

# Reducing false alarm rates in observer-based distributed fault detection schemes by analyzing moving averages<sup>\*</sup>

Stefano Attuati<sup>\*</sup> Marcello Farina<sup>\*\*</sup> Francesca Boem<sup>\*\*\*</sup>  
Thomas Parisini<sup>\*\*\*\*,†,‡</sup>

<sup>\*</sup> e-mail: [stefano.attuati@gmail.com](mailto:stefano.attuati@gmail.com).

<sup>\*\*</sup> Dipartimento di Elettronica, Informazione e Bioingegneria,  
Politecnico di Milano, Italy (e-mail: [marcello.farina@polimi.it](mailto:marcello.farina@polimi.it))

<sup>\*\*\*</sup> Department of Electronic and Electrical Engineering, University  
College London, UK (e-mail: [f.boem@ucl.ac.uk](mailto:f.boem@ucl.ac.uk))

<sup>\*\*\*\*</sup> Department of Electrical and Electronic Engineering, Imperial  
College London, UK (e-mail: [t.parisini@imperial.ac.uk](mailto:t.parisini@imperial.ac.uk))

<sup>†</sup> KIOS Research and Innovation Centre of Excellence, University of  
Cyprus

<sup>‡</sup> Department of Engineering and Architecture, University of Trieste,  
Italy

**Abstract:** In this paper, we first analyze the possible limitations of a model-based fault detection method grounded on a partition-based distributed Luenberger observer. The corresponding fault detection test consists of comparing, for each time instant, the output prediction error with a suitable bound, computed analytically in a distributed and scalable way. As a result, we highlight the presence of an often restrictive tradeoff between false-alarm and missed-detection rates. To overcome this significant drawback, we resort to a method based on the analysis of moving averages of residuals. Tests on an academic case study show the effectiveness of this approach.

*Keywords:* Model-based distributed fault detection, partition-based observers.

## 1. INTRODUCTION

One of the most important tasks for computers supervising complex process plants is the detection and diagnosis of faults. As stated in (Isermann, 2005), advanced methods for supervision, fault detection, and fault diagnosis become increasingly important for the improvement of reliability, safety, and efficiency of many technical processes. This holds especially for safety-related processes like aircrafts, trains, cars, power and chemical plants.

Among the system monitoring problems we focus on fault detection (FD), (Gertler, 1998; Basseville and Nikiforov, 1998) and, specifically, we consider a model-based approach (Isermann, 2005). In particular we use observers to estimate the state of the (nominal) system, and analyze the properties of the output variable estimation error to test if it fits the expected requirements in a non-faulty scenario. This leads to statistical tests that guarantee that a fault is present or absent with a prescribed probability. The use of model-based approaches can become practically prohibitive (concerning both the related computational burden and the required amount of information to be

transmitted) in complex scenarios, when large-scale systems are involved, possibly composed by a number of cooperating and coupled subsystems (Samad and Parisini, 2011). Indeed, in this paper we deal with large-scale linear discrete-time systems, characterized by the interconnection of  $M$  subsystems. Each subsystem is described by

$$x_i(k+1) = A_{ii}x_i(k) + \sum_{j \neq i} A_{ij}x_j(k) + w_i(k) \quad (1a)$$

$$y_i(k) = C_i x_i(k) + v_i(k) \quad (1b)$$

where  $x_i(k) \in \mathbb{R}^{n_i}$  is the state and  $y_i(k) \in \mathbb{R}^{p_i}$  is the output of the subsystem. Signals  $w_i(k) \in \mathbb{R}^{n_i}$  and  $v_i(k) \in \mathbb{R}^{p_i}$  are zero-mean white noises, for  $i = 1, \dots, M$ , and  $\mathbb{E}[w_i(k)w_j^T(k)] = Q_i \delta_{ij}$ ,  $\mathbb{E}[v_i(k)v_j^T(k)] = R_i \delta_{ij}$ ,  $\mathbb{E}[w_i(k)v_j^T(h)] = 0$  for all  $i, j = 1, \dots, M$  and  $h, k \geq 0$ . In the above notation  $\delta_{ij}$  is the *Kronecker delta* function, i.e.  $\delta_{ij} = 1$  if  $i = j$  and  $\delta_{ij} = 0$  if  $i \neq j$ .

As discussed in the surveys (Sijs et al., 2008; Farina et al., 2010b), two main classes of estimation techniques for distributed schemes are currently under investigation, both generally referred to as *distributed* state-estimation algorithms. The first class of problems (studied, e.g., in (Olfati-Saber, 2007)) concerns the case where the full state of the overall system (i.e., the overall state  $x = \text{col}(x_1, \dots, x_M)$ ) is estimated by each subsystem: following this approach each local observer needs to access the information re-

<sup>\*</sup> This work has been partially supported by European Union's Horizon 2020 research and innovation programme under grant agreement No 739551 (KIOS CoE). This work has also been conducted as part of the research project *Stability and Control of Power Networks with Energy Storage* (STABLE-NET) which is funded by the RCUK Energy Programme (contract no: EP/L014343/1)

garding the overall system. The other class of algorithms focuses on the estimation of a part of the state vector (i.e., on  $x_i$  solely, for each  $i = 1, \dots, M$ ) gathering information only from the neighboring subsystems: the latter approach is called *partition-based estimation* (studied, e.g., in (Stanković et al., 2009; Farina et al., 2010a; Farina and Carli, In press; Schneider et al., 2015)) and enjoys many advantages regarding both the communication and computational loads.

Some works have been devoted to partition-based fault detection and control system fault recovery in the past, e.g., (Shames et al., 2011; Zhang and Zhang, 2012; Reppa et al., 2015; Lan and Patton, 2016; Blanke et al., 2016; Gupta and Puig, 2016; Rivero et al., 2016; Boem et al., 2017; Lauricella et al., 2017). In this paper, we consider a partition-based Luenberger estimator (denoted PLF) previously proposed in (Boem et al., 2016). Besides providing reliable state estimation with granted properties, this algorithm also provides a consistent estimate of the prediction error covariance matrix (in the sense specified in (Uhlmann, 2003)), which can be used in order to design suitable threshold values for FD purposes. In particular, in (Boem et al., 2016) the FD test consists of comparing, for each time instant, the output prediction error with a suitable bound. In this paper, we analyze, using a simple academic example, some merits and limitations of this approach, the performances of which have been also extensively tested on a more complex simulation example using a Power Networks application in (Boem et al., 2016). Indeed, we highlight the presence of an often restrictive tradeoff between the false-alarm (FA) and the missed-detection rates: indeed, as the FA decreases the threshold value increases, and too large thresholds may lead to unacceptably large missed-detection rates. Therefore, we explore a possible solution to this issue, exploiting to a method described in (Gertler, 1998), based on the analysis of moving averages of residual values, rather than single time points. The aim is to show how moving averages can be exploited to indeed improve the performance of the FD scheme; notably, theoretical results will be provided to provide analytical estimates of the probabilistic properties of the moving averages of residuals. To keep the analysis simple, we consider Gaussian-type noises.

In Section 2, we briefly describe the method proposed in (Boem et al., 2016). In Section 2.4 we show some limitations of this method and we propose a solution in Section 3. The comparison between the two methods is shown in Section 4 on the academic example described below. Some conclusions are drawn in Section 5.

**Academic example.** An academic simple example will be used throughout the paper to analyze in a simple way the results obtained using different FD approaches. Consider the nominal system (1). To force a non-zero steady state, we add a constant known input, by modifying equation (1a) as follows:  $x_i(k+1) = A_{ii}x_i(k) + \sum_{j \neq i} A_{ij}x_j(k) + B_{ii}\bar{u} + w_i(k)$ . We set

$$A_{ii} = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & -0.9 \end{bmatrix}, \quad A_{ij} = \begin{bmatrix} 0.05 & 0.025 \\ 0.025 & 0.05 \end{bmatrix}$$

for all  $i, j = 1, \dots, M$  with  $i \neq j$ . Also,  $B_{ii} = [1 \ 0]^T$  and  $C_i = [1 \ 1]$  for all  $i = 1, \dots, M$ . We set  $\bar{u} = 5$ ; the noise

signals  $w_i(k)$  and  $v_i(k)$  have covariances  $Q_i = \text{diag}(1, 1)$  and  $R_i = 1$  for all  $i = 1, \dots, M$ . We simulate two types of faults acting on subsystem 1: additive and multiplicative faults on the state equation. For simulating them, the first state equation is replaced by the following faulty one

$$x_1(k+1) = (1 + f_{\text{sim}}^{\text{mult}}(k))A_{11}x_1(k) + \sum_{j \neq 1} A_{1j}x_j(k) + B_{11}\bar{u} + B_{f,1}f_{\text{sim}}^{\text{add}}(k) + w_1(k)$$

where  $B_{f,1} = [1 \ 0]^T$ . The faults are assumed to be persistent, i.e.,  $f_{\text{sim}}^{\text{add}}(k) = \bar{f}\text{step}(k - T_{\text{fault}}^{\text{add}})$  and  $f_{\text{sim}}^{\text{mult}}(k) = \bar{f}\text{step}(k - T_{\text{fault}}^{\text{mult}})$ , where  $\bar{f}$  may vary in order to test different fault sizes and where  $T_{\text{fault}}^{\text{add}}$  and  $T_{\text{fault}}^{\text{mult}}$  denote the instants when the additive and multiplicative, respectively, faults occur.

Since the proposed method is distributed and scalable, as explained in (Boem et al., 2016), the complexity linearly depends on the number of neighboring subsystems. Hence, increasing the number of subsystems in the simulation example does not improve the significance of the results. For the sake of simplicity, we show the case with  $M = 2$  subsystems.

**Notation.** Regarding model (1), we define the set of predecessors of subsystem  $i$  as  $\mathcal{N}_i = \{j | A_{ij} \neq 0\}$  and the set of successors of  $i$  as  $\mathcal{S}_i = \{j | i \in \mathcal{N}_j\}$ . It is also useful to define the sets of strict predecessors and successors as  $\tilde{\mathcal{N}}_i = \mathcal{N}_i \setminus \{i\}$  and  $\tilde{\mathcal{S}}_i = \mathcal{S}_i \setminus \{i\}$ .

Given a stochastic variable  $x$ , we denote with  $\mathbb{E}[x]$  its expected value. The symbols  $\geq$  and  $>$  are also used to denote positive semi-definite matrices and positive definite matrices, respectively. The cardinality of a set  $\mathcal{N}$  is denoted with  $|\mathcal{N}|$  and a square matrix  $A$  is Schur stable if its spectral radius is strictly smaller than one.

## 2. OBSERVER AND PARTITION-BASED FAULT DETECTION

### 2.1 The partition-based Luenberger state estimator

The partition-based Luenberger Filter used in this paper is a distributed estimation scheme, proposed in (Boem et al., 2016), of the type

$$\begin{aligned} \hat{x}_i(k+1) &= \sum_{j \in \mathcal{N}_i} \{A_{ij}\hat{x}_j(k) + L_{ij}(y_j(k) - C_j\hat{x}_j(k))\} \\ \hat{y}_i(k) &= C_i\hat{x}_i(k) \end{aligned} \quad (2)$$

where the static gains  $L_{ij}$  are computed (see Section 2.3) to guarantee stability of the state estimate  $e_i(k) = x_i(k) - \hat{x}_i(k)$ . Note that the scheme is distributed, since the information required at any time instant  $k$  by the  $i$ -th state estimator consists of the state estimation  $\hat{x}_j(k)$  and the measurement  $y_j(k)$  collected by the neighbors of  $i$ , i.e.,  $j \in \mathcal{N}_i$ .

### 2.2 FD method (a): single residual testing

The FD algorithm proposed in (Boem et al., 2016) is based on the analysis of the residual

$$r_i(k) = y_i(k) - \hat{y}_i(k) \quad (3)$$

which is computed locally by the  $i$ -th local diagnoser using data coming from the local state estimator and from the

subsystem output.

The general idea is to compare  $r_i(k)$ , at each time step, with a suitable threshold computed based on the covariance of  $r_i(k)$ , i.e.,

$$\Sigma_i(k) = \mathbb{E}[r_i(k)r_i(k)^T]$$

Note that the variance  $\sigma_{i,l}^2(k)$  of  $r_{i,l}(k)$ , i.e., the  $l$ -th entry of  $r_i(k)$ , is the  $l$ -th diagonal element of  $\Sigma_i(k)$ . The proposed FD scheme relies on the fact that, in view of the Gaussianity of  $r_{i,l}(k)$  (which follow from the Gaussianity of  $w_j(k)$ ,  $v_j(k)$ ,  $j = 1, \dots, M$  and from the linearity of both the state and observer equations (1) and (2), respectively), for any  $p \in (0, 1]$ , we can define a scalar  $\alpha \geq 0$  such that

$$P(|r_{i,j}(k)|/|\sigma_{i,l}(k)| \geq \alpha) = p \quad \text{i.e.} \quad \int_{-\alpha}^{\alpha} f(x)dx = 1 - p \quad (4)$$

where  $f$  is the probability density function of a zero mean Gaussian variable with unitary variance. From this, it follows that (see also (Gertler, 1998))  $|r_{i,l}(k)| > \rho_{i,l}(k) = \alpha\sigma_{i,l}(k)$  with probability  $p$  if in nominal (i.e., non-faulty) conditions. This criterion is used for detecting faults, and the following rule is applied at any time instant.

$$\begin{cases} \text{if } |r_{i,l}(k)| < \rho_{i,l}(k), \text{ then no fault is detected} \\ \text{otherwise fault is detected} \end{cases}$$

It is therefore clear that the probability  $p$  corresponds to the FA rate and will be denoted  $p_{\text{FA}}$  for better clarity. In the following, the previously explained method will be denoted *detection testing method (a)*.

### 2.3 Approximation of the residual variance

The key ingredient for the definition of the FD approach discussed in Section 2.2 is the variance  $\Sigma_i(k)$  of the  $i$ -th residual  $r_i(k)$ . The simplest method for obtaining an approximation of it consists of computing the sampled covariance from available data. However, this approach has some possible drawbacks: (i) it is consistent only under the stationarity assumption, i.e., only if data are collected in steady state conditions; (ii) it is reliable only if a very large number of samples are available; (iii) it is correct only if data are collected in healthy state. In view of this, this method can be ineffective in transient conditions (e.g., in non-autonomous systems, during set-point changes) and can entail significant detection delays. Also, it can lead to serious detection problems if the data are collected, unintentionally, in a faulty or non-nominal setting.

In this section we summarize the novel approach proposed in (Boem et al., 2016), which allows to compute (in a distributed and computationally scalable way) a (possibly conservative) upper bound to  $\Sigma_i(k)$ . To this regard note that, if an upper bound to the variance is used in place of the actual one, the threshold computed for a given  $p_{\text{FA}}$  is an upper bound to the tight threshold value. As a result, in non-faulty conditions  $|r_{i,l}(k)| \geq \rho_{i,l}(k)$  with probability smaller than  $p_{\text{FA}}$ . This entails conservativity of the proposed FD method. Indeed, obtaining a larger threshold means that there will be less FAs but, on the other hand, with a smaller probability to detect faults. Next, this trade-off will be better analyzed using the discussed academic example.

Note that, from (1) and (2),  $r_i(k) = C_i e_i(k) + v_i(k)$ , and

$$\Sigma_i(k) = C_i \Pi_{ii}(k) C_i^T + R_i \quad (5)$$

where  $\Pi_{ii}(k) = \mathbb{E}[e_i(k)e_i(k)^T]$ . The  $i$ -th local state prediction error, in view of (1) and (2), evolves according to

$$e_i(k+1) = \sum_{j \in \mathcal{N}_i} \{(A_{ij} - L_{ij}C_j)e_j(k) - L_{ij}v_j(k)\} + w_i(k) \quad (6)$$

Since the subsystems are interconnected with each other, the evolution of  $e_i(k)$ , as well as that of its variance  $\Pi_{ii}(k)$ , cannot be determined in a purely local way. To consider the overall system, we define the collective state prediction error as  $\mathbf{e}(k) = (e_1(k), \dots, e_M(k))$ , while the full noise vectors are  $\mathbf{w}(k) = (w_1(k), \dots, w_M(k))$  and  $\mathbf{v}(k) = (v_1(k), \dots, v_M(k))$ , with covariances  $\mathbf{Q} = \text{diag}(Q_1, \dots, Q_M)$  and  $\mathbf{R} = \text{diag}(R_1, \dots, R_M)$ , respectively. In view of (6), the overall error dynamics is described by

$$\mathbf{e}(k+1) = (\mathbf{A} - \mathbf{L}\mathbf{C})\mathbf{e}(k) - \mathbf{L}\mathbf{v}(k) + \mathbf{w}(k) \quad (7)$$

where  $\mathbf{C} = \text{diag}(C_1, \dots, C_M)$  and

$$\mathbf{A} = \begin{bmatrix} A_{11} & \dots & A_{1M} \\ \vdots & \ddots & \vdots \\ A_{M1} & \dots & A_{MM} \end{bmatrix}$$

and, similarly,  $\mathbf{L}$  is the block matrix having the  $(i, j)$ -th element equal to  $L_{ij}$  for  $i = 1, \dots, M$  and  $j = 1, \dots, M$ .

Let us now define  $\mathbf{F} = \mathbf{A} - \mathbf{L}\mathbf{C}$ . The covariance matrix of the collective estimation error is defined by  $\mathbf{\Pi}(k) := \mathbb{E}[\mathbf{e}(k)\mathbf{e}^T(k)]$  and obeys the recursive equation

$$\mathbf{\Pi}(k+1) = \mathbf{F}\mathbf{\Pi}(k)\mathbf{F}^T + \mathbf{L}\mathbf{R}\mathbf{L}^T + \mathbf{Q} \quad (8)$$

The evolution of  $\mathbf{\Pi}(k)$  can be characterized only in a centralized and non-scalable way. However, in (Boem et al., 2016) a method has been proposed to compute, in a distributed and scalable fashion, a block-diagonal upper bound to  $\mathbf{\Pi}(k)$ , at each time instant. Indeed, we define matrices  $B_i(k)$ ,  $i = 1, \dots, M$  in a recursive way as

$$B_i(k+1) = Q_i + \sum_{j \in \mathcal{N}_i} [(\tilde{A}_{ij} - L_{ij}\tilde{C}_j)B_j(k)(\tilde{A}_{ij} - L_{ij}\tilde{C}_j)^T + L_{ij}\tilde{R}_jL_{ij}^T] \quad (9)$$

where  $\tilde{A}_{ij} = \sqrt{\zeta_i}A_{ij}$ ,  $\tilde{C}_i = \sqrt{\zeta_i}C_i$ ,  $\tilde{R}_i = \zeta_i R_i$  and  $\zeta_i = |\mathcal{S}_i|$ , for all  $i, j = 1, \dots, M$ .

In (Boem et al., 2016) it is proved that  $B_i(k)$  can be used as an upper bound to  $\Pi_{ii}(k)$ , for all  $i = 1, \dots, M$  and for all  $k \geq 1$ , as stated in the following theorem.

*Theorem 1.* If one sets  $\text{diag}(B_1(1), \dots, B_M(1)) \geq \mathbf{\Pi}(1)$  then, for all  $i = 1, \dots, M$  and for all  $k \geq 1$ , it holds that  $B_i(k) \geq \Pi_{ii}(k)$ .  $\square$

Finally, using (5), we define the analytical upper bound to  $\Sigma_i(k)$  as

$$\Sigma_i^B(k) = C_i B_i(k) C_i^T + R_i \quad (10)$$

To complete the picture of the presented method for covariance evaluation, in (Boem et al., 2016) a condition is provided, that guarantees that  $B_i(k)$  remain bounded for all  $k$  and, at the same time, stability of the estimation error dynamics. More specifically, this condition consists of the Schur stability of the following matrix.

$$\mathbb{F} = \tilde{F} \odot \tilde{F} = \begin{bmatrix} \tilde{F}_{11} \otimes \tilde{F}_{11} & \dots & \tilde{F}_{1M} \otimes \tilde{F}_{1M} \\ \vdots & \ddots & \vdots \\ \tilde{F}_{M1} \otimes \tilde{F}_{M1} & \dots & \tilde{F}_{MM} \otimes \tilde{F}_{MM} \end{bmatrix} \quad (11)$$

where, for all  $i, j$ ,  $\tilde{F}_{ij} = (\tilde{A}_{ij} - L_{ij}\tilde{C}_j)$  and the matrix  $\tilde{F}$  is the matrix whose blocks are  $\tilde{F}_{ij}$ . Also,  $\odot$  denotes the Khatri-Rao product, while  $\otimes$  denotes the Kronecker product (see (Horn and Johnson, 2012)).

#### 2.4 Tests on the FD method (a)

In this section we show the results of Montecarlo tests performed on the academic example introduced in Section 1. For each experiment,  $N_m = 1000$  Montecarlo runs are performed. For the implementation of the algorithm, the static gains used in the PLF have been computed in order to achieve the Schur stability of matrix  $\mathbb{F}$ , granting the stability of the estimation and the convergence of matrices  $B_i$  as stated in Section 2.3.

The covariance term has been approximated using the analytical method described in Section 2.3 and, for comparison, using the sampled covariance. In line with this, the following cases are tested.

I) PLF-empirical: we set  $\Sigma_i(k) = \tilde{\Sigma}_i^{\text{EMP}}$ , where

$$\tilde{\Sigma}_i^{\text{EMP}} = \frac{1}{N} \sum_{k=1}^N r_i(k)r_i(k)^T \quad (12)$$

II) PLF-analytical: we set  $\Sigma_i(k) = \Sigma_i^B(k)$  for all  $k$ .

The covariance  $\tilde{\Sigma}_i^{\text{EMP}}$  is obtained by preliminarily simulating the system in nominal conditions and using  $N = 10000$  available samples.

Different scenarios are simulated varying the fault type (i.e., additive and multiplicative), fault amplitude (i.e., small and large, as better specified in the following), and different guaranteed FA rates (i.e.,  $p_{\text{FA}} = 0.0002$  and  $p_{\text{FA}} = 0.01$ ). In all simulations, fault detection starts at time  $T = 41$ , while the faults occur at time  $T = 100s$ . The fault sizes are: (i) for additive faults, small ones have amplitude  $\tilde{f} = 2$ , while large faults have amplitude  $\tilde{f} = 5$ ; (ii) for multiplicative faults, small ones have amplitude  $\tilde{f} = 0.015$ , while large faults have amplitude  $\tilde{f} = 0.04$ .

The plots shown in Figures 1, 2 denote the evolution of the *cumulative fault detection rate*  $R_{\text{FD}}(k) = N_{\text{FD}}(k)/N_m$ , where  $N_{\text{FD}}(k)$  is the number of Montecarlo runs in which the corresponding fault detection threshold has been passed up to the time instant  $k$ . The slope of such a line, for both subsystems, in the interval  $k = [41, 100]$  (i.e., before the occurrence of the fault) is proportional to the FA rate. From time  $T = 100$  on, for subsystem 1 we note that the rate of detection increases significantly due to the fault occurrence while, for subsystem 2, the effect of the fault occurrence on the rate of detection is not apparent. As expected, the rate of detection is smaller when the analytical upper bound proposed in (Boem et al., 2016) is used with respect to the empirical one.

Clearly, large faults can be promptly detected by the proposed method for both values of  $p_{\text{FA}}$ . However, while a small value  $p_{\text{FA}} = 0.0002$  is required for reducing the number of systems for which false alarms are detected, it corresponds to a threshold value  $\rho_1(k)$  that is too large to properly and promptly detect small-amplitude faults. On the other hand, to well detect small-amplitude faults, it is necessary to increase the false-alarm rate  $p_{\text{FA}}$ , that results unacceptable in a dynamic scenario.

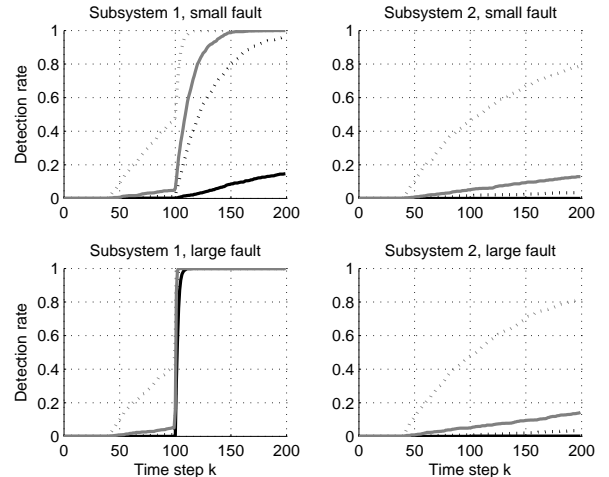


Fig. 1. Cumulative detection rates  $R_{\text{FD}}(k)$  with additive faults on subsystem 1 at  $T = 100$ . Solid lines are obtained with PLF-analytical, while dotted lines are obtained with PLF-empirical. Black lines are obtained with  $p_{\text{FA}} = 0.0002$ , while grey lines are obtained with  $p_{\text{FA}} = 0.01$ .

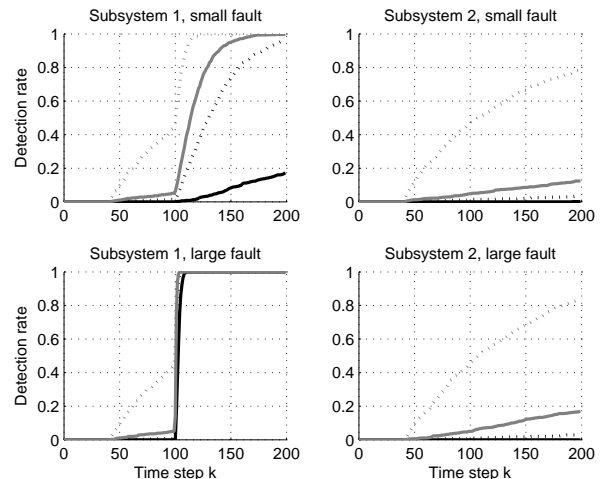


Fig. 2. Cumulative detection rates  $R_{\text{FD}}(k)$  with multiplicative faults on subsystems 1 at  $T = 100$ . Solid lines are obtained with PLF-analytical, while dotted lines are obtained with PLF-empirical. Black lines are obtained with  $p_{\text{FA}} = 0.0002$ , while grey lines are obtained with  $p_{\text{FA}} = 0.01$ .

### 3. FD METHOD (B): SLIDING WINDOW AVERAGES.

As shown in Section 2.4, the analysis of a single residual  $r_{i,l}(k)$  at each time instant can lead to unsatisfactory FD results in case of small-amplitude faults. The goal of this section is to show how data related to the residual time series can be exploited in order to improve the performance of the FD scheme in terms of false alarm rate and conservativity of the bounds and that the probabilistic properties of such data series can be estimated in an analytical and distributed way.

### 3.1 Moving averages of residuals

The strategy described in this section consists of testing the average of the residual values computed over the sliding window. Specifically, we define  $\bar{r}_{i,l}^{(m)}(k) = \frac{1}{m} \sum_{j=0}^{m-1} r_{i,l}(k-j)$ . Note that, in view of the assumptions introduced in the previous sections,  $\bar{r}_{i,l}^{(m)}(k)$  is a Gaussian variable with zero mean and variance  $(\sigma_{i,l}^{(m)}(k))^2 = \mathbb{E}[(\bar{r}_{i,l}^{(m)}(k))^2]$ . As such, it can be considered as a single observation, and the corresponding threshold can be computed similarly to the one computed in Section 2.2, i.e.,

$$\begin{cases} \text{if } |\bar{r}_{i,l}^{(m)}(k)| < \bar{\rho}_{i,l}^{(m)}(k), \text{ then no fault is detected} \\ \text{otherwise fault is detected} \end{cases}$$

where  $\bar{\rho}_{i,l}^{(m)}(k) = \alpha \bar{\sigma}_{i,l}^{(m)}(k)$ .

The use of the average over  $m$  values rather than single values is justified by the fact that  $\bar{\sigma}_{i,l}^{(m)}(k) \leq \sigma_{i,l}(k)$ , leading to the following advantages, with respect to the FD method (a): (i) we can in principle obtain the same FA rate (i.e., the same value of  $\alpha$ ) with a smaller threshold, which allows to detect faults having smaller amplitude; (ii) we can reduce the FA rate without reducing the missed detection rate, i.e., without raising the threshold value.

### 3.2 Analytical approximation of the average covariance

In this section we show how to compute analytically an upper bound to  $\bar{\sigma}_{i,l}^{(m)}(k)$  in a distributed and scalable way. Denote with  $(\sigma_{i,l}^B(k))^2$  the  $l$ -th diagonal entry of matrix  $\Sigma_i^B(k)$ . Considering that

$$(\bar{\sigma}_{i,l}^{(m)}(k))^2 = \frac{1}{m^2} \sum_{j,h=0}^{m-1} \gamma_{i,l}(k-j, k-h) \quad (13)$$

where  $\gamma_{i,l}(k-l, k-h) = \mathbb{E}[r_{i,l}(k-j)r_{i,l}(k-h)]$ , we need to find a reliable, although possibly conservative, upper bound to  $\gamma_{i,l}(k-j, k-h)$ , for all  $j, h = 0, \dots, m$ . This is provided by the following proposition.

*Proposition 1.* It holds that both  $\gamma_{i,l}(k-j, k-h) \leq \bar{\gamma}_{i,l}^{B,1}(k-j, k-h)$  and, for  $j \neq h$ ,  $\gamma_{i,l}(k-j, k-h) \leq \bar{\gamma}_{i,l}^{B,2}(k-j, k-h)$ , where

$$\bar{\gamma}_{i,l}^{B,1}(k-j, k-h) = \frac{1}{2}((\sigma_{i,l}^B(k-j))^2 + (\sigma_{i,l}^B(k-h))^2) \quad (14a)$$

$$\begin{aligned} \bar{\gamma}_{i,l}^{B,2}(k-j, k-h) &= \frac{1}{2} \|C_{i,l}\| \|\text{diag}(\mathbf{B}(k - \max(h, j)))\| \\ &+ C_{i,l} B_i(k - \max(h, j)) C_{i,l}^T \mathbf{1}_n \|\mu \lambda^{|h-j|}\| \\ &+ \|C_{i,l}\| \|\mathbf{L}_i^c R_{i,l}^c\| \|\mu \lambda^{|h-j|-1}\| \end{aligned} \quad (14b)$$

and where  $\mathbf{L}_i^c$  is the  $i$ -th block column of  $\mathbf{L}$  and  $R_{i,l}^c$  is the  $l$ -th column of matrix  $R_i$ . Scalars  $\mu > 0$ ,  $\lambda \in [0, 1)$  are defined in such a way that  $\|\mathbf{F}^j\| \leq \mu \lambda^j$  for all  $j > 0$ .  $\square$

In view of Proposition 1 we define the upper bound to  $(\bar{\sigma}_{i,l}^{(m)}(k))^2$  as

$$(\bar{\sigma}_{i,l}^{(m),B}(k))^2 = \frac{1}{m^2} \sum_{j,h=0}^{m-1} \bar{\gamma}_{i,l}^B(k-j, k-h)$$

If  $j = h$   $\bar{\gamma}_{i,l}^B(k-j, k-h) = \bar{\gamma}_{i,l}^{B,1}(k-j, k-h)$  but, if  $j \neq h$ ,

$$\begin{aligned} \bar{\gamma}_{i,l}^B(k-j, k-h) &= \\ \min\{\bar{\gamma}_{i,l}^{B,1}(k-j, k-h), \bar{\gamma}_{i,l}^{B,2}(k-j, k-h)\} \end{aligned} \quad (15)$$

Note that the simplest, yet theoretically sound, approximation that one can use is  $\bar{\gamma}_{i,l}^{B,1}$  for all  $j, h$ . This would lead to the conservative approximation

$$(\bar{\sigma}_{i,l}^{(m),B}(k))^2 = \frac{1}{m} \sum_{j=0}^{m-1} (\sigma_{i,l}^B(k-j))^2$$

which, however, does not allow to play with the tradeoff between FA guaranteed rate and missed detection rate. On the other hand, (15) allows to reduce the conservativity of the approximation, since it accounts for the fact that, since the process  $r_{i,l}(k)$  is stationary (in view of the stability properties of the distributed state estimator), the correlation between  $r_{i,l}(k-j)$  and  $r_{i,l}(k)$  asymptotically tends to zero as  $j$  increases.

### Proof of Proposition 1.

The first inequality can be proven considering that

$$\begin{aligned} \gamma_{i,l}(k-j, k-h) &= \mathbb{E}[r_{i,l}(k-j)r_{i,l}(k-h)] \\ &\leq \mathbb{E}\left[\frac{1}{2}(r_{i,l}(k-j)^2 + r_{i,l}(k-h)^2)\right] \\ &= \frac{1}{2}(\sigma_{i,l}^2(k-j) + \sigma_{i,l}^2(k-h)) \end{aligned} \quad (16)$$

and that  $\sigma_{i,l}^2(k) \leq (\sigma_{i,l}^B(k))^2$  for all  $k \geq 1$ .

The second inequality is proven considering that (if  $h > j$  without loss of generality)  $\gamma_{i,l}(k-j, k-h) = \mathbf{C}_{i,l} \mathbb{E}[\mathbf{e}(k-j)(\mathbf{C}_{i,l} \mathbf{e}(k-h) + v_{i,l}(k-h))]$ , where  $v_{i,l}(k-h)$  is the  $l$ -th element of  $v_i(k-h)$  and  $\mathbf{C}_{i,l}$  is the line of  $\mathbf{C}$  which corresponds with the  $l$ -th entry of output  $y_i(k)$ . Iterating (7),  $\mathbf{e}(k-j) = \mathbf{F}^{h-j} \mathbf{e}(k-h) + \sum_{s=0}^{h-j-1} \mathbf{F}^{h-j-(s+1)} (\mathbf{w}(k-h+s) - \mathbf{L}\mathbf{v}(k-h+s))$ , and  $\gamma_{i,l}(k-j, k-h) = \mathbf{C}_{i,l} \mathbf{F}^{h-j} \mathbf{\Pi}(k-h) \mathbf{C}_{i,l}^T - \mathbf{C}_{i,l} \mathbf{F}^{h-j-1} \mathbf{L} \mathbb{E}[\mathbf{v}(k-h)v_{i,l}(k-h)]$ . It follows that  $\gamma_{i,l}(k-j, k-h) \leq \|\mathbf{C}_{i,l}\| \|\mathbf{F}^{h-j}\| \|\mathbf{\Pi}(k-h) \mathbf{C}_{i,l}^T\| + \|\mathbf{C}_{i,l}\| \|\mathbf{F}^{h-j-1}\| \|\mathbf{L}_i^c R_{i,l}^c\|$ .

In view of the block-diagonality of  $\mathbf{C}$ ,  $\mathbf{\Pi}(k-h) \mathbf{C}_{i,l}^T = (\Pi_{1i}(k-h) C_{i,l}^T, \dots, \Pi_{Mi}(k-h) C_{i,l}^T)$ , where  $C_{i,l}$  is the  $l$ -th row of  $C_i$ . The proof can be completed considering that, for all  $s = 1, \dots, M$ ,  $\Pi_{si}(k-h) C_{i,l}^T = \mathbb{E}[e_s(k-h)(C_{i,l} e_i(k-h))] = (\mathbb{E}[e_{s,1}(k-h)(C_{i,l} e_i(k-h))], \dots, \mathbb{E}[e_{s,n_s}(k-h)(C_{i,l} e_i(k-h))]) \leq \frac{1}{2}(\text{var}(e_{s,1}(k-h)) + \text{var}((C_{i,l} e_i(k-h)))) + \dots + \text{var}(e_{s,n_s}(k-h)) + \text{var}((C_{i,l} e_i(k-h)))$ .

## 4. TESTS ON THE PROPOSED FD METHODS

In this section, results of Montecarlo simulation tests on the academic example are illustrated. For each experiment,  $N_m = 1000$  Montecarlo runs have been performed, with a similar set-up with respect to the one in Section 2.4. Here we test and compare methods (a) and (b), the latter with both  $m = 20$  and  $m = 40$ . Only the value  $p_{\text{FA}} = 0.0002$  is considered here. For the FD method (b), the covariance term has been approximated using the analytical method described in Section 3.2 and, for comparison, using the sampled covariance. We denote:

- PLF( $m$ )-empirical: we set  $(\sigma_{i,l}^{(m)}(k))^2 = (\sigma_{i,l}^{(m),\text{EMP}}(k))^2$ , where

$$(\sigma_{i,l}^{(m),\text{EMP}}(k))^2 = \frac{1}{N-m+1} \sum_{k=m}^N \bar{r}_{i,l}^{(m)}(k) \bar{r}_{i,l}^{(m)}(k)^T \quad (17)$$

- PLF( $m$ )-analytical: we set  $(\sigma_{i,l}^{(m)}(k))^2 = (\sigma_{i,l}^{(m),B}(k))^2$  for all  $m$  and  $k$ .

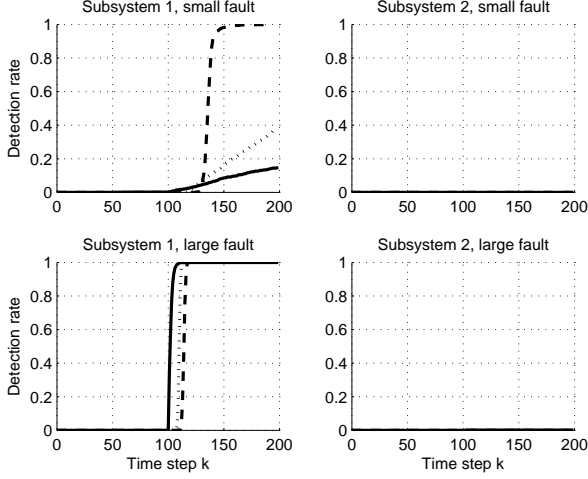


Fig. 3. Cumulative detection rates  $R_{\text{FD}}(k)$  with additive faults on subsystem 1 at  $T = 100$  and  $p_{\text{FA}} = 0.0002$ . Solid lines: PLF-analytical; dotted lines: PLF(20)-analytical; dashed lines: PLF(40)-analytical.

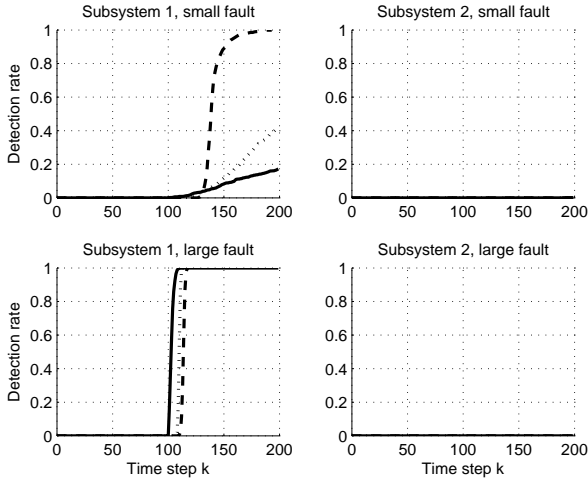


Fig. 4. Cumulative detection rates  $R_{\text{FD}}(k)$  with multiplicative faults on subsystem 1 at  $T = 100$  and  $p_{\text{FA}} = 0.0002$ . Solid lines: PLF-analytical; dotted lines: PLF(20)-analytical; dashed lines: PLF(40)-analytical.

The plots shown in Figures 3-6 show the cumulative fault detection rates  $R_{\text{FD}}(k)$  obtained in the described scenarios. We can draw the following considerations.

- The use of the FD method (b) allows to reduce the threshold values, for constant guaranteed FA rates. This makes it possible to detect, in a efficient and prompt way, faults of small amplitude, while guaranteeing acceptably small FA rates. The drawback of this approach consists of the fact that faults (espe-

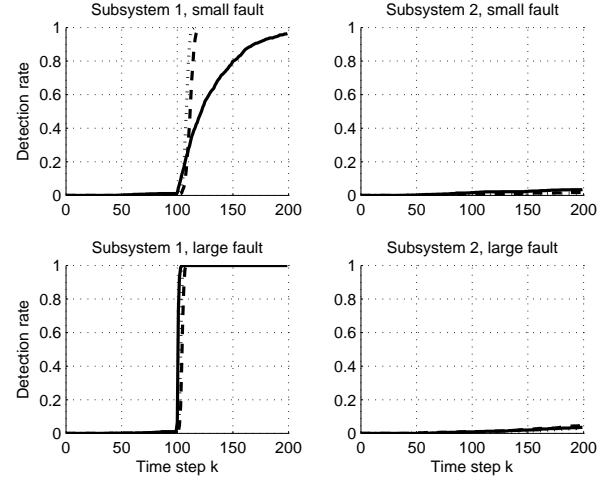


Fig. 5. Cumulative detection rates  $R_{\text{FD}}(k)$  with additive faults on subsystem 1 at  $T = 100$  and  $p_{\text{FA}} = 0.0002$ . Solid lines: PLF-empirical; dotted lines: PLF(20)-empirical; dashed lines: PLF(40)-empirical.

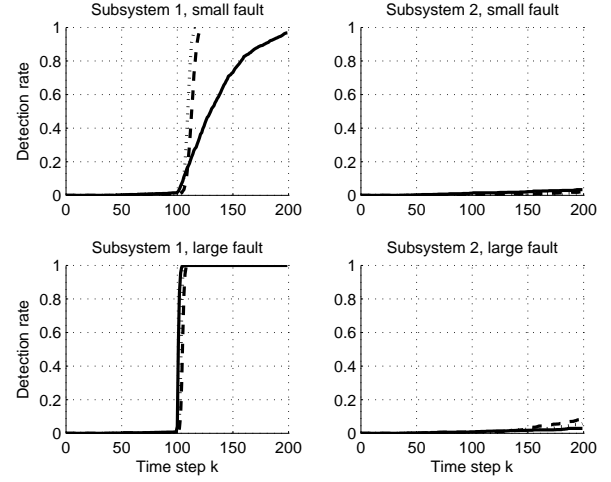


Fig. 6. Cumulative detection rates  $R_{\text{FD}}(k)$  with multiplicative faults on subsystem 1 at  $T = 100$  and  $p_{\text{FA}} = 0.0002$ . Solid lines: PLF-empirical; dotted lines: PLF(20)-empirical; dashed lines: PLF(40)-empirical.

cially with small amplitude) may be detected with some delay, consisting of the window length  $m$ .

- The covariance matrix approximations proposed in this paper allow to obtain conservative, but yet very satisfactory, fault detection performances. The use of the proposed approach is therefore quite promising.

A final remark is due. Additional simulation tests (not shown here for lack of space) have been performed in case of step-wise sensor faults (consisting of both additive and multiplicative perturbations in the output equation (1b)). In this case the FD method (b) discussed in this paper reveals not effective, with both empirical and analytically-obtained covariances. The reason of this fact, in our opinion, is that step-wise sensor faults have an instantaneous effect on  $r_i(t)$  (well detectable with single-point residual testing methods), which is asymptotically attenuated thanks to the observer filtering action, making methods based on averaged data less effective.

This consideration will pave the way to combined schemes using the two approaches at the same time, that can be used, in a fault isolation scenario, to distinguish between process and sensor faults.

## 5. CONCLUSIONS

In this paper we have highlighted (with reference to (Boem et al., 2016)) that FD approaches based on tests on single residuals may be prone to a restrictive tradeoff between false alarm and missed detection rates. To overcome this, we can resort to an approach, described in (Gertler, 1998), based on the analysis of moving averages of residuals. In this paper we provide a way to compute, in a distributed, analytical, and scalable way, the corresponding thresholds using the tools developed in (Boem et al., 2016).

Future research efforts will be devoted to the extension of the proposed approach to non-Gaussian distributions resorting to the two-sided Chebishev inequality and to the analysis of the feasibility and effectiveness of other approaches that focus on sliding windows of residual values. Finally, the developed method will be evaluated on more complex and realistic case studies.

## REFERENCES

- Basseville, M. and Nikiforov, I.V. (1998). *Detection of abrupt changes - Theory and applications*. Prentice-Hall, Inc.
- Blanke, M., Kinnaert, M., Lunze, J., and Staroswiecki, M. (2016). Distributed fault diagnosis and fault-tolerant control. In *Diagnosis and Fault-Tolerant Control*, 467–518. Springer.
- Boem, F., Carli, R., Farina, M., Ferrari-Trecate, G., and Parisini, T. (2016). Scalable monitoring of interconnected stochastic systems. In *IEEE 55th Conference on Decision and Control*, 1285–1290.
- Boem, F., Ferrari, R., Keliris, C., Parisini, T., and Polycarpou, M. (2017). A distributed networked approach for fault detection of large-scale systems. *IEEE Transactions on Automatic Control*, 62(1), 18–33.
- Farina, M. and Carli, R. (In press). Partition-based distributed kalman filter with plug and play features. *IEEE Transactions on Control of Network Systems*.
- Farina, M., Ferrari-Trecate, G., and Scattolini, R. (2010a). Moving-horizon partition-based state estimation of large-scale systems. *Automatica*, 46(5), 910–918.
- Farina, M., Scattolini, R., Garcia, J., Espinosa, J., and Rawlings, J. (2010b). Report on the state of the art in distributed state and variance estimation, and on preliminary results on disturbance modelling for distributed systems. Technical report, HD-MPC EU Project Deliverable D5.1. URL [http://www.ict-hd-mpc.eu/deliverables/hd\\_mpc.D\\_5\\_1.pdf](http://www.ict-hd-mpc.eu/deliverables/hd_mpc.D_5_1.pdf).
- Gertler, J.J. (1998). *Fault Detection and Diagnosis in engineering Systems*.
- Gupta, V. and Puig, V. (2016). Distributed fault diagnosis using minimal structurally over-determined sets: Application to a water distribution network. In *3rd Conference on Control and Fault-Tolerant Systems (SysTol)*, 811–818. IEEE.
- Horn, R. and Johnson, C.R. (2012). *Matrix Analysis*. Cambridge University Press.
- Isermann, R. (2005). Model-based fault-detection and diagnosis - status and applications. *Annual Reviews in Control*, 29(1), 71 – 85.
- Lan, J. and Patton, R. (2016). Decentralized fault estimation and fault-tolerant control for large-scale interconnected systems: An integrated design approach. In *UKACC 11th International Conference on Control*, 1–6. IEEE.
- Lauricella, M., Farina, M., Schneider, R., and Scattolini, R. (2017). A distributed fault detection and isolation algorithm based on moving horizon estimation. *IFAC-PapersOnLine*, 50(1), 15259 – 15264.
- Olfati-Saber, R. (2007). Distributed kalman filtering for sensor networks. In *46th IEEE Conference on Decision and Control*, 5492–5498.
- Reppa, V., Polycarpou, M.M., and Panayiotou, C.G. (2015). Distributed sensor fault diagnosis for a network of interconnected cyberphysical systems. *IEEE Transactions on Control of Network Systems*, 2(1), 11–23.
- Riverso, S., Boem, F., Ferrari-Trecate, G., and Parisini, T. (2016). Plug-and-play fault detection and control-reconfiguration for a class of nonlinear large-scale constrained systems. *IEEE Transactions on Automatic Control*, 61(12), 3963–3978.
- Samad, T. and Parisini, T. (2011). Systems of systems. In T. Samad and A.M. Annaswamy (eds.), *The Impact of Control Technology*, 175–183. IEEE Control Systems Society.
- Schneider, R., Hannemann-Tamás, R., and Marquardt, W. (2015). An iterative partition-based moving horizon estimator with coupled inequality constraints. *Automatica*, 61, 302 – 307.
- Shames, I., Teixeira, A.M., Sandberg, H., and Johansson, K.H. (2011). Distributed fault detection for interconnected second-order systems. *Automatica*, 47(12), 2757–2764.
- Sijs, J., Lazar, M., Bosh, P.V.D., and Papp, Z. (2008). An overview of non-centralized Kalman filters. *IEEE Conference on Decision and Control*, 739–744.
- Stanković, S., Stanković, M., and Stipanović, D. (2009). Consensus based overlapping decentralized estimation with missing observations and communication faults. *Automatica*, 45, 1397 – 1406.
- Uhlmann, J.K. (2003). Covariance consistency methods for fault-tolerant distributed data fusion. *Information Fusion*, 4, 201–215.
- Zhang, X. and Zhang, Q. (2012). Distributed fault diagnosis in a class of interconnected nonlinear uncertain systems. *International Journal of Control*, 85(11), 1644–1662.