1^2 and σ_2^2 , respectively, are independent. $d_k(n)$, a sequence of ones and zeros, which is an independently and identically distributed (i.i.d) Bernoulli random process with occurrence probabilities, $Pr(d_k(n) = 1) = p$. It is necessary to note that the variance σ_2^2 is chosen to be much larger than σ_1^2 , so that a large impulse will appear when $d_k(n) = 1$.

3. Brief Background of the Maximum Correntropy Criterion

MCC has been successfully and widely applied in adaptive filtering [29,30]. Correntropy is generalized to measure the similarity of two random variables. The correntropy is defined as [28]:

$$V(X, Y) = E[\kappa(X, Y)] = \int \kappa(x, y) dF_{x,y}(x, y) \quad (4)$$

where E denotes the expectation operator, $F_{x,y}(x,y)$ is the joint distribution of the two variables and $\kappa(\cdot, \cdot)$ is a Mercer kernel. In practice, the joint distribution $F_{x,y}(x,y)$ is unavailable, and the number of data we know is limited. In these cases, the correntropy can be estimated as the sample mean:

$$V(X, Y) = E[\kappa(X, Y)] \approx \frac{1}{N} \sum_{i=1}^N \kappa(x_i, y_i) \quad (5)$$

where N is the sample number.

In this paper, the most popular Gaussian kernel [32] is applied in correntropy, and it can be expressed as:

$$\kappa(x, y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{e^2}{2\sigma^2}\right) \quad (6)$$

where $e = x - y$ is the error and σ is the kernel size. According to the Gaussian kernel, the instantaneous MCC cost is given by [29]:

$$J_{MCC}(n) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{e^2(n)}{2\sigma^2}\right) \quad (7)$$

MCC has some desirable advantages. For example, it is almost bounded for any distribution; it is also a local similarity measure and is robust to outliers. Based on these favorable advantages, we derive the distributed diffusion MCC-based power estimation algorithm in the following section.

4. Distributed Diffusion Maximum Correntropy Criterion-Based Power Estimation and Spectrum Detection

4.1. Derivation of the Distributed DMCC-Based Power Estimation Algorithm

According to the energy model (2), each CU receives the signal transmitted by the PU and estimates its power. As the channel gain is different, the estimated power at each CU differs. When the channel condition at CU k is poor, the power is low. On the contrary, if the CU has a good channel condition, it will have a high power performance. In this article, CUs cooperate to estimate a common parameter P^o , the average power of all CUs.

$$P^o = \frac{1}{K} \sum_{k=1}^K E[|x_k(n)|^2] = S \frac{1}{K} \sum_{k=1}^K |a_k|^2 + \sigma_v^2 \quad (8)$$

According to the MCC cost (7), the global cost function for each CU can be expressed as:

$$J^{glob}(P) = \sum_{k=1}^K \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2} |x_k(n)|^2 - P\right)^2 \quad (9)$$

The optimal solution is obtained by minimizing (9).

In this paper, every CU is interested in estimating the average power P^o in a distributed manner, where the CU gets the available information from its neighbors. The distributed manner does not need any central processing unit, which improves the robustness and stability of the algorithm. The local cost function [33] of the DMCC for each CU is defined as:

$$J_k^{loc}(P) = \sum_{l \in \mathcal{N}_k} \alpha_{l,k} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2} |x_l(n)|^2 - P\right)^2 \quad (10)$$

where \mathcal{N}_k is the neighborhood set of CU k , $\{\alpha_{l,k}\}$ is a set of nonnegative coefficients, which satisfy the following conditions:

$$\alpha_{l,k} \geq 0, \sum_{l=1}^K \alpha_{l,k} = 1, \text{ and } \alpha_{l,k} = 0 \text{ if } l \notin \mathcal{N}_k \quad (11)$$

This means that for every CU k , the sum of the coefficients $\{\alpha_{l,k}\}$ is one. We collect the entries $\{\alpha_{l,k}\}$ into a $K \times K$ matrix \mathbf{A} .

The derivative of (10) is:

$$\nabla J_k^{loc}(P) = \frac{1}{\sigma^2} \sum_{l \in \mathcal{N}_k} \alpha_{l,k} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2} \left| |x_l(n)|^2 - P \right|^2\right) \left(|x_l(n)|^2 - P \right) \quad (12)$$

We take the steepest descent method to yield:

$$\begin{aligned} P_k(n+1) &= P_k(n) + \eta_k \nabla J_k^{loc}(P) \\ &= P_k(n) + \frac{\eta_k}{\sigma^2} \sum_{l \in \mathcal{N}_k} \alpha_{l,k} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2} \left| |x_l(n)|^2 - P \right|^2\right) \left(|x_l(n)|^2 - P \right) \end{aligned} \quad (13)$$

where η_k is the step size. For the sake of simplicity, we set $\mu_k = \frac{\eta_k}{\sqrt{2\pi}\sigma^3}$ ($k = 1, 2, \dots, K$) as the new step size. Therefore we have:

$$P_k(n+1) = P_k(n) + \mu_k \sum_{l \in \mathcal{N}_k} \alpha_{l,k} \exp\left(-\frac{1}{2\sigma^2} \left| |x_l(n)|^2 - P \right|^2\right) \left(|x_l(n)|^2 - P \right) \quad (14)$$

Next, we can obtain the intermediate estimates of each CU by:

$$\psi_k(n) = \sum_{l \in \mathcal{N}_k} \beta_{l,k} P_l(n) \quad (15)$$

where $\psi_k(n)$ represents an intermediate estimate for CU k at instant n . The non-negative element $\beta_{l,k}$ defines if the estimate from CU l (including CU k) is available for CU k . They satisfy the conditions:

$$\beta_{l,k} \geq 0, \sum_{l=1}^K \beta_{l,k} = 1, \text{ and } \beta_{l,k} = 0 \text{ if } l \notin \mathcal{N}_k \quad (16)$$

We collect the entries $\{\beta_{l,k}\}$ into a $K \times K$ matrix \mathbf{B} . With the intermediate estimates, the CUs update the estimates by (14):

$$\theta_k(n) = \psi_k(n) + \mu_k \sum_{l \in \mathcal{N}_k} \alpha_{l,k} \exp\left(-\frac{1}{2\sigma^2} \left| |x_l(n)|^2 - \psi_k(n) \right|^2\right) \left(|x_l(n)|^2 - \psi_k(n) \right) \quad (17)$$

The coefficients $\{\alpha_{l,k}\}$ decide which CUs should share their measurements with CU k . At last, each CU combines the estimates as:

$$P_k(n+1) = \sum_{l \in \mathcal{N}_k} \gamma_{l,k} \theta_l(n) \quad (18)$$

where the coefficients $\{\gamma_{l,k}\}$ are similar to $\{\beta_{l,k}\}$, and they represent whether CUs should share their intermediate estimates $\theta_l(n)$ with CU k . We collect the entries $\{\gamma_{l,k}\}$ into a $K \times K$ matrix \mathbf{R} .

There are detailed descriptions of the selection of the weights $\beta_{l,k}$, $\alpha_{l,k}$ and $\gamma_{l,k}$ in [34]. We can see that Equation (17) is similar to those in [33,34]. The only difference is that it contains an extra scaling factor $\exp\left(-\frac{1}{2\sigma^2} \left| |x_l(n)|^2 - \psi_k(n) \right|^2\right)$, which is an exponential function of the error. When a large noise occurs, this factor is close to zero, which endows the DMCC method with robustness and significantly improves the adaptation performance in impulsive noise.

4.2. ATC DMCC-Based Power Estimation

There are mainly two different schemes (including the adapt-then-combine (ATC) scheme and the combine-then-adapt (CTA) scheme) for the diffusion estimation [33,34]. The ATC scheme first utilizes the adaptive algorithm to update the local estimates and then combines the estimates. The CTA scheme, however, has a reverse order. As the learning performances of the two versions of DMCC-based algorithms are almost the same [34], we only discuss the ATC DMCC in this paper.

According to the adapt-then-combine scheme, one can obtain the following ATC DMCC method for the power estimation by combining (17) and (18):

$$\begin{cases} \theta_k(n) = P_k(n) + \mu_k \sum_{l \in \mathcal{N}_k} \alpha_{l,k} \exp\left(-\frac{1}{2\sigma^2} \left| |x_l(n)|^2 - P_k(n) \right|^2\right) \left(|x_l(n)|^2 - P_k(n) \right) \\ P_k(n+1) = \sum_{l \in \mathcal{N}_k} \gamma_{l,k} \theta_l(n) \end{cases} \quad (19)$$

The ATC DMCC estimation algorithm consists of two parts, an information exchange step and a combination step. In the information exchange step, every CU utilizes the information $\{x_l(n)\}$ from its neighbors to update the estimate $\theta_k(n)$; while in the combination step, each CU combines the estimates from its neighbors to obtain the intermediate estimate $P_k(n+1)$.

For less information communications, we set $A = I$. No information exchange is performed in the first part, so the ATC DMCC (19) reduces to:

$$\begin{cases} \theta_k(n) = P_k(n) + \mu_k \exp\left(-\frac{1}{2\sigma^2} \left| |x_k(n)|^2 - P_k(n) \right|^2\right) \left(|x_k(n)|^2 - P_k(n) \right) \\ P_k(n+1) = \sum_{l \in \mathcal{N}_k} \gamma_{l,k} \theta_l(n) \end{cases} \quad (20)$$

4.3. ATC DMCC-Based Robust Spectrum Sensing

We summarize (20) together with energy detection as Algorithm 1.

Algorithm 1 ATC DMCC-Based Robust Spectrum Sensing.

Start with $P_k(0) = P(0)$ for each CU. Choose proper coefficients $\gamma_{l,k}$, μ and σ .

for every time instant $n \geq 1$ do

for every CU $k = 1, \dots, K$, do

1. Power estimation:

Adaption:

$$\theta_k(n) = P_k(n) + \mu_k \exp\left(-\frac{1}{2\sigma^2} \left| |x_k(n)|^2 - P_k(n) \right|^2\right) \left(|x_k(n)|^2 - P_k(n) \right)$$

Combination:

$$P_k(n+1) = \sum_{l \in \mathcal{N}_k} \gamma_{l,k} \theta_l(n)$$

2. Detection decision:

$$H_0 : P_k(n+1) < \lambda \text{ or } H_1 : P_k(n+1) > \lambda$$

The threshold λ is described in detail in Subsection 5.3.

end for

end for

5. Performance Analysis

In this section, we study the performance analysis of the proposed algorithm. The mean performance and the variance of the performance of the proposed algorithm are analyzed first. Then, we study the energy detection performance. In order to facilitate the analysis, the following assumptions are to be adopted.

Assumption 1. All input signals $x_k(n)$ are spatially and temporally independent.

Assumption 2. The error nonlinearity $r_k(n) = \exp\left(-\frac{1}{2\sigma^2}\left|x_l(n)\right|^2 - \sum_{i \in \mathcal{N}_k} \beta_{k,i} P_i(n)\right)^2$ is independent of the input signal $x_k(n)$.

Strictly speaking, Assumption 2 does not accord with this fact because $r_k(n)$ is a function of error. However, this function can be considered as a variable step size term.

Because of the information exchange amongst CUs, the current estimates will affect their update. Therefore, in view of this dependence between CUs, we study the performance of the whole network. The proposed DMCC algorithm can be expressed as:

$$P_k(n+1) = \sum_{l \in \mathcal{N}_k} \beta_{k,l} P_l(n) + \mu_k \sum_{i \in \mathcal{N}_k} \alpha_{i,k} r_k(n) \left(|x_l(n)|^2 - \sum_{l \in \mathcal{N}_k} \beta_{k,l} P_l(n) \right) \tag{21}$$

When $A = I$, the algorithm will reduce to a simple version as:

$$P_k(n+1) = \sum_{l \in \mathcal{N}_k} \beta_{k,l} P_l(n) + \rho_k(n) \left(|x_k(n)|^2 - \sum_{l \in \mathcal{N}_k} \beta_{k,l} P_l(n) \right) \tag{22}$$

where $\rho_k(n) = \mu_k r_k(n)$ as a new step size factor.

Furthermore, for more convenience, some other new variables are needed. Then, we stack the local ones into global variables as follows:

$$\mathbf{P}(n) = \text{col} \{P_1(n), P_2(n), \dots, P_K(n)\} \tag{23}$$

$$\mathbf{\Psi}(n) = \text{diag} \{\rho_1(n), \rho_2(n), \dots, \rho_K(n)\} \tag{24}$$

$$\mathbf{X}(n) = \text{col} \left\{ |x_1(n)|^2, |x_2(n)|^2, \dots, |x_K(n)|^2 \right\} \tag{25}$$

We define extra matrix \mathbf{U} , which contains the step size parameters as follows

$$\mathbf{U} = \text{diag} (\mu_1, \mu_2, \dots, \mu_K) \tag{26}$$

According to the above new variables above, we remodel the update equations to represent the global network:

$$\mathbf{P}(n+1) = \mathbf{B}\mathbf{P}(n) + \mathbf{\Psi}(n) [\mathbf{X}(n) - \mathbf{B}\mathbf{P}(n)] \tag{27}$$

where $\mathbf{\Psi}(n) = \mathbf{U}\mathbf{R}(n)$ is a diagonal matrix, and $\mathbf{R}(n)$ is defined by:

$$\mathbf{R}(n) = \text{diag} \left\{ \exp\left(-\frac{1}{2\sigma^2}\left|x_1(n)\right|^2 - \sum_{l \in \mathcal{N}_1} \beta_{1,l} P_l(n)\right)^2, \exp\left(-\frac{1}{2\sigma^2}\left|x_2(n)\right|^2 - \sum_{l \in \mathcal{N}_2} \beta_{2,l} P_l(n)\right)^2, \dots, \exp\left(-\frac{1}{2\sigma^2}\left|x_K(n)\right|^2 - \sum_{l \in \mathcal{N}_K} \beta_{K,l} P_l(n)\right)^2 \right\} \tag{28}$$

Through the above equations, we can derive the mean performance and the variance performance.

5.1. Mean Performance

According to (27), we can rewrite the recursion as follows:

$$\mathbf{P}(n+1) = [\mathbf{I} - \mathbf{\Psi}(n)]\mathbf{B}\mathbf{P}(n) + \mathbf{\Psi}(n)\mathbf{X}(n) \tag{29}$$

Taking the expectation on both sides of (30), we have:

$$E[\mathbf{P}(n+1)] = E[\mathbf{I} - \mathbf{\Psi}(n)] \mathbf{B} E[\mathbf{P}(n)] + E[\mathbf{\Psi}(n) \mathbf{X}(n)] \quad (30)$$

We employ Assumption 2 to infer that the matrix $\mathbf{\Psi}(n)$ is independent of the matrix $\mathbf{X}(n)$, and then, we have:

$$E[\mathbf{P}(n+1)] = \mathbf{B} [E[\mathbf{I} - E[\mathbf{\Psi}(n)]]] E[\mathbf{P}(n)] + E[\mathbf{\Psi}(n)] E[\mathbf{X}(n)] \quad (31)$$

From (31), the mean is stable if and only if the eigenvalues of matrix $\mathbf{B} [E[\mathbf{I} - E[\mathbf{\Psi}(n)]]]$ satisfy the following condition:

$$|\lambda_{\max}(\mathbf{B} [E[\mathbf{I} - E[\mathbf{\Psi}(n)]]])| = |\lambda_{\max}(\mathbf{B}\mathbf{Z})| < 1 \quad (32)$$

where $\mathbf{Z} = E[\mathbf{I} - E[\mathbf{\Psi}(n)]]$, and the maximum eigenvalue of a matrix is denoted by $\lambda_{\max}(\cdot)$. Because of the relation $\|\mathbf{B}\mathbf{Z}\| \leq \|\mathbf{B}\| \|\mathbf{Z}\|$ and $\|\mathbf{B}\| = 1$, we derive $|\lambda_{\max}(\mathbf{B}\mathbf{Z})| \leq |\lambda_{\max}(\mathbf{Z})|$. The algorithm will be stable if $|\lambda_{\max}(\mathbf{Z})| < 1$, so we have:

$$|\lambda_{\max}(\mathbf{Z})| = |\lambda_{\max}(\mathbf{I} - E[\mathbf{\Psi}(n)])| = |\lambda_{\max}(\mathbf{I} - \mu_k E[\mathbf{R}(n)])| < 1 \quad (33)$$

Thus, the step size satisfies:

$$|1 - \mu_k E[r_k(n)]| < 1 \quad (34)$$

We further derive:

$$0 < \mu_k < \frac{2}{E[r_k(n)]}, k = 1, 2, \dots, K \quad (35)$$

Therefore, the algorithm will be stable if the step size is in the bound of (35).

It is necessary to note that the condition of (35) is similar to those in [17,18]; the only difference is the extra term $E[r_k(n)]$, which is the expectation of the error nonlinearity introduced by MCC.

5.2. Variance Performance

We denote the covariance of the estimate as $Cov(\mathbf{P}(n))$, which is defined as:

$$Cov(\mathbf{P}(n)) = E[\|\mathbf{P}(n) - E[\mathbf{P}(n)]\|_{\Sigma}^2] \quad (36)$$

Substituting (29) and (30) into (36), we have:

$$Cov(\mathbf{P}(n+1)) = E[\|[\mathbf{I} - \mathbf{\Psi}(n)] \mathbf{X} \mathbf{P}(n) + \mathbf{\Psi}(n) \mathbf{X}(n) - [\mathbf{I} - E[\mathbf{\Psi}(n)]] \mathbf{B} E[\mathbf{P}(n)] - E[\mathbf{\Psi}(n)] E[\mathbf{X}(n)]\|_{\Sigma}^2] \quad (37)$$

We can see that when the kernel size is large, the elements in $\mathbf{\Psi}(n) = \mu_k \mathbf{R}(n)$ are very small, and the variation of $\mathbf{\Psi}(n)$ is also very small. Therefore, we can consider that $\mathbf{\Psi}(n) = E[\mathbf{\Psi}(n)]$. Thus, we have:

$$Cov(\mathbf{P}(n+1)) = E[\|\mathbf{B}(\mathbf{I} - E[\mathbf{\Psi}(n)]) (\mathbf{P}(n) - E[\mathbf{P}(n)]) + E[\mathbf{\Psi}(n)] (\mathbf{X}(n) - E[\mathbf{X}(n)])\|_{\Sigma}^2] \quad (38)$$

Considering the fact that $\mathbf{P}(n)$ and observation vector $\mathbf{X}(n)$ are independent, the covariance recursion can be shown as:

$$Cov(\mathbf{P}(n+1)) = \mathbf{B} (\mathbf{I} - E[\mathbf{\Psi}(n)]) E[\mathbf{P}(n) - E[\mathbf{P}(n)]] (\mathbf{I} - E[\mathbf{\Psi}(n)]) \mathbf{B}^T + E[\mathbf{\Psi}(n)] E[\mathbf{X}(n) - E[\mathbf{X}(n)]] E[\mathbf{\Psi}(n)] \quad (39)$$

Therefore, we have:

$$\text{Cov}(\mathbf{P}(n+1)) = \mathbf{B}(\mathbf{I} - E[\Psi(n)]) \text{Cov}(\mathbf{P}(n)) (\mathbf{I} - E[\Psi(n)]) \mathbf{B}^T + E[\Psi(n)] \text{Cov}(\mathbf{X}(n)) E[\Psi(n)] \tag{40}$$

This is the transient behavior of the network. Although (40) does not explicitly show the variance performance, it is in fact subsumed in the weighting matrix $\mathbf{B}(\mathbf{I} - E[\Psi(n)])$, which varies for each iteration. However, the effect of the algorithm on the performance is clearly shown in (40).

5.3. Detection Performance Analysis

In order to derive the probability of false alarm (P_f) and detection (P_d) for the proposed algorithm, the probability density function (PDF) of the test statistic $P_k(n)$ under both hypotheses H_0 and H_1 needs to be evaluated.

As mentioned earlier, the DMCC method is very robust to outlier points. Thus, when an impulsive noise occurs, the factor $\exp\left(-\frac{1}{2\sigma^2} \left| |x_k(n)|^2 - \Psi_k(n) \right|^2\right)$, which can be seen as the weight of the error $|x_k(n)|^2 - \Psi_k(n)$, is close to zero, so the impulsive noise does not lead to a large estimate. Therefore, when $K = 1$, we can regard the test statistic of energy detection $P_k(n)$ as a chi-square distributed random variable with $2N$ degrees of freedom. The test statistic $P_k(n)$ consists of a lot of identically distributed variables. According to the central limit theorem (CLT), when the number of samples is large enough, the chi-square distribution is approximated by a Gaussian distribution [35]. However, when $K > 1$, the test statistic $P_k(n)$ in the case of hypothesis H_1 consists of a sum of a various independent, but not identically distributed variables. In this case, we apply the Lyapunov CLT [36]; when the number of samples N is large enough, $P_k(n)$ is a Gaussian approximation. The simulation result (Figure 1), shows that this approximation is reasonable. In this case, when $P_k(n)$ is a Gaussian approximation, the formulas for the P_f and P_d of the energy detector can be derived under the Neyman–Pearson criterion [37].

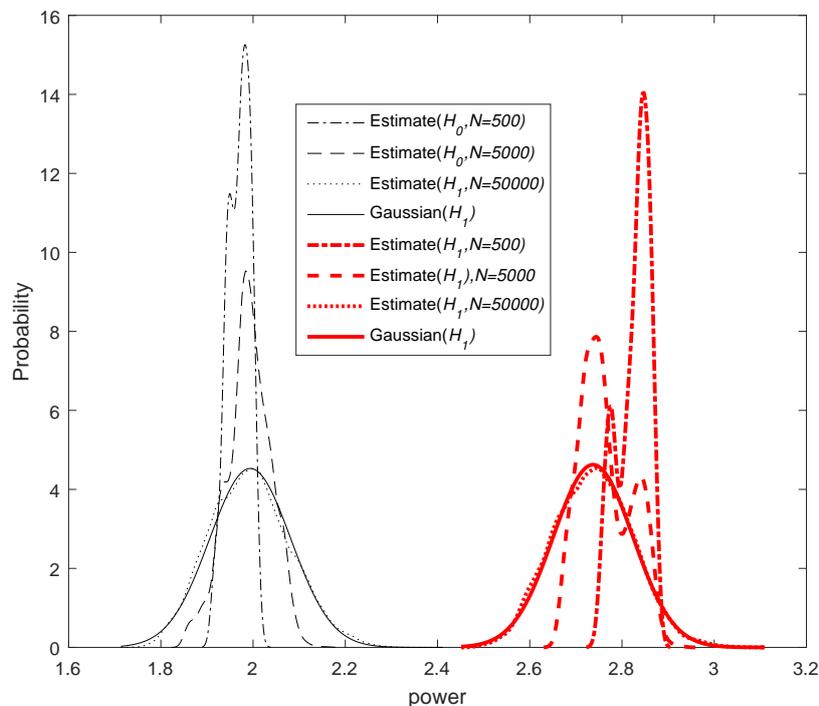


Figure 1. PDF of the power estimation and the Gaussian distribution.

By taking the previous results into account, the approximate formulas for the recursive performance are derived. The probability of false alarm P_f under hypothesis H_0 is given as follows:

$$P_f = Q \left(\frac{\lambda - E(P_k(n) | H_0)}{\sqrt{\text{Var}(P_k(n) | H_0)}} \right) \quad (41)$$

where λ is the threshold, and $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(-\frac{t^2}{2}\right) dt$.

Similarly, the probability of detection P_d under hypothesis H_1 is given as follows:

$$P_d = Q \left(\frac{\lambda - E(P_k(n) | H_1)}{\sqrt{\text{Var}(P_k(n) | H_1)}} \right) \quad (42)$$

The sensing threshold of the detector can be calculated from (41) by the predefined value of P_f . Thus:

$$\lambda = Q^{-1}(P_f) \sqrt{\text{Var}(P_k(n) | H_0)} + E(P_k(n) | H_0) \quad (43)$$

The threshold is determined by the noise power and the predefined false alarm probability P_f .

6. Simulation Results

In this section, the performance of the proposed algorithm is simulated. In order to verify the performance, we consider a distributed network with $K = 20$ CUs. The network topology is generated as the realization of the random geometric graph model (shown in Figure 2). The parameters for measurement Gaussian mixture noise are $\sigma_1^2 = 1$, $\sigma_2^2 = 100\sigma_1^2$ and $P_r(d_k(n) = 1) = p = 0.01$. We examine the learning performance of the new algorithm by the global average mean-square error, $MSE = \frac{1}{K} \sum_{k=1}^K E[|P^0 - P_k(n)|^2]$. We compare the performance of the new power estimation algorithm with the diffusion algorithm called ATC DLMS [18].

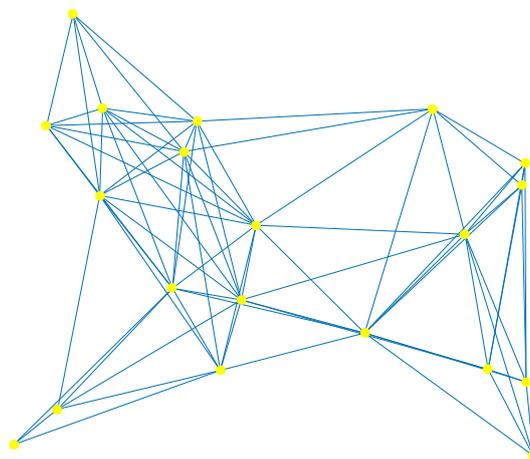


Figure 2. Network topology with $K = 20$ cognitive users (CUs).

6.1. Performance of Power Estimation Comparison among the New Method and Other Algorithms

First of all, we investigate the power estimation algorithm and the comparison among the new method (ATC DMCC) and ATC DLMS in [18]. The channel gain is assumed to be constant and generated by a standard normal distribution: $a_k \sim N(0, 1)$. All algorithms use the same channel gain. To guarantee the almost same initial convergence rate, we set the step sizes at 0.001 and 0.0004 for the mentioned ATC DMCC and ATC DLMS, respectively. The kernel size is chosen as 50 for ATC DMCC algorithm. Furthermore, all the CUs receive $N = 4 \times 10^4$ samples, and the PU signal is

absent during the first half of the samples $n = 1, 2, \dots, 2 \times 10^4$ and present in the other half of samples range. Under both detection hypotheses, all CUs are disturbed by the same Gaussian mixture noise. The combination weights $\gamma_{l,k}$ are calculated by the averaging rule: [38]

$$\gamma_{l,k} = \begin{cases} \frac{1}{n_k}, & \text{if } k \neq l \text{ are neighbors or } k = l \\ 0, & \text{otherwise} \end{cases} \quad (44)$$

where n_k is the degree of CU k , which means CU k has n_k neighbors. All parameters are set by scanning for the best results. All the simulation results are obtained by taking the ensemble average of the network over 100 independent Monte Carlo runs.

Figure 3 shows the performance curves in terms of power estimation with different signal to noise ratios (SNR = -5 dB, 0 dB, 5 dB). Figure 3a shows the convergence curves in terms of the power estimation. One can observe that the ATC DMCC algorithm works well in Gaussian mixture noise, while the ATC DLMS algorithm fluctuates significantly. As one can see from the results, the proposed ATC DMCC algorithm has a much better performance in convergence rate and accuracy compared with the ATC DLMS algorithm. The results confirm that the proposed algorithm shows a significant improvement in robust performance in impulsive noise environments. Figure 3b shows the convergence curves in terms of MSE, and Figure 3c shows the steady-state MSEs at each CU k . As expected, the ATC DMCC algorithm performs better than the ATC DLMS algorithm.

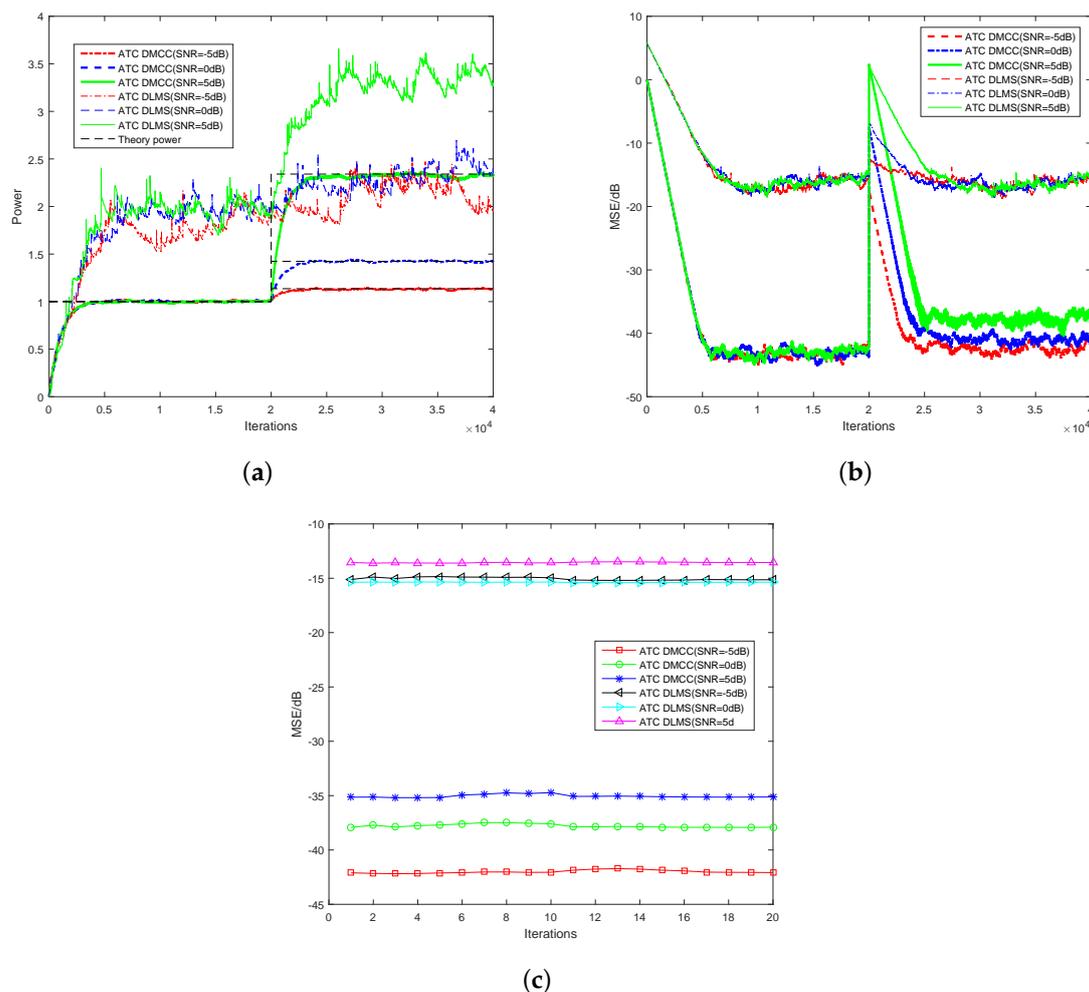


Figure 3. Performance in terms of power estimation with different SNRs. ATC, adaptation to combination; DMCC, diffusion maximum correntropy criterion. (a) Convergence curves in terms of power estimation; (b) Convergence curves in terms of MSE; (c) MSE at steady-state for the 20th CU.

Secondly, we compare the learning performance of the ATC DLMS algorithm with that of the proposed ATC DMCC under different kernel sizes σ . The kernel sizes of ATC DMCC are selected at $\sigma = 10, 50, 500, 1000, 5000$, respectively. The SNR is chosen as 5 dB. The other parameters for the algorithms keep the same as those in the previous simulation. It is known that kernel size is a key parameter for the proposed ATC DMCC algorithm. When the kernel size is $\sigma \rightarrow \infty$, the ATC DMCC algorithm degenerates into an ATC DLMS algorithm whose robustness is poor. From Figure 4, we can see that when the kernel size $\sigma = 5000$ (which is considered to be pretty large), the power estimation of the proposed ATC DMCC is as poor as ATC DLMS. On the other hand, the smaller the kernel size we choose, the stronger robustness is obtained. However, the difference between power estimates (being absent or present under the PU signal) is smaller, as well. This is because if the kernel size is small, the scaling factor $\exp\left(-\frac{1}{2\sigma^2} \left| |x_I(n)|^2 - \psi_k(n) \right|^2\right)$ in (17) will suppress the impulsive noise more effectively; the power estimate on the other hand will also be small. After consideration of the difference between power estimations (under H_0 and H_1) and the robustness of the proposed ATC DMCC, the kernel size is selected at $\sigma = 50$ for the following simulations.

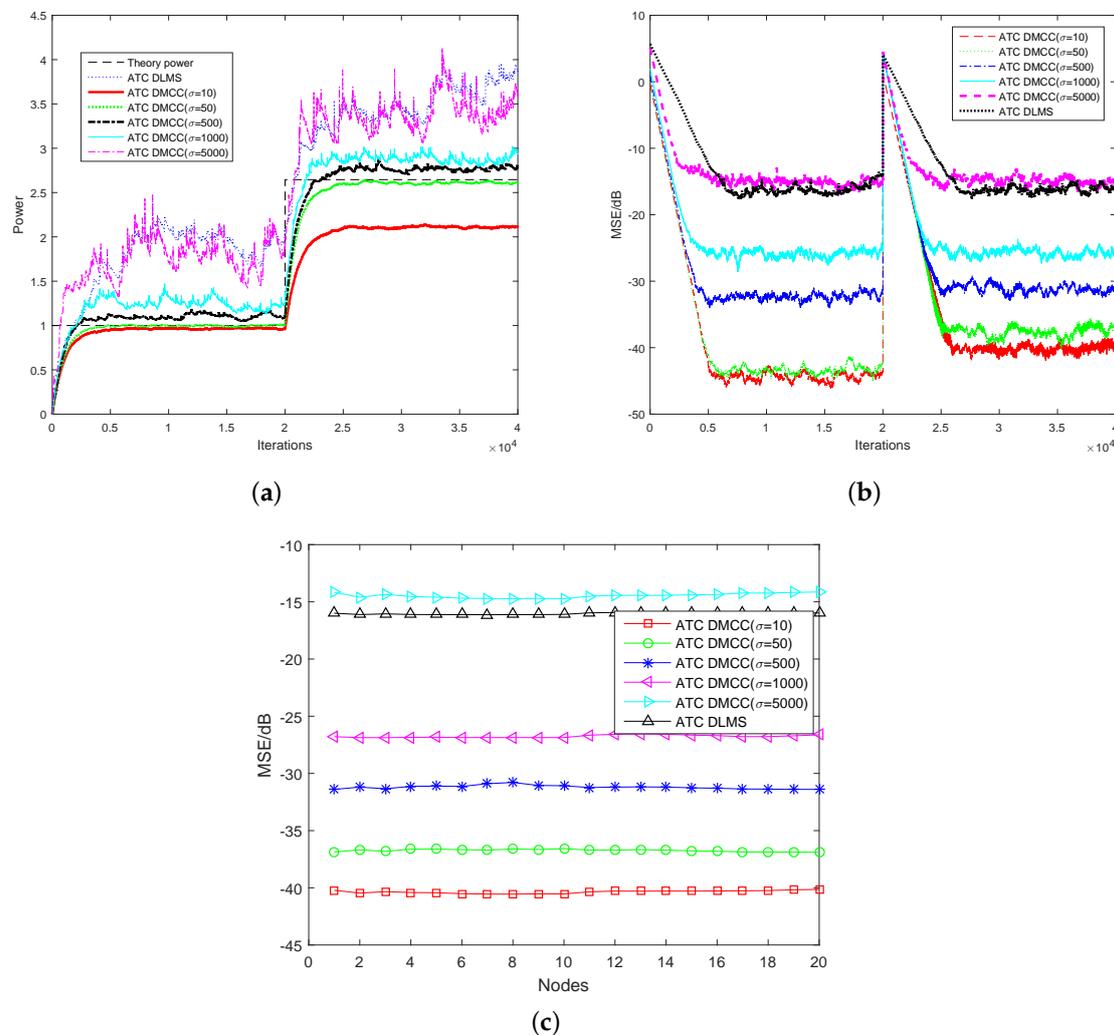


Figure 4. Performance in terms of power estimation with different σ . (a) Convergence curves in terms of power estimation; (b) Convergence curves in terms of MSE; (c) MSE at steady-state for the 20th CU.

6.2. Probability of Detection Comparison among ATC DMCC and ATC DLMS

Next, the probability of detection of the proposed robust power estimation algorithm is investigated. Then, we simulate the comparison among the new method and ATC DLMS. In the following simulations, we also set the network size $K = 20$ CUs. Figure 5 is the simulation result of ATC DMCC and ATC DLMS. Specifically, Figure 5a shows the receiver operating characteristic (ROC) curves of both schemes with different noise power levels (SNR = -15 dB, -10 dB, -5 dB). Figure 5b is the detection result with different false alarm probabilities ($P_f = 0.001, 0.01, 0.1$). As represented in Figure 5, the proposed algorithm has a much better performance than the ATC DLMS algorithm in non-Gaussian noise. This is because of the fact that the new method has the ability to restrain the impulse noise.

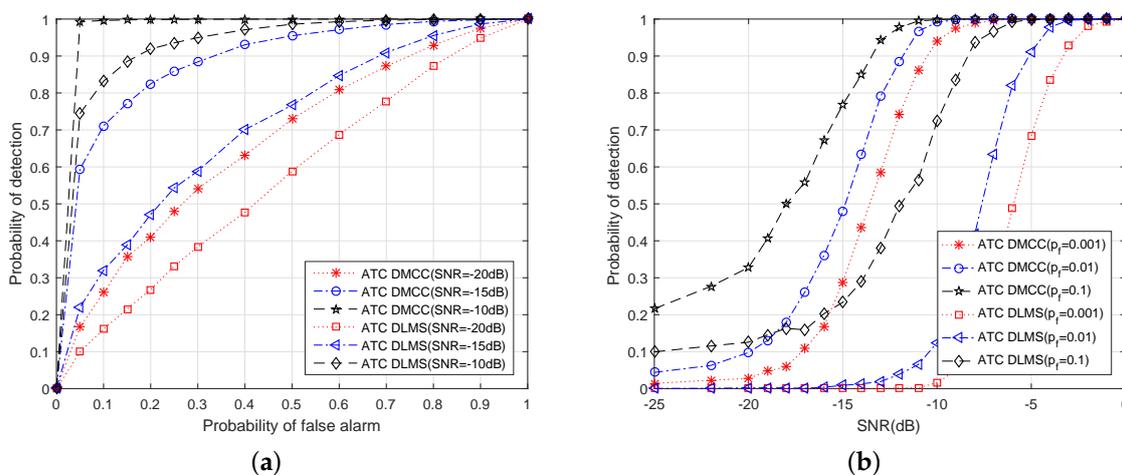


Figure 5. The probability of detection, ATC DMCC vs. ATC DLMS. (a) Description of what is contained in the first panel; (b) Description of what is contained in the second panel.

6.3. Probability of Detection of ATC DMCC with Different Network Sizes

Lastly, the performances of five different network sizes ($K = 1, 3, 7, 13, 20$) are simulated. The desired probability of false alarm is chosen as $P_f = 0.001$. In order to better display the simulation results, we set the step size at 0.004 for ATC DMCC. The comparison among the estimated and theoretical results of P_d is simulated. In the experiment, the threshold of the energy detector is calculated by using (43) for each CU under the detection hypothesis. The theoretical P_d is obtained by using (42). Figure 6 shows the detection performance curves of the proposed algorithm with different network sizes. One can observe that the P_d increases as the number of CUs increases. We can also see that the detection performance is the worst when $K = 1$. In this case, there is not much that can be done to improve the P_d if only one CU is used to detect the spectrum. For $K = 3, K = 7$ and $K = 13$, the detection probability increases due to the distributed estimation. However, when the network size is large enough (i.e., $K = 13$), the P_d increases a little as the number of CUs K increases. Meanwhile, we can see that the simulated and the theoretical P_d are almost the same.

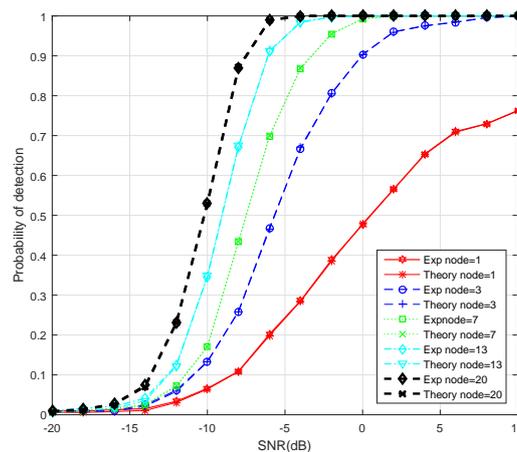


Figure 6. The probability of the detection of ATC DMCC with different CU numbers.

7. Conclusions

In this paper, we propose a version of the DMCC-based robust spectrum sensing scheme, namely the ATC DMCC algorithm, for impulsive noise. The new algorithm shows strong robustness against impulsive disturbance as MCC is very effective at handling non-Gaussian noise with large outliers. Mean and variance convergence analyses have been carried out. We also theoretically analyzed the detection performance of the new method. The performance of the proposed distributed ATC DMCC-based spectrum sensing algorithm has been compared with the ATC DLMS-based spectrum sensing [18]. Simulation results illustrate that the proposed algorithm performs very well in non-Gaussian noise environments. It can be concluded that the ATC DMCC method can achieve better performance than its predecessor (ATC DLMS) in impulsive noise.

Acknowledgments: This work was supported by the National Natural Science Foundation of China (No. 61531013), the National Science and Technology Major Project of China (No. 2018ZX03001016) and the Research Fund of Ministry of Education-China Mobile (MCM20150102).

Author Contributions: Xiguang Xu and Hua Qu modeled and analyzed the problem. Jihong Zhao, Feiyu Yan and Weihua Wang proposed and implemented the result. Xiguang Xu wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mitola, J.; Maguire, G.Q. Cognitive radio: Making software radios more personal. *IEEE Pers. Commun.* **1999**, *6*, 13–18.
2. Haykin, S. Cognitive radio: Brain-empowered wireless communications. *IEEE J. Sel. Areas Commun.* **2005**, *23*, 201–220.
3. Mitola, J. Cognitive radio architecture evolution. *Proc. IEEE* **2009**, *97*, 626–641.
4. Digham, F.F.; Alouini, M.S.; Simon, M.K. On the Energy Detection of Unknown Signals Over Fading Channels. *IEEE Trans. Commun.* **2007**, *55*, 21–24.
5. Ma, L.; Li, Y.; Demir, A. Matched filtering assisted energy detection for sensing weak primary user signals. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Kyoto, Japan, 25–30 March 2012; pp. 3149–3152.
6. Kim, K.; Akbar, I.A.; Bae, K.K.; Um, J.; Spooner, C.M.; Reed, J.H. Cyclostationary Approaches to Signal Detection and Classification in Cognitive Radio. In Proceedings of the IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, Dublin, Ireland, 18–20 April 2007; pp. 212–215.
7. Akyildiz, I.F.; Lo, B.F.; Balakrishnan, R. Cooperative spectrum sensing in cognitive radio networks: A survey. *Phys. Commun.* **2011**, *4*, 40–62.

8. Yucek, T.; Arslan, H. A survey of spectrum sensing algorithms for cognitive radio applications. *IEEE Commun. Surv. Tutor.* **2009**, *11*, 116–130.
9. Letaief, K.B.; Zhang, W. Cooperative Communications for Cognitive Radio Networks. *Proc. IEEE* **2009**, *97*, 878–893.
10. Xu, X.; Qu, H.; Zhao, J.; Chen, B. Cooperative spectrum sensing in cognitive radio networks with kernel least mean square. In Proceedings of the IEEE International Conference on Information Science and Technology, Changsha, China, 24–26 April 2015; pp. 574–578.
11. Rodger, J.A. Toward reducing failure risk in an integrated vehicle health maintenance system: A fuzzy multi-sensor data fusion Kalman filter approach for IVHMS. *Expert Syst. Appl.* **2012**, *39*, 9821–9836.
12. Thuc, K.X.; Insoo, K. Cooperative spectrum sensing using Kalman filter based adaptive fuzzy system for cognitive radio networks. *Ksii Trans. Internet Inf. Syst.* **2012**, *6*, 287–304.
13. Mishra, S.M.; Sahai, A.; Brodersen, R.W. Cooperative Sensing among Cognitive Radios. In Proceedings of the IEEE International Conference on Communications, London, UK, 8–12 June 2015; pp. 1658–1663.
14. Li, Z.; Yu, R.F.; Huang, M. A Distributed Consensus-Based Cooperative Spectrum-Sensing Scheme in Cognitive Radios. *IEEE Trans. Veh. Technol.* **2010**, *59*, 383–393.
15. Zhang, W.; Guo, Y.; Liu, H.; Chen, Y.; Wang, Z.; Mitola, J. Distributed consensus-based weight design for cooperative spectrum sensing. *IEEE Trans. Parallel Distrib. Syst.* **2015**, *26*, 54–64.
16. Wu, R. Distributed spectrum sensing using belief propagation framework. In Proceedings of the IEEE International Inter-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support, San Antonio, TX, USA, 3–6 March 2014; pp. 137–143.
17. Ainomae, A.; Trump, T.; Bengtsson, M. CTA diffusion based recursive energy detection. In Proceedings of the Latest Trends in Circuits, System, Signal Processing and Automatic Control, Salerno, Italy, 3–5 June 2014; pp. 38–47.
18. Ainomae, A.; Trump, T.; Bengtsson, M. Distributed diffusion LMS based energy detection. In Proceedings of the International Congress on Ultra Modern Telecommunications and Control Systems and Workshops, St. Peterburg, Russia, 6–8 October 2014; pp. 176–183.
19. Ainomae, A.; Trump, T.; Bengtsson, M. Distributed largest eigenvalue detection. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, New Orleans, LA, USA, 5–9 March 2017; pp. 3519–3523.
20. Middleton, D. Non-Gaussian noise models in signal processing for telecommunications: New methods and results for class A and class B noise models. *IEEE Trans. Inf. Theory* **1999**, *45*, 1129–1149.
21. Blackard, K.L.; Rappaport, T.S.; Bostian, C.W. Measurements and models of radio frequency impulsive noise for indoor wireless communications. *IEEE J. Sel. Areas Commun.* **1993**, *11*, 991–1001.
22. Sanchez, M.G.; De Haro, L.; Ramon, M.C.; Mansilla, A.; Ortega, C.M.; Oliver, D. Impulsive noise measurements and characterization in a UHF digital TV channel. *IEEE Trans. Electromagn. Compat.* **1999**, *41*, 124–136.
23. Nassar, M.; Gulati, K.; Sujeeth, A.K.; Aghasadeghi, N.; Evans, B.L.; Tinsley, K.R. Mitigating near-field interference in laptop embedded wireless transceivers. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV, USA, 31 March–4 April 2008; pp. 1–12.
24. Zhu, X.; Zhu, Y.; Bao, Y.; Zhu, W.; Zhu, W. A p th order moment based spectrum sensing for cognitive radio in the presence of independent or weakly correlated Laplace noise. *Signal Process.* **2017**, *173*, 109–123.
25. Margoosian, A.; Abouei, J.; Plataniotis, K.N. An accurate kernelized energy detection in Gaussian and non-Gaussian/impulsive noises. *IEEE Trans. Signal Process.* **2015**, *63*, 5621–5636.
26. Karimzadeh, M.; Rabiei, A.M.; Olfat, A. Soft-limited polarity-coincidence-array spectrum sensing in the presence of non-Gaussian noise. *IEEE Trans. Veh. Technol.* **2017**, *66*, 1418–1427.
27. Zhu, X.; Champagne, B.; Zhu, W.P. Rao test based cooperative spectrum sensing for cognitive radios in non-Gaussian noise. *Signal Process.* **2014**, *97*, 183–194.
28. Liu, W.; Pokharel, P.P.; Principe, J.C. Correntropy: Properties and applications in non-Gaussian signal processing. *IEEE Trans. Signal Process.* **2007**, *55*, 5286–5298.
29. Chen, B.; Wang, J.; Zhao, H.; Zheng, N.; Principe, J.C. Convergence of a fixed-point algorithm under maximum correntropy criterion. *IEEE Signal Process. Lett.* **2015**, *22*, 1723–1727.
30. Ma, W.; Chen, B.; Duan, J.; Zhao, H. Diffusion maximum correntropy criterion algorithms for robust distributed estimation. *Digit. Signal Process.* **2016**, *58*, 10–19.

31. Panigrahi, T.; Panda, G.; Mulgrew, B. Error saturation nonlinearities for robust incremental LMS over wireless sensor networks. *ACM Trans. Sens. Netw.* **2014**, *11*, 1–20.
32. Liu, W.; Pokharel, P.P.; Principe, J.C. The kernel least-mean-square algorithm. *IEEE Trans. Signal Process.* **2008**, *56*, 543–554.
33. Lopes, C.G.; Sayed, A.H. Diffusion least-mean squares over adaptive networks: Formulation and performance analysis. *IEEE Trans. Signal Process.* **2008**, *56*, 3122–3136.
34. Takahashi, N.; Yamada, I.; Sayed, A.H. Diffusion least-mean squares with adaptive combiners: Formulation and performance analysis. *IEEE Trans. Signal Process.* **2010**, *58*, 4795–4810.
35. Papoulis, A.; Pillai, S. *Probability, Random Variables and Stochastic Processes*, 4th ed.; McGraw Hill: New York, NY, USA, 2002.
36. Billingsley, P. *Probability and Measure*, 3rd ed.; Wiley: New York, NY, USA, 1995.
37. Blum, R.S. Necessary conditions for optimum distributed sensor detectors under the Neyman-Pearson criterion. *IEEE Trans. Inf. Theory* **1996**, *42*, 990–994.
38. Sayed, A.H. Diffusion adaptation over networks. *Acad. Press Libr. Signal Process.* **2013**, *3*, 323–453, arXiv:1205.4220.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).