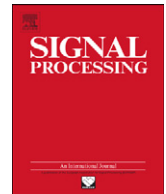




ELSEVIER

Contents lists available at SciVerse ScienceDirect

## Signal Processing

journal homepage: [www.elsevier.com/locate/sigpro](http://www.elsevier.com/locate/sigpro)

# Consensus based distributed change detection using Generalized Likelihood Ratio methodology<sup>☆</sup>

Nemanja Ilić<sup>a,\*</sup>, Srdjan S. Stanković<sup>a</sup>, Miloš S. Stanković<sup>b</sup>, Karl Henrik Johansson<sup>b</sup>

<sup>a</sup> Faculty of Electrical Engineering, University of Belgrade, 11000 Belgrade, Serbia

<sup>b</sup> School of Electrical Engineering, Royal Institute of Technology, 100-44 Stockholm, Sweden

## ARTICLE INFO

## Article history:

Received 4 June 2011

Received in revised form

6 January 2012

Accepted 7 January 2012

## Keywords:

Sensor networks

Distributed change detection

Generalized Likelihood Ratio

Consensus

Convergence

## ABSTRACT

In this paper a novel distributed algorithm derived from the Generalized Likelihood Ratio is proposed for real time change detection using sensor networks. The algorithm is based on a combination of recursively generated local statistics and a global consensus strategy, and does not require any fusion center. The problem of detection of an unknown change in the mean of an observed random process is discussed and the performance of the algorithm is analyzed in the sense of a measure of the error with respect to the corresponding centralized algorithm. The analysis encompasses asymmetric constant and randomly time varying matrices describing communications in the network, as well as constant and time varying forgetting factors in the underlying recursions. An analogous algorithm for detection of an unknown change in the variance is also proposed. Simulation results illustrate characteristic properties of the algorithms including detection performance in terms of detection delay and false alarm rate. They also show that the theoretical analysis connected to the problem of detecting change in the mean can be extended to the problem of detecting change in the variance.

© 2012 Elsevier B.V. All rights reserved.

## 1. Introduction

One of the typical tasks of *sensor networks*, which is in the focus of many researchers, is *distributed detection*, e.g., [1,2]. The classical multi-sensor distributed detection schemes require the existence of a *fusion center*, which collects relevant information from all the sensors and where the final decision is made. In [3] distributed detection has been broadly divided into *three classes*, where the aforementioned parallel architecture with a

fusion center represents the first class. Removal of a global fusion center brings, in principle, many advantages, consisting of increased reliability and reduced communication requirements, in spite of a certain loss of performance with respect to the optimal centralized system. The second class includes some recent attempts to apply *consensus techniques* to the distributed detection problem in order to eliminate the need for a fusion center [4]. However, the dynamic agreement process is introduced *after all data had been collected*, implying inapplicability to real time change detection problems. Namely, two detection phases are assumed: the sensing phase, where each sensor collects observations over a period of time, and the communication phase, where sensors subsequently run the consensus algorithm to fuse their local statistics.

The third class of distributed detection algorithms assumes that both the sensing and the communication phase occur in parallel, at the same time step. This class is mostly linked to the concept of “running consensus”,

<sup>☆</sup> The material in this paper is partially presented in Proceedings of the 19th Mediterranean Conference on Control and Automation, pp. 1170–1175.

\* Corresponding author. Tel.: +381 11 337 0150;

fax: +381 11 324 8681.

E-mail addresses: nemiliexp@yahoo.com, nemili@etf.rs (N. Ilić), stankovic@etf.rs (S.S. Stanković), milsta@kth.se (M.S. Stanković), kallej@kth.se (K.H. Johansson).

which has been introduced in the algorithms proposed and discussed in [5,6], assuming a consensus scheme with symmetric consensus matrices. An analysis of such algorithms based on the large deviations theory has been presented in [3]. An algorithm that combines minimum-variance distributed estimation (based on the so-called diffusion) with Neyman–Pearson detection has been proposed in [7]. In [8], a running consensus algorithm has been proposed for solving the quickest detection problem, based on the CUSUM (cumulative sum) statistic [9]. It represents a powerful practical tool for real time change detection, but it contains a nonlinearity used in the resetting rule of the algorithm, implying difficulties in the theoretical analysis of the algorithm. In [10], a novel class of distributed consensus-based real time change detection algorithms has been proposed, based on a combination of recursive geometric moving average control charts [9] with a consensus algorithm. Along with its inherent tracking capability, it introduces a more general setting of asymmetric consensus matrices. However, it assumes, as all of the aforementioned algorithms lying in the third class, that the parameter value after change is known.

In this paper, as a continuation of the work in [10], two new algorithms are proposed for distributed detection of *unknown changes* in (a) the mean and (b) the variance of a piecewise stationary random process, while monitoring the environment using a sensor network. Both algorithms have recursive forms derived from the expressions for the Generalized Likelihood Ratio (GLR) statistics for hypothesis testing, where the hypothesis  $H_0$  corresponds to the constant known parameter value before change, and the hypothesis  $H_1$  to the unknown parameter value after change. In [11] a window-truncated version of the GLR statistic for sequential multiple hypothesis testing which does not allow recursive structure has been proposed. Herein a constant forgetting factor in the derived recursions is introduced, resulting in algorithms belonging to the class of moving average control charts, applicable to the on-line change detection problem [9] (abrupt changes from  $H_0$  to  $H_1$ ). The obtained recursive form is structurally similar to the one discussed in [10], but with a much more complex innovation term. It is to be emphasized that the GLR is taken here as a starting point in the derivation of the algorithm in order to circumvent the restrictions inherent to the approach in [10], and to allow tracking of unknown parameter jumps. Furthermore, following [10], a dynamic consensus scheme is introduced, and algorithms which asymptotically provide nearly equal behavior of all the nodes are obtained, *i.e.*, any node can be selected for testing the decision variable w.r.t. a pre-specified threshold.

The derived algorithm for *change detection in the mean* is analyzed theoretically for both constant and randomly time varying asymmetric consensus matrices characterizing the network. The analysis is focused on the error between the generated distributed decision variables and the corresponding centralized statistics. The aforementioned complexity of the innovation term makes the analysis more complicated than the one from [10]. Moreover, it has been found to be necessary to introduce novel performance criteria. It is shown that under hypothesis  $H_1$

the ratio of the norm of the mean square error matrix and the mean square value of the centralized decision variable is bounded in the case of constant consensus matrices by  $K_1^1(1-\alpha)^2$ , where  $0 < \alpha < 1$  is the forgetting factor of the algorithm, while in the case of random consensus matrices it is bounded by  $K_2^1(1-\alpha)$ , where  $K_1^1$  and  $K_2^1$  are finite constants. Under hypothesis  $H_0$ , it is shown that the aforementioned ratio is bounded in the case of constant consensus matrices by  $K_1^0(1-\alpha)$ , while in the case of random consensus matrices it is bounded by  $K_2^0$ , where  $K_1^0$  and  $K_2^0$  are finite constants. In the case of time varying forgetting factors (behaving like  $t/(t+1)$ ), corresponding to the initial hypothesis testing problem, the corresponding bounds are also found, following the analogy between  $t^{-1}$  and the term  $1-\alpha$  from the constant forgetting factor case. A number of simulation results are given as an illustration of the characteristic properties of the proposed algorithm, including detection performance in terms of detection delay and false alarm rate.

The algorithm for *change detection in the variance* is designed similarly as the change in the mean algorithm, starting from the derivation of a recursive form of the GLR. Since the obtained innovation term in the recursions is very difficult to analyze, properties of the change in the variance algorithm are analyzed by means of simulation, showing that, qualitatively, all the results of the analysis connected to the change in the mean case hold also for the detection of the change in the variance.

The outline of the paper is as follows. Section 2 begins with local recursive algorithm derived from the GLR connected to the change in the mean case (Section 2.1). A novel distributed change detection scheme based on a consensus algorithm is given (Section 2.2), as well as an analysis of the error between the statistics generated by the proposed algorithm and the corresponding centralized scheme (for both constant and time varying forgetting factors—Sections 2.3 and 2.4, respectively). A change in the variance detection algorithm is proposed in Section 3 while Section 4 deals with some illustrative simulation examples.

## 2. Recursive distributed detection of change in the mean

### 2.1. Local recursions

Assume that we have a sensor network containing  $n$  nodes, in which the measurement signal of the  $i$ -th node is given by

$$y_i(t) = \theta_i + \epsilon_i(t), \quad (1)$$

where  $\epsilon_i(t) \sim N(0, \sigma_i^2)$ ,  $i = 1, \dots, n$ , are mutually independent iid processes. At first, consider a binary hypothesis problem, where the goal of the  $i$ -th node is to discriminate between the hypothesis  $H_0^i$  that  $\theta_i = \theta_i^0 = 0$  and the hypothesis  $H_1^i$  that  $\theta_i = \theta_i^1 \neq 0$ . In the case when  $\theta_i^1$ ,  $i = 1, \dots, n$ , is not a priori known, it is possible to apply the GLR methodology for hypothesis testing and to obtain the following local statistics based on  $N$  successive measurements [9,12]

$$s_i^1(N) = \max_{\theta_i^1} \sum_{t=1}^N \log \frac{p_{\theta_i^1}(y_i(t))}{p_{\theta_i^0}(y_i(t))} = \frac{N}{2} \bar{y}_i(N)^2 \sigma_i^{-2}, \quad (2)$$

where  $\bar{y}_i(N) = (1/N) \sum_{t=1}^N y_i(t)$ .

Calculation of  $s_i^l(N)$  can be performed on-line, recursively. Introducing  $t$  for current time, we obtain, using [12], the following basic recursion for the local decision function

$$s_i^l(t+1) = \frac{t}{t+1} s_i^l(t) + \frac{\sigma_i^{-2}}{t+1} \left[ (t+1) \bar{y}_i(t+1) - \frac{1}{2} y_i(t+1) \right] y_i(t+1), \quad (3)$$

where  $\bar{y}_i$  is also generated recursively by

$$\bar{y}_i(t+1) = \frac{t}{t+1} \bar{y}_i(t) + \frac{1}{t+1} y_i(t+1), \quad \bar{y}_i(0) = 0. \quad (4)$$

## 2.2. Centralized and consensus based recursive algorithm

The global centralized decision function for the whole sensor network, which should make distinction between the hypothesis  $H_0 : \theta_i = \theta_i^0 = 0, i = 1, \dots, n$ , and the hypothesis  $H_1 : \theta_i = \theta_i^1 \neq 0, i = 1, \dots, n$ , is defined as a sum of the local statistics given in (2).<sup>1</sup> After neglecting the second term in the brackets at the right hand side of (3), we obtain the following recursion for the centralized decision function:

$$s_c(t+1) = \frac{t}{t+1} s_c(t) + \sum_{i=1}^n \sigma_i^{-2} \bar{y}_i(t+1) y_i(t+1), \quad s_c(0) = 0. \quad (5)$$

The statistics given in (3) and (5) can distinguish between the two hypotheses, but cannot track parameter changes. Therefore, we introduce an approximation which replaces  $t/(t+1)$  by a constant  $\alpha$  close to one (which acts as a forgetting factor), in order to address the change detection problem. Namely, our goal is to detect a change from the hypothesis  $H_0$  to the hypothesis  $H_1$ , which occurs simultaneously at all sensors at unknown time  $t_0$  (it is also possible to assume that the change occurs for a non-empty subset of the network nodes [10]). Denoting

$$x_i(t) = \bar{y}_i(t) y_i(t), \quad (6)$$

where

$$\bar{y}_i(t+1) = \alpha \bar{y}_i(t) + (1-\alpha) y_i(t+1), \quad \bar{y}_i(0) = 0, \quad (7)$$

the centralized decision function now becomes

$$s_c(t+1) = \alpha s_c(t) + \sum_{i=1}^n w_i x_i(t+1), \quad s_c(0) = 0, \quad (8)$$

where  $w_i$  are nonnegative weights, equal to  $\sigma_i^{-2}$  in (5). Note that the obtained centralized decision function (8) is essentially one variant of the geometric moving average algorithm [9] with non-normalized weights, in which the application of the GLR results into a specific form of the function  $x_i$ , allowing tracking of unknown parameter jumps. For the sake of convenience, we shall further adopt that the weights are normalized in such a way that  $\sum_{i=1}^n w_i = 1$ ; accordingly, in (8) we introduce  $w_i = \sigma_i^{-2} (\sum_{i=1}^n \sigma_i^{-2})^{-1}$ . The global detection procedure is based on testing the decision function  $s_c(t)$  with respect to an appropriately chosen threshold  $\lambda_c > 0$ , so that a change is detected when  $s_c(t)$

exceeds  $\lambda_c$ . Notice that the algorithm requires a *fusion center*. It is to be noticed that it is also possible to adopt  $x_i(t) = \sigma_i^{-2} \bar{y}_i(t) y_i(t)$ , resulting in equal weights  $w_i = n^{-1}$ ; this represents a special case of the above setting.

The aim of this paper is to propose a distributed change detection algorithm which *does not require a fusion center* and in which the output of any preselected node can be used as a representative of the whole network and tested w.r.t. a pre-specified *common* threshold. The basic assumption is that the nodes of the network are connected in accordance with a time varying directed graph represented by a weighted adjacency matrix  $C(t) = [c_{ij}(t)]_{n \times n}$ , satisfying  $c_{ij}(t) \geq 0, i \neq j$  and  $c_{ii}(t) > 0, i, j = 1, \dots, n$  ( $c_{ij}(t)$  represents the communication gain from the node  $j$  to the node  $i$ ). We shall assume, additionally, that matrices  $C(t)$  are row-stochastic, random, iid and statistically independent from the sequences  $\{x_i(t)\}, i = 1, \dots, n$ .

We propose the following algorithm for generating the vector decision function  $s(t) = [s_1(t) \dots s_n(t)]^T$  for the whole network:

$$s(t+1) = \alpha C(t) s(t) + C(t) x(t+1), \quad s(0) = 0, \quad (9)$$

where  $x(t) = [x_1(t) \dots x_n(t)]^T$ . The algorithm is derived from the consensus based state and parameter estimation algorithms proposed in [13,14]; it is also similar to the detection algorithm based on “running consensus” proposed in [5,6,8]. Notice that the matrix  $C(t)$  performs for each node “convexification” of the neighboring states and enforces in such a way consensus between the nodes. After achieving  $s_i(t) \approx s_j(t), i, j = 1, \dots, n$ , change detection can be done by testing  $s_i(t)$  for any  $i$  with respect to the same  $\lambda_c$  as in the case of (8), provided (9) achieves a good approximation of  $s_c(t)$  generated by (8).

In order to implement the proposed algorithm it is necessary to set the communication gains in  $C(t)$  in accordance with the communication structure constraints resulting from the availability of communication links. We shall assume, in general, that  $C(t)$  is realized at each discrete time instant  $t$  as  $C^{(k)}$  with probability  $p_k, k = 1, \dots, N, N < \infty, \sum_{k=1}^N p_k = 1$  (the case of constant gains simply follows as a special case). The realization matrices  $C^{(k)} = [c_{ij}^{(k)}]_{n \times n}, k = 1, \dots, N, i, j = 1, \dots, n$ , will be assumed to be constant nonnegative row stochastic matrices, satisfying  $c_{ii}^{(k)} > 0, i = 1, \dots, n$ , so that we have

$$\bar{C} = E\{C(t)\} = \sum_{k=1}^N C^{(k)} p_k. \quad (10)$$

This formal setting obviously encompasses the asynchronous asymmetric gossip algorithm with one message at a time, various types of synchronous asymmetric gossip algorithms, as well as communication faults. We shall not be concerned here with concrete ways of generating the realizations of  $C^{(k)}$ : our further analysis is applicable to any preselected technical setting satisfying the adopted network model.

We shall assume further that

- (A1)  $\bar{C}$  has the eigenvalue 1 with algebraic multiplicity 1;
- (A2)  $\lim_{i \rightarrow \infty} \bar{C}^i = \mathbf{1} \mathbf{w}^T$ .

<sup>1</sup> It can be easily shown that the corresponding vector-valued GLR is in a form of a sum of the local GLRs connected to the individual nodes.

The first assumption is related to the *a priori* given topology of the underlying multi-agent network, implying that the graph associated with  $\bar{C}$  has a spanning tree and that  $\bar{C}^i$  converges to a nonnegative row stochastic matrix with equal rows when  $i$  tends to infinity, e.g., [15,16]. Assumption (A2) establishes a formal connection between the algorithm (9) and the centralized (8), implying that the realization matrices  $C^{(k)}$ , the corresponding probabilities  $p_k$  and the weight vector  $w$  are connected by the relation

$$w^T \bar{C} = w^T \sum_{k=1}^N C^{(k)} p_k = w^T. \quad (11)$$

For an *a priori* given vector  $w$ , according to the requirements resulting from the selected centralized detector (8), Eq. (11) should be solved for  $C^{(k)}$  and  $p_k$ . It is a nonlinear equation, which can be solved in practice by adopting one set of parameters (probabilities  $p_k$ , for example) and solving the linear programming problem for the remaining set of parameters (parameters in  $C^{(k)}$ ), or *vice versa* [17]. Notice that in the case of the asynchronous randomized gossip algorithm with one communication at a time,  $C^{(k)}$  is characterized by only one scalar parameter; in general,  $C^{(k)}$  is characterized by more parameters satisfying the given constraints. It is to be emphasized that solving (11) in the special case when all  $w_i = n^{-1}$  results in symmetric average consensus matrices  $\bar{C}$  when the communication links allow such a structure; otherwise, we have an asymmetric  $\bar{C}$ , satisfying (11). The related literature covers only the symmetric case [5,6,8,18]; the asymmetric case has been treated in [10,17].

### 2.3. Analysis of the consensus based algorithm

The theoretical analysis given in this section will be concerned with the relationship between the proposed consensus based algorithm (9) and the centralized (8) taken as a reference. Our goal is to show that the proposed algorithm generates statistics that are (sufficiently) close to the centralized statistics. Theoretical analysis of the performance of the proposed algorithm in terms of standard detection performance measures—detection and false alarm rate and detection delay assumes the knowledge about the distributions of the generated statistics. It is very difficult and beyond the scope of this paper to obtain these distributions, having in mind that we are dealing with a combination of consensus dynamics with the dynamics of a variant of geometric moving average algorithm. However, the aforementioned performance measures will be discussed in detail via simulations in Section 4.

The error vector between the states of the consensus based algorithm and the centralized scheme is defined as  $e(t) = s(t) - \mathbf{1}s_c(t)$ ,

$$(12)$$

where  $\mathbf{1} = [1 \dots 1]^T$ . Iterating (9) and (8) back to the zero initial conditions, we get

$$s(t) = \sum_{i=0}^{t-1} \alpha^i \varphi(t-1, t-i) x(t-i), \quad (13)$$

where  $\varphi(i, j) = C(i) \dots C(j)$ ,  $i \geq j$ , and

$$s_c(t) = \sum_{i=0}^{t-1} \alpha^i w^T x(t-i), \quad (14)$$

wherefrom

$$e(t) = \sum_{i=0}^{t-1} \alpha^i [\varphi(t-1, t-i) - \mathbf{1}w^T] x(t-i). \quad (15)$$

From (15) we obtain directly

$$E\{e(t)\} = \sum_{i=0}^{t-1} \alpha^i (\bar{C} - \mathbf{1}w^T)^{i+1} m = \sum_{i=0}^{t-1} \alpha^i \tilde{C}^{i+1} m, \quad (16)$$

where  $m = E\{x(t)\}$  and  $\tilde{C} = \bar{C} - \mathbf{1}w^T$ , having in mind that, under (A2), we have  $(\bar{C} - \mathbf{1}w^T)^i = \bar{C}^i - \mathbf{1}w^T$ . Obviously,  $s(t)$  is a biased estimator of  $\mathbf{1}s_c(t)$  when  $m \neq \mu \mathbf{1}$ , where  $\mu$  is a given scalar, having in mind that  $\tilde{C}m = 0$  for  $m = \mu \mathbf{1}$ .

Calculating  $m = [E\{x_1(t)\} \dots E\{x_n(t)\}]^T$  we obtain from (6), (7) and (1)

$$E\{x_i(t)\} = (1-\alpha) \sum_{j=0}^{t-1} \alpha^j y(t-i)y_j(t) \approx \theta_i^2 + (1-\alpha)\sigma_i^2, \quad (17)$$

where we used the approximation (which will be used throughout the remainder of this paper) that for  $t$  sufficiently large we have  $1 - \alpha^t \approx 1$ .

By Assumptions (A1) and (A2), it follows that  $\bar{C}$  and  $\mathbf{1}w^T$  have the same eigenvectors. Therefore,  $\bar{C}$  has the same eigenvalues as  $\tilde{C}$ , except for the eigenvalue 1 of  $\bar{C}$  which is replaced by the eigenvalue 0 of  $\tilde{C}$ . Having in mind that  $c_{ii} > 0$ ,  $i = 1, \dots, n$ , it follows that the modules of all the eigenvalues of  $\tilde{C}$  are strictly less than 1 [15]. We denote  $\max_i \{|\lambda_i(\tilde{C})|\} = \lambda_M < 1$ . Now we can see that

$$\|E\{e(t)\}\| \leq \sum_{i=0}^{t-1} \alpha^i \|\tilde{C}^{i+1}\| \|m\| \leq \frac{k\lambda_M \|m\|}{1-\alpha\lambda_M} < \frac{k\lambda_M \|m\|}{1-\lambda_M}, \quad (18)$$

having in mind that  $\|\tilde{C}^i\| \leq k\lambda_M^i$  for any matrix norm, where  $k$  is an appropriately chosen constant, and that  $\lambda_M < 1$ . A comparison with the properties of an analogous algorithm presented in [10] should be made, where the upper limit of  $\|E\{e(t)\}\|$  is proportional to  $1-\alpha$  under both hypotheses.

However, the obtained quality of approximating the centralized solution can be more adequately expressed by normalizing  $\|E\{e(t)\}\|$  by the mathematical expectation of the centralized decision variable itself. In this case we readily obtain that under both hypotheses

$$\frac{\|E\{e(t)\}\|}{E\{s_c(t)\}} \leq K(1-\alpha). \quad (19)$$

where  $K < \infty$ , having in mind that  $E\{s_c(t)\} \approx w^T m / (1-\alpha)$ . Under hypothesis  $H_1$ , the mean of the centralized statistics grows as  $1/(1-\alpha)$  when  $\alpha$  approaches 1, while the upper limit of the error mean remains constant; under hypothesis  $H_0$ , the mean of the centralized statistics remains constant and independent of  $\alpha$ , while the error mean decreases linearly as  $1-\alpha$  (having in mind that under  $H_0$  we have that  $m \sim \mathbf{1}$ ).

A more complete insight into the quality of approximation can be obtained from an analysis of the mean

square error matrix

$$Q(t) = E\{e(t)e(t)^T\}. \quad (20)$$

The following lemma serves as a prerequisite.

**Lemma 1.** *The covariance function  $r_i(\tau) = E\{(x_i(t) - m_i)(x_i(t + \tau) - m_i)\}$  for algorithm (5) satisfies*

$$\sum_{\tau=0}^{\infty} |r_i(\tau)| \leq K_1; \quad i = 1, \dots, n, \quad 0 < K_1 < \infty. \quad (21)$$

**Proof.** Starting from (6) we have

$$\begin{aligned} r_i(\tau) &= E\{(\bar{y}_i(t)y_i(t) - m_i)(\bar{y}_i(t+\tau)y_i(t+\tau) - m_i)\} \\ &= E\left\{ \left( (1-\alpha) \sum_{j=0}^{t-1} \alpha^j (\theta_i^2 + \theta_i(\epsilon_i(t) + \epsilon_i(t-j)) \right. \right. \\ &\quad \left. \left. + \epsilon_i(t)\epsilon_i(t-j) - (\theta_i^2 + (1-\alpha)\sigma_i^2) \right) \right. \\ &\quad \left( (1-\alpha) \sum_{k=0}^{t+\tau-1} \alpha^k (\theta_i^2 + \theta_i(\epsilon_i(t+\tau) + \epsilon_i(t+\tau-k)) \right. \\ &\quad \left. \left. + \epsilon_i(t+\tau)\epsilon_i(t+\tau-k) - (\theta_i^2 + (1-\alpha)\sigma_i^2) \right) \right\} \\ &= E\left\{ (1-\alpha)^2 \sum_{j=0}^{t-1} \alpha^j \theta_i(\epsilon_i(t) + \epsilon_i(t-j)) \right. \\ &\quad \left. \times \sum_{k=0}^{t+\tau-1} \alpha^k \theta_i(\epsilon_i(t+\tau) + \epsilon_i(t+\tau-k)) \right\} + \delta_{\tau,0} r_{\epsilon\epsilon}, \quad (22) \end{aligned}$$

where  $r_{\epsilon\epsilon}$  is a part of  $r_i(\tau)$  connected to the mathematical expectation of the product of the terms  $(1-\alpha)(\sum_{j=0}^{t-1} \alpha^j \epsilon_i(t)\epsilon_i(t-j) - \sigma_i^2)$  and  $(1-\alpha)(\sum_{k=0}^{t+\tau-1} \alpha^k \epsilon_i(t+\tau) + \epsilon_i(t+\tau-k) - \sigma_i^2)$  which is non-zero for  $\tau = 0$  and  $k = j$ ,

$$\begin{aligned} r_{\epsilon\epsilon} &= (1-\alpha)^2 \left( E\left\{ \epsilon_i^4(t) + \sum_{j=1}^{t-1} \alpha^{2j} \epsilon_i^2(t)\epsilon_i^2(t-j) \right\} - \sigma_i^4 \right) \\ &\approx (1-\alpha)^2 \left( 2\sigma_i^4 + \frac{\alpha^2}{1-\alpha^2} \sigma_i^4 \right) \\ &= (1-\alpha)\sigma_i^4 \frac{2-\alpha^2}{1+\alpha}. \quad (23) \end{aligned}$$

Since  $r_i(\tau) = r_i(-\tau)$ , we can see that for  $\tau > 0$  we have non-zero terms in the remaining terms of (22) only in the cases when  $k = \tau$  and  $k = \tau + j$ ; for  $\tau = 0$  we have non-zero terms not only in the cases when  $k = 0$  and  $k = j$  but also in the case when  $j = 0$ , together with the term connected to  $\theta_i^2 \epsilon_i^2(t)$  which is non-zero for all  $j$  and  $k$ . Therefore, we obtain the following expression for  $r_i(\tau)$  (for  $\tau \geq 0$ ):

$$\begin{aligned} r_i(\tau) &= (1-\alpha)^2 E\left\{ \sum_{j=0}^{t-1} \alpha^j \theta_i^2 (\alpha^\tau \epsilon_i^2(t) + \alpha^{\tau+j} \epsilon_i^2(t-j)) \right\} + \delta_{\tau,0} (r_{\epsilon\epsilon} + r_\epsilon) \\ &\approx (1-\alpha)^2 \theta_i^2 \sigma_i^2 \left( \frac{1}{1-\alpha} + \frac{1}{1-\alpha^2} \right) \alpha^\tau + \delta_{\tau,0} (r_{\epsilon\epsilon} + r_\epsilon) \\ &= (1-\alpha)\theta_i^2 \sigma_i^2 \frac{2+\alpha}{1+\alpha} \alpha^\tau + \delta_{\tau,0} (r_{\epsilon\epsilon} + r_\epsilon), \quad (24) \end{aligned}$$

where

$$r_\epsilon = (1-\alpha)^2 E\left\{ \sum_{k=0}^{t-1} \alpha^k (\theta_i^2 \epsilon_i^2(t) + \sum_{j=0}^{t-1} \alpha^j \theta_i^2 \epsilon_i^2(t)) \right\}$$

$$\approx (1-\alpha)\theta_i^2 \sigma_i^2 + \theta_i^2 \sigma_i^2. \quad (25)$$

Having in mind that  $0 < \alpha < 1$  we have that

$$r_i(\tau) < (1-\alpha)\theta_i^2 \sigma_i^2 \kappa_1 \alpha^\tau + \delta_{\tau,0} ((1-\alpha)\sigma_i^4 \kappa_2 + (1-\alpha)\theta_i^2 \sigma_i^2 + \theta_i^2 \sigma_i^2), \quad (26)$$

where  $\kappa_1$  and  $\kappa_2$  are constants that do not depend on  $\alpha$  (e.g.,  $\kappa_1 = \kappa_2 = 2$ ). Therefore, (21) is satisfied under both hypotheses. More precisely, we have under hypothesis  $H_1$  that

$$\sum_{\tau=0}^{\infty} |r_i(\tau)| < \theta_i^2 \sigma_i^2 (\kappa_1 + 1) + (1-\alpha)(\sigma_i^4 \kappa_2 + \sigma_i^2 \theta_i^2) < K_1 < \infty, \quad (27)$$

where  $K_1$  is a constant that does not depend on  $\alpha$  (e.g.,  $K_1 = \theta_i^2 \sigma_i^2 (\kappa_1 + 1) + (\sigma_i^4 \kappa_2 + \sigma_i^2 \theta_i^2)$ ) while under hypothesis  $H_0$  we have only one non-zero term:

$$\sum_{\tau=0}^{\infty} |r_i(\tau)| < (1-\alpha)\sigma_i^4 \kappa_2 \leq K_0(1-\alpha) < \infty, \quad (28)$$

where  $K_0$  is a constant that does not depend on  $\alpha$ .  $\square$

**Theorem 1.** *Let Assumptions (A1) and (A2) hold, and let*

$$J(t) = \frac{\|Q(t)\|_\infty}{E\{s_c(t)^2\}}.$$

*Then, under hypothesis  $H_1$ , in the case of constant consensus matrices,*

$$J(t) \leq K_1^1 (1-\alpha)^2,$$

*while in the case of random consensus matrices*

$$J(t) \leq K_2^1 (1-\alpha);$$

*under hypothesis  $H_0$ , in the case of constant consensus matrices,*

$$J(t) \leq K_1^0 (1-\alpha),$$

*while in the case of random consensus matrices*

$$J(t) \leq K_2^0,$$

*where  $K_1^1, K_2^1, K_1^0, K_2^0 < \infty$  are constants that do not depend on  $\alpha$  and  $\|A\|_\infty = \max_i \sum_j |a_{ij}|$ , where  $A = [a_{ij}]$  is a given matrix.*

**Proof.** First, we shall obtain a lower bound for the variance of the centralized statistics:

$$\begin{aligned} \text{var}\{s_c(t)\} &= E\left\{ \left( \sum_{j=0}^{t-1} \alpha^j w^T (x(t-j) - m) \right)^2 \right\} \\ &= \sum_{j=0}^{t-1} \alpha^j \sum_{k=0}^{t-1} \alpha^k w^T \tilde{R}_{jk} w, \quad (29) \end{aligned}$$

where

$$\tilde{R}_{jk} = \text{diag}\{r_1(j-k), \dots, r_n(j-k)\}. \quad (30)$$

From (23)–(25) we can also obtain lower bounds for  $r_i(\tau)$ , namely

$$r_i(\tau) > (1-\alpha)\kappa_3 \alpha^{|\tau|} + \delta_{\tau,0} ((1-\alpha)\kappa_4 + \kappa_5), \quad (31)$$

where  $\kappa_3$ ,  $\kappa_4$  and  $\kappa_5$  are constants that do not depend on  $\alpha$  (e.g.,  $\kappa_3 = \frac{3}{2} \min_i \theta_i^2 \sigma_i^2$ ,  $\kappa_4 = \min_i (\frac{1}{2} \sigma_i^4 + \theta_i^2 \sigma_i^2)$  and  $\kappa_5 = \min_i \theta_i^2 \sigma_i^2$ ). Therefore, under hypothesis  $H_1$



$$\begin{aligned} \text{var}\{s_c(t)\} &> \sum_{j=0}^{t-1} \alpha^j \sum_{k=0}^{t-1} \alpha^k (1-\alpha) \alpha^{|j-k|} \sum_{i=1}^n w_i^2 \kappa_3 \\ &+ \sum_{j=0}^{t-1} \alpha^{2j} \left( (1-\alpha) \sum_{i=1}^n w_i^2 \kappa_4 + \sum_{i=1}^n w_i^2 \kappa_5 \right). \end{aligned} \quad (32)$$

Analyzing the first sum in (32) we have

$$\begin{aligned} \sum_{j=0}^{t-1} \alpha^j \sum_{k=0}^{t-1} \alpha^k \alpha^{|j-k|} &= \sum_{j=0}^{t-1} \alpha^j \left( \sum_{k=0}^{j-1} \alpha^k \alpha^{j-k} + \sum_{k=j}^{t-1} \alpha^k \alpha^{k-j} \right) \\ &\approx \sum_{j=0}^{t-1} \left( j \alpha^{2j} + \frac{\alpha^{2j}}{1-\alpha^2} \right) \approx \frac{2}{(1-\alpha^2)^2}. \end{aligned} \quad (33)$$

Therefore, we finally obtain that under hypothesis  $H_1$

$$\begin{aligned} \text{var}\{s_c(t)\} &> \frac{2(1-\alpha)}{(1-\alpha^2)^2} \sum_{i=1}^n w_i^2 \kappa_3 \\ &+ \frac{1}{1-\alpha^2} \left( (1-\alpha) \sum_{i=1}^n w_i^2 \kappa_4 + \sum_{i=1}^n w_i^2 \kappa_5 \right) > \kappa_6 (1-\alpha)^{-1}, \end{aligned} \quad (34)$$

where  $\kappa_6$  is a constant that does not depend on  $\alpha$  (e.g.,  $\kappa_6 = \frac{1}{2} \sum_{i=1}^n w_i^2 \kappa_5$ ).

Calculation of the lower bound for the variance of the centralized statistics is simpler under hypothesis  $H_0$  (using the fact that  $r_i(\tau) > \delta_{\tau,0} (1-\alpha) \kappa_7$ , where  $\kappa_7 \neq \kappa_7(\alpha)$ , e.g.,  $\kappa_7 = \frac{1}{2} \min_i \sigma_i^4$ ):

$$\text{var}\{s_c(t)\} > \sum_{j=0}^{t-1} \alpha^{2j} (1-\alpha) \sum_{i=1}^n w_i^2 \kappa_7 > \kappa_8, \quad (35)$$

where  $\kappa_8 \neq \kappa_8(\alpha)$  (e.g.,  $\kappa_8 = \frac{1}{2} \sum_{i=1}^n w_i^2 \kappa_7$ ).

Having in mind that  $E\{s_c(t)\} \approx w^T(m/(1-\alpha))$  we obtain that under hypothesis  $H_1$

$$E\{s_c(t)^2\} = E\{s_c(t)\}^2 + \text{var}\{s_c(t)\} \geq m_1 (1-\alpha)^{-2}, \quad (36)$$

while under hypothesis  $H_0$

$$E\{s_c(t)^2\} \geq m_0, \quad (37)$$

where  $m_1, m_0 < \infty$  do not depend on  $\alpha$ .

It is to be noticed that it is possible to find, in a similar way as above, that the upper bounds for the variance of the centralized statistics have the same form as the lower bounds (34) and (35), but with different constants. Therefore, under  $H_1$  the variance of the centralized statistics grows as  $\alpha$  is getting closer to 1 ( $\kappa_{H_1}^l < (1-\alpha) \text{var}\{s_c(t)\} < \kappa_{H_1}^u$ ), while under  $H_0$  it remains within a constant interval ( $\kappa_{H_0}^l < \text{var}\{s_c(t)\} < \kappa_{H_0}^u$ ).

Further, consider an arbitrary deterministic  $n$ -vector  $y$  and analyze the quadratic form  $y^T Q(t) y$  under hypothesis  $H_1$ .

In the case of constant consensus matrices we have that  $Q(t) = Q_1(t) + Q_2(t)$ , in which

$$Q_1(t) = \Phi(t)^T \tilde{R}(t) \Phi(t) \quad (38)$$

and

$$Q_2(t) = \Phi(t)^T m_X(t) m_X(t)^T \Phi(t), \quad (39)$$

where  $\Phi(t) = [\alpha^{t-1} \tilde{C}^t; \alpha^{t-2} \tilde{C}^{t-1}; \dots; \alpha^0 \tilde{C}]^T$ ,  $\tilde{R}(t) = R(t) - m_X(t) m_X(t)^T$ ,  $R(t) = E\{X(t) X(t)^T\}$ ,  $X(t) = [x(1)^T \dots x(t)^T]^T$  and  $m_X(t) = E\{X(t)\}$ .

Analyzing first  $y^T Q_1(t) y$ , we conclude that  $\tilde{R}(t) = [\tilde{R}_{ij}]$ ,  $i, j = 1, \dots, t$ , where  $\tilde{R}_{ij}$  are constant  $n \times n$  block matrices defined as (30) and that

$$\lambda_{\max}(\tilde{R}(t)) \leq \|\tilde{R}(t)\|_{\infty} \leq K_1 < \infty \quad (40)$$

because of the absolute summability of the covariance functions.

Coming back to (38), we realize further that the expression  $y^T \Phi(t)^T \Phi(t) y$  is in the form of a sum of terms containing  $y^T \tilde{C}^i \tilde{C}^{iT} y$ ,  $i = 1, \dots, t$ . Having in mind that the modules of all the eigenvalues of  $\tilde{C}$  are strictly less than 1, we have now that  $\|y^T \tilde{C}^i \tilde{C}^{iT} y\| \leq k \lambda_M^{2i} \|y\|^2$ , where  $k < \infty$ ,  $i = 1, \dots, t$  and  $\lambda_M = \max_i \{|\lambda_i(\tilde{C})|\} < 1$ .

Therefore, we have

$$y^T Q_1(t) y \leq k' K_1 \sum_{i=0}^{t-1} \alpha^{2i} \lambda_M^{2(i+1)} \|y\|^2 \leq k' K_1 \frac{\lambda_M^2}{1-\lambda_M^2} \|y\|^2 \leq k_1^1 \|y\|^2, \quad (41)$$

where  $k_1^1 < \infty$  does not depend on  $\alpha$ , while analyzing  $Q_2(t)$  we find that

$$y^T Q_2(t) y \leq \left( \sum_{i=0}^{t-1} \alpha^i \|\tilde{C}^{i+1}\| \|m\| \right)^2 \|y\|^2 \leq k'' \left( \frac{\lambda_M}{1-\lambda_M} \right)^2 \|y\|^2 \leq k_2^1 \|y\|^2, \quad (42)$$

where  $k_2^1 < \infty$  does not depend on  $\alpha$ .

In the case of random consensus matrices the mean square error matrix is decomposed as  $Q(t) = Q_3(t) + Q_4(t)$ , where

$$Q_3(t) = E\{E_x\{e(t)e(t)^T\} - E_x\{e(t)\}E_x\{e(t)\}^T\} \quad (43)$$

and

$$Q_4(t) = E\{E_x\{e(t)\}E_x\{e(t)\}^T\}, \quad (44)$$

$E_x\{\cdot\}$  denoting the conditional expectation given the  $\sigma$ -algebra generated by  $\{C(t)\}$ .

We obtain, in analogy with (38) and (39), that

$$Q_3(t) = E\{\tilde{\Phi}(t)^T \tilde{R}(t) \tilde{\Phi}(t)\}, \quad (45)$$

where  $\tilde{\Phi}(t) = [\alpha^{t-1}(\varphi(t-1, 0) - \mathbf{1}w^T); \alpha^{t-2}(\varphi(t-1, 1) - \mathbf{1}w^T); \dots; \alpha^0(\varphi(t-1, t-1) - \mathbf{1}w^T)]^T$  and

$$Q_4(t) = E\{\tilde{\Phi}(t)^T m_X(t) m_X(t)^T \tilde{\Phi}(t)\}. \quad (46)$$

Analyzing the term connected to  $Q_3(t)$  we use (40) directly as a consequence of independence between  $\{x(t)\}$  and  $\{C(t)\}$  and realize that we are concerned here with the expression

$$E\{\tilde{\Phi}(t)^T \tilde{\Phi}(t)\} = \sum_{j=0}^{t-1} D(t-1, j) \alpha^{2(t-j-1)}, \quad (47)$$

where  $D(t-1, j) = E\{(\varphi(t-1, j) - \mathbf{1}w^T)(\varphi(t-1, j) - \mathbf{1}w^T)^T\}$ . Based on the result from [10] that norm of the matrices  $D(t-1, j)$ ,  $j = 0, \dots, t-1$  has a finite upper bound that does not depend on  $\alpha$  we obtain that

$$y^T Q_3(t) y \leq m' K_1 \sum_{i=0}^{t-1} \alpha^{2i} \|y\|^2 \leq k_3^1 (1-\alpha)^{-1} \|y\|^2, \quad (48)$$

where  $k_3^1 < \infty$  does not depend on  $\alpha$ , while the term  $y^T Q_4(t)y$  can be analyzed analogously. We use the fact that  $E\{\tilde{\Phi}(t)^T m_X(t) m_X(t)^T \tilde{\Phi}(t)\} \leq 2\alpha^{2(t-1)} E\{(\varphi(t-1, 0) - \mathbf{1}w^T)mm^T(\varphi(t-1, 0) - \mathbf{1}w^T)^T\} + \dots + 2\alpha^{2 \cdot 0} E\{(\varphi(t-1, t-1) - \mathbf{1}w^T)mm^T(\varphi(t-1, t-1) - \mathbf{1}w^T)^T\}$  and obtain that

$$y^T Q_4(t)y \leq m'' \sum_{i=0}^{t-1} \alpha^{2i} \|m\|^2 \|y\|^2 \leq k_4^1 (1-\alpha)^{-1} \|y\|^2, \quad (49)$$

where  $k_4^1 < \infty$  does not depend on  $\alpha$ .

Consequently, by choosing  $y = e_i$ , where  $e_i$  denotes the  $n$ -vector of zeros with only the  $i$ -th entry equal to one, one obtains that in the case of constant consensus matrices  $Q_{ii}(t) \leq k_{12}^1$ , where  $k_{12}^1 < \infty$ ,  $i = 1, \dots, n$ . Furthermore,  $|Q_{ij}(t)| \leq \max_i Q_{ii}(t)$ , having in mind elementary properties of positive semidefinite matrices. In the case of random consensus matrices, we have that  $\max_{i,j} Q_{ij}(t) \leq k_{34}^1 (1/(1-\alpha))$ , where  $k_{34}^1 < \infty$ . Dividing the mean square error matrices by the mean square value of the centralized decision variable (36) we obtain the result.

Under hypothesis  $H_0$  we have that constant  $K_1$  from (40) depends on  $\alpha$ , namely,  $K_1 \sim 1-\alpha$ , so that the inequalities connected to the quadratic forms (41) and (48) should be multiplied by  $1-\alpha$ . Moreover, under  $H_0$ , the mean of  $x(t)$  shows a similar behavior,  $m \sim 1-\alpha$ , so that the inequalities connected to the quadratic forms (42) and (49) should be multiplied by  $(1-\alpha)^2$ . Therefore, we have in the case of constant consensus matrices

$$y^T Q(t)y \leq k_1^0 (1-\alpha) \|y\|^2 + k_2^0 (1-\alpha)^2 \|y\|^2 < k_{12}^0 (1-\alpha) \|y\|^2, \quad (50)$$

while in the case of random consensus matrices

$$y^T Q(t)y \leq k_3^0 \|y\|^2 + k_4^0 (1-\alpha) \|y\|^2 < k_{34}^0 \|y\|^2. \quad (51)$$

Thus, the result.  $\square$

#### 2.4. Time varying forgetting factor

The recursive algorithms (8) and (9) with constant forgetting factor  $\alpha$  represent essentially tracking algorithms, aimed at coping with abrupt parameter changes [9]. It is also interesting to analyze the case of time varying forgetting factor corresponding to the hypothesis testing problem to see the analogy between  $1-\alpha$  and  $t^{-1}$  (following the methodology from [10]).

**Theorem 2.** *Let in (8) and (9) the forgetting factor be in the form  $\alpha(t+1) = t/(t+1)$  and let Assumptions (A1) and (A2) hold. Then, under hypothesis  $H_1$ , in the case of constant consensus matrices*

$$J(t) = O(t^{-2}),$$

while in the case of random consensus matrices

$$J(t) = O(t^{-1});$$

under hypothesis  $H_0$ , in the case of constant consensus matrices

$$J(t) = O(t^{-1}),$$

while in the case of random consensus matrices

$$J(t) = O(1).$$

**Proof.** First we obtain an expression for the centralized statistics

$$s_c(t) = \sum_{i=0}^{t-1} \frac{t-i}{t} w^T x(t-i), \quad (52)$$

having in mind that  $(t-1)/t \cdot (t-2)/(t-1) \cdot \dots \cdot (t-i)/(t-i+1) = (t-i)/t$ . It is straightforward to show that  $E\{x(t)\} = O(1)$  under hypothesis  $H_1$  and that  $E\{x(t)\} = O(t^{-1})$  under hypothesis  $H_0$ . Similarly as in (36) and (37) it can be shown that in the case of constant consensus matrices  $E\{s_c(t)^2\} = O(t^2)$ , while in the case of random consensus matrices  $E\{s_c(t)^2\} = O(1)$  (notice the analogy between  $1-\alpha$  and  $1/t$ ).

We have now the following expression for the error:

$$e(t) = \sum_{i=0}^{t-1} \frac{t-i}{t} \tilde{C}^{i+1} x(t-i). \quad (53)$$

Applying the line of thought of Theorem 1 regarding hypothesis  $H_1$ , we can obtain for constant consensus matrices, similarly as in (38), the following expression:

$$y^T Q_1(t)y = y^T \Psi(t)^T \tilde{R}(t) \Psi(t)y, \quad (54)$$

where  $\Psi(t) = [\frac{1}{t} \tilde{C}^t; \frac{2}{t} \tilde{C}^{t-1}; \dots; \tilde{C}]$ . Proceeding like in the proof of Theorem 1, we obtain

$$y^T Q_1(t)y \leq k' K_1 \sum_{i=0}^{t-1} \left(1 - 2\frac{i}{t} + \frac{i^2}{t^2}\right) \lambda_M^{2(i+1)} \|y\|^2 = O(1) \|y\|^2, \quad (55)$$

where we used Kronecker's lemma (e.g., [19]) to obtain

$$\lim_{t \rightarrow \infty} \sum_{i=0}^{t-1} \left(2\frac{i}{t} + \frac{i^2}{t^2}\right) \lambda_M^{2(i+1)} = 0. \quad (56)$$

An analogous reasoning can be applied to the term  $Q_2(t)$  from (39) to show that  $y^T Q_2(t)y = O(1) \|y\|^2$ .

In the case of random consensus matrices, one obtains, proceeding like in Theorem 1,

$$y^T Q_3(t)y \leq m' K_1 \sum_{i=0}^{t-1} \left(1 - 2\frac{i}{t} + \frac{i^2}{t^2}\right) \|y\|^2 = O(t) \|y\|^2. \quad (57)$$

Analogously, one can show that  $y^T Q_4(t)y = O(t) \|y\|^2$ .

Under hypothesis  $H_0$  inequalities connected to the terms  $Q_1(t)$  and  $Q_3(t)$  should be multiplied by  $t^{-1}$ , because  $K_1 \sim t^{-1}$ ; the inequalities connected with the terms  $Q_2(t)$  and  $Q_4(t)$  should be multiplied by  $t^{-2}$  because  $m \sim t^{-1}$ , and therefore their influence can be neglected compared to the terms  $Q_1(t)$  and  $Q_3(t)$ . Similarly as in Theorem 1 we obtain the result.  $\square$

### 3. Distributed recursive detection of change in the variance

Assume, without loss of generality, that we have the following zero-mean system model:

$$y_i(t) = \epsilon_i(t), \quad (58)$$

where the hypothesis  $H_0^i$  is that  $\epsilon_i(t) \sim N(0, (\sigma_i^0)^2)$  and the hypothesis  $H_1^i$  that  $\epsilon_i(t) \sim N(0, (\sigma_i^1)^2)$ ;  $\{\epsilon_i(t)\}$  under each hypothesis are supposed to be mutually independent iid processes. In the case when  $(\sigma_i^1)^2$  is not a priori known, the application of the GLR methodology for hypothesis testing leads to the following statistics based on  $N$  successive measurements [9,12]:

$$s_i^l(N) = \max_{\sigma_i^1} \sum_{t=1}^N \log \frac{p_{\sigma_i^1}(y_i(t))}{p_{\sigma_i^0}(y_i(t))} \\ = N \log \frac{\sigma_i^0}{\bar{\sigma}_i(N)} + \frac{1}{2(\sigma_i^0)^2} \sum_{t=1}^N y_i(t)^2 - \frac{N}{2}, \quad (59)$$

where  $\bar{\sigma}_i(N)^2 = (1/N) \sum_{t=1}^N y_i(t)^2$ .

Introducing  $t$  for current time, we derive, similarly as in (3), the following basic local recursions for calculating  $s_i^l(t)$ :

$$s_i^l(t+1) = \frac{t}{t+1} s_i^l(t) + \left(1 - \frac{1}{2(t+1)}\right) \log \frac{(\sigma_i^0)^2}{\bar{\sigma}_i(t+1)^2} \\ + \frac{1}{2} \left( \frac{t}{t+1} \frac{1}{(\sigma_i^0)^2} - \left(\frac{t}{t+1}\right)^2 \frac{1}{\bar{\sigma}_i(t+1)^2} \right) y_i(t+1)^2 \\ + \frac{1}{2(\sigma_i^0)^2} (\bar{\sigma}_i(t+1)^2 - (\sigma_i^0)^2). \quad (60)$$

For  $t$  sufficiently large, we introduce the approximations  $1/(t+1) \ll 1$  and  $t/(t+1) \approx 1$  connected to innovation terms, and, after replacing  $t/(t+1)$  by  $\alpha$  close to 1, we finally obtain the following recursion for on-line change detection:

$$s_i^l(t+1) = \alpha s_i^l(t) + \log \frac{(\sigma_i^0)^2}{\bar{\sigma}_i(t+1)^2} + \frac{1}{2} \left( \frac{1}{(\sigma_i^0)^2} - \frac{1}{\bar{\sigma}_i(t+1)^2} \right) y_i(t+1)^2 \\ + \frac{1}{2(\sigma_i^0)^2} (\bar{\sigma}_i(t+1)^2 - (\sigma_i^0)^2), \quad (61)$$

where  $\bar{\sigma}_i(t+1)^2$  is generated recursively by

$$\bar{\sigma}_i(t+1)^2 = \alpha \bar{\sigma}_i(t)^2 + (1-\alpha) y_i(t+1)^2. \quad (62)$$

Adopting the general approach from [6,10] that the centralized statistics is defined as a sum of the local statistics (given in (61)) and denoting  $\log((\sigma_i^0)^2 / (\bar{\sigma}_i(t+1)^2)) + \frac{1}{2}((1/(\sigma_i^0)^2) - (1/\bar{\sigma}_i(t+1)^2)) y_i(t+1)^2 + (1/2(\sigma_i^0)^2) (\bar{\sigma}_i(t+1)^2 - (\sigma_i^0)^2)$  as  $x_i(t+1)$ , we come to the same form of the centralized (8) and distributed algorithm (9), as in the case of detecting change in the mean. Obviously, these algorithms should now use equal normalized weights  $w_i = 1/n$ ,  $i = 1, \dots, n$ . Complexity of the expression for  $x_i(t+1)$  (recursively generated  $\bar{\sigma}_i(t+1)^2$  in the denominator, correlated with  $y_i(t+1)^2$ , plus the logarithmic term) makes any theoretical analysis regarding statistical properties of  $x_i^*(t)$  very difficult. An analysis connected to the centralized and distributed statistics is even more difficult, so that the properties of the change in the variance detection algorithm will be analyzed in the next section by means of simulation.

One can simplify calculation in the recursions by replacing  $x_i(t)$  with  $x_i^*(t) = \log(\sigma_i^0 / \bar{\sigma}_i(t)) + \frac{1}{2}((1/(\sigma_i^0)^2) - (1/\bar{\sigma}_i(t)^2)) y_i(t)^2$ . It can be shown that the mathematical

expectation of the term  $x_i^*(t)$  (assuming that  $\alpha$  is sufficiently close to 1, so that  $\bar{\sigma}_i(t)^2$  has converged to  $\sigma_i^1$ ) has the same sign as  $x_i(t)$ , but with smaller ordinates.

## 4. Simulation results

### 4.1. Change in the mean

Let us consider a sensor network with  $n=10$  nodes, where the means  $\theta_i^1$  (unknown to the designer of the detection scheme) are randomly taken from the interval  $(0,1)$ , and the variances  $\sigma_i^2$  randomly taken from the interval  $[0.5,1.5]$ ; it is assumed that  $\theta_i^0 = 0$  in the case of no change,  $i=1, \dots, n$ . Communication gains are obtained by solving Eq. (11) for both constant and time varying consensus matrices under the constraints that the consensus matrices are row stochastic and possess a pre-defined structure (places of zeros). The assumed network topology corresponds to the modified Geometric Random Graph in which the nodes represent randomly spatially distributed agents (in this case within a square area), and they are connected if their distance is less than some predetermined threshold (in this case half of the side of the square, see, e.g., [18]), resulting in an initially undirected graph. The modification is that roughly 10% of the original two-way communications are made to be one-way. It is highly likely that one-way communications arise in practice when working with sensor networks. The weight vector components are chosen as  $w_i = \sigma_i^{-2} (\sum_{i=1}^n \sigma_i^{-2})^{-1}$  (see Section 2.2). In the case of random consensus matrices the asymmetric asynchronous ‘‘gossip’’ algorithm with one communication at a time is assumed. The values of the elements of the realizations of the consensus matrices corresponding to communicating nodes are taken to be 0.5, so that (11) is solved for the probabilities of individual realizations, see [17].

Fig. 1 shows, for comparison, one typical realization of the centralized decision function (8) for  $\alpha = 0.9$  and  $\alpha = 0.99$ , together with the corresponding realizations obtained at one randomly selected node in the network for constant and random consensus matrices (one component of (9)). The moment of change is chosen to be  $t=500$ . In addition, in Fig. 2 the mean  $\pm$  one standard deviation of the global decision function is represented by dashed lines, together with the decision function of one randomly selected node (solid line), using 1000 realizations. It can be seen that the means and the variances of both centralized and distributed statistics increase with  $\alpha$  getting closer to 1 under the hypothesis  $H_1$ , and that they remain within a constant interval under  $H_0$ .

Fig. 3 (left, solid line) illustrates the dependence of the error between the proposed algorithm and the corresponding centralized solution on the forgetting factor  $\alpha$  under the hypothesis  $H_1$  (see Theorem 1 from Section 2.3). For the above network with 10 nodes, the ratio of the mean square error for one randomly selected node and the mean square value of the centralized statistics at  $t=1000$  is calculated using 1000 Monte Carlo runs, as a function of  $(1-\alpha)^2$  in the case of constant consensus matrices and of  $(1-\alpha)$  in the case of random consensus matrices. Fig. 4 (left, solid line) illustrates the dependence



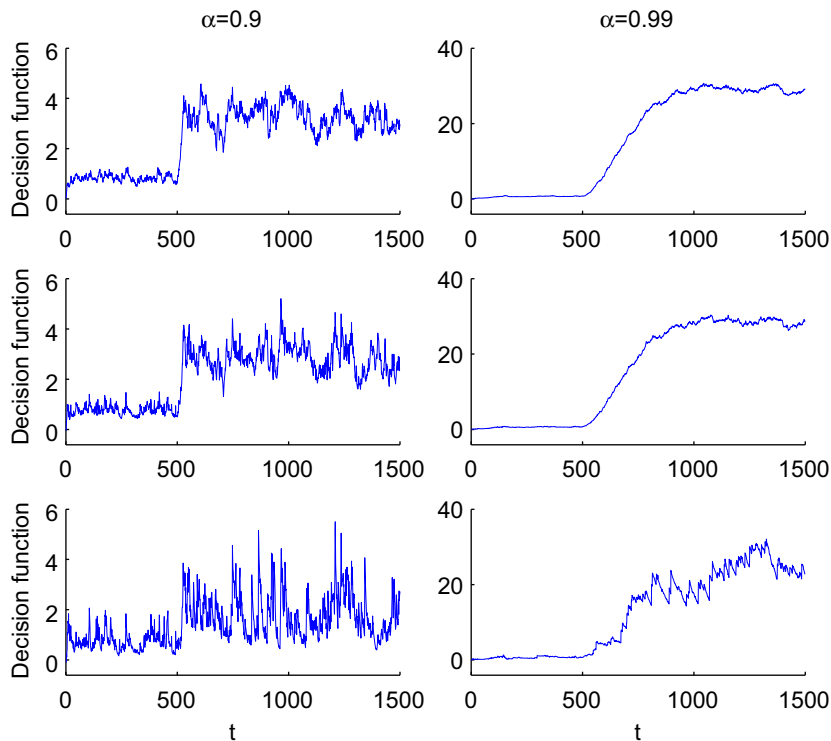


Fig. 1. Realizations of decision functions: centralized strategy (top), constant consensus matrices (middle), random consensus matrices (bottom).

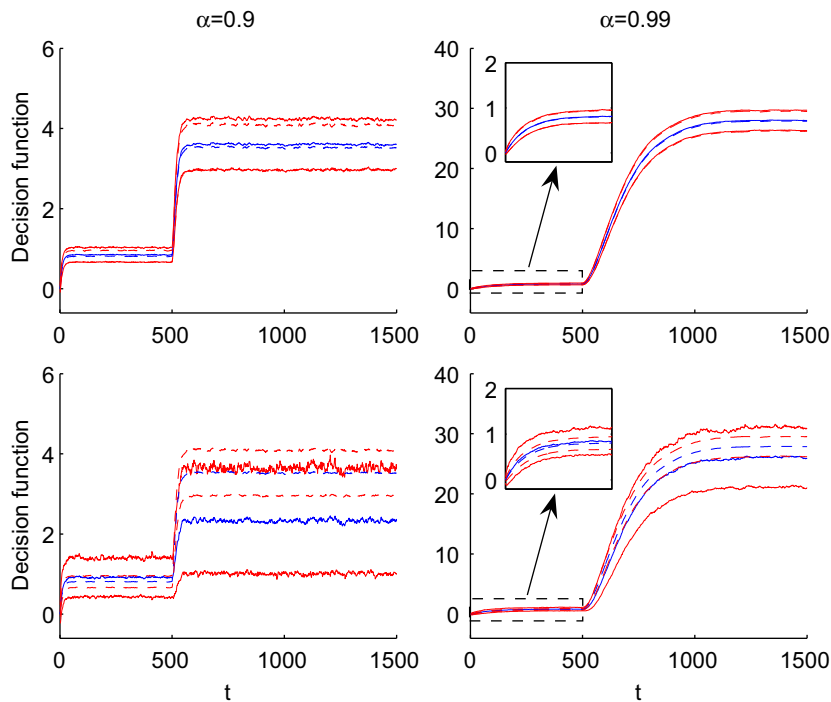
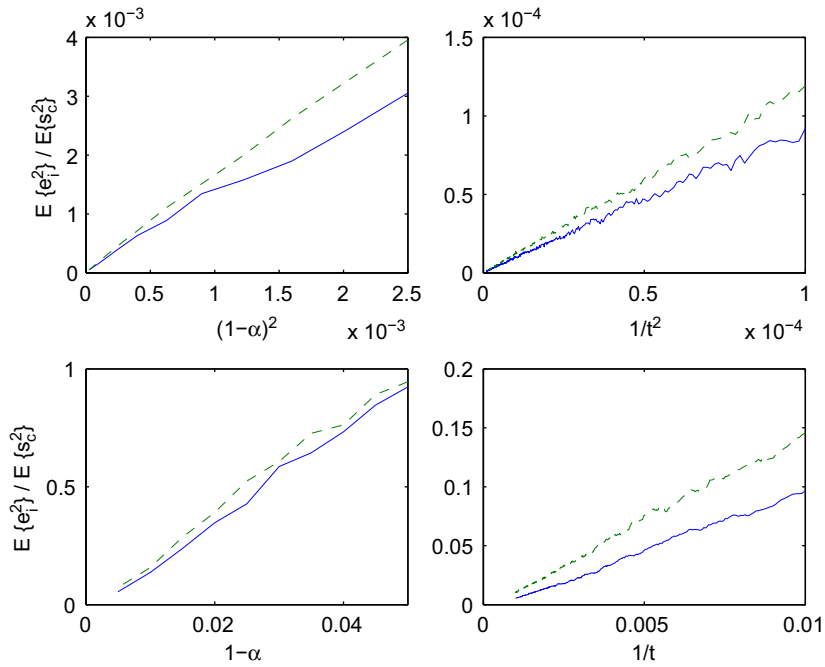


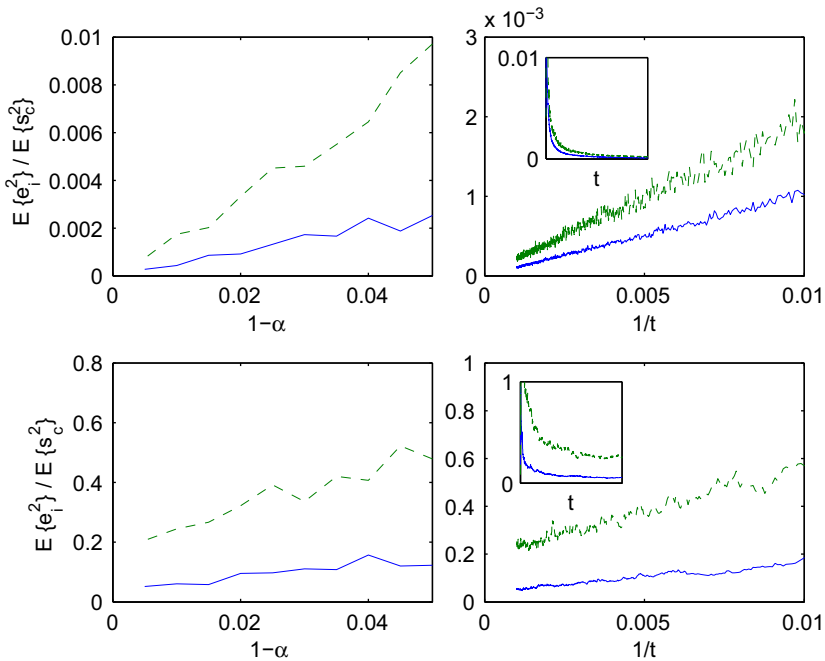
Fig. 2. Means  $\pm$  one standard deviation for decision functions: centralized strategy (dashed lines), proposed algorithm (solid lines); constant consensus matrices (up), random consensus matrices (down).

of the error on the forgetting factor  $\alpha$  under the hypothesis  $H_0$ : the aforementioned ratio is calculated as a function of  $(1-\alpha)$  for both cases of constant and random

consensus matrices. The results of Theorem 1 are clearly justified, since the obtained curves are approximately linear.



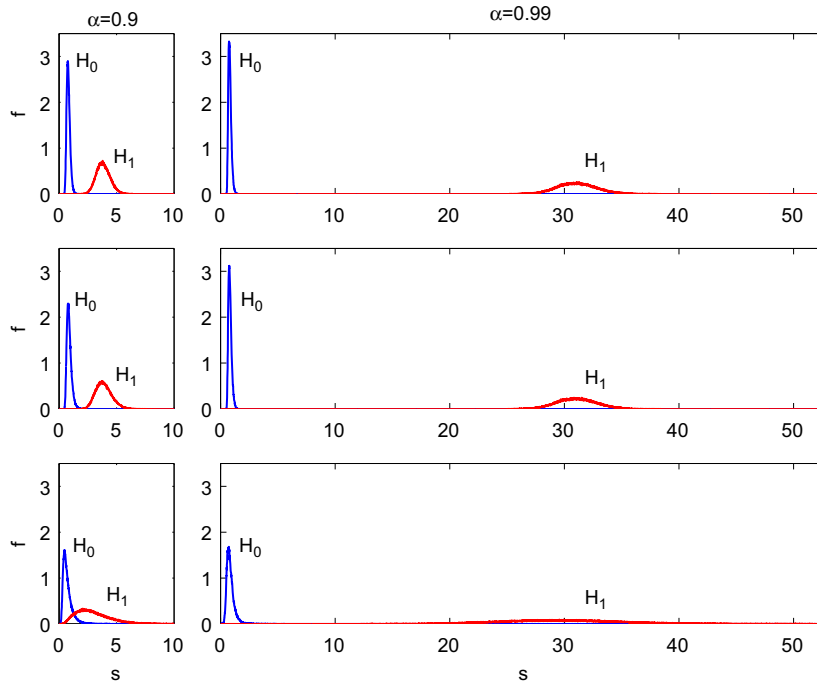
**Fig. 3.** Ratio of the mean square error and the mean square value of the centralized statistics under  $H_1$ : constant consensus matrices (top), random C (bottom); change in the mean (solid line), change in the variance (dashed line); constant forgetting factor (left), time varying forgetting factor (right).



**Fig. 4.** Ratio of the mean square error and the mean square value of the centralized statistics under  $H_0$ : constant consensus matrices (top), random C (bottom); change in the mean (solid line), change in the variance (dashed line); constant forgetting factor (left), time varying forgetting factor (right).

As the first step in the evaluation of the proposed algorithm in terms of the detection performance, distributions of the generated statistics under both hypotheses are estimated using  $\sim 10^5$  time samples. Estimated distributions for one randomly selected node are shown in

Fig. 5. As can be seen, choosing  $\alpha$  closer to 1 results in a greater separation of the statistics under the two hypotheses. Higher dispersion of the statistics in the case of random consensus matrices is a result of the chosen communication strategy (one one-way communication



**Fig. 5.** Estimated distributions of the statistics under the hypothesis  $H_0$  and  $H_1$ : centralized strategy (top), constant consensus matrices (middle), random consensus matrices (bottom);  $\alpha = 0.9$  (left),  $\alpha = 0.99$  (right).

per time sample); communication strategies that assume exchange of more information per time sample would give statistics that are closer to the statistics generated in the case of constant consensus matrices.

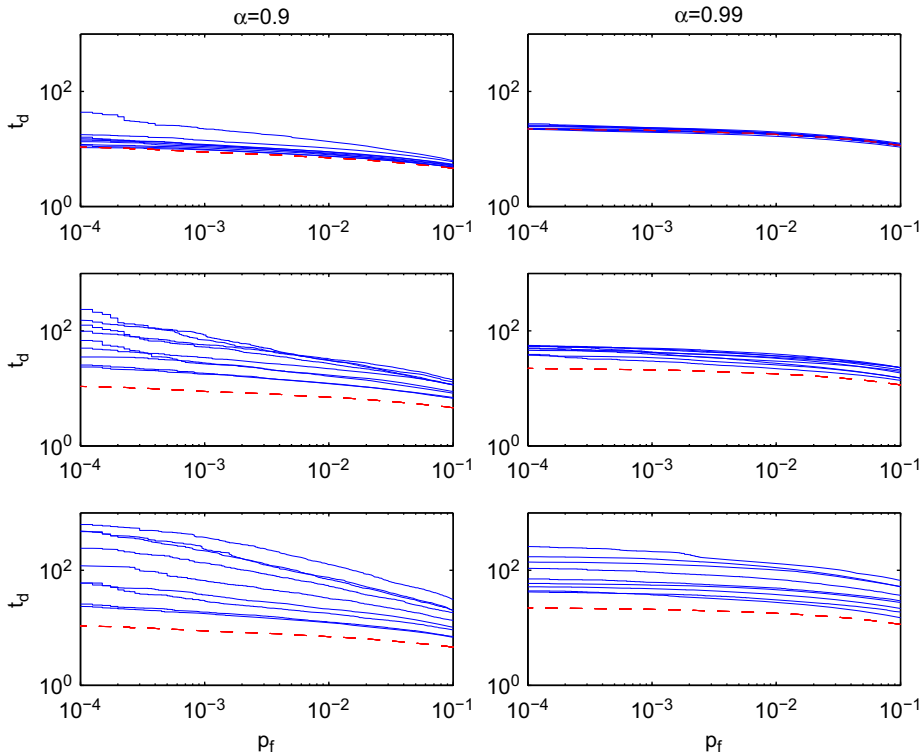
Choosing  $\alpha$  closer to 1 results in the appearance of an interval for the threshold values where the probability of false alarm is practically equal to 0 and the probability of detection is equal to 1. However, with  $\alpha$  increasing, the detection delay increases as well, so that the detection delay becomes more important than probability of detection. Fig. 6 shows detection delay versus false alarm rate, obtained by testing the statistics over a range of detection thresholds, for all the nodes in the network (detection delay is averaged over 500 Monte Carlo runs). Along with the centralized case and the cases of constant and random consensus matrices, the case of completely decentralized local detectors is also shown. It can be seen that the introduction of the consensus scheme significantly improves the detection performance of the algorithm w.r.t. the local case (lower detection delay for a given false alarm rate), even when working with very sparse random consensus matrices. In order to see which value of  $\alpha$  is to be chosen, some characteristic values of detection delays are taken from Fig. 6 and shown in Table 1. As we claim that any node can be chosen as a final decision maker, the value of interest is the maximal delay across the nodes. Obviously, for a relatively high value of the probability of false alarm, choosing smaller  $\alpha$  is a better solution, while for low false alarm rates  $\alpha$  close to 1 is the right choice, which makes that the detection delay among all the nodes for the probability of false alarm of  $10^{-4}$  does not exceed  $10^2$  even in the case of random consensus

matrices (its value is 55.83, while the worst delay among the nodes without consensus is 258.6).

Fig. 7 shows probability of detection versus probability of false alarm for all the nodes in the network. Cases of constant and random consensus matrices, as well as the local case are shown. The probability of detection in the centralized case for all possible values of the probability of false alarm is equal to 1 for both  $\alpha = 0.9$  and  $\alpha = 0.99$ . It could be concluded from the figure that for the cases of both constant and random consensus matrices choosing  $\alpha$  sufficiently close to 1 results in superb efficiency of the proposed algorithm. However, this is misleading in the context of change detection, having in mind the aforementioned increase of the detection delay with the increase of  $\alpha$ ; this is the reason why the focus of the above analysis is not on the probability of detection but on the detection delay.

#### 4.2. Time varying forgetting factor

As an illustration of the analysis of the error between the proposed algorithm and the corresponding centralized solution with time varying forgetting factors connected to Theorem 2 from Section 2.4, Fig. 3 (right) and Fig. 4 (right) show the time evolution of the error. The aforementioned network was used and the ratio of the mean square error for one randomly selected node and the mean square value of the centralized statistics under the hypothesis  $H_1$  is calculated using 1000 Monte Carlo runs as a function of  $t^{-2}$  in the case of constant consensus matrices and of  $t^{-1}$  in the case of random consensus matrices. The same ratio was calculated under the hypothesis  $H_0$  as a function of  $t^{-1}$  for both cases



**Fig. 6.** Detection delay versus probability of false alarm for all the nodes: constant consensus matrices (top), random consensus matrices (middle), no consensus (bottom); centralized strategy (dashed line).

**Table 1**

Detection delay (minimum, mean and maximum across all the nodes in the network) for probability of false alarm  $p_f = 10^{-4}$  and  $p_f = 10^{-1}$  and for  $\alpha = 0.9$  and  $\alpha = 0.99$ .

Algorithm	$p_f = 10^{-4}$						$p_f = 10^{-1}$					
	$\alpha = 0.9$			$\alpha = 0.99$			$\alpha = 0.9$			$\alpha = 0.99$		
	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
Centr.	10.77			22.37			4.66			11.7		
C	10.58	16.7	43.47	21.92	23.78	27.39	4.59	5.27	6.34	10.68	11.72	12.7
$C(t)$	23.51	84.36	236.9	36.3	47.73	55.83	5.92	9.88	14.22	12.83	18.58	23.13
Local	23.5	213.5	625.6	37.89	98.42	258.6	6.15	14.34	31.06	15.09	33.78	66.82

of constant and random consensus matrices. It is also calculated as a function of  $t$  to show that the error converges to a constant greater than zero in the case of random consensus matrices unlike the case of constant consensus matrices where it converges to zero as  $t^{-1}$ . The results of [Theorem 2](#) are clearly justified.

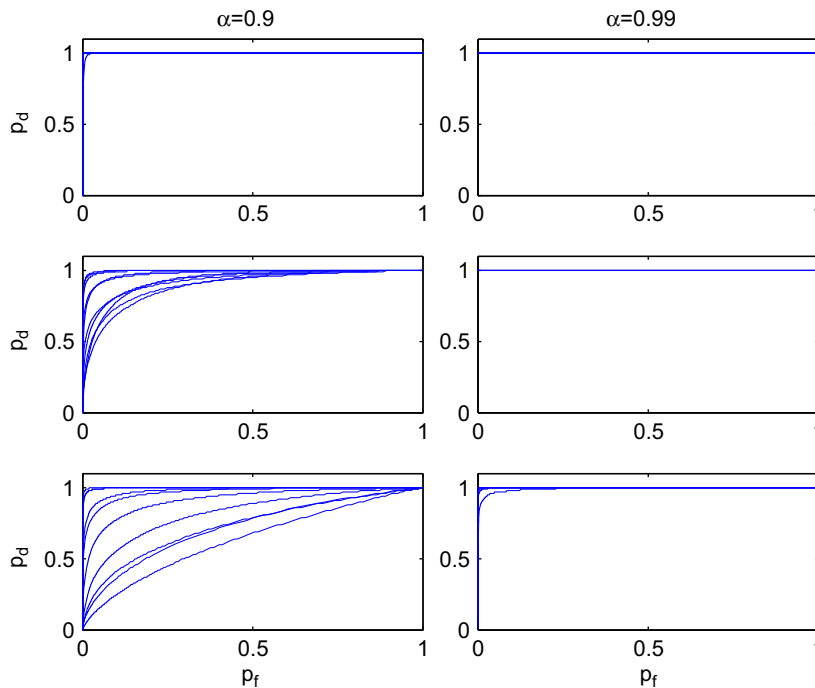
#### 4.3. Change in the variance

A similar network as the one described above is used, where  $(\sigma_i^1)^2$  (unknown to the designer of the detection scheme) are randomly taken from the interval  $(0.5, 1]$  and  $(\sigma_i^0)^2$  randomly taken from the interval  $(0, 0.5]$ . Communication gains are obtained by solving Eq. (11), with the weight vector components  $w_i = 1/n$ ,  $i = 1, \dots, n$ . Similarly

as above, an analysis of the error under hypotheses  $H_1$  and  $H_0$  is given in [Figs. 3 and 4](#), respectively (dashed lines), confirming that all the theoretical results from [Section 2.3](#) connected to the change in the mean hold qualitatively also for the detection algorithm of the change in the variance.

#### 5. Conclusion

In this paper a novel distributed algorithm derived from the Generalized Likelihood Ratio is proposed for real time change detection using sensor networks, with the idea to overcome the basic limitations of the approach in [\[10\]](#) and allow tracking of unknown parameter changes. The algorithm is based on a combination of recursively



**Fig. 7.** Probability of detection versus probability of false alarm for all the nodes: constant consensus matrices (top), random consensus matrices (middle), no consensus (bottom).

generated local statistics, having a specific form following from the GLR, and a global consensus strategy, like in [10]. Problems of detection of an unknown change in either the mean or the variance of a piece-wise stationary stochastic process are discussed. Performance of the proposed detection algorithm for change in the mean is analyzed in the sense of a measure of the error with respect to the corresponding centralized algorithm. The given analysis represents the central point in the paper, since it is much more complex than the one from [10], due to the specific form of the algorithm. It is shown that the statistics generated by the proposed algorithm can be made sufficiently close to the centralized solution. It is also shown that the generated statistics exhibit different behavior w.r.t. the forgetting factor  $\alpha$  under the two hypotheses. Namely, the first and the second moment of the recursively generated statistics grow as  $\alpha$  approaches 1 under hypothesis  $H_1$ , while under  $H_0$  they remain bounded within a constant interval. Therefore, the algorithm with  $\alpha$  close to 1 exhibits high performance in terms of low false alarm rate and high detection rate. However, with  $\alpha$  increasing, the detection delay also increases, so that a careful compromise should be made. Simulation results provide a detailed analysis of this phenomenon, indicating an adequate choice of  $\alpha$ . They also show that all the analysis connected to the change in the mean detection problem holds also for the problem of detecting a change in the variance.

Further work can be aimed at practical aspects of the implementation of the proposed algorithm in sensor networks deployed for different purposes, when it is desirable to avoid the existence of a fusion center. The algorithm can

also be directly applied to decentralized fault detection and isolation (FDI) schemes at the stage of distributed residual evaluation.

## References

- [1] J.F. Chamberland, V. Veeravalli, Decentralized detection in sensor networks, *IEEE Transactions on Signal Processing* 50 (2003) 407–416.
- [2] R. Vishwanathan, P. Varshney, Distributed detection with multiple sensors. Part I. Fundamentals, *Proceedings of the IEEE* 85 (1997) 54–63.
- [3] D. Bajović, D. Jakovetić, J. Xavier, B. Sinopoli, J.M.F. Moura, Distributed detection via Gaussian running consensus: large deviations asymptotic analysis, *IEEE Transactions on Signal Processing* (2011).
- [4] E. Franco, R. Olfati-Saber, T. Parisini, M.M. Polycarpou, Distributed fault diagnosis using sensor networks and consensus based filters, in: *Proceedings of the 45th IEEE CDC Conference*, 2006.
- [5] P. Braca, S. Marano, V. Matta, Enforcing consensus while monitoring the environment in wireless sensor networks, *IEEE Transactions on Signal Processing* 56 (2008) 3375–3380.
- [6] P. Braca, S. Marano, V. Matta, P. Willett, Asymptotic optimality of running consensus in testing binary hypotheses, *IEEE Transactions on Signal Processing* 58 (2010) 814–825.
- [7] F.S. Cattivelli, A.H. Sayed, Distributed detection over adaptive networks using diffusion adaptation, *IEEE Transactions on Signal Processing* 59 (2011) 1917–1932.
- [8] P. Braca, S. Marano, V. Matta, P. Willett, Consensus-based page's test in sensor networks, *Signal Processing* 91 (2011) 919–930.
- [9] M. Basseville, L.V. Nikiforov, *Detection of Abrupt Changes: Theory and Applications*, Prentice Hall, 1993.
- [10] S.S. Stanković, N. Ilić, M.S. Stanković, K.H. Johansson, Distributed detection based on a consensus algorithm, *IEEE Transactions on Signal Processing* 59 (12) (2011) 5686–5697.
- [11] T.L. Lai, *Sequential Multiple Hypothesis Testing and Efficient Fault Detection-Isolation in Stochastic Systems*, Technical Report, Stanford University, 1998.



- [12] S.X. Ding, Model Based Fault Diagnosis Techniques—Design Schemes, Algorithms and Tools, Springer Verlag, 2008.
- [13] S.S. Stanković, M.S. Stanković, D.M. Stipanović, Consensus based overlapping decentralized estimation with missing observations and communication faults, *Automatica* 45 (2009) 1397–1406.
- [14] S.S. Stanković, M.S. Stanković, D.M. Stipanović, Decentralized parameter estimation by consensus based stochastic approximation, *IEEE Transactions on Automatic Control* 56 (2011) 531–543.
- [15] W. Ren, R. Beard, Consensus seeking in multi-agent systems using dynamically changing interaction topologies, *IEEE Transactions on Automatic Control* 50 (2005) 655–661.
- [16] R. Olfati-Saber, A. Fax, R. Murray, Consensus and cooperation in networked multi-agent systems, *Proceedings of the IEEE* 95 (2007) 215–233.
- [17] N. Ilić, S.S. Stanković, Communication gains design in a consensus based distributed change detection algorithm, in: *Proceedings of the 8th European Workshop on Advanced Control and Diagnosis*, 2010.
- [18] S. Boyd, A. Ghosh, B. Prabhakar, D. Shah, Randomized gossip algorithms, *IEEE Transactions on Information Theory* 52 (2006) 2508–2530.
- [19] H.F. Chen, *Stochastic Approximation and its Applications*, Kluwer Academic, Dordrecht, The Netherlands, 2002.