

Article

Diffusion Logarithm-Correntropy Algorithm for Parameter Estimation in Non-Stationary Environments over Sensor Networks

Limei Hu ^{1,†}, Feng Chen ^{1,2,†,*}, Shukai Duan ² and Lidan Wang ²

¹ College of Electronic and Information Engineering, School of Mathematics and Statistics, Southwest University, Chongqing 400715, China; hlm0903@email.swu.edu.cn

² Key Laboratory of Nonlinear Circuits and Intelligent Information Processing, and College of Electronic and Information Engineering, Southwest University, and Chongqing Collaborative Innovation Center for Brain Science, Chongqing 400715, China; duansk@swu.edu.cn (S.D.); ldwang@swu.edu.cn (L.W.)

* Correspondence: fengchenmit@swu.edu.cn

† Current address: Chongqing Collaborative Innovation Center for Brain Science, Southwest University, Chongqing 400715, China.

Received: 12 August 2018; Accepted: 8 October 2018; Published: 10 October 2018

Abstract: This paper considers the parameter estimation problem under non-stationary environments in sensor networks. The unknown parameter vector is considered to be a time-varying sequence. To further promote estimation performance, this paper suggests a novel diffusion logarithm-correntropy algorithm for each node in the network. Such an algorithm can adopt both the logarithm operation and correntropy criterion to the estimation error. Moreover, if the error gets larger due to the non-stationary environments, the algorithm can respond immediately by taking relatively steeper steps. Thus, the proposed algorithm achieves smaller error in time. The tracking performance of the proposed logarithm-correntropy algorithm is analyzed. Finally, experiments verify the validity of the proposed algorithmic schemes, which are compared to other recent algorithms that have been proposed for parameter estimation.

Keywords: non-stationary; sensor networks; parameter estimation; diffusion logarithm-correntropy algorithm; tracking performance

1. Introduction

Sensor networks are useful tools for disaster relief management, target localization and tracking, and environment monitoring [1–4]. Distributed parameter estimation plays an essential role in sensor networks [5–7]. The objective of the parameter estimation is to estimate some essential parameters from noisy observation measurements through cooperation between nodes. Moreover, distributed strategies are of great significance to solve the problem of parameter estimation in sensor networks, due to their robustness against imperfections, low complexity, and low power demands.

Among these distributed schemes, in the incremental strategy [8], a cyclic path is defined over the nodes and data are processed in a cyclic manner through the network until optimization is achieved. However, determining a cyclic path that runs across all nodes is generally a challenging (NP-hard) task to perform. In the consensus strategy [9], vanishing step sizes are used to ensure that nodes can reach consensus and converge to the same optimizer in steady-state. In the diffusion strategy, information is processed locally and simultaneously at all nodes. The processed data are diffused through a real-time sharing mechanism that ripples through the network continuously [10,11]. The diffusion strategies are particularly attractive because they are robust [12–15], flexible, and fully distributed compared with incremental and consensus strategies, so we adopt diffusion strategies in this paper.

Most prior literature is mainly concerned with the case where nodes estimate the parameter vector collaboratively in the stationary case over sensor networks [10,16]. However, in the real world, the non-stationary case is normal. In this work, we mainly consider the parameter estimation in the non-stationary case, where the parameter is always time-varying. The observation data are nonlinear and non-Gaussian, since the data may be disturbed by changing communication links or outliers under the non-stationary environments.

Inspired by the differentiability and mathematical tractability of logarithm functions, we introduce the logarithm function as the error cost function [17]. Moreover, the correntropy criterion is a nonlinear measure of similarity between two random variables [18], which is a robust optimality criterion has been successfully used in the field of non-Gaussian signal processing. To make the error cost function more suitable for non-stationary environments, we propose a diffusion signal processing framework with a logarithm-correntropy cost function to solve the parameter estimation problem, which can elegantly and gradually adjust the cost function in its optimization based on the error amount.

A. Related Works

The tracking behavior of a wide range of adaptive networks under non-stationary conditions was thoroughly investigated in [19–22]. In stationary conditions, based on the p norm error criterion, a diffusion minimum average p -power (dLMP) was proposed to estimate the parameters in wireless sensor networks [23]. To estimate the mean-square weight deviations under the zero-mean stationary measurement noise, the proportionate-type normalized least mean square algorithms were proposed in [24]. The diffusion normalized least-mean-square algorithm (dNLMS) was proposed for parameter estimation in a distributed network [25], and the variable step size of the dNLMS algorithm was obtained by minimizing the mean-square deviation to achieve fast convergence rate. The gradient-descent total least-squares (dTLS) algorithm is a stochastic-gradient adaptive filtering algorithm that compensates for error in both input and output data [26]. The steady-state analysis of gradient-descent total least-squares was inspired by the energy-conservation-based approach to the performance analysis of adaptive filters. When measurement noise involves impulsive interference, Ni, Chen, and Chen [27] designed a diffusion sign-error LMS (dSE-LMS) to solve the parameter estimation. The tracking performance of a variable step-size diffusion LMS algorithm is considered in non-stationary environment [28], but this research did not get the closed-form expression of steady-state mean-square deviation (MSD) or excess mean-square error (EMSE) of the network. Consequently, the theory and simulation do not match well. To date, the performance of distributed estimation algorithms has been predominantly studied under stationary conditions. However, the performances of these algorithms may degrade in non-stationary environments.

To find the optimal adaptation step sizes over the networks, Abdolee, Vakilian, and Champagne [29] formulated a constrained nonlinear optimization problem and solved it through a log-barrier Newton algorithm in an iterative manner. By using the optimal step size at each node, the performance of diffusion least-mean squares (DLMS) could be improved in non-stationary signal environments. Compared with this research, the proposed algorithm can respond immediately by taking relatively steeper steps when the error gets larger, and as a result, the new algorithm can perform well in non-stationary environments without finding the optimal step size at each node.

B. Our Contributions and Organization

To further promote estimation performance in non-stationary environments over sensor networks, a novel algorithm needs to be designed. In this paper, the random-walk model is introduced for non-stationary environments. We proposed the logarithm-correntropy algorithm for parameter estimation in sensor networks under the non-stationary environments. This algorithm can adopt both the logarithm operation and correntropy criterion to the estimation error. Moreover, if the error gets larger due to the non-stationary environments, the algorithm can respond immediately by taking relatively steeper steps. Thus, the proposed algorithm achieves smaller error in time. The tracking

performance of the proposed algorithm was analyzed. Simulation results are presented to evaluate the proposed algorithm.

The rest of this paper is organized as follows. In Section 2, we describe the estimation problem in a non-stationary environment. Section 3 introduces the adapt-then-combine (ATC) diffusion logarithmic-correntropy algorithm. In Section 4, the tracking performance analysis of the proposed algorithm is presented. Simulation results are presented in Section 5. Finally, conclusions are drawn in Section 6.

Notation: In what follows, let bold letters denote random variables and non-bold letters represent their realizations. Operators $(\cdot)^T$ and $E[\cdot]$ denote transposition and expectation, respectively. \mathbf{I}_m denotes an $m \times m$ identity matrix. $\mathbf{1}$ is an $N \times 1$ all-unity vector. $|\cdot|$ is the absolute value of a scalar.

2. Estimation Problem in a Non-Stationary Environment

Consider a network with N nodes (sensors) deployed to observe some physical phenomena and specific events in a special environment. It is fundamentally necessary to consider and analyze parameter estimation under non-stationary conditions with the intent of employing them for practical applications. One challenge confronted in real-world applications is the non-stationary nature of the underlying parameters. For this purpose, a data model with a varying parameter is required. In this paper, we use the random walk model in [19] to depict the non-stationary condition.

Assumption 1. (Random Walk Model): The parameter vector varies based on the following model:

$$w_i^* = w_{i-1}^* + \eta_i, \quad (1)$$

where w_{i-1}^* is a random variable with a constant mean, where i is the time index. η_i is a zero-mean random sequence with a covariance matrix R_η .

Assumption 2. The sequence η_i is independent of $u_{k,i}$ and $n_{k,i}$ for all k and i .

Assumption 3. The initial conditions w_{-1}^* are independent of all $d_{k,i}$, $u_{k,i}$, $n_{k,i}$, and η_i .

At every time i , every node k can only exchange information with the nodes from its neighborhoods N_k (including node k itself), and takes a scalar measurement $d_{k,i}$ according to:

$$d_{k,i} = (w_i^*)^T u_{k,i} + n_{k,i}, \quad (2)$$

where $u_{k,i}$ denotes the $M \times 1$ random regression input signal vector and we assume $I > M$, $n_{k,i}$ is the Gaussian noise with zero mean and variance $\sigma_{n,k}^2$. The problem is to estimate an $M \times 1$ unknown varying vector w_i^* at each node k from collected measurements. The objective of the network is to search for all unknown variable w and find the best estimation w^* at the end by minimizing the MSE cost function in a distributed manner as follows:

$$J_k(w) = E \left| d_{k,i} - w^T u_{k,i} \right|^2. \quad (3)$$

The cost function of the global network can be described as:

$$\min_{(w)} J^{global}(w) = \sum_{k=1}^N E \left| d_{k,i} - w^T u_{k,i} \right|^2. \quad (4)$$

The optimization problem in Equation (3) can be solved by the diffusion strategies proposed in [30,31]. In these strategies, the estimate for each node is generated through a fixed combination

strategy, which refers to giving different weights to the estimation of k 's neighbors to minimize the local function as follows:

$$J_k^{local}(w) = \sum_{l \in N_k} c_{lk} E \left| d_{l,i} - w^T u_{l,i} \right|, \quad (5)$$

where c_{lk} is the combination coefficient. For simplicity and good performance, we use the Metropolis rule in our work. The description of the Metropolis rule is:

$$c_{l,k} = \begin{cases} \frac{1}{\max(n_k, n_l)}, & \text{if } l \in N_k \setminus k, \\ 1 - \sum_{l \in N_k \setminus k} c_{lk}, & \text{if } l = k, \\ 0, & \text{if } l \notin N_k, \end{cases} \quad (6)$$

where n_k is the degree of node k (the number of nodes connected to node k). The combining coefficients c_{lk} also satisfy the following conditions: $\sum_{l \in N_k \cup k} c_{lk} = 1$ and $c_{lk} = 0$ if $l \notin N_k$, $C \mathbb{I} = \mathbb{I}$, $\mathbb{I}^T C = \mathbb{I}^T$, where C is an $N \times N$ matrix with non-negative real entries $\{c_{lk}\}$.

3. Diffusion Logarithmic-Correntropy Algorithm

In the non-stationary case, the parameter is always time-varying. We propose a new logarithmic-correntropy method to solve the parameter estimation problem. In order to solve Equation (3), since nodes in sensor networks have access to the observed data, we can take advantage of node cooperation by introducing a distributed diffusion learning manner.

In this paper, we are inspired from the recent developments in the information theoretic learning (ITL) related to the "logarithmic cost function" and the "correntropy"-based approaches [17,32]. The logarithmic function is differentiable, which makes it mathematically tractable. We introduce the logarithmic function as an efficient cost function in the adaptive algorithm. In this framework, we introduce an error cost function using the logarithmic function given by:

$$J(e_{k,i}) = F(e_{k,i}) - \frac{1}{\alpha} \ln(1 + \alpha F(e_{k,i})), \quad (7)$$

where $\alpha > 0$ is a small systemic parameter and $F(e_{l,i})$ is a conventional cost function of the estimation error $e_{k,i}$ on each node k . The estimation error is $e_{k,i} = d_{k,i} - w^T u_{k,i}$. In this paper, we introduce the correntropy criterion to formulate the conventional cost function $F(\cdot)$. The correntropy is a similarity measure based on the ITL criterion. Given two random variables X and Y , the corresponding correntropy between them can be defined by [33]:

$$V(X, Y) = E[k_\sigma(X - Y)] = \int k_\sigma(x - y) dF_{XY}(x, y), \quad (8)$$

where $k_\sigma(\cdot)$ is a continuous, symmetric, positive-definite function with bandwidth σ , also called the Mercer kernel. E is an expectation operator. The joint distribution function of X and Y is $F_{XY}(x, y)$. The Gaussian kernel is mainly concerned in this paper.

$$k_\sigma(x - y) = \exp\left(-\frac{(x - y)^2}{2\sigma^2}\right) \quad (9)$$

For each node k , based on the correntropy criterion, the instantaneous conventional cost function $F(e_{k,i})$ is:

$$F(e_{k,i}) = k_\sigma(e_{k,i}) = \exp\left(-\frac{(d_{k,i} - w^T u_{k,i})^2}{2\sigma^2}\right) = \exp\left(-\frac{(e_{k,i})^2}{2\sigma^2}\right). \quad (10)$$

In non-stationary conditions over networks, the communication among nodes is subject to link noise, and it is natural that the observation vectors are affected by noise. The total least squares (TLS) method for estimation can have desirable performance by reducing the noise effect from both the observation vector and the data matrix [34]. We briefly explain the TLS method as follows:

Consider the linear parameter estimation problem $\mathbf{Ax} \approx \mathbf{b}$, where \mathbf{A} is the data matrix, \mathbf{b} is the observation vector, and \mathbf{x} is the unknown parameter vector. The least squares (LS) approach considers that the observation vector is noisy while the data matrix is noiseless. However, the total least squares (TLS) approach considers that both the observation vector and the data matrix are noisy [35]. The LS approach seeks the estimate of the unknown parameter vector \mathbf{x} by minimizing a sum of squared residuals expressed by:

$$\min_x (\mathbf{Ax} - \mathbf{b})^2, \quad (11)$$

while the TLS approach minimizes a sum of weighted squared residuals expressed by:

$$\min_x \frac{(\mathbf{Ax} - \mathbf{b})^2}{(\mathbf{x})^2 + 1}. \quad (12)$$

From the matrix algebra viewpoint, the total least squares (TLS) approach is a refinement of the LS method when there are errors in both the observation vector and the data matrix. Inspired by the desirable features of the TLS method, to make the logarithm-correntropy method more suitable for non-stationary environments, we rewrite the conventional cost function as:

$$\tilde{F}(e_{k,i}) = \exp\left(-\frac{(d_{k,i} - w^T u_{k,i})^2}{2\sigma^2(w^2 + 1)}\right) = \exp\left(-\frac{(e_{k,i})^2}{2\sigma^2(w^2 + 1)}\right). \quad (13)$$

To demonstrate the superiority of the proposed logarithm-correntropy method, we introduce different stochastic cost functions, such as the least mean square cost e^2 and absolute difference cost $|e|$. Figure 1 compares these cost functions with the proposed cost function logarithm-correntropy (e). It can be observed that the proposed cost function logarithm-correntropy (e) is less sensitive to tiny interference on the error, and shows comparable steepness for quite large error interference. Furthermore, this new logarithm-correntropy cost function benefits from mapping the original input space into a potential higher-dimensional "feature space", where linear methods can be employed. Particularly, if the error gets larger due to the non-stationary environments, the algorithm can respond immediately by taking relatively steeper steps. Thus, the proposed algorithm achieves smaller error in time and takes more gradual steps in space.

Given the data model, all nodes can observe data generated by the data model in Equation (2). It is natural to expect collaboration between nodes to be beneficial for a distributed sensor network. This means that neighbor nodes can share information with each other as permitted by the network topology. Therefore, according to Equations (7) and (13), we define the global cost function so that all nodes in the sensor network can be adapted in a distributed manner, then the new global function can be built as follows:

$$\begin{aligned} J^{global}(w) &= \sum_{k=1}^{N_k} \left(\tilde{F}(e_{k,i}) - \frac{1}{\alpha} \ln(1 + \alpha \tilde{F}(e_{k,i})) \right) \\ &= \sum_{k=1}^{N_k} \left(\exp\left(-\frac{(e_{k,i})^2}{2\sigma^2(w^2+1)}\right) + \frac{1}{\alpha} \ln\left(1 + \alpha \exp\left(-\frac{(e_{k,i})^2}{2\sigma^2(w^2+1)}\right)\right) \right). \end{aligned} \quad (14)$$

To develop the distributed diffusion logarithm-correntropy algorithm in non-stationary environments over sensor networks, we can build the following new diffusion Logarithm-Correntropy Algorithm (dLCA) local cost function at every node k as:

$$\begin{aligned}
 J_k^{local}(w_{k,i}) &= \sum_{l \in N_k} c_{lk} \left(\tilde{F}(e_{l,i}) - \frac{1}{\alpha} \ln(1 + \alpha \tilde{F}(e_{l,i})) \right) \\
 &= \sum_{l \in N_k} c_{lk} H(e_{l,i}) \\
 &= \sum_{l \in N_k} c_{lk} \left(\exp\left(-\frac{(e_{l,i})^2}{2\sigma^2((w_{l,i})^2+1)}\right) + \frac{1}{\alpha} \ln\left(1 + \alpha \exp\left(-\frac{(e_{l,i})^2}{2\sigma^2((w_{l,i})^2+1)}\right)\right) \right),
 \end{aligned} \tag{15}$$

where $w_{k,i}$ is the local estimate obtained by node k at time i , $e_{l,i} = d_{l,i} - (w_{l,i})^T u_{l,i}$ is the estimation error at node l , l denotes any neighbor node of node k , $H(e_{l,i}) = \tilde{F}(e_{l,i}) - \frac{1}{\alpha} \ln(1 + \alpha \tilde{F}(e_{l,i}))$, and the $c_{l,k}$ denote combining coefficients, which also is subjected to the Metropolis rule.

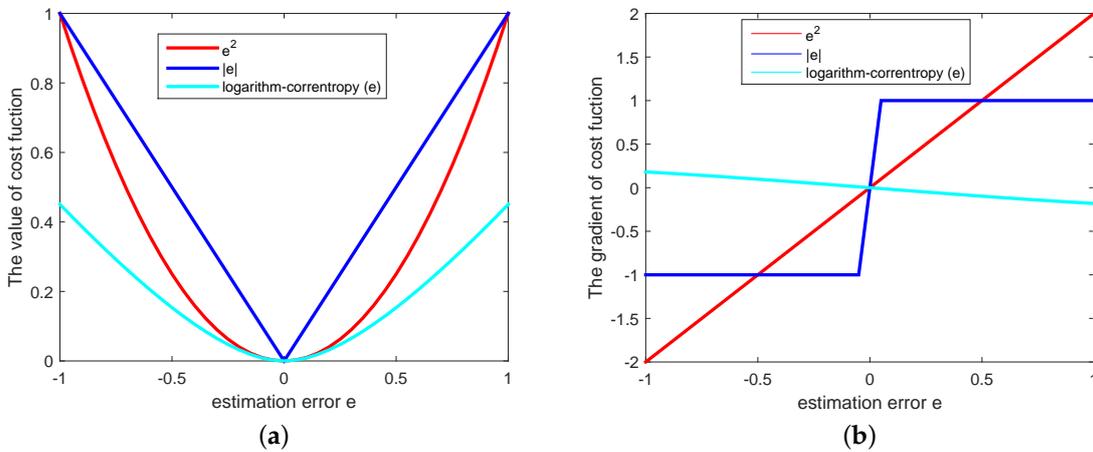


Figure 1. (a) The value of cost function; (b) The gradient of cost error function.

To reach the minimum w_i^* , it is a natural thought to use the steepest-descent method. Taking the derivative of Equation (15), we have

$$\frac{\partial J_k^{local}(w_{k,i})}{\partial w_{k,i}} = \sum_{l \in N_k} c_{lk} \frac{\partial H(e_{l,i})}{\partial w_{l,i}} = \sum_{l \in N_k} c_{lk} \left(\tau_{l,i} + \frac{\tau_{l,i}}{1 + \alpha \zeta_{l,i}} \right), \tag{16}$$

where $\zeta_{l,i} = \exp\left(-\frac{(e_{l,i})^2}{2\sigma^2((w_{l,i-1})^2+1)}\right)$ and $\tau_{l,i} = \left(\exp\left(-\frac{(e_{l,i})^2}{2\sigma^2((w_{l,i-1})^2+1)}\right) \cdot \frac{e_{l,i}(u_{l,i}+d_{l,i}w_{l,i})}{\sigma^2((w_{l,i-1})^2+1)^2} \right)$.

Since nodes in the sensor networks have access to all observed data, we can take advantage of node cooperation by introducing a diffusion strategy to estimate the parameter $w_{k,i}$ in a fully distributed manner. This paper concerns the Adapt-then-Combine (ATC) scheme of the diffusion strategy. As Figure 2 shows, in the ATC scheme, nodes in networks combine information from their immediate neighbors firstly, and then employ updates by the following steps:

(1) Adaptation: In order to obtain an intermediate estimate, we introduce a step-size parameter μ . Each node updates its current estimate for the true parameter value by taking steepest-descent method. We can obtain an intermediate estimate $\varphi_{k,i}$ as follows:

$$\varphi_{k,i} = \varphi_{k,i-1} - \mu \frac{\partial H(e_{k,i})}{\partial w_{k,i-1}} = \varphi_{k,i-1} - \mu \left(\tau_{l,i} + \frac{\tau_{l,i}}{1 + \alpha \zeta_{l,i}} \right). \tag{17}$$

(2) Combination: This step is also called the diffusion step, to obtain a new estimate, each node aggregates its own intermediate estimate from all its neighbor nodes as follows:

$$w_{k,i} = \sum_{l \in N_k} c_{lk} \varphi_{l,i}. \tag{18}$$

For the purpose of clarity, we summarize the procedures of the diffusion Logarithm-Correntropy Algorithm (dLCA) (Algorithm 1) as follows:

Algorithm 1: diffusion Logarithm-Correntropy Algorithm

Initialize: Start with $\{w_{l,-1} = 0\}$ for all l , initialize $w_{k,0}$ for each node k , step-size μ , and cooperative coefficients c_{lk} . Set $\alpha > 0, \sigma > 0$.

for $t = 1 : T$

for each node k :

Adaptation.

$\tilde{\zeta}_{l,i} = \exp\left(-\frac{(e_{l,i})^2}{2\sigma^2((w_{l,i-1})^2+1)}\right)$

$\tau_{l,i} = \left(\exp\left(-\frac{(e_{l,i})^2}{2\sigma^2((w_{l,i-1})^2+1)}\right) \cdot \frac{e_{l,i}(u_{l,i}+d_{l,i}w_{l,i})}{\sigma^2((w_{l,i-1})^2+1)^2}\right)$

$\varphi_{k,i} = \varphi_{k,i-1} - \mu \frac{\partial H(e_{k,i})}{\partial w_{k,i-1}} = \varphi_{k,i-1} - \mu \left(\tau_{l,i} + \frac{\tau_{l,i}}{1+\alpha\tilde{\zeta}_{l,i}}\right)$

Communication.

 Transmit the intermediate $\varphi_{k,i}$ to all neighbors in N_k .

Combination.

$w_{k,i} = \sum_{l \in N_k} c_{lk} \varphi_{l,i}$

N_k are the neighbor nodes of node k in the communication subnetwork.

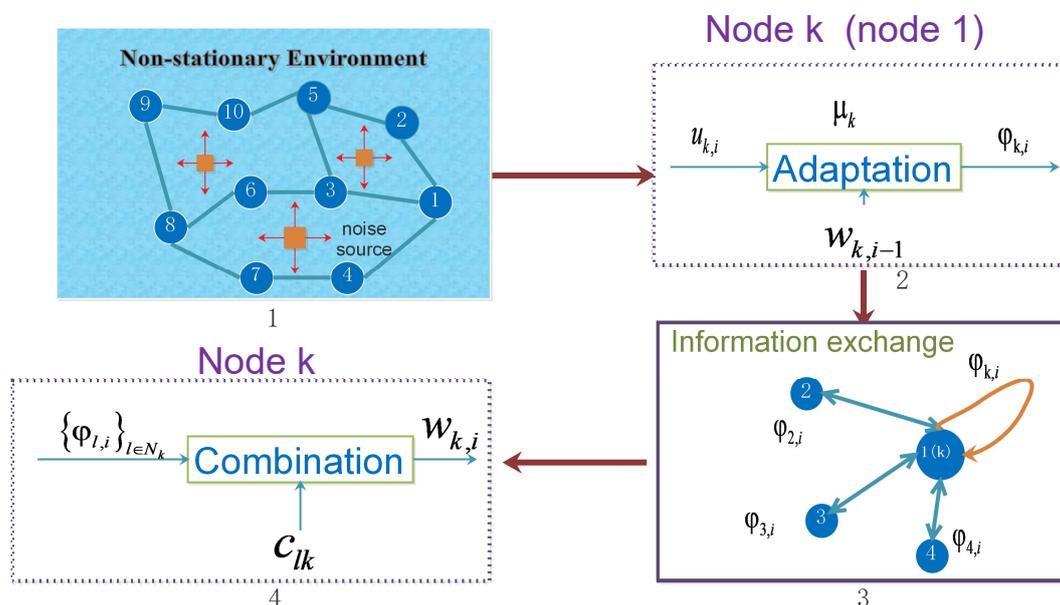


Figure 2. Adapt-then-Combine (ATC) diffusion strategies. Step 1 depicts a sensor network working in a non-stationary environment. In the adaptation stage 2, each node is using observed data $(u_{k,i}, d_{k,i})$ to update its intermediate estimate $\varphi_{k,i}$. Step 3 shows the information exchanging process between nodes. In the combination stage 4, each node collects the intermediate estimates from its neighbors.

4. Tracking Performance Analysis

The tracking performance of the proposed diffusion logarithm-correntropy algorithm is analyzed in this section. The convergence condition is first studied with $\tilde{\mathbf{w}}_i$, defined as the error signal, which is a time-varying parameter under the random walk model:

$$\tilde{\mathbf{w}}_i \triangleq \mathbf{w}_i^* - \mathbf{w}_{k,i}. \quad (19)$$

It has been proven that subtracting w_i^* from both sides of the update procedure on a node and then taking the expectation value leads to the following relation under stationary conditions in [10]:

$$\mathbb{E}\tilde{\mathbf{w}}_i = \left[I - \mu \sum_{k=1}^{N_k} R_{u,k} \right] \mathbb{E}\{\tilde{\mathbf{w}}_{i-1}\}. \quad (20)$$

Then, considering the Assumptions 2 and 3 of the random sequence $\{\eta_i\}$, we observe that w_i^* has a constant mean and hence $E[w_i^*] = E[w_{i-1}^*]$ under the relation in Equation (1). Taking the expectation value leads to the following relation under non-stationary conditions in Equation (1). We obtain

$$\mathbb{E}\tilde{\mathbf{w}}_i = \left[I - \mu \sum_{k=1}^{N_k} R_{u,k} \right] \mathbb{E}\{\tilde{\mathbf{w}}_{i-1} + \boldsymbol{\eta}_i\}. \quad (21)$$

In Equation (21), $\boldsymbol{\eta}_i$ is a zero-mean variable sequence with covariance matrix R_{η} . Our purpose is to achieve mean square deviation (MSD) and excess mean square error (EMSE) for each node, which are defined as:

$$MSD_k = \mathbb{E} \|\tilde{\mathbf{w}}_{k,\infty}\|_I^2. \quad (22)$$

$$EMSE_k = \mathbb{E} \|\tilde{\mathbf{w}}_{k,\infty}\|_{R_{u,k}}^2. \quad (23)$$

In the proposed algorithm in Equation (17), the error signals can be defined as follows:

$$\tilde{\boldsymbol{\psi}}_{k,i} \triangleq \mathbf{w}_i^* - \boldsymbol{\psi}_{k,i}, \quad (24)$$

$$\mathbf{e}_{k,i} \triangleq \mathbf{d}_{k,i} - \mathbf{u}_{k,i} \tilde{\boldsymbol{\psi}}_{k-1,i}. \quad (25)$$

In the non-stationary case with $w_i^* = w_{i-1}^* + \eta_i$, based on the definition in Equation (15), $J_k^{local}(w)$ is twice continuous differentiable when $w \neq 0$. Then we obtain the Hessian matrix of $J_k^{local}(w)$, which is defined as $\nabla_w^2 J_k^{local}(w)$.

From Lemma 1 and Theorem 1 in [12], the bound Hessian is: $\lambda_{k,\min} I_M \leq \nabla_w^2 J_k^{local} \leq \lambda_{k,\max} I_M$ and $0 \leq \lambda_{k,\min} \leq \lambda_{k,\max}$.

Equations (16)–(18) cause gradient error. The error recursion is then given by

$$\tilde{\boldsymbol{\varphi}}_{k,i} = [I_M - \mu \mathbf{H}_{k,i-1}] \tilde{\boldsymbol{w}}_{k,i-1} - \mu n_{k,i}, \quad (26)$$

$$\tilde{\boldsymbol{w}}_{k,i} = \sum_{l \in N_k} \mathbf{c}_{lk} \tilde{\boldsymbol{\varphi}}_{l,i}. \quad (27)$$

In Equation (26), as a positive-definite random matrix, $\mathbf{H}_{k,i-1}$ is defined as

$$\mathbf{H}_{k,i-1} \triangleq \int_0^1 \nabla_w^2 J_k^{local}(w^* - t \cdot \tilde{\boldsymbol{w}}_{k,i-1}) dt. \quad (28)$$

Applying Jensen's inequality to Equation (27), the variance of $\tilde{\boldsymbol{w}}_{k,i}$ is bounded by

$$E\|\tilde{w}_{k,i}\|^2 \leq \sum_{l \in N_k} c_{lk} E\|\tilde{\varphi}_{l,i}\|^2, \tag{29}$$

where $\|\cdot\|^2$ is a convex function and represents the squared Euclidean norm.

Integrating both sides of Equation (26), we achieve

$$E\|\tilde{\varphi}_{k,i}\|^2 = E\left(\|\tilde{w}_{k,i-1}\|_{\Sigma_{k,i-1}}^2\right) + \mu^2 E\|n_{k,i}\|^2, \tag{30}$$

$$\Sigma_{k,i-1} \triangleq (I_M - \mu H_{k,i-1})^T (I_M - \mu H_{k,i-1}) + \mu H_{k,i-1}^T H_{k,i-1}. \tag{31}$$

It follows from the bound Hessian and Equation (28) that

$$0 \leq \Sigma_{k,i-1} \leq \tau_k^2 I_M, \tag{32}$$

where

$$\tau_k^2 \triangleq \max \left\{ (1 - \mu_k \lambda_{k,\max}(R_{u,k}))^2, (1 - \mu_k \lambda_{k,\min}(R_{u,k}))^2 \right\} + \mu^2 \lambda_{k,\max}^2(R_{u,k}). \tag{33}$$

According to [12], substituting Equation (32) into Equation (30), we get

$$E\|\tilde{\varphi}_{k,i}\|^2 \leq \left(\tau_k^2 + \alpha \mu_k^2\right) E\|\tilde{w}_{k,i-1}\|^2 + \mu^2 \sigma_{n,k}^2, \tag{34}$$

where $\alpha \geq 0$ is a constant. The global MSD is introduced, which leads to

$$\tilde{w}_i \triangleq \text{col} \left\{ E\|\tilde{w}_{1,i}\|^2, E\|\tilde{w}_{2,i}\|^2, \dots, E\|\tilde{w}_{N,i}\|^2 \right\}. \tag{35}$$

We collect the $\{c_{lk}\}$ into $N \times N$ matrices C_i , such that $C = E[C_i]$. The C_i is left-stochastic, that is, $C_i^T \mathbf{1}_N = \mathbf{1}_N$, $\mathbf{1}_N$ means the $N \times 1$ all one vector. From Equations (29) and (35), it holds that

$$\tilde{w}_i \preceq C^T \Gamma \tilde{w}_{i-1} + C^T \Xi \mathbf{1}_N, \tag{36}$$

where \preceq denotes element-wise ordering and

$$\Gamma \triangleq \text{diag} \left\{ \tau_1^2 + \alpha \mu^2, \dots, \tau_N^2 + \alpha \mu^2 \right\}, \tag{37}$$

$$\Xi \triangleq \text{diag} \left\{ \mu^2 \sigma_{n,1}^2, \dots, \mu^2 \sigma_{n,N}^2 \right\}. \tag{38}$$

In order to ensure the stability of the proposed algorithm in the mean sense, according to Theorem 1 (*mean – square stability*) in Reference [12], it should hold that

$$\mu_k < \min \left\{ \frac{2\lambda_{k,\max}(R_{u,k})}{\lambda_{k,\max}^2(R_{u,k}) + \alpha}, \frac{2\lambda_{k,\min}(R_{u,k})}{\lambda_{k,\min}^2(R_{u,k}) + \alpha} \right\}. \tag{39}$$

Since $\lambda_{k,\min}^2(R_{u,k}) + \lambda_{k,\max}^2(R_{u,k}) \geq \lambda_{k,\min}^2(R_{u,k})$. As $i \rightarrow \infty$, which indicates

$$\lim_{i \rightarrow \infty} \|\tilde{w}_i\|_\infty \leq \frac{\|\Xi\|_\infty}{1 - \|\Gamma\|_\infty} = \frac{\max_k (\mu^2 \sigma_{n,k}^2)}{1 - \max_k (\tau_k^2 + \alpha \mu^2)}, \tag{40}$$

where $\|\cdot\|_\infty$ is the l_∞ norm. When the step-sizes $\{\mu\}$ are sufficiently small, we can further yield the conclusion that

$$\lim_{i \rightarrow \infty} \|\tilde{w}_i\|_\infty \leq \frac{\sigma_{n,k}^2}{2 \min_{1 \leq k \leq N} (\lambda_{k,\min})} \mu_{\max}^2. \tag{41}$$

According to the bound in Equation (41), if step-sizes $\{\mu\}$ are sufficiently small, the MSD of each node is $E\|\tilde{\omega}_k(i)\|^2$, which can become sufficiently small.

5. Simulation Results

In this section, to verify the performance of the proposed diffusion logarithm-correntropy algorithm, we considered a network consisting of 20 nodes and 50 communication links. The topology is shown in Figure 3. The sensor nodes were randomly deployed in an area of 100×100 and the communication distance between nodes was set as 35. All results below were averaged over 150 independent Monte Carlo simulations with randomly generated samples.

In this simulation part, firstly, the performance of the proposed algorithm was verified in a non-stationary environment over sensor networks and the communication links were ideal. In Figure 4, the regression inputs $\{\mathbf{u}_k(i)\}$ are independent identically distributed (i.i.d.), which are zero-mean Gaussian vectors with covariance matrices $\mathbf{R}_{u,k} = \sigma_{u,k}^2 \mathbf{I}_M$, and the $\sigma_{u,k}^2$ is the input variance. The background noises $\{\mathbf{n}_k(i)\}$ are drawn independently of the regressors and are i.i.d. The unknown parameter vector w_i^* is time-varying, as Figure 5 shows. The fixed step-size $\mu = 0.002$ is used in the simulations.

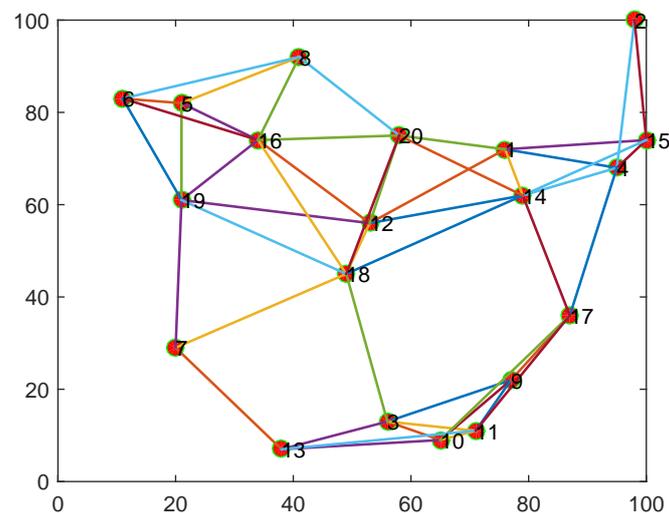


Figure 3. The topology of sensor network with 20 nodes.

The MSD learning curves are plotted in Figure 6. It shows that the proposed dLCA algorithm obtained the fastest convergence rate when compared with the dSE-LMS, dLMP, dTLS, and dNLMS algorithms. It also shows that the dLCA algorithm could achieve relatively good performance in terms of the network MSD. The proposed algorithm had relatively smaller MSD than the mentioned algorithms. From these simulation results, it can be seen that diffusion logarithm-correntropy algorithm exhibited better tracking ability in non-stationary environments than the existing classical algorithms.

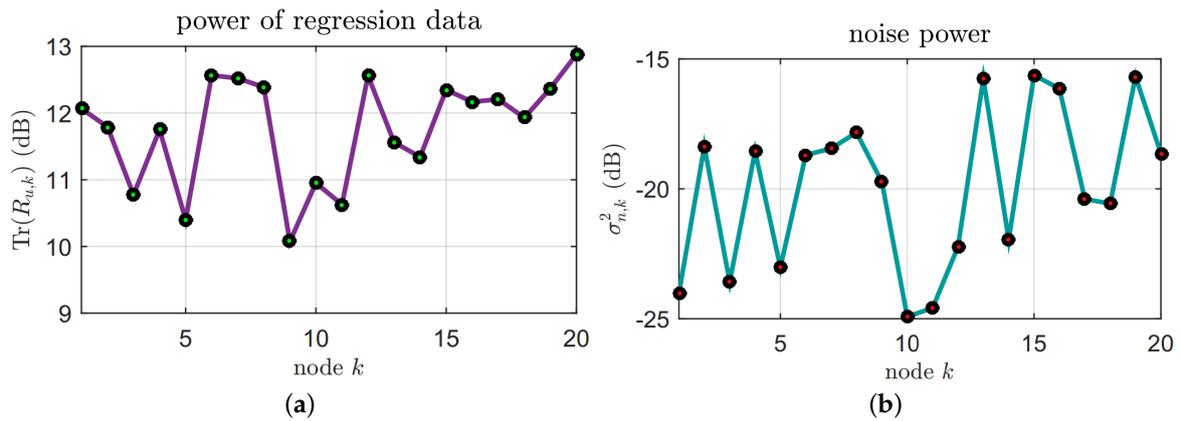


Figure 4. (a) Regressor statistics; (b) Noise variances.

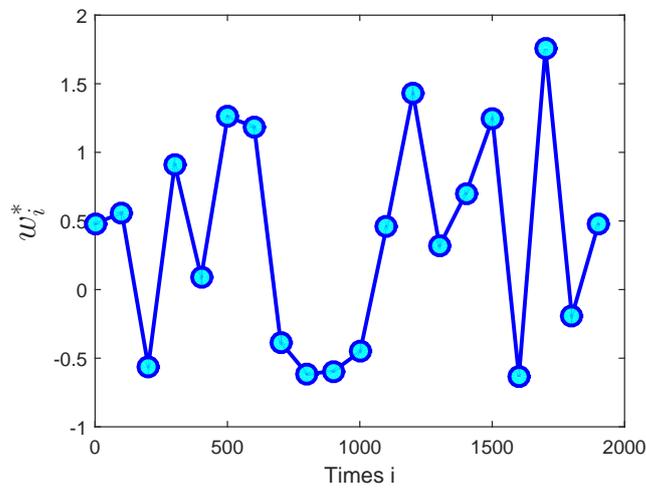


Figure 5. The desired time-varying vector, w_i^* .

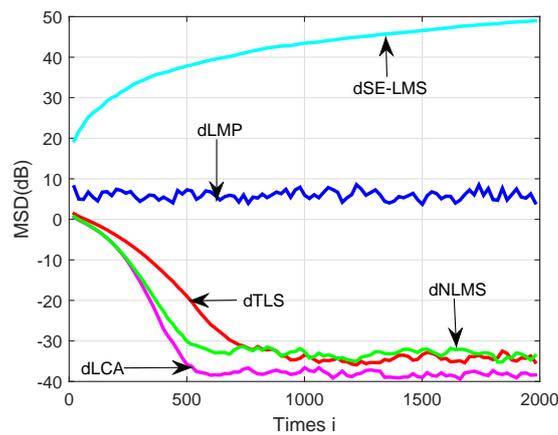


Figure 6. A comparison of simulated MSD learning curves in a non-stationary environment over sensor networks for the diffusion sign-error least-mean-square (dSE-LMS), diffusion minimum average p-power (dLMP), gradient-descent total least-squares (dTLS), diffusion normalized least-mean-square algorithm (dNLMS), and diffusion logarithm-correntropy algorithm (dLCA) algorithms.

Figure 7 compares the steady-state EMSE performances of related algorithms on each node in the sensor networks. It can be observed that a large difference was observed at some nodes that achieved low EMSE. By averaging over 150 experiments and over 50 time samples after convergence, the steady-state EMSE values were obtained. The proposed algorithm captured a better trend of the steady-state performance than other algorithms.

Secondly, to further simulate the non-stationary scenarios in sensor networks, the link was assumed to change at time 4000. The unknown parameter vector w_i^* was time-varying with link changing, as Figure 8 shows.

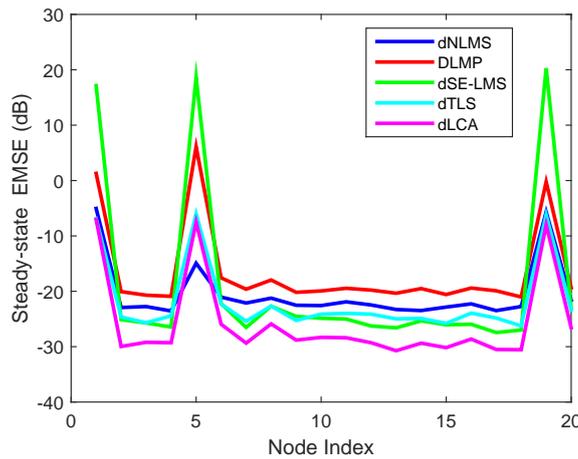


Figure 7. Estimated accuracy comparison in terms of excess mean-square error (EMSE) on each node for the dSE-LMS, dLMP, dTLS, dNLMS, and dLCA algorithms.

From the simulation results shown in Figure 9, in non-stationary environments over sensor networks with links changing, the diffusion logarithm-correntropy algorithm had smaller MSD than other related algorithms, such as dSE-LMS, dLMP, dTLS, and dNLMS algorithms. It further shows that the proposed dLCA algorithm had better tracking ability in non-stationary environments.

Finally, we compared the simulated network MSD curves with theoretical results under Equation (41) in Figure 10. One can see that theoretical network MSD curves of the proposed algorithm showed good match with its simulated MSD curves.

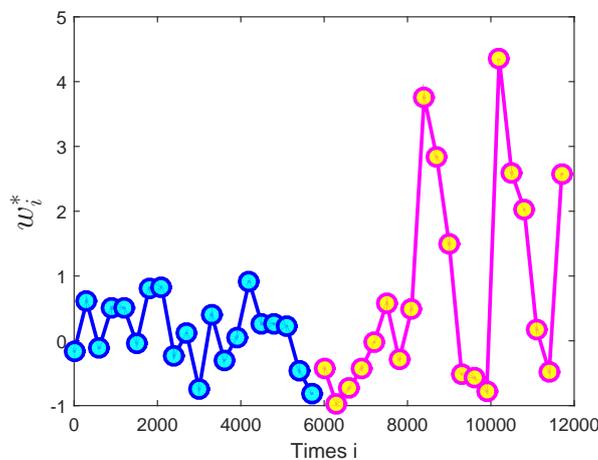


Figure 8. The desired time-varying vector with link changing, w_i^* .

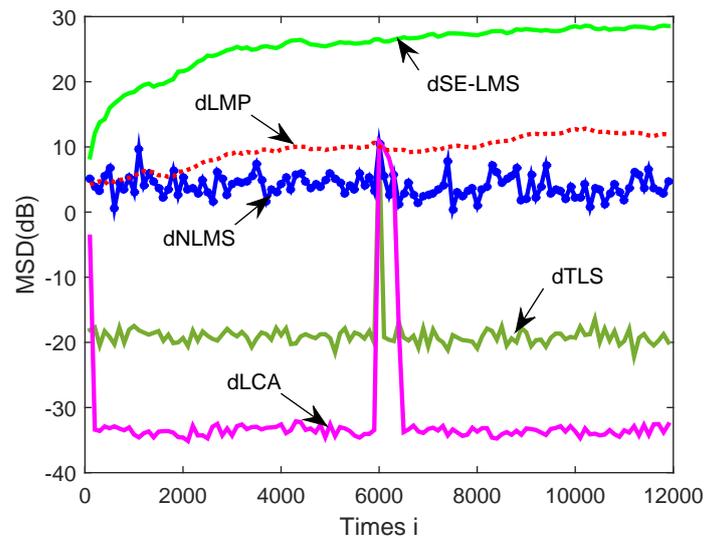


Figure 9. A comparison of simulated MSD learning curves of the global network for the dSE-LMS, dLMP, dTLS, dNLMS, and dLCA algorithms in non-stationary environments over sensor networks with links changing.

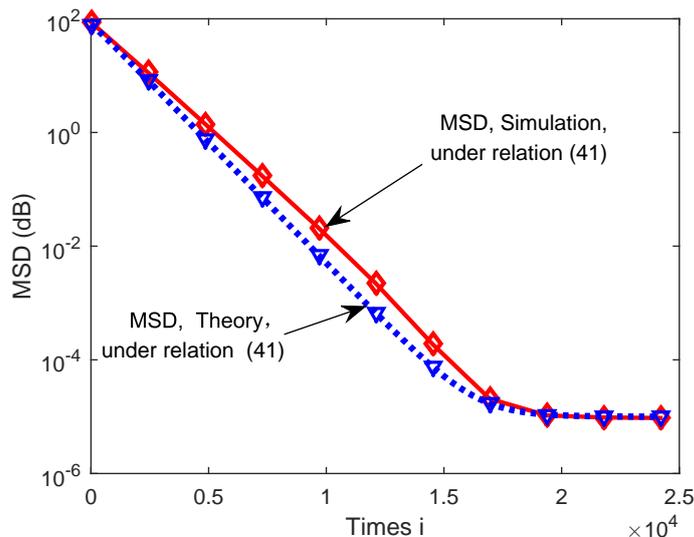


Figure 10. Theoretical and simulated MSD curves of the proposed dLCA algorithm under Equation (41).

6. Conclusions

To solve the problem of parameter estimation in non-stationary environments over sensor networks, each node in the sensor networks was equipped with the logarithm-correntropy cost function. The proposed algorithm can gradually adjust the cost function in its optimization based on the estimation error amount. We investigated the tracking behavior of the proposed algorithm under non-stationary conditions. Furthermore, the simulations were implemented in the non-stationary environments, where the parameters were time-varying with link changing. Simulation experiments were conducted to verify the analytical results, and illustrated that the proposed algorithm outperformed existing algorithms, such as dSE-LMS, dLMP, dTLS, and dNLMS algorithms.

Author Contributions: Data curation, F.C.; Funding acquisition, S.D.; Project administration, L.H.; Software, L.H.; Supervision, S.D. and L.W.; Writing—original draft, L.H.; Writing—review and editing, F.C. and L.W.

Funding: This work was supported in part by the Doctoral Fund of Southwest University (No. SWU113067), the Fundamental Research Funds for the Central Universities (Grant Nos. XDJK2017B053, XDJK2017D176, XDJK2017D180, XDJK2017D181), Chongqing Research Program of Basic Research and Frontier Technology (No. cstc2017jcyjAX0265) and Program for New Century Excellent Talents in University (Grant No. [2013] 47).

Acknowledgments: Authors would like to thank the editor and the anonymous reviewers for their constructive comments that improved the quality of the paper.

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

References

1. Safdarian, A.; Fotuhi-Firuzabad, M.; Lehtonen, M. A distributed algorithm for managing residential demand response in smart grids. *IEEE Trans. Ind. Inf.* **2014**, *10*, 2385–2393. [[CrossRef](#)]
2. Talebi, S.P.; Kanna, S.; Mandic, D.P. A Distributed Quaternion Kalman Filter with Applications to Smart Grid and Target Tracking. *IEEE Trans. Signal Inf. Process. Netw.* **2016**, *2*, 477–488. [[CrossRef](#)]
3. Harris, N.; Cranny, A.; Rivers, M. Application of distributed wireless chloride sensors to environmental monitoring: Initial results. *IEEE Trans. Instrum. Meas.* **2016**, *4*, 736–743. [[CrossRef](#)]
4. Sayed, A.H. *Adaptive Filters*; John Wiley and Sons: Hoboken, NJ, USA, 2008.
5. Tan, T.H.; Gochoo, M.; Chen, Y.F.; Hu, J.J.; Chiang, J.Y.; Chang, C.S.; Hsu, J.C. Ubiquitous emergency medical service system based on wireless biosensors, traffic information, and wireless communication technologies: Development and evaluation. *Sensors* **2017**, *17*, 202. [[CrossRef](#)] [[PubMed](#)]
6. Jiang, P.; Xu, Y.; Liu, J. A Distributed and Energy-Efficient Algorithm for Event K-Coverage in Underwater Sensor Networks. *Sensors* **2017**, *17*, 186. [[CrossRef](#)] [[PubMed](#)]
7. Kang, X.; Huang, B.; Qi, G. A Novel Walking Detection and Step Counting Algorithm Using Unconstrained Smartphones. *Sensors* **2018**, *18*, 297. [[CrossRef](#)] [[PubMed](#)]
8. Lopes, C.G.; Sayed, A.H. Incremental adaptive strategies over distributed networks. *IEEE Trans. Signal Process.* **2007**, *55*, 4064–4077. [[CrossRef](#)]
9. Soatti, G.; Nicoli, M.; Savazzi, S. Consensus-Based Algorithms for Distributed Network-State Estimation and Localization. *IEEE Trans. Signal Inf. Process. Netw.* **2017**, *3*, 430–444. [[CrossRef](#)]
10. Liu, Y.; Li, C.; Zhang, Z. Diffusion sparse least-mean squares over networks. *IEEE Trans. Signal Process.* **2012**, *60*, 4480–4485.
11. Abdolee, R.; Vakilian, V. An Iterative Scheme for Computing Combination Weights in Diffusion Wireless Networks. *IEEE Wirel. Commun. Lett.* **2017**, *6*, 510–513. [[CrossRef](#)]
12. Chen, J.; Sayed, A.H. Diffusion adaptation strategies for distributed optimization and learning over networks. *IEEE Trans. Signal Process.* **2012**, *60*, 4289–4305. [[CrossRef](#)]
13. Chen, F.; Shao, X. Broken-motifs Diffusion LMS Algorithm for Reducing Communication Load. *Signal Process.* **2017**, *133*, 213–218. [[CrossRef](#)]
14. Chouvardas, S.; Slavakis, K.; Theodoridis, S. Adaptive robust distributed learning in diffusion sensor networks. *IEEE Trans. Signal Process.* **2011**, *10*, 4692–4707. [[CrossRef](#)]
15. Chen, F.; Shi, T.; Duan, S.; Wang, L. Diffusion least logarithmic absolute difference algorithm for distributed estimation. *Signal Process.* **2018**, *142*, 423–430. [[CrossRef](#)]
16. Sayed, A.H. *Fundamentals of Adaptive Filtering*; Wiley: Hoboken, NJ, USA, 2003.
17. Sayin, M.O.; Vanli, N.D.; Kozat, S.S. A Novel Family of Adaptive Filtering Algorithms Based on the Logarithmic Cost. *IEEE Trans. Signal Process.* **2014**, *62*, 4411–4424. [[CrossRef](#)]
18. Chen, B.; Xing, L.; Liang, J.; Zheng, N.; Principe, J.C. Steady-state mean-square error analysis for adaptive filtering under the maximum correntropy criterion. *IEEE Signal Process. Lett.* **2014**, *21*, 880–884.
19. Nosrati, H.; Shamsi, M.; Taheri, S.M.; Sedaagh, M.H. Adaptive networks under non-stationary conditions: Formulation, performance analysis, and application. *IEEE Trans. Signal Process.* **2015**, *63*, 4300–4314. [[CrossRef](#)]
20. Predd, J.B.; Kulkarni, S.B.; Poor, H.V. Distributed learning in wireless sensor networks. *IEEE Signal Process. Mag.* **2006**, *23*, 56–69. [[CrossRef](#)]
21. Bertsekas, D.P.; Tsitsiklis, J.N. Gradient convergence in gradient methods with errors. *SIAM J. Optim.* **2000**, *10*, 627–642. [[CrossRef](#)]

22. Arablouei, R.; Dogancay, K. Adaptive Distributed Estimation Based on Recursive Least-Squares and Partial Diffusion. *IEEE Trans. Signal Process.* **2014**, *14*, 3510–3522. [[CrossRef](#)]
23. Wen, F. Diffusion least-mean P-power algorithms for distributed estimation in alpha-stable noise environments. *Electron. Lett.* **2013**, *49*, 1355–1356. [[CrossRef](#)]
24. Wagner, K.; Doroslovacki, M. Proportionate-type normalized least mean square algorithms with gain allocation motivated by mean-square-error minimization for white input. *IEEE Trans. Signal Process.* **2011**, *59*, 2410–2415. [[CrossRef](#)]
25. Jung, S.M.; Seo, J.H.; Park, P. A variable step-size diffusion normalized least-mean-square algorithm with a combination method based on mean-square deviation. *Circuits Syst. Signal Process.* **2015**, *34*, 3291–3304. [[CrossRef](#)]
26. Arablouei, R.; Werner, S. Analysis of the gradient-descent total least-squares adaptive filtering algorithm. *IEEE Trans. Signal Process.* **2014**, *62*, 1256–1264. [[CrossRef](#)]
27. Ni, J.; Chen, J.; Chen, X. Diffusion sign-error LMS algorithm: Formulation and stochastic behavior analysis. *Signal Process.* **2016**, *128*, 142–149. [[CrossRef](#)]
28. Zhao, X.; Tu, S.; Sayed, A.H. Diffusion adaptation over networks under imperfect information exchange and non-stationary data. *IEEE Trans. Signal Process.* **2012**, *60*, 3460–3475. [[CrossRef](#)]
29. Abdolee, R.; Vakilian, V.; Champagne, B. Tracking Performance and Optimal Step-Sizes of Diffusion LMS Algorithms in Nonstationary Signal Environment. *IEEE Trans. Control Netw. Syst.* **2016**, *5*, 67–78. [[CrossRef](#)]
30. Cattivelli, F.S.; Sayed, A.H. Diffusion LMS Strategies for Distributed Estimation. *IEEE Trans. Signal Process.* **2010**, *58*, 1035–1048. [[CrossRef](#)]
31. 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. [[CrossRef](#)]
32. Chen, B.; Xing, L.; Zhao, H.; Zheng, N.; Pri, J.C. Generalized correntropy for robust adaptive filtering. *IEEE Trans. Signal Process.* **2016**, *64*, 3376–3387. [[CrossRef](#)]
33. Chen, B.; Xing, L.; Xu, B.; Zhao, H.; Principe, J.C. Insights into the Robustness of Minimum Error Entropy Estimation. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *3*, 1–7. [[CrossRef](#)] [[PubMed](#)]
34. Rahman, M.D.; Yu, K.B. Total least squares approach for frequency estimation using linear prediction. *IEEE Trans. Acoust. Speech Signal Process.* **1987**, *35*, 1440–1454. [[CrossRef](#)]
35. Markovsky, I.; Van Huffel, S. Overview of total least-squares methods. *Signal Process.* **2007**, *87*, 2283–2302. [[CrossRef](#)]



© 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/>).