Automatica 93 (2018) 211–223

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Distributed Pareto-optimal state estimation using sensor networks☆

Francesca Boem [a,*], Yilun Zhou [b], Carlo Fischione [c], Thomas Parisini [d,e,f]

[a] Department of Electronic and Electrical Engineering, University College, London, UK
[b] Huawei Technologies, China
[c] Automatic Control Department, KTH Royal Institute of Technology, Stockholm, Sweden
[d] Department of Electrical and Electronic Engineering, Imperial College, London, UK
[e] Department of Engineering and Architecture, University of Trieste, Italy
[f] KIOS Research and Innovation Centre of Excellence, University of Cyprus, Cyprus

ARTICLE INFO

Article history:
Received 18 January 2017
Received in revised form 15 December 2017
Accepted 27 February 2018

Keywords:
Distributed
State estimation
Prediction
Sensor
Networks
Optimal estimation

ABSTRACT

A novel model-based dynamic distributed state estimator is proposed using sensor networks. The estimator consists of a filtering step – which uses a weighted combination of information provided by the sensors – and a model-based predictor of the system's state. The filtering weights and the model-based prediction parameters jointly minimize – at each time-step – the bias and the variance of the prediction error in a Pareto optimization framework. The simultaneous *distributed* design of the filtering weights and of the model-based prediction parameters is considered, differently from what is normally done in the literature. It is assumed that the weights of the filtering step are in general unequal for the different state components, unlike existing consensus-based approaches. The state, the measurements, and the noise components are allowed to be individually correlated, but no probability distribution knowledge is assumed for the noise variables. Each sensor can measure only a subset of the state variables. The convergence properties of the mean and of the variance of the prediction error are demonstrated, and they hold both for the global and the local estimation errors at any network node. Simulation results illustrate the performance of the proposed method, obtaining better results than state of the art distributed estimation approaches.

1. Introduction

One of the fundamental applications of sensor networks is to estimate and track the state of targets or processes that are evolving in the sensing field. Useful in many monitoring scenarios, such as for example, target tracking and environment and agriculture monitoring, in sensor networks the estimations have to be distributed at each sensor node. In this paper,[1] we address the problem of distributed state estimation and prediction over sensor networks in a multi-objective optimization framework.

Given their importance, distributed estimators have been the subject of many investigations in the area of networked control (see, as example, Christofides, Scattolini, de la Pena, & Liu, 2013; Farina, Ferrari-Trecate, & Scattolini, 2010; Garin & Schenato, 2010) and distributed fault diagnosis (Boem, Ferrari, Parisini, & Polycarpou, 2011; Franco, Olfati-Saber, Parisini, & Polycarpou, 2006), among others. Generally, in these papers it is assumed that distributed estimation works according to the following procedure (Ding, Wang, & Shen, 2014; Garin & Schenato, 2010): each node in the network locally estimates the state of a common dynamic system; then, it communicates measurements and estimates only to neighboring nodes, and filters the measurements by taking a linear combination of its own and neighboring's measurements and predictions; finally, each node uses the current

☆ This work has been partially supported by the 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) funded by the RCUK Energy Programme (contract no: EP/L014343/1). The material in this paper was partially presented at the 51st IEEE Conference on Decision and Control, December 10-13, 2012, Maui, Hawaii, the 12th European Control Conference, July 17-19 2013, Zurich, Switzerland, and the 14th annual European Control Conference, July 15-17, 2015, Linz, Austria. This paper was recommended for publication in revised form by Associate Editor Erik Weyer under the direction of Editor Torsten Söderström.

* Corresponding author.
E-mail addresses: francesca.l.boem@gmail.com (F. Boem), zhouyilun1989@gmail.com (Y. Zhou), carlofi@kth.se (C. Fischione), t.parisini@gmail.com (T. Parisini).

https://doi.org/10.1016/j.automatica.2018.03.071
0005-1098/

[1] See Boem, Xu, Fischione, and Parisini (2012, 2013, 2015) for some preliminary results: in Boem et al. (2012) signal estimation is considered, while in Boem et al. (2015) the entire state is assumed to be measurable by each node.

filtered measurements to implement a model-based predictor, smoothing the previous prediction error. However, there are several aspects in this general procedure that have not yet been fully considered.

(1) The first important aspect pertains the number of accessible states. Due to geographic nodes distributions, technological constraints, etc., it can happen that, although the overall network observes the entire state, each node measures only a subset of the variables forming the overall state. We refer to this case as *partially-measurable state*. Most of the existing results have been obtained under the assumption of complete measurement information, thereby bringing much conservatism in applications (see Ding et al., 2014 for a survey about distributed filtering over sensor networks). However, if the state components are correlated, then a node could still in principle perform an accurate estimation of the state components it has not directly access to. How to perform estimation and prediction of the overall state at each node, despite partial measurement and incomplete information, has been investigated, for example, in Khan, Kar, Jadbabaie, and Moura (2010), Stanković, Stanković, and Stipanović (2009), Wu, Jia, Johansson, and Shi (2013) and Zhang, Feng, and Yu (2012).

(2) A second important aspect concerns the possible presence of bias in the measurements, an important aspect often neglected in existing approaches. Due to measurement errors, model uncertainties, and message losses, the estimates are in practice affected by bias. The bias leads to unknown statistical distribution of the estimation error. If the bias of the estimators is not considered, it may grow unbounded. Nevertheless, the performance criterion of the estimators in the literature is essentially based only on the variance of the estimation error (see Caballero-Águila, García-Garrido, & Linares-Pérez, 2014; Meng & Chen, 2014; Speranzon, Fischione, Johansson, & Sangiovanni-Vincentelli, 2008; Yang, Chen, Wang, & Shi, 2014 as examples), which leads to poor performance of the distributed estimation process when biases are present. Therefore, when designing distributed estimators, we face at least two indicators of the quality of the estimators: the mean of the estimation error (bias) and the variance. To the best of our knowledge, in this paper we present the first approach in which these two indicators are *simultaneously* taken into account. Specifically, we simultaneously minimize both the mean (the bias) and the variance of the global prediction error by posing a multi-objective Pareto optimization problem that can be solved in a distributed way by each sensor without a centralized coordination.

(3) A third important aspect is related to the fact that in the literature the filtering phase and the prediction phase are designed independently, e.g., Alriksson and Rantzer (2006), Carli, Chiuso, Schenato, and Zampieri (2008) and Stanković et al. (2009) for the sake of tractability and ease of implementation. This separation may lead to suboptimal solutions. Instead, in the paper the filtering weights and model-based prediction parameters are allowed to be time-varying and are jointly optimized by each sensor at each step, thus paving the way to improved prediction schemes compared to the state of the art.

(4) A fourth important aspect is the instantaneous performance compared to the asymptotic one. Although distributed estimators may asymptotically perform well, in the transient bias and variance of the estimation error may take on unacceptably large values. In the proposed approach – even if we show that the asymptotic convergence is achieved like in well-known Kalman-based approaches (see Olfati-Saber, 2009) – the bias and the variance of the prediction error are jointly optimized at each time-step thus showing good instantaneous performance. Convex sufficient conditions to guarantee asymptotic convergence of the estimation error mean are derived. Furthermore, the proposed approach only assumes the knowledge of the mean and variance of the process and measurement noises without need of any further assumption on their probability distributions.

To sum up, the proposed distributed estimation technique is characterized by the following main features:

(1) only a subset of the state variables are measured at each node;
(2) the mean and the variance of the estimation error are jointly minimized via Pareto optimization;
(3) the filtering and the prediction steps are jointly designed;
(4) optimized performance at each estimation step and asymptotic convergence of the estimation error mean;
(5) knowledge of mean and variance of process and measurement noises only is required.

In the following, we further elaborate on the original contributions brought about by these characteristics with respect to the literature.

1.1. State of the art

Distributed Kalman Filtering is an active area of research, see, e.g., Ding et al. (2014) and Mahmoud and Khalid (2013), where a survey about distributed filtering methods over sensor networks and distributed Kalman filtering methods are presented, respectively. Unlike Kalman filtering approaches (such as Olfati-Saber, 2009), in our study no Gaussian assumptions on the probability distribution of the measurement and modeling noises are made. Instead, we assume knowledge of the mean and covariance matrix of the noise components, without these being necessarily Gaussian. When the estimation problem we are considering is solved by a centralized approach, the Kalman filter is optimal under Gaussian assumption on the noises, and represents the best linear filter also when disturbances are non-Gaussian (Davis, 1977). However, the scenario we are facing is more challenging because we consider a distributed case, where the prediction is computed locally without the coordination of a central agent, differently from Deshmukh, Natarajan, and Pahwa (2014).

Besides distributed Kalman filtering, roughly two different approaches have been proposed to the problem of distributed state estimation and prediction. First, the approaches based on diffusion strategies, such as the ones proposed in Cattivelli and Sayed (2010) and Speranzon et al. (2008), where the diffusion of the local estimations across neighbors is applied after incremental update. These are in contrast with the second approach: consensus strategies, used, e.g., in Spanos, Olfati-Saber, and Murray (2005), where consensus is applied to obtain average observations or estimations at each filtering step. Finally, Kalman-Consensus filtering approaches have been designed (see Olfati-Saber, 2009 as example) with the objectives of estimating the state of the system and reaching a consensus with neighboring estimator agents on the estimate.

In this paper, we consider a multi-objective optimization case and we simultaneously take into account both the mean and the variance of the prediction error. This is in contrast to Caballero-Águila et al. (2014), Carli et al. (2008), Meng and Chen (2014), Speranzon et al. (2008) and Yang et al. (2014), where only minimum variance solution is studied, and from Stanković et al. (2009), where the consensus parameters minimize the steady-state mean-square prediction error. In Mitra and Sundaram (2016), Park and Martins (2017) and Wang and Morse (2017) distributed observers are designed for the case where the state is only partially observable by each sensor, but the estimation weights are designed to guarantee convergence and state omniscience properties, not optimizing bias and error variance features. In Khan et al. (2010), the considered problem for distributed estimation is similar, dealing with the design of the consensus parameters and local innovation gains to optimize a different performance criterion. Differently from the proposed method, in Khan et al. (2010) a special case is considered allowing to reformulate the problem so to obtain a

scalar gain estimator and a single-objective optimization problem with a single scalar constant decision variable is then analyzed. On the other hand, in this paper we consider at each time-step a multi-objective optimization, where the decision variables are time-varying matrices. Compared to Alriksson and Rantzer (2006), Carli et al. (2008) and Olfati-Saber (2009), we do not assume that the distribution of the disturbance is known. We only assume to know the mean and variance of such disturbance, differently from Battistelli and Chisci (2014), where it is necessary to know the PDFs of measurement noises and modeling uncertainties.

Finally, an important aspect of our study is that we consider the multi-dimensional state estimation (unlike Carli et al., 2008 that is for the scalar case), by taking into account correlations between the different components of the state and of the noises, which is a major analysis challenge.

The rest of the paper is organized as follows. In Section 2, we introduce the problem formulation. We describe the distributed dynamic estimation method in Section 3. We derive some convergence conditions in Section 3.3 and we then formulate in Section 4 the multi-objective optimization problem to select the filtering-coefficients and the prediction parameters. The optimal solution is derived in Section 4.2. Finally, simulation results and concluding remarks are provided in Sections 6 and 7.

Notation. Given a stochastic variable x, we represent as $\mathbb{E}x$ its expected value. By $\mathbf{1}_s$ and I_s we denote the vector $(1, \ldots, 1)^\top$ and the identity matrix with appropriate size s, respectively. Given a matrix M, we denote $\mathrm{diag}(M)$ the vector collecting the diagonal elements of M. Finally, $\otimes$ denotes the Kronecker product and the operator $\circ$ represents the component-by-component product.

2. Problem formulation

We aim at computing the one-step ahead prediction $\hat{x}(t+1)$ of the state of a linear stochastic system described as

$$x(t+1) = Ax(t) + w(t), \tag{1}$$

where t is the discrete time, $x \in \mathbb{R}^m$ denotes the state vector and $w \in \mathbb{R}^m$ is modeled as $w(t) = \bar{w}(t) + \tilde{w}(t)$, with $\bar{w}(t)$ denoting a known time-varying bias possibly including some known non-linearities, while $\tilde{w}(t)$ models uncertainties and process noises.

Assumption 1. We assume that $\tilde{w}(t)$ is a zero-mean process with covariance matrix $\Sigma_w(t)$. ◁

The state prediction $\hat{x}(t+1)$ is computed in a distributed way by a *sensor network*, made of n nodes, that monitors system (1) by taking measurements at each time-instant t. More specifically, each sensor knows the model (1) of the monitored system and may measure directly some, possibly not all, state components, so that, for each sensor node i, with $i = 1, \ldots, n$, we have

$$y_i(t) = C_i x(t) + v_i(t), \tag{2}$$

where $y_i \in \mathbb{R}^{p_i}$, with $p_i \leq m$, denotes the measurements vector taken by sensor i, $v_i \in \mathbb{R}^{p_i}$ is a zero-mean measurement noise, with Σ_{v_i} as covariance matrix, and $C_i \in \mathbb{R}^{p_i \times m}$ is the output matrix, each row having a single element equal to 1 in correspondence with the measured state component, and 0 otherwise. Since each sensor may measure directly one or more components of the state vector, the matrix C_i is defined so that $C_i \mathbf{1}_m = \mathbf{1}_{p_i}$.

Assumption 2. The measurement noise v_i and process disturbances w are not correlated, for each $i = 1, \ldots, n$. ◁

Each node of the network exchanges measurements and predictions only with neighboring nodes. The associated communication network is modeled as a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of the nodes and $\mathcal{E}$ is the set of the edges connecting communicating nodes. We denote $\mathcal{N}_i = \{j \in \mathcal{V} : (j, i) \in \mathcal{E}\} \cup \{i\}$ the set of neighbors of node $i \in \mathcal{V}$ plus the node itself.

Considering the entire sensor network, the global measurement output equation becomes

$$y(t) = Cx(t) + v(t),$$

with $y,\ v \in \mathbb{R}^{p_E}$, $C \in \mathbb{R}^{p_E \times m}$, $p_E = \sum_{j \in 1,\ldots,n} p_j$, where vectors and matrices collect all the local vectors y_j and v_j and matrices C_j, $j = 1, \ldots, n$, of the sensor network. The following further assumption is needed

Assumption 3. The graph $\mathcal{G}$ is partitioned into strongly connected sub-networks of sensors $\tilde{\mathcal{G}} \subseteq \mathcal{G}$, characterized by the output equation $\tilde{y}(t) = \tilde{C}x(t) + \tilde{v}(t)$, where $\tilde{y},\ \tilde{v} \in \mathbb{R}^{\tilde{p}}$, $\tilde{C} \in \mathbb{R}^{\tilde{p} \times m}$, $\tilde{p} = \sum_{j \in \tilde{\mathcal{G}}} p_j$, collect in a column all the local vectors y_j and v_j and matrices C_j, respectively, for $j \in \tilde{\mathcal{G}}$. For each strongly connected sub-network $\tilde{\mathcal{G}}$, we assume that $\tilde{C}$ satisfies the following conditions: $\tilde{C}\mathbf{1}_m = \mathbf{1}_{\tilde{p}}$ and $\tilde{C}^\top \mathbf{1}_{\tilde{p}} \geq \mathbf{1}_m$. ◁

The meaning of Assumption 3 is: each row of $\tilde{C}$ has one entry equal to 1 all the other entries being set to "0", i.e. each measurement refers to a state component, and each state component is directly measured by at least one node in each strongly connected sub-network.

Assumption 3 is a sufficient condition for global observability from each strongly connected sub-network of sensors $\tilde{\mathcal{G}} \subset \mathcal{G}$, i.e. it implies that each pair $A, \tilde{C}$ is completely observable. Note that this is less restrictive than requiring local observability from each sensor; in fact, in general the pairs (A, C_i) may be not observable for all i. This assumption allows to write a simple expression for the filtering estimation error (6) and to analytically formulate and solve the optimization Problem in Section 4.

3. Distributed state dynamic estimation

In the proposed state estimator, each node i of the sensor network implements a two steps dynamic estimator: a filtering-merging step and a prediction step.

First, by communicating with neighboring nodes, the estimator at the ith node filters the measurement noise in a consensus-like fashion by computing a linear combination of its own and neighboring available measurements and predictions:

$$\bar{x}_i(t) = \sum_{j \in \mathcal{N}_i} k_{i,j}(t)\hat{x}_j(t) + \sum_{j \in \mathcal{N}_i} h_{i,j}(t) C_j^\top y_j(t), \tag{3}$$

where $k_{i,j}(t)$ and $h_{i,j}(t) \in \mathbb{R}^{m \times m}$ are time-varying filter weights diagonal matrices that we intend to design. The term $C_j^\top$ guarantees that the available measurements are used for each node. By communicating with neighbors, each node shares its available information and spreads it through the network. The objective of this first step is for each node to reduce its own measurement uncertainty, *without the use of centralized coordination*. Sensors are allowed to communicate only once per time-step.

After the filtering-merging step, each node implements a model-based one-step-ahead prediction, using filtered measurements (obtained from the first step):

$$\hat{x}_i(t+1) = A\bar{x}_i(t) + \bar{w}(t) + \lambda_i'(t)(\hat{x}_i(t) - \bar{x}_i(t)), \tag{4}$$

where $\lambda_i'(t) \in \mathbb{R}^{m \times m}$ is a matrix collecting time-varying filter parameters that will be designed later on in the paper. The term $\lambda_i'(t)(\hat{x}_i(t) - \bar{x}_i(t))$ represents a local correction of the previous

prediction error, taking into account the system dynamics. The estimates are initialized with $\bar{x}_i(0) = C_i^\top y_i(0)$ and $\hat{x}_i(0) = \bar{x}_i(0)$.

To facilitate the analysis, a compact vector form is introduced considering all the nodes in the sensor network:

$$\begin{aligned} \bar{x}(t) &= K(t)\hat{x}(t) + H(t)C_E^\top y(t) \\ \hat{x}(t+1) &= A_E\bar{x}(t) + \bar{w}_E(t) + \lambda'(t)(\hat{x}(t) - \bar{x}(t)), \end{aligned} \tag{5}$$

where $\bar{x}, \hat{x} \in \mathbb{R}^{mn\times 1}$ and $y \in \mathbb{R}^{p_E\times 1}$ are column vectors collecting the local vectors $\bar{x}_i$, $\hat{x}_i$ and y_i, respectively, with $i = 1, \ldots, n$; A_E is a diagonal block matrix, with each block on the diagonal equal to A, and $\bar{w}_E$ is a column vector of appropriate dimension, where the process disturbance vector $\bar{w}$ is repeated n times. $K(t)$ and $H(t)$ are block matrices, where each (i, j)th block, with $i = 1, \ldots, n$ and $j = 1, \ldots, n$, is a diagonal matrix. The (i, j)th block is a null matrix if i and j are not neighboring nodes, whereas it collects the coefficients with which the ith nodes weights measurements or estimates components developed by the jth node, if they are neighbors. C_E and λ' are block matrices having the matrices C_i, λ_i' respectively, on the diagonal, with $i = 1, \ldots, n$.

As already mentioned, the main goal of the paper is to devise a local design principle of the time-varying weights $H(t)$, $K(t)$ and $\lambda'(t)$ in order to minimize bias and variance of the global prediction error at each time-instant t. In particular, we aim at establishing whether the filtering weights $H(t)$ and $K(t)$ can be designed separately with respect to the prediction weight $\lambda'(t)$. We address this issue in the following subsections.

3.1. Estimation error

Let us introduce the filtering error $\bar{e}(t) = \bar{x}(t) - x_E(t)$ and the prediction error $\hat{E}(t) = \hat{x}(t) - x_E(t)$, with x_E being a column vector of appropriate length, where the process state vector x is repeated n times. The following condition is introduced for ease of computation:

Condition 3.1. $(K(t) + H(t)C_E^\top C_E)\mathbf{1}_{mn} = \mathbf{1}_{mn}$. ◁

Thanks to Assumption 3, Condition 3.1 is not restrictive since $C_E^\top C_E$ is a diagonal matrix with ones and zeros on the diagonal. The filtering estimation error can be computed as

$$\bar{e}(t) = K(t)\hat{E}(t) + H(t)C_E^\top v(t), \tag{6}$$

where we use Condition 3.1 so that $H(t)C_E^\top C_E x_E(t) + K(t)x_E(t) = x_E(t)$. For the sake of simplicity, we express the prediction parameter as $\lambda'(t) = A_E\lambda(t)$, with $\lambda(t)$ being a diagonal matrix. Consequently,

$$\begin{aligned} \hat{E}(t+1) = {} & A_E\left[(I - \lambda(t))K(t) + \lambda(t)\right]\hat{E}(t) \\ & + A_E(I - \lambda(t))H(t)C_E^\top v(t) - \tilde{w}_E(t). \end{aligned} \tag{7}$$

Let us compute mean and variance of the global estimation and prediction errors at time t. The expected values are

$$\mathbb{E}\bar{e}(t) = K(t)\mathbb{E}\hat{E}(t), \tag{8}$$

$$\mathbb{E}\hat{E}(t+1) = A_E\left[(I - \lambda(t))K(t) + \lambda(t)\right]\mathbb{E}\hat{E}(t). \tag{9}$$

The covariance matrices can be computed, with respect to the stochastic variable $\bar{e}(t)$ at time t, as

$$\begin{aligned} & \mathbb{E}\left[(\bar{e}(t) - \mathbb{E}\bar{e}(t))(\bar{e}(t) - \mathbb{E}\bar{e}(t))^\top\right] \\ & = K(t)\Gamma_{\hat{E}}(t)K(t)^\top + H(t)C_E^\top\Sigma_v(t)C_E H(t)^\top, \end{aligned} \tag{10}$$

where $\Gamma_{\hat{E}}(t) = \mathbb{E}\left[(\hat{E}(t) - \mathbb{E}\hat{E}(t))(\hat{E}(t) - \mathbb{E}\hat{E}(t))^\top\right]$ and $\Sigma_v(t) = \mathbb{E}\left[(v(t) - \mathbb{E}v(t))(v(t) - \mathbb{E}v(t))^\top\right]$, noting that the covariance between the two vectors of stochastic variables $\hat{E}(t)$ and $v(t)$ computed at time t is $\mathrm{Cov}(\hat{E}(t), v(t)) = 0$, since $\hat{E}(t)$ depends on the state $x(t)$ (not depending on $v(t)$) and the prediction $\hat{x}(t)$ which is a deterministic quantity at time t, being computed at time $t-1$ based on the measurements and estimates available at time $t-1$. We can thus compute the covariance matrix $\Gamma_{\hat{E}}(t+1)$ as

$$\begin{aligned} & \mathbb{E}\left[(\hat{E}(t+1) - \mathbb{E}\hat{E}(t+1))(\hat{E}(t+1) - \mathbb{E}\hat{E}(t+1))^\top\right] \\ & = W_1(t)\Gamma_{\hat{E}}(t)W_1(t)^\top + W_2(t)\Sigma_v(t)W_2(t)^\top + \Sigma_{w_E}(t), \end{aligned} \tag{11}$$

with $W_1(t) = A_E\left[(I - \lambda(t))K(t) + \lambda(t)\right]$,

$W_2(t) = A_E(I - \lambda(t))H(t)C_E^\top$,

$\Sigma_{w_E}(t) = \mathbb{E}\left[(w_E(t) - \mathbb{E}w_E(t))(w_E(t) - \mathbb{E}w_E(t))^\top\right]$,

being $\mathrm{Cov}(\hat{E}(t), v(t)) = 0$, $\mathrm{Cov}(\hat{E}(t), \tilde{w}_E(t)) = 0$ and $\mathrm{Cov}(v(t), \tilde{w}_E(t)) = 0$, owing to Assumption 2.

3.2. Local estimation and prediction errors

In this section, we show that the global estimation and prediction errors can be computed in a distributed way. In the following we derive the local expressions for the estimation and prediction errors, which are needed to design the optimal estimator in a distributed way. Each node computes *locally* a filtered estimation and a model-based prediction as follows:

$$\begin{aligned} \bar{x}_i(t) &= \kappa_i(t)\hat{x}_i^{\mathrm{reg}}(t) + \eta_i(t)C_{iE}^\top y_i^{\mathrm{reg}}(t) \\ \hat{x}_i(t+1) &= A\bar{x}_i(t) + \bar{w}(t) + A\lambda_i(t)(\hat{x}_i(t) - \bar{x}_i(t)), \end{aligned} \tag{12}$$

where $\hat{x}_i^{\mathrm{reg}}$ and y_i^{reg} are two column vectors collecting the prediction and the measurements vectors respectively available at node i, ordered according to their indexes. Moreover, $\kappa_i(t)$ and $\eta_i(t)$ are the time-varying block matrices corresponding to the non-zero matrices related to the ith node of matrices $K(t)$ and $H(t)$ respectively; $\lambda_i(t)$ is a diagonal matrix collecting the local components of matrix $\lambda(t)$. C_{iE} is a block matrix having on the diagonal the matrices C_j, with $j \in \mathcal{N}_i$.

Let us now define the local filtering error $\bar{e}_i(t) = \bar{x}_i(t) - x(t)$ and the local prediction error $\hat{E}_i(t) = \hat{x}_i(t) - x(t)$. They can be computed similarly as their global form as

$$\bar{e}_i(t) = \kappa_i(t)\hat{\epsilon}_i(t) + \eta_i(t)C_{iE}^\top v_{\epsilon_i}(t), \tag{13}$$

$$\begin{aligned} \hat{E}_i(t+1) = {} & A(I - \lambda_i(t))\kappa_i(t)\hat{\epsilon}_i(t) + A\lambda_i(t)\hat{E}_i(t) \\ & + A(I - \lambda_i(t))\eta_i(t)C_{iE}^\top v_{\epsilon_i}(t) - \tilde{w}(t), \end{aligned} \tag{14}$$

where $\hat{\epsilon}_i$ and v_{ϵ_i} collect the prediction error and the measurement noise vectors, respectively, related to the measurements available at node i, both ordered following their indexes. Now, the expressions of the bias and the variance for the local estimation and prediction errors can be computed: $\mathbb{E}\bar{e}_i(t) = \kappa_i(t)\mathbb{E}\hat{\epsilon}_i(t)$ and

$$\mathbb{E}\hat{E}_i(t+1) = A[(I - \lambda_i(t))\kappa_i(t) + \lambda_i^0(t)]\mathbb{E}\hat{\epsilon}_i(t), \tag{15}$$

being $\lambda_i^0(t)$ a $m \times mN_i$ block matrix having the sub-matrix $\lambda_i(t)$ in the block position corresponding to the ith index in the neighboring set $\mathcal{N}_i$, and all the other blocks components equal to 0. A cumbersome algebra gives the expression of the variance as

$$\begin{aligned} & \mathbb{E}[(\hat{E}_i(t+1) - \mathbb{E}\hat{E}_i(t+1))(\hat{E}_i(t+1) - \mathbb{E}\hat{E}_i(t+1))^\top] \\ & = W_{1i}(t)\Gamma_{\hat{\epsilon}_i}(t)W_{1i}(t)^\top + W_{2i}(t)\Sigma_{v_{\hat{\epsilon}_i}}W_{2i}(t)^\top + \Sigma_w(t), \end{aligned} \tag{16}$$

where

$$W_{1i}(t) = A[(I - \lambda_i(t))\kappa_i(t) + \lambda_i^0(t)], \tag{17}$$

$$W_{2i}(t) = A(I - \lambda_i(t))\eta_i(t)C_{iE}^\top, \tag{18}$$

$$\Gamma_{\hat{\epsilon}_i}(t) = \mathbb{E}[(\hat{\epsilon}_i(t) - \mathbb{E}\hat{\epsilon}_i(t))(\hat{\epsilon}_i(t) - \mathbb{E}\hat{\epsilon}_i(t))^\top] \tag{19}$$

and $\Sigma_{v_{\hat{\epsilon}_i}}$ is the measurement noise covariance matrix, including correlations between neighboring sensors.

3.3. Stability of the estimation error

In this subsection, we show that the mean of the global prediction error given by (9) converges to zero under some local sufficient conditions on the time-varying estimation weights[2] κ_i, η_i and λ_i.

Proposition 1. *Consider the global prediction error mean given by (9). The following local conditions are sufficient to guarantee its asymptotic stability. For each lth row of each ith block row of the global matrices K, H and λ, with $i = 1, \ldots, n$ and $l = 1, \ldots, m$:*

$$\sum_{j=1}^{n} \left|k_{i,j}^l\right| < \frac{1}{\|A\|_\infty} \quad \textit{AND} \tag{20}$$

– if $1/\|A\|_\infty > 1$,

$$\frac{-\frac{1}{\|A\|_\infty} + \sum_{j=1}^n \left|k_{i,j}^l\right|}{1 + \sum_{j=1}^n \left|k_{i,j}^l\right|} < \lambda_i^l < \frac{\frac{1}{\|A\|_\infty} + \sum_{j=1}^n \left|k_{i,j}^l\right|}{1 + \sum_{j=1}^n \left|k_{i,j}^l\right|}; \tag{21}$$

– if $1/\|A\|_\infty = 1$,

$$\frac{-\frac{1}{\|A\|_\infty} + \sum_{j=1}^n \left|k_{i,j}^l\right|}{1 + \sum_{j=1}^n \left|k_{i,j}^l\right|} \leq \lambda_i^l < 1; \tag{22}$$

– if $1/\|A\|_\infty < 1$,

$$\frac{-\frac{1}{\|A\|_\infty} + \sum_{j=1}^n \left|k_{i,j}^l\right|}{1 + \sum_{j=1}^n \left|k_{i,j}^l\right|} < \lambda_i^l < 1 - \frac{1 - \frac{1}{\|A\|_\infty}}{1 - \sum_{j=1}^n \left|k_{i,j}^l\right|}. \tag{23}$$

Proof. Eq. (9) represents the dynamics of a linear time-varying system. Asymptotic stability is ensured (Bauer, Premaratne, & Duran, 1993) if there exists a finite $k > 0$ such that

$$\left\| \prod_{t=1}^{k} A_E\left[(I - \lambda(t))K(t) + \lambda(t)\right] \right\|_\infty < 1. \tag{24}$$

This is implied by the satisfaction, at each t, of the condition $\|A_E\left[(I - \lambda(t))K(t) + \lambda(t)\right]\|_\infty < 1$. Recall that $\lambda(t)$ and $I - \lambda(t)$ are two diagonal matrices; by definition $K(t)$ is a block matrix, where each block is a diagonal matrix; therefore also $(I-\lambda(t))K(t)+\lambda(t)$ is a block matrix with all diagonal blocks. Using the submultiplicative property of the norm, the convergence condition is satisfied if, for each lth row of each ith block row of matrix $(I - \lambda(t))K(t) + \lambda(t)$, with $i = 1, \ldots, n$ and $l = 1, \ldots, m$, we have

$$\left|1 - \lambda_i^l\right| \sum_{j=1}^{n} \left|k_{i,j}^l\right| + \left|\lambda_i^l\right| < \frac{1}{\|A\|_\infty}, \tag{25}$$

since $\|A_E\|_\infty = \|A\|_\infty$. The analysis of (25) in all possible different scenarios concerning the value of λ_i^l immediately leads to prove the statement of the proposition. □

Note that different less conservative convergence conditions may be derived, but the sufficient conditions proposed in Proposition 1 have the notable advantage that they can be computed in a distributed way, using only local information available at each time t and so they can be used for the on-line computation of the time-varying prediction weights.

[2] For the sake of notation simplicity, we omit the dependence on t of the matrices.

4. Filtering weights and prediction parameters optimization

The goal of the proposed distributed method is to predict the state minimizing the bias and variance of the global prediction error at each time-instant. To do that, we propose that each sensor at each step computes the local optimal time-varying weights by solving a multi objective optimization problem, where the first objective is the squared bias, given in (15), and the second objective is the variance of the estimation error given in (16). Since these two objectives are convex[3] in the decision variables (namely the filtering coefficients and the prediction parameter), then we can consider the trace of the multi-objective optimization problem (Boyd & Vandenberghe, 2004) and pose the following Pareto optimization problem:

$$\min_{\kappa_i, \eta_i, \lambda_i} \ \text{tr}\left[\rho_i B_i^2 + (1 - \rho_i)V_i\right] \tag{26a}$$

$$\text{s.t. } (\kappa_i(t) + \eta_i(t)C_{iE}^\top C_{iE})\mathbf{1}_{m_i} = \mathbf{1}_m, \tag{26b}$$

$$\text{Eq. (20)} \wedge ((21) \vee (22) \vee (23)), \tag{26c}$$

where $m_i = mN_i$, $0 \leq \rho_i \leq 1$ is the Pareto parameter, $B_i = \mathbb{E}\hat{E}_i(t+1)$ is the prediction error bias given in (15), $V_i = \mathbb{E}[(\hat{E}_i(t+1) - \mathbb{E}\hat{E}_i(t+1))(\hat{E}_i(t+1) - \mathbb{E}\hat{E}_i(t+1))^\top]$ is the variance of the prediction error given in (16). The first constraint is the local Condition 3.1; the other constraints are the convergence conditions derived in Section 3.3, depending on matrix A norm.

Since the mean and the variance of the local prediction error can be computed in a distributed way as shown in Section 3.2, then the minimization of mean and variance of the local prediction error implies the minimization of mean and variance of the global prediction error, noting that the trace of the global objective function is equivalent to the sum of the local objective functions. Note that (21), (22), (23) are non convex expressions which complicate the derivation of the solution for what concerns the use of necessary optimality conditions. We now analyze problem (26) to obtain a convex approximation problem that can be solved in an optimal way.

4.1. Constraints approximation

We start by analyzing the convergence condition (20). Since the absolute value would make the problem more difficult to solve, we substitute (20) with the following more restrictive conditions:

$$\begin{cases} \kappa_i^\top(t)\mathbf{1}_m \geq \mathbf{0}_{m_i} \\ \kappa_i(t)\mathbf{1}_{m_i} < \dfrac{1}{\|A\|_\infty}\mathbf{1}_m. \end{cases} \tag{27}$$

Moreover, as regards the expressions (21), (22), (23), we note that if (20) holds, then

$$\frac{-\frac{1}{\|A\|_\infty} + \sum_{j=1}^n \left|k_{i,j}^l\right|}{1 + \sum_{j=1}^n \left|k_{i,j}^l\right|} < 0;$$

if, in addition, $1/\|A\|_\infty > 1$, then

$$\frac{\frac{1}{\|A\|_\infty} + \sum_{j=1}^n \left|k_{i,j}^l\right|}{1 + \sum_{j=1}^n \left|k_{i,j}^l\right|} > 1$$

while if $1/\|A\|_\infty < 1$, then

$$0 < 1 - \frac{1 - \frac{1}{\|A\|_\infty}}{1 - \sum_{j=1}^n \left|k_{i,j}^l\right|} < 1.$$

[3] The objective functions in (26) and (30) are convex because they are sum of quadratic terms and matrices M_i, M_i^{loc} and S_i are positive definite. Convexity of quadratic functions is discussed in Boyd and Vandenberghe (2004).

We use the above relations to approximate the sufficient convergence conditions for the definition of convex constraints. Using the element-wise inequality, we obtain the following conditions, implying Eqs. (21), (22), (23):
– if $1/\|A\|_\infty > 1$,

$$0 \leq \lambda_i \leq I, \tag{28}$$

– if $1/\|A\|_\infty \leq 1$,

$$0 \leq \lambda_i < I. \tag{29}$$

We then formulate a convex approximation of (26):

$$\begin{aligned}
&\min_{\kappa_i(t),\eta_i(t),\lambda_i(t)} \operatorname{tr}\,[A(I-\lambda_i(t))\kappa_i(t)M_i(t,\rho_i)\kappa_i(t)^\top \cdot \\
&(I-\lambda_i(t))^\top A^\top + A\lambda_i(t)M_i(t,\rho_i)^{\text{loc}}\lambda_i(t)^\top A^\top \\
&+A(I-\lambda_i(t))\eta_i(t)C_{iE}^\top S_i(t,\rho_i)C_{iE}\eta_i(t)^\top(I-\lambda_i(t))^\top A^\top \\
&+(1-\rho_i)\Sigma_w(t)]
\end{aligned} \tag{30a}$$

$$\text{s. t. } (\kappa_i(t)+\eta_i(t)C_{iE}^\top C_{iE})\mathbf{1}_{m_i} = \mathbf{1}_m, \tag{30b}$$

$$\kappa_i^\top(t)\mathbf{1}_m \geq \mathbf{0}_{m_i} \tag{30c}$$

$$\kappa_i(t)\mathbf{1}_{m_i} < \frac{1}{\|A\|_\infty}\mathbf{1}_m \tag{30d}$$

$$\lambda_i(t)\mathbf{1}_m + \varphi 1_m \leq \mathbf{1}_m, \tag{30e}$$

$$\lambda_i(t)\mathbf{1}_m \geq \mathbf{0}_m, \tag{30f}$$

where we have rewritten the objective function of (26) by using the derived expressions (15) and (16) for the bias and the variance of the prediction error and the definitions of W_{1i} and W_{2i} given in (17) and (18), and by substituting the following expressions (that are data of the problem or can be computed on-line using samples) in order to highlight the dependencies on the decision variables $\kappa_i(t)$, $\eta_i(t)$ and $\lambda_i(t)$:

$$M_i(t,\rho_i) = \rho_i \mathbb{E}\hat{\epsilon}_i(t)\mathbb{E}\hat{\epsilon}_i^\top(t) + (1-\rho_i)\Gamma_{\hat{\epsilon}_i}(t),$$

$$M_i(t,\rho_i)^{\text{loc}} = \rho_i \mathbb{E}\hat{E}_i(t)\mathbb{E}\hat{E}_i^\top(t) + (1-\rho_i)\Gamma_{\hat{E}_i}(t),$$

$$S_i(t,\rho_i) = (1-\rho_i)\Sigma_{v_{\hat{\epsilon}_i}}(t).$$

Moreover, φ is a small positive constant, where the last two constraints are equivalent to Eq. (28), if $\varphi = 0$, or (29) if $\varphi \neq 0$, depending on matrix A norm. It is important to note that problem (26) and problem (30) share the same cost function and the first constraint, but problem (30) is constrained by more conservative conditions, which are anyway sufficient conditions for the convergence of the local prediction error mean. This gives an optimization problem which is tractable. In fact, problem (30) is convex since the objective function has a quadratic form and M_i, M_i^{loc} and S_i are positive definite matrices, and the constraints are convex. Coherently, we can use Lagrangian duality to solve problem (30).

Let us introduce the dual variables ξ_{i1}, ξ_{i2}, ξ_{i4} and ν_i, which are $m \times 1$ vectors, and the $m_i \times 1$ vector ξ_{i3}. Then we have the following result:

Lemma 1. *Consider optimization problem (30). Let ϵ_1 be a positive constant. Then, the optimal values of the primal ($\kappa_i(t)$, $\eta_i(t)$ and $\lambda_i(t)$) and dual (ξ_{i1}, ξ_{i2}, ξ_{i3}, ξ_{i4} and ν_i) variables satisfy the following conditions:*

$$(\kappa_i(t)+\eta_i(t)C_{iE}^\top C_{iE})\mathbf{1}_{m_i} - \mathbf{1}_m = 0_m, \tag{31}$$

$$\lambda_i(t)\mathbf{1}_m + \varphi 1_m - \mathbf{1}_m \leq 0_m, \tag{32}$$

$$-\lambda_i(t)\mathbf{1}_m \leq 0_m, \tag{33}$$

$$\xi_{i1}^\top(t)(\lambda_i(t)\mathbf{1}_m + \varphi 1_m - \mathbf{1}_m) = 0 \quad \xi_{i1}(t) \geq 0, \tag{34}$$

$$\xi_{i2}^\top(t)(-\lambda_i(t)\mathbf{1}_m) = 0 \quad \xi_{i2}(t) \geq 0, \tag{35}$$

$$-\kappa_i(t)\mathbf{1}_m \leq 0_m, \tag{36}$$

$$\kappa_i(t)\mathbf{1}_{m_i} - \frac{1}{\|A\|_\infty}\mathbf{1}_m + \epsilon_1 \leq 0_m, \tag{37}$$

$$\xi_{i3}^\top(t)(-\kappa_i(t)\mathbf{1}_m) = 0 \quad \xi_{i3}(t) \geq 0, \tag{38}$$

$$\xi_{i4}^\top(t)(\kappa_i(t)\mathbf{1}_{m_i} - \frac{1}{\|A\|_\infty}\mathbf{1}_m + \epsilon_1) = 0 \quad \xi_{i4}(t) \geq 0, \tag{39}$$

$$\begin{aligned}
&2[M_i(t,\rho_i)\kappa_i(t)^\top(I-\lambda_i(t))^\top A^\top A(I-\lambda_i(t))] \circ (\mathbf{1}\otimes I) \\
&+ [\mathbf{1}_{m_i}\nu_i^\top] \circ (\mathbf{1}\otimes I) + [\mathbf{1}_{m_i}\xi_{i4}^\top] \circ (\mathbf{1}\otimes I) \\
&- [\xi_{i3}(t)\mathbf{1}_m^\top] \circ (\mathbf{1}\otimes I) = 0,
\end{aligned} \tag{40}$$

$$\begin{aligned}
&2[C_{iE}^\top S_i(t,\rho_i)C_{iE}\eta_i(t)^\top(I-\lambda_i(t))^\top A^\top A(I-\lambda_i(t))] \\
&\circ (\mathbf{1}\otimes I) + [C_{iE}^\top C_{iE}\mathbf{1}_{m_i}\nu_i^\top] \circ (\mathbf{1}\otimes I) = 0,
\end{aligned} \tag{41}$$

$$\begin{aligned}
&2[[\kappa_i(t)M_i(t,\rho_i)\kappa_i(t)^\top + \eta_i(t)C_{iE}^\top S_i(t,\rho_i)C_{iE}\eta_i(t)^\top \\
&+ M_i(t,\rho_i)^{\text{loc}}]\lambda_i(t)A^\top A \\
&- [\kappa_i(t)M_i(t,\rho_i)\kappa_i(t)^\top + \eta_i(t)C_{iE}^\top S_i(t,\rho_i)C_{iE}\eta_i(t)^\top]A^\top A] \\
&\circ I + [\mathbf{1}_m\xi_{i1}^\top(t)] \circ I - [\mathbf{1}_m\xi_{i2}^\top(t)] \circ I = 0.
\end{aligned} \tag{42}$$

Proof. Problem (30) is convex. Therefore the Karush–Kuhn–Tucker conditions are both necessary and sufficient for optimality. In Eqs. (31)–(33) we derive the canonical form of the constraints. Since $\kappa_i(t)$, $\eta_i(t)$ and $\lambda_i(t)$ are composed by diagonal blocks, Eqs. (40)–(42) are obtained by using matrix derivatives for diagonal and symmetric matrices (Minka, 2001; Petersen & Pedersen, 2006): given block-diagonal matrix κ, positive semidefinite matrix M and matrix A,

$$\frac{d\,\operatorname{tr}(A\kappa M\kappa^T A^T)}{d\,\kappa} = 2(M\kappa^T A^T A) \circ (\mathbf{1}\otimes I). \tag{43}$$

This concludes the proof. □

4.2. The optimal weights for the approximated problem

We analyze the KKT conditions to derive the optimal values for the decision variables $\kappa_i(t)$, $\eta_i(t)$ and $\lambda_i(t)$. We need the following preliminary result:

Lemma 2. *Consider square matrices M, A and X with dimension $n \times n$, where A is symmetric and X is diagonal. Then,* $\operatorname{diag}(MXA) = (M \circ A)\operatorname{diag}(X)$.

Proof. Omitted due to space constraints. □

We use this lemma in the following proposition to solve Eqs. (40)–(42). We define the $m_i \times 1$ vector $\kappa_i^{\text{vec}}(t) = \kappa_i^\top(t)\mathbf{1}_m$, collecting all the decision variables of the diagonals of $\kappa_i(t)$ on a column vector. Similarly, we define $\eta_i^{\text{vec}}(t) = \eta_i^\top(t)\mathbf{1}_m$, $\lambda_i^{\text{vec}}(t) = \lambda_i^\top(t)\mathbf{1}_m$ and $\nu_i^E(t) = [\mathbf{1}_{m_i}\nu_i^\top] \circ (\mathbf{1}\otimes I)\mathbf{1}_m$, collecting N_i times the vector ν_i. We provide the optimal solution for problem (30).

Proposition 2. *The optimal solution for problem (30), for each node i at each time step, is given by $\xi_{i_1} = \xi_{i_2} = \xi_{i_3} = \xi_{i_4} = 0$,*

$$\kappa_i^{\text{vec}}(t,\rho_i) = -\left(2M_i(t,\rho_i) \circ (\mathbf{1}_{N_i}\mathbf{1}_{N_i}^T \otimes D_i)\right)^{-1}\nu_i^E(t,\rho_i), \tag{44a}$$

$$\begin{aligned}
\eta_i^{\text{vec}}(t,\rho_i) = &-\left(2C_{iE}^\top S_i(t,\rho_i)C_{iE} \circ (\mathbf{1}_{N_i}\mathbf{1}_{N_i}^T \otimes D_i)\right)^{-1}C_{iE}^\top \\
&C_{iE}\nu_i^E(t,\rho_i),
\end{aligned} \tag{44b}$$

$$\nu_i(t,\rho_i) = -2(I_{m\times m_i}((M_i(t,\rho_i)\circ\mathbf{D}_i)^{-1} + C_{iE}^\top C_{iE} \left(C_{iE}^\top S_i(t,\rho_i)C_{iE}\circ\mathbf{D}_i\right)^{-1} C_{iE}^\top C_{iE})I_{m\times m_i}^T)^{-1}\mathbf{1}_m\,, \tag{44c}$$

$$\lambda_i^{\text{vec}}(t,\rho_i) = (2(M_i^{\text{loc}}(t,\rho_i) + \kappa_i M_i(t,\rho_i)\kappa_i^\top(t,\rho_i) + \eta_i(t,\rho_i)C_{iE}^\top S_i C_{iE}\eta_i^\top(t,\rho_i))\circ(A^TA))^{-1} (\text{diag}(2(\kappa_i(t,\rho_i)M_i\kappa_i^\top(t,\rho_i) + \eta_i(t,\rho_i)C_{iE}^\top S_i(t,\rho_i)C_{iE}\eta_i^\top(t,\rho_i))A^TA))\,, \tag{44d}$$

with $D_i(t,\rho_i) := (I-\lambda_i(t,\rho_i))^TA^TA(I-\lambda_i(t,\rho_i))$, *if* $\kappa_i(t,\rho_i)$ *and* $\lambda_i(t,\rho_i)$ *so computed in addition satisfy* (32), (33), (36), (37).

Proof. In the following we omit the dependence on time and on the ρ_i Pareto parameter. We observe that Eq. (42) is equivalent to

$$\text{diag}\left(2(M_i^{\text{loc}} + \kappa_i M_i\kappa_i^T + \eta_i C_{iE}^\top S_i C_{iE}\eta_i^T)\lambda_i A^TA\right) = \text{diag}\left(2(\kappa_i M_i\kappa_i^T + \eta_i C_{iE}^\top S_i C_{iE}\eta_i^T)A^TA\right) + \xi_{i_2} - \xi_{i_1}\,. \tag{45}$$

By using the result from Lemma 2, we have

$$\lambda_i^{\text{vec}} = \left(2(M_i^{\text{loc}} + \kappa_i M_i\kappa_i^T + \eta_i C_{iE}^\top S_i C_{iE}\eta_i^T)\circ(A^TA)\right)^{-1} \left(\text{diag}\left(2(\kappa_i M_i\kappa_i^T + \eta_i C_{iE}^\top S_i C_{iE}\eta_i^T)A^TA\right) + \xi_{i_2} - \xi_{i_1}\right)\,. \tag{46}$$

Let us now analyze KKT conditions in Eqs. (40) and (41). We find a solution with $\xi_{i_3} = 0$ and $\xi_{i_4} = 0$. Let us note that Eq. (40) holds if and only if it holds

$$2(M_i\kappa_i^T(I-\lambda_i)^TA^TA(I-\lambda_i))\circ(\mathbf{1}\otimes I)\,\mathbf{1}_m = -\mathbf{1}_{m_i}\nu_i^T\circ(\mathbf{1}\otimes I)\,\mathbf{1}_m = -\nu_i^E\,. \tag{47}$$

Let us denote $\kappa_i = [\kappa_{i_1},\ldots,\kappa_{i_{N_i}}]$, where κ_{i_j} is the (j)th $m\times m$ diagonal block of κ_i, with $j=1,\ldots,N_i$. Similarly we consider $M_{i_{pq}}$, which is the $m\times m$ block of matrix M_i. Then, we use the block operations and we achieve

$$\begin{cases} 2(\sum_{k=1}^{N_i} M_{i_{1k}}\kappa_{i_k}D_i)\circ I\,\mathbf{1}_m = -\nu_i \\ \qquad\qquad\vdots \qquad\qquad , \\ 2(\sum_{k=1}^{N_i} M_{i_{N_ik}}\kappa_{i_k}D_i)\circ I\,\mathbf{1}_m = -\nu_i \end{cases}$$

with $D_i := (I-\lambda_i)^TA^TA(I-\lambda_i)$. Thus by using block by block the result in Lemma 2: and remembering that by definition κ_i^{vec} is the vector that collects all the elements from κ_i, we obtain

$$\kappa_i^{\text{vec}} = -\left(2M_i\circ(\mathbf{1}_{N_i}\mathbf{1}_{N_i}^T\otimes D_i)\right)^{-1}\nu_i^E\,,$$

and similarly

$$\eta_i^{\text{vec}} = -\left(2C_{iE}^\top S_i C_{iE}\circ(\mathbf{1}_{N_i}\mathbf{1}_{N_i}^T\otimes D_i)\right)^{-1}C_{iE}^\top C_{iE}\nu_i^E\,,$$

as in the statement of the proposition. According to the first KKT condition, we have

$$(\kappa_i + \eta_i C_{iE}^\top C_{iE})\mathbf{1}_{m_i} = \mathbf{1}_{N_i}^T\otimes I(\kappa_i^{\text{vec}} + C_{iE}^\top C_{iE}\eta_i^{\text{vec}}) = \mathbf{1}_m\,, \tag{48}$$

which implies that

$$\nu_i = -2(I_{m\times m_i}((M_i\circ\mathbf{D}_i)^{-1} + C_{iE}^\top C_{iE}\left(C_{iE}^\top S_i C_{iE}\circ\mathbf{D}_i\right)^{-1}C_{iE}^\top C_{iE})I_{m\times m_i}^T)^{-1}\mathbf{1}_m\,,$$

where $I_{m\times m_i}$ denotes $\mathbf{1}_{N_i}^T\otimes I_m$, and $\mathbf{D}_i = \mathbf{1}_{N_i}\mathbf{1}_{N_i}^T\otimes D_i$. Thus we have obtained the optimal value of the dual variable ν_i and, by substituting the result in Eqs. (44a) and (44b), we obtain the optimal values of κ_i and η_i depending on λ_i. This is an optimal solution for the Pareto optimization problem (30), if, once computed the solutions on-line, in addition they satisfy the constraints (32), (33), (36) and (37), which concludes the proof. □

The solution given by the previous proposition, is the optimal solution of problem (30), the approximated convex version of problem (26). In order to use the result in Proposition 2, we propose a computational method (Algorithm 1) to find at each step the time-varying optimal values κ_i^*, η_i^* and λ_i^*. In order to guarantee constraints satisfaction, we project at each iteration the computed values of κ_i and λ_i in the corresponding sets defined by the constraints. This method can be used with any matrix A norm, since the convergence conditions constraints are not active.

Algorithm 1 Optimal weights computation

Set $\lambda_i^+ = I_m/2, \epsilon$
$I_{m\times m_i} = \mathbf{1}_{N_i}^T\otimes I_m$
repeat
  $\lambda_i = \lambda_i^+$
  $\mathbf{D}_i = \mathbf{1}_{N_i}\mathbf{1}_{N_i}^T\otimes(I_m-\lambda_i)^TA^TA(I_m-\lambda_i)$
  $\nu_i = -2[I_{m\times m_i}((M_i\circ\mathbf{D}_i)^{-1} + C_{iE}^\top C_{iE}\left(C_{iE}^\top S_i C_{iE}\circ\mathbf{D}_i\right)^{-1}C_{iE}^\top C_{iE})I_{m\times m_i}^T]^{-1}\mathbf{1}_m$ (Eq. (44c))
  $\kappa_i^{\text{vec}} = \mathbf{P}_K\left(-(2M_i\circ\mathbf{D}_i)^{-1}\mathbf{1}_{N_i}\otimes I_m\,\nu_i\right)$
  $\eta_i^{\text{vec}} = -\left(2C_{iE}^\top S_i C_{iE}\circ\mathbf{D}_i\right)^{-1}\mathbf{1}_{N_i}\otimes I_m\,C_{iE}^\top C_{iE}\nu_i$ (Eq. (44b))
  $R_i = \kappa_i M_i\kappa_i^T + \eta_i C_{iE}^\top S_i C_{iE}\eta_i^T$
  $\lambda_i^{\text{vec}} = \mathbf{P}_\Lambda\left((M_i^{\text{loc}} + R_i)\circ(A^TA)\right)^{-1}\text{diag}\left(R_iA^TA\right)$
  $\lambda_i^+ = \text{diag}(\lambda_i^{\text{vec}})$
until $|\lambda_i - \lambda_i^+| \le \epsilon$
return $\kappa_i(t) = \text{diag}(\kappa_i^{\text{vec}})$, $\eta_i(t) = \text{diag}(\eta_i^{\text{vec}})$, and $\lambda_i(t) = \lambda_i^+$

Proposition 3. *Alg. 1 converges to the optimal solutions* κ_i^*, η_i^* *and* λ_i^*.

Proof. In this proof, we first show that Alg. 1 converges, then we show that it converges to the optimal solution. Let us note that the optimization problem in Eq. (30) fulfills Slack's conditions, which implies that the optimal solution have finite optimal value. Furthermore, the two iterative sub-optimization steps in Alg. 1 ensure that the intermediate solutions improve (decrease) the current optimal value. Thus it is sure that Alg. 1 converges. Moreover, let us recall that the cost function is convex on all the variables. Suppose that Alg. 1 converges to a point $(\kappa_i', \eta_i', \lambda_i')$ different from $(\kappa_i^*, \eta_i^*, \lambda_i^*)$. Then we have $f(\kappa_i', \eta_i', \lambda_i') \le f(\kappa_i^*, \eta_i^*, \lambda_i') \le f(\kappa^*, \eta_i^*, \lambda_i^*)$, where $f(\kappa_i, \eta_i, \lambda_i)$ is the cost function. Since κ_i^*, η_i^* and λ_i^* are the optimal solutions, κ_i', η_i' and λ_i' must be equal to κ_i^*, η_i^* and λ_i^* respectively, which completes the proof. □

We now provide an analytical solution for problem (30).

Proposition 4. *The solution for problem* (30), *for each node i, is given by* $\xi_{i_2} \neq 0, \xi_{i_1} = \xi_{i_3} = \xi_{i_4} = \lambda_i = 0$,

$$\kappa_i^{\text{vec}}(t,\rho_i) = -\left(2M_i(t,\rho_i)\circ(\mathbf{1}_{N_i}\mathbf{1}_{N_i}^T\otimes D_i)\right)^{-1}\nu_i^E(t,\rho_i)\,, \tag{49a}$$

$$\eta_i^{\text{vec}}(t,\rho_i) = -\left(2C_{iE}^\top S_i(t,\rho_i)C_{iE}\circ(\mathbf{1}_{N_i}\mathbf{1}_{N_i}^T\otimes D_i)\right)^{-1} C_{iE}^\top C_{iE}\nu_i^E(t,\rho_i)\,, \tag{49b}$$

$$\nu_i(t,\rho_i) = -2(I_{m\times m_i}((M_i(t,\rho_i)\circ\mathbf{D}_i)^{-1} + C_{iE}^\top C_{iE}\left(C_{iE}^\top S_i(t,\rho_i)C_{iE}\circ\mathbf{D}_i\right)^{-1}C_{iE}^\top C_{iE})I_{m\times m_i}^T)^{-1}\mathbf{1}_m\,, \tag{49c}$$

where $I_{m\times m_i} := \mathbf{1}_{N_i}^T\otimes I_m$, $D_i := A^TA$ *and* $\mathbf{D}_i = \mathbf{1}_{N_i}\mathbf{1}_{N_i}^T\otimes D_i$ *if* $\kappa_i(t,\rho_i)$ *so computed satisfies the constraints* (36)–(37), *depending on the norm of matrix A.*

Proof. The proof follows the procedure used for Proposition 2. We then impose $\xi_{i_2} \neq 0$, $\xi_{i_3} = 0$, $\xi_{i_4} = 0$. We observe that the KKT conditions hold by imposing $\xi_{i_2} \neq 0$ only if λ_i is null because of Eq. (35) and therefore $\xi_{i_1} = 0$ because of Eq. (34). Therefore, we

obtain the expressions in closed form for κ_i and η_i by substituting $\lambda_i = 0$ in Eqs. (44a), (44b) and (44c). □

This analytical solution is given to provide a simpler solution to (30) than the computational solution given in Proposition 2. We cannot guarantee in general that the solution proposed in Proposition 4 satisfies the convergence conditions and that it is optimal at each time step. Moreover, in the analytical solution given by the previous proposition, the convergence conditions constraint (33) is active. Since in problem (30) we have imposed more restrictive conditions than the ones obtained in Section 3.3 for (26), so the proposed solution could be sub-optimal w.r.t. problem (26). In Section 6, the solutions proposed in Propositions 2 and 4 are compared.

Finally, we propose here a computational solution of the original non convex problem (26), using the original constraints (20), (21), (22) or (23).

Proposition 5. *A solution for problem (26), for each node i, is given by (44a), (44b), (44c), (44d), with* $D_i(t, \rho_i) := (I - \lambda_i(t, \rho_i))^T A^T A(I - \lambda_i(t, \rho_i))$, *if* $\kappa_i(t, \rho_i)$ *and* $\lambda_i(t, \rho_i)$ *so computed in addition satisfy the convergence constraints (20) and (21), or (22) or (23).*

Proof. Problem (26) is non convex so the KKT conditions cannot be used as sufficient conditions for optimality. Nevertheless, KKT conditions are necessary conditions also in the non convex case. We can then compute the KKT conditions for problem (26). These are equal to the conditions obtained in Lemma 1, with exception of Eqs. (32)–(39), which change according to the different constraints, and Eqs. (40) and (40), where the final terms are slightly different, but continue to depend linearly on the dual variables. By imposing the dual variables all null so that the constraints related to the convergence conditions are all non active, we can follow the same procedure as in proof of Proposition 2. This is a solution for the Pareto optimization problem (26), if, once computed the solutions, in addition they satisfy the convergence constraints. □

The obtained solution can be a local minimum or a saddle point for Problem (26). The results obtained by Algorithm 1 are optimal for the approximated problem (30) and represent a local solution for the original problem (26), if in addition they satisfy the original non convex constraints (21), or (22) or (23), depending on A norm.

4.3. On the separability of the consensus and prediction

In the previous subsection, we have proposed two methods to obtain the values of the optimal time-varying weights for the proposed distributed dynamical estimator. A question that may rise concerns the possible separability of the consensus and prediction.

Consider the optimization problem (30). The separate design of the consensus-filtering weights κ_i and η_i and of the model-based prediction parameters λ_i leads to suboptimal solutions. In fact, from Proposition 2, it follows that the optimal value of λ_i depends on the optimal values of κ_i and η_i, and vice versa. Therefore, the computation of the solution of two separate optimization problems would certainly lead to a suboptimal solution unless additional assumption on the optimization problem are introduced. The proposed joint optimization approach may lead to improved solutions.

In this respect, in Section 6 simulation results show that choosing a constant fixed value for λ_i is in general suboptimal, since at some time steps the parameters and weights values proposed in Proposition 4 with $\lambda_i = 0$ do not satisfy the convergence constraints (36)–(37), and so the solution is not optimal. Even considering only error variance minimization, thus choosing $\rho_i = 0$ in the proposed method, the choice of the filtering weights and prediction parameters has to be jointly designed. This is an important result, since in many works of the state of the art (see Alriksson & Rantzer, 2006; Carli et al., 2008; Cattivelli & Sayed, 2010, as examples) the two steps (filtering and prediction) are designed separately, thus possibly causing suboptimal performances.

4.4. Bounds and stability of the mean and of the variance of the prediction error

Basing on the results obtained in the previous subsection, it is possible to derive a time-varying bound on the prediction bias. For the sake of simplicity, the result is derived using the optimal weights in Proposition 4, but it can easily be extended to the case in Proposition 2.

Proposition 6. *The mean of the global prediction error vector can be bounded as follows, with* $\gamma < 1$*:*

$$\left\|\mathbb{E}\hat{E}(t+1)\right\|_\infty \leq \gamma^t \left\|(A_E C_E^\top C_E - I)x_E(0)\right\|_\infty . \tag{50}$$

Proof. By considering the global prediction bias in Eq. (9) and being $\lambda_i = 0, \ \forall i$ due to Proposition 4, we have

$$\mathbb{E}\hat{E}(t+1) = A_E K(t)\mathbb{E}\hat{E}(t).$$

Let us then consider the infinity norm, representing the maximum value of the bias vector, and apply the submultiplicative property. Since, thanks to the convergence conditions (21), $\|K(t)\|_\infty < 1/\|A_E\|_\infty$ and $1/\|A_E\|_\infty > 1$, we can say that there exists a scalar $\gamma < 1$ for which, for each t

$$\gamma'(t) = \|A_E\|_\infty \|K(t)\|_\infty \leq \gamma .$$

We can then write that

$$\left\|\mathbb{E}\hat{E}(t+1)\right\|_\infty \leq \prod_{h=0}^{t} \gamma'(h) \left\|\mathbb{E}\hat{E}(0)\right\|_\infty \leq \gamma^t \left\|\mathbb{E}\hat{E}(0)\right\|_\infty .$$

Finally, by noting that the algorithm is initialized so that $\hat{x}_i(0) = AC_i^\top y_i(0)$, we have $\mathbb{E}\hat{E}(0) = (A_E C_E^\top C_E - I)x_E(0)$, thus obtaining the statement of the proposition. □

This result proves that at each time step the bias is bounded and confirms its convergence to zero. We have then the following result.

Proposition 7. *The sequence of the variance terms of the global prediction error* $\Gamma_{\hat{E}}(t+1)$ *in Eq. (11) is bounded and converges.*

Proof. In Anderson and Moore (1981) (Theorem 4.3), it is stated that if a time-varying system $x(t+1) = F(t)x(t)$ is exponentially stable, and if the matrices $F(t)$ and $G(t)$ are bounded, then there exists a unique bounded nonnegative definite matrix sequence satisfying the following relation:

$$P(t+1) = F(t)P(t)F(t)^\top + G(t)G(t)^\top . \tag{51}$$

We note that Eq. (11) has the form of Eq. (51), by substituting $P(t) := \Gamma_{\hat{E}}(t)$, $F(t) := A_E K(t)$ and $Q(t) := G(t)G(t)^\top = W_2(t)\Sigma_v(t)W_2(t)^\top + \Sigma_{w_E}(t)$. We have then to demonstrate that the time-varying system $x(t+1) = F(t)x(t)$ is exponentially stable. In Proposition 6 we have shown that $\|F(t)\|_\infty \leq \|A_E\|_\infty \|K(t)\|_\infty \leq \gamma < 1$. Therefore, $\|x(t+1)\| = \|F(t)x(t)\| \leq \gamma^t \|x(0)\|$. For $t \to \infty$, the sequence converges to zero. Moreover, in Sichitiu and Bauer (2001) it is proved that, given a system $x(t+1) = F(t)x(t)$, with $A(t) \in \text{conv}(A_1, \ldots, A_N)$, it is exponentially stable iff $\exists$ a sufficiently large integer k such that $\left\|A_{t_1} \ldots A_{t_k}\right\| \leq \gamma < 1 \ \forall (t_1, \ldots, t_k) \in \{1, \ldots, N\}^k$, where $\text{conv}(A_1, \ldots, A_N)$ is the convex matrix polyhedron of the set of constant matrices $\{A_1, \ldots, A_N\}$ and $\|\cdot\|$ is any

vector induced matrix norm. In our case, the hypothesis is satisfied. Then the exponentially stability of the system $x(t+1) = F(t)x(t)$ is demonstrated and we use the above result in Anderson and Moore (1981) to prove the statement of the proposition. Furthermore, by iterating Eq. (51), we obtain

$$P(t) = \Phi_{t,0}\Sigma_v(0)\Phi_{t,0}^\top + \sum_{l=0}^{t-1} \Phi_{t,l+1}Q(l)\Phi_{t,l+1}^\top,$$

where $\Phi_{t,l} := F(t-1)\dots F(l)$, with $l < t$, and $\Phi_{t,t} := I$, which concludes the proof. □

5. Implementation of the distributed prediction algorithm

In this section, we address how to implement the proposed distributed prediction method. Each node has to implement Algorithm 2. In the literature, it is common to determine the best value of ρ_i by building the Pareto trade-off curve and selecting the "knee-point" of this curve, that is, choosing ρ_i^* such that B_i and V_i, computed with the optimal values $\kappa_i^*(\rho_i^*)$, $\eta_i^*(\rho_i^*)$ and $\lambda_i^*(\rho_i^*)$, are $V_i = B_i^2$. This can be obtained by solving the following further problem:

$$\min_{\rho_i}(V_i(\kappa_i^*(\rho_i), \eta_i^*(\rho_i), \lambda_i^*(\rho_i^*)) - B_i^2(\kappa_i^*(\rho_i), \eta_i^*(\rho_i), \lambda_i^*(\rho_i^*)))^2.$$

This problem is highly non-linear. Numerical methods can be used to compute the optimal value. In this and previous papers, we tested different approaches for the definition of the Pareto parameter. It is possible to choose it locally, using the Nelder–Mead simplex algorithm as described in Lagarias, Reeds, Wright, and Wright (1998), to minimize the cost function $(1-\rho_i)V_i + \rho_i B_i^2$ with the values of parameters and weights obtained at the previous step. Note that the values of V_i and B_i are functions of ρ_i. Once the parameter ρ_i has been set, the optimal weights $\kappa_i(t)$, $\eta_i(t)$ and $\lambda_i(t)$ are computed using the result in Proposition 4 or the result in Proposition 2 (following Algorithm 1) and so the local prediction of the state $x_i(t+1)$ can be obtained. The nodes are not required to on-line solve the optimization problem (26) formulated in the previous sections. After that, the values of the estimates of $\hat{\Gamma}_i(t)$ and $\hat{m}_{\epsilon_i}(t)$ can be updated using new signal estimates and measurements samples (see Boem et al., 2012; Speranzon et al., 2008 for details).

5.1. Computational complexity

The computational complexity of the proposed distributed estimator is given mainly by two components: the computational complexity of matrices inverse and that needed for the estimation of the covariance matrix. The computation of matrices inverse has complexity $O((m \times |\mathcal{N}_i|)^3)$ and is required to compute the optimal weights $\kappa_i(t)$, $\eta_i(t)$ and $\lambda_i(t)$. In the simulation, the typical number of iterations $\mathrm{N_{Iter1}}$ of Algorithm 1 at each step is less than 10. The computation of the covariance matrix is required to compute the approximate estimates $\hat{\Gamma}_i(t)$ and $\hat{\Lambda}_i(t)$: the complexity is $O(\mathrm{Table_{size}}\log(\mathrm{Table_{size}}))$, where the $\mathrm{Table_{size}}$ is the size of a look-up table used to speed up the computation of a quadratically constrained least-square problem (Speranzon et al., 2008). We set $\mathrm{Table_{size}} = 100$.

6. Simulation results

In this section, simulation results are given with the purpose of illustrating the analysis. We consider two different examples.

Algorithm 2 Prediction algorithm for node i

$t := 0$
$\hat{m}_{\epsilon_i}(0) := 0$
$\hat{\Gamma}_i(0) := \Sigma_{v_i}$
$\rho_i := 0.5$
$\bar{x}_i(0) := C_i^\top y_i(0)$
$\hat{x}_i(1) := A\bar{x}_i(0)$
repeat
  $N_i := |\mathcal{N}_i|$
  $t := t+1$
  Collect predictions $\hat{\mathbf{x}}_i(t) := (\hat{x}_{i_1}(t), \dots, \hat{x}_{i_{N_i}}(t))^T$ where $\{i_1, \dots, i_{N_i}\} \in \mathcal{N}_i$
  Collect measurements $\mathbf{y}_i(t) := (C_{i_1}^\top y_{i_1}(t), \dots, C_{i_{N_i}}^\top y_{i_{N_i}}(t))^T$ where $\{i_1, \dots, i_{N_i}\} \in \mathcal{N}_i$
  Compute ρ_i
  $S_i := (1-\rho_i)\Sigma_{v_{\epsilon_i}}(t)$
  $M_i := \rho_i(\hat{m}_{\epsilon_i}(t)\hat{m}_{\epsilon_i}^\top(t)) + (1-\rho_i)\hat{\Gamma}_i(t)$
  Compute $\kappa_i(t)$
  Compute $\eta_i(t)$
  Compute $\lambda_i(t)$
  $\bar{x}_i(t) = \kappa_i(t)\hat{\mathbf{x}}_i(t)(t) + \eta_i(t)\mathbf{y}_i(t)(t)$
  $\hat{x}_i(t+1) = A\bar{x}_i(t) + \bar{w}(t) + \lambda_i(t)(A\hat{x}_i(t) - A\bar{x}_i(t))$
  $\hat{\epsilon}_i := \frac{\hat{\mathbf{x}}_i}{1+\nu} - \frac{\nu\mathbf{1}^\top\hat{\mathbf{x}}_i + (1+\nu)\mathbf{1}^\top\mathbf{y}_i}{N_i(1+2\nu)(1+\nu)}\mathbf{1}$
  $\hat{m}_{\epsilon_i}(t) := \frac{t-1}{t}\hat{m}_{\epsilon_i}(t-1) + \frac{1}{t}\hat{\epsilon}_i(t)$
  $\hat{\Gamma}_i(t) := \frac{t-1}{t}\hat{\Gamma}_i(t-1) + \frac{1}{t}(\hat{\epsilon}_i(t) - \hat{m}_{\epsilon_i}(t))(\hat{\epsilon}_i(t) - \hat{m}_{\epsilon_i}(t))^\top$
until forever

6.1. Example 1

In the first case, we consider a linear system, whose dynamics are described by the matrix:

$$A = \begin{pmatrix} 0.8 & 0.1 & 0 & 0.05 \\ 0.1 & 0.55 & 0 & 0.2 \\ 0.04 & 0 & 0.7 & 0.2 \\ 0 & 0.15 & 0.3 & 0.5 \end{pmatrix}.$$

To manage the non-completely measurable case, we assume only 2 or 3 states can be measured by each node. We then randomly choose the matrix C on the basis of this assumption and the conditions in Section 2.

We analyzed the state of the art and we selected the following prediction methods for comparison[4] :

E_{CKF}: Centralized steady-state Kalman predictor. The estimator collects all the measurements y_i available at the entire sensor network and the related output matrices C_i, for all $i = 1, \dots, n$ and implements the classical steady-state Kalman predictor.
E_{DKF}: Decentralized Kalman filter as in Olfati-Saber (2007).
E_{DE1}: Distributed state estimation approach as in Battistelli and Chisci (2014). The estimator is adopted by formulating consensus on probability density functions of the states.
E_{DE2}: Distributed state estimation with event triggered communication protocols as in Meng and Chen (2014)
E_{p1}: The proposed estimator using the computational solution proposed in Algorithm 1.

[4] The distributed estimation methods in Caballero-Águila et al. (2014) and Yang et al. (2014) consider a non-completely measured state case, but require the assumption of a strongly connected sensor network and so we cannot use them for comparison.

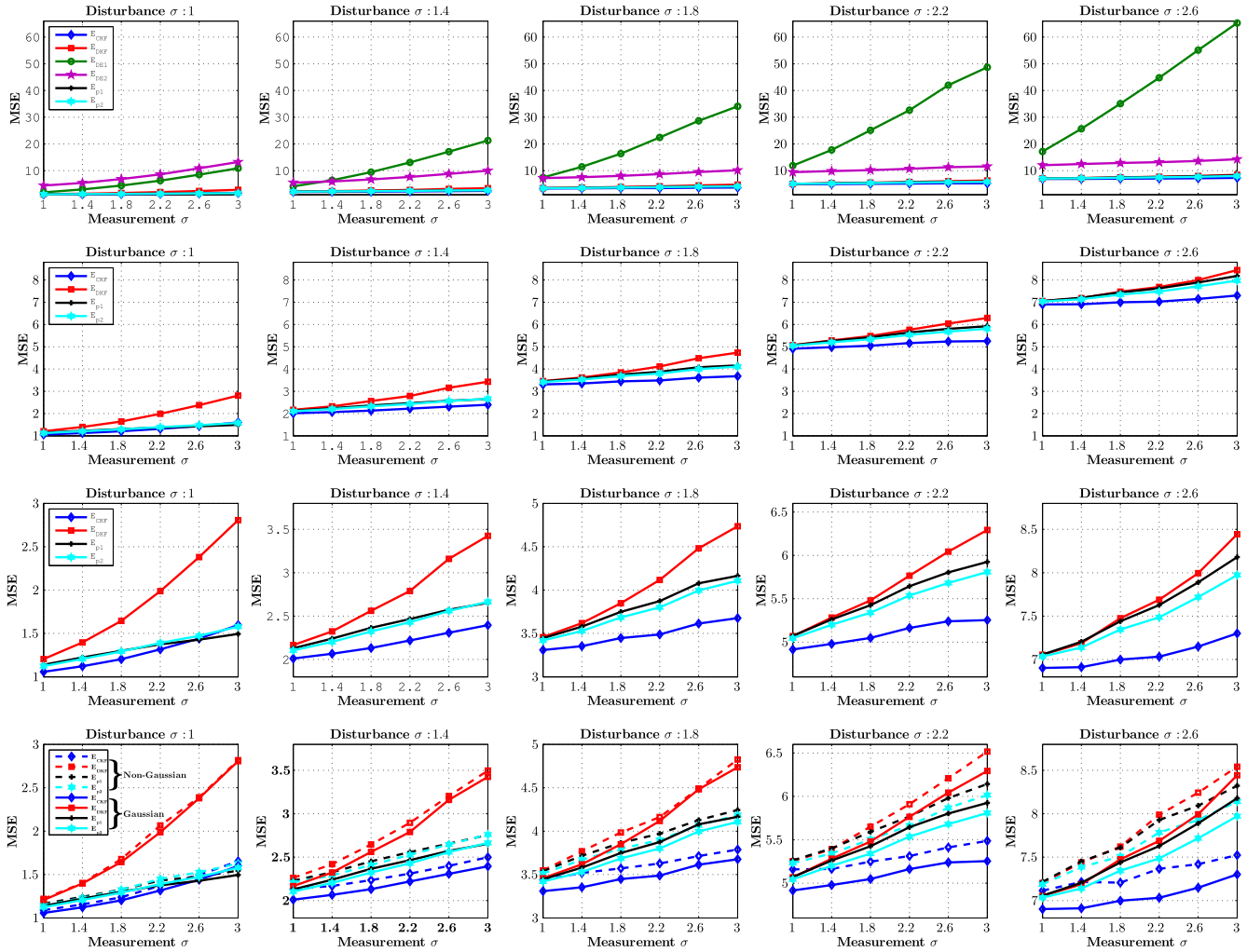


Fig. 1. Performance comparison in terms of MSE of the six considered estimation methods, with different standard deviation values for the measurement noises and the process disturbance. The first three rows consider the Gaussian case. In the fourth row, the Gaussian case is represented by solid lines, while the dashed lines show the non-Gaussian noise case performance.

E_{p2}: The proposed estimator using the analytic solution in Proposition 4.

The use of the centralized Kalman predictor is included to provide a baseline performance index and, for the sake of simplicity, the *steady-state* Kalman predictor has been considered. This latter predictor, in specific scenarios may show suboptimal performance during the transient when compared to time-varying algorithms like the proposed one. For all the methods, we compare the one-step ahead prediction error.

A 15-nodes network is obtained by distributing the nodes randomly over a squared area of size $N/2$ and the graph by letting two nodes communicate if their relative distance is less than $1.7\sqrt{N}$. We investigate the performance of the proposed methods with different noise probability distributions and different standard deviations for the measurement and modeling disturbance noises. We repeat the experiment 80 times with different random network topologies for each scenario. After many simulation experiments, we decided to fix $\rho_i = 0.5$ for each node i in the simulations, in order to reduce the computational complexity, since the obtained performances were similar to the knee-point method described in Section 5. The evaluated performance metric in one experiment is the mean square error of the predictions at each node, that is then averaged over all the nodes of the network. We then average this value over 80 experiments for each noise scenario (in the figures we use the term MSE to refer to this final average).

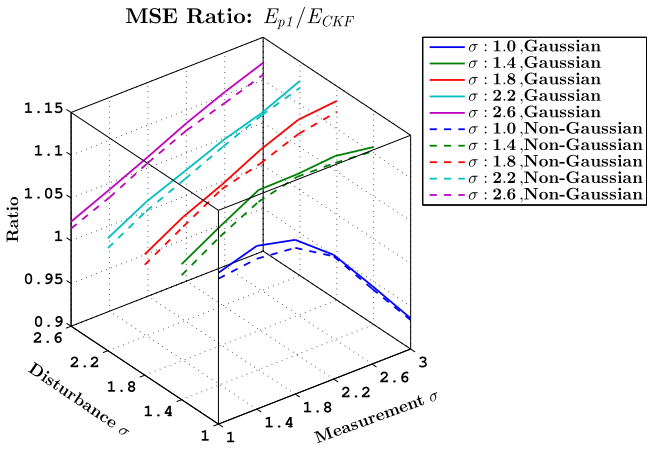


Fig. 2. The distance between the performance of the proposed approach E_{p1} and the Centralized Steady-state Kalman Predictor in terms of ratio of respective MSE values for the 30 different noise scenarios, both in the Gaussian and in the non-Gaussian case.

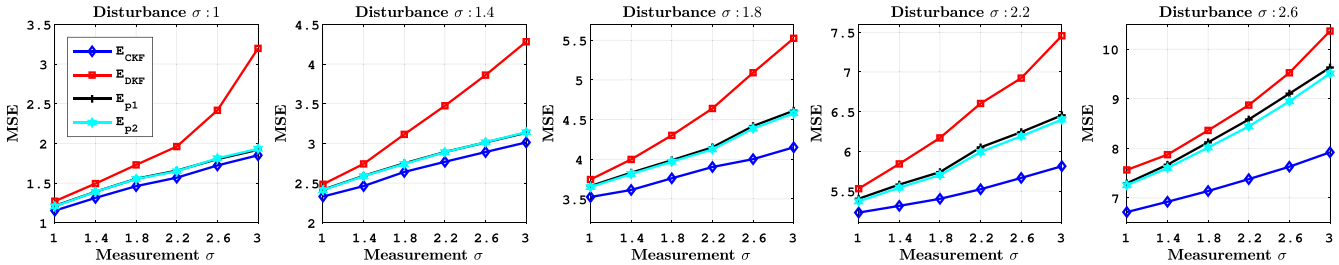


Fig. 3. Performance comparison in terms of MSE of the considered estimation methods, with different standard deviation values for the measurement noises and the process disturbance.

We firstly consider Gaussian white noise for both the measurement noise and system disturbance. The results are presented in the first three rows of Fig. 1. Each figure shows the performance of the considered prediction methods with different noise scenarios, letting varying the standard deviation of the measurement noise from 1 to 3 on the x axis and the standard deviation of the disturbance noise from 1 to 2.6 (different figures on the same row). In the first row, all the six considered methods are illustrated. The second row shows the performances of only four of the considered methods in order to appreciate, with the different scale, how the prediction error increases as the standard deviation values increase. The third row is needed to properly compare the performances of the four best methods, using different scales for each figure. We see that the proposed estimators always have better results than all the other distributed methods, in all the considered scenarios. Compared with the centralized steady-state Kalman predictor, which is optimal in the Gaussian case compared to steady-state linear estimators, the error of the proposed algorithms is lower than that of the centralized Kalman predictor in a single scenario, when the standard deviation of measurement noise is far larger than that of the process disturbance (as it is possible to see in Fig. 1, third row, first column). This is due to the fact that the time-varying proposed algorithm optimizes the used information at each step, also in the transient.

We then tested our method on a more challenging scenario with non-Gaussian distribution of the noises (see fourth row in Fig. 1). We consider the Tri-Gaussian noise, which is a non-Gaussian noise introduced in Hendeby and Gustafsson (2005), both for the measurement and the process noises.

From Fig. 1, last row, we see the performance of the best four estimation methods in each of the 30 noise scenarios, for both the probability distributions (Gaussian and non-Gaussian). Obviously in the Gaussian case all the methods perform better than in the non-Gaussian case. It is anyway interesting to see that the distance between our method and the centralized Kalman predictor performance decreases in the non-Gaussian case, as shown in Fig. 2, where the ratios between E_{p1} MSE and CKF MSE are illustrated for all the considered noise scenarios. We have similar results for E_{p2}.

We can see from simulation results that the two proposed methods have similar performance. Sometimes the analytical solution E_{p2} has better results than E_{p1}. This is due to computational convergence problems: the convergence of the proposed algorithm is guaranteed but the convergence time is unknown and in the simulations we had a maximum number of iterations.

6.2. Example 2

In this second example, in order to show the effectiveness of the proposed approach also for non-asymptotically stable systems, we consider a network of $N = 20$ sensor nodes monitoring a system representing a moving object on a plane, as described in Franco et al. (2006). The dynamics of the system can be represented as: $x(t+1) = Ax(t) + \xi(t)$, where

$$A = \begin{bmatrix} 1 & \delta & 0 & 0 & 0 & 0 \\ 0 & 1-\frac{\delta\mu}{m} & \frac{\delta}{m} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \delta & 0 \\ 0 & 0 & 0 & 0 & 1-\frac{\delta\mu}{m} & \frac{\delta}{m} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

where $\delta = 0.1$ s is the sampling time, $m = 0.75$ kg is the mass of the vehicle and $\mu = 0.15$ is the friction coefficient; the process noise $\xi(t)$ is a zero-mean Gaussian noise with $\sigma_\xi^2 = 10^{-4}\text{diag}(1, \ldots, 1)$. As in Franco et al. (2006), the state vector is initialized as $x(0) = \text{col}[0, 0, 0.1, 0, 0, 0.1]$. In Fig. 3 we show the performance of the best four estimation methods applied to this second simulation example for 30 different Gaussian noise scenarios. We see that the proposed estimators always have better results than the other distributed methods, also in this scenario.

7. Concluding remarks

In this paper, we proposed a novel distributed prediction method for dynamic systems using sensor networks, able to minimize both the mean and the variance of the prediction error. The optimal filtering weights and prediction parameters were computed locally at each step by each sensor node. We do not require Gaussian distribution for the noises. The state may be not entirely measured by each node. We showed that the filtering weights and the prediction parameters have to be jointly optimized.

As a future work, the adoption of the proposed work for distributed fault diagnosis purposes will be considered. Furthermore, we will consider time-varying network topologies and unreliable communication networks affected by delays and packet losses.

Acknowledgment

The authors wish to thank Yuzhe Xu (Electrical Engineering, KTH Royal Institute of Technology, Stockholm, Sweden) for useful feedback and suggestions.

References

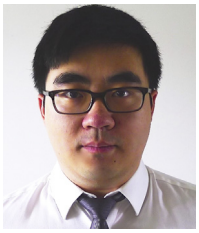
Alriksson, P., & Rantzer, A. (2006). Distributed Kalman filtering using weighted averaging. In *17th Int. Symposium on Mathematical Theory of Networks and Systems* (pp. 1–60).

Anderson, B. D. O., & Moore, J. B. (1981). Detectability and stabilizability of time-varying discrete-time linear systems. *SIAM Journal on Control and Optimization, 19*(1), 20–32.

Battistelli, G., & Chisci, L. (2014). Kullback–Leibler average, consensus on probability densities, and distributed state estimation with guaranteed stability. *Automatica, 50*(3), 707–718.

Bauer, P. H., Premaratne, K., & Duran, J. (1993). A necessary and sufficient condition for robust asymptotic stability of time-variant discrete systems. *IEEE Transactions on Automatic Control*, *38*(9), 1427–1430.
Boem, F., Ferrari, R. M. G., Parisini, T., & Polycarpou, M. M. (2011). Distributed fault detection and isolation of continuous-time nonlinear systems. *European Journal of Control*, *5–6*, 603–620.
Boem, F., Xu, Y., Fischione, C., & Parisini, T. (2012). A distributed estimation method for sensor networks based on Pareto optimization. In *51th IEEE Conference on Decision and Control* (pp. 775–781).
Boem, F., Xu, Yuzhe, Fischione, C., & Parisini, T. (2013). Distributed fault detection using sensor networks and Pareto estimation. In *European Control Conference* (pp. 932–937).
Boem, F., Xu, Yuzhe, Fischione, C., & Parisini, T. (2015). A distributed Pareto-optimal dynamic estimation method. In *European Control Conference* (pp. 3678–3685).
Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge University Press.
Caballero-Águila, R., García-Garrido, I., & Linares-Pérez, J. (2014). Information fusion algorithms for state estimation in multi-sensor systems with correlated missing measurements. *Applied Mathematics and Computation*, *226*, 548–563.
Carli, R., Chiuso, A., Schenato, L., & Zampieri, S. (2008). Distributed Kalman filtering based on consensus strategies. *IEEE Journal on Selected Areas in Communications*, *26*(4), 622–633.
Cattivelli, F. S., & Sayed, A. H. (2010). Diffusion strategies for distributed Kalman filtering and smoothing. *IEEE Transactions on Automatic Control*, *55*(9), 2069–2084.
Christofides, P. D., Scattolini, R., de la Pena, D. M., & Liu, J. (2013). Distributed model predictive control: A tutorial review and future research directions. *Computers & Chemical Engineering*, *51*, 21–41.
Davis, M. H. A. (1977). Linear estimation and stochastic control. London.
Deshmukh, S., Natarajan, B., & Pahwa, A. (2014). State estimation over a lossy network in spatially distributed cyber-physical systems. *IEEE Transactions on Signal Processing*, *62*(15), 3911–3923.
Ding, D., Wang, Z., & Shen, B. (2014). Recent advances on distributed filtering for stochastic systems over sensor networks. *International Journal of General Systems*, *43*(3–4), 372–386.
Farina, M., Ferrari-Trecate, G., & Scattolini, R. (2010). Distributed moving horizon estimation for linear constrained systems. *IEEE Transactions on Automatic Control*, *55*(11), 2462–2475.
Franco, E., Olfati-Saber, R., Parisini, T., & Polycarpou, M. M. (2006). Distributed fault diagnosis using sensor networks and consensus-based filters. In *45th IEEE Conference on Decision and Control* (pp. 386–391).
Garin, F., & Schenato, L. (2010). A survey on distributed estimation and control applications using linear consensus algorithms. In *Networked control systems, Vol. 406* (pp. 75–107). Springer.
Hendeby, G., & Gustafsson, F. (2005). *Fundamental filtering limitations in linear non-Gaussian systems*. Linkoping University Electronic Press, Report.
Khan, U. A., Kar, S., Jadbabaie, A., & Moura, J. M. F. (2010). On connectivity, observability, and stability in distributed estimation. In *49th IEEE Conference on Decision and Control* (pp. 6639–6644). IEEE.
Lagarias, J. C., Reeds, J. A., Wright, M. H., & Wright, P. E. (1998). Convergence properties of the nelder-mead simplex method in low dimensions. CiteSeerX - Scientific Literature Digital Library and Search Engine [http://citeseerx.ist.psu.edu/oai2] (United States).
Mahmoud, M. S., & Khalid, H. M. (2013). Distributed Kalman filtering: a bibliographic review. *IET Control Theory & Applications*, *7*(4).
Meng, X., & Chen, T. (2014). Optimality and stability of event triggered consensus state estimation for wireless sensor networks. In *American Control Conference* (pp. 3565–3570).
Minka, T. P. (2001). Old and new matrix algebra useful for statistics.
Mitra, A., & Sundaram, S. (2016). An approach for distributed state estimation of lti systems. In *54th Annual Allerton Conference on Communication, Control, and Computing* (pp. 1088–1093). IEEE.
Olfati-Saber, R. (2007). Distributed Kalman filtering for sensor networks. In *46th IEEE Conference on Decision and Control* (pp. 5492–5498).
Olfati-Saber, R. (2009). Kalman-consensus filter: Optimality, stability, and performance. In *48th IEEE Conference on Decision and Control, and 28th Chinese Control Conference* (pp. 7036–7042).
Park, S., & Martins, N. C. (2017). Design of distributed lti observers for state omniscience. *IEEE Transactions on Automatic Control*, *62*(2), 561–576.
Petersen, K. B., & Pedersen, M. S. (2006). *The matrix cookbook*. Technical University of Denmark.
Sichitiu, M. L., & Bauer, P. H. (2001). Stability of discrete time-variant linear delay systems and applications to network control. In *8th IEEE International Conference on Electronics, Circuits and Systems, Vol. 2* (pp. 985–989).
Spanos, D. P., Olfati-Saber, R., & Murray, R. M. (2005). Approximate distributed Kalman filtering in sensor networks with quantifiable performance. In *4th International Symposium on Information Processing in Sensor Networks* (pp. 133–139).
Speranzon, A., Fischione, C., Johansson, K. H., & Sangiovanni-Vincentelli, A. (2008). A distributed minimum variance estimator for sensor networks. *IEEE Journal on Selected Areas in Communications*, *26*(4), 609–621.
Stanković, S. S., Stanković, M. S., & Stipanović, D. M. (2009). Consensus based overlapping decentralized estimation with missing observations and communication faults. *Automatica*, *45*(6), 1397–1406.
Wang, L., & Morse, A. S. (2017). A distributed observer for a time-invariant linear system. In *American Control Conference* (pp. 2020–2025). IEEE.
Wu, J., Jia, Q. S., Johansson, K. H., & Shi, L. (2013). Event-based sensor data scheduling: Trade-off between communication rate and estimation quality. *IEEE Transactions on Automatic Control*, *58*(4), 1041–1046.
Yang, W., Chen, G., Wang, X., & Shi, L. (2014). Stochastic sensor activation for distributed state estimation over a sensor network. *Automatica*, *50*(8), 2070–2076.
Zhang, W.-A., Feng, G., & Yu, L. (2012). Multi-rate distributed fusion estimation for sensor networks with packet losses. *Automatica*, *48*(9), 2016–2028.

Francesca Boem received the M.Sc. degree (cum laude) in Management Engineering in 2009 and the Ph.D. degree in Information Engineering in 2013, both from the University of Trieste, Italy. She was Post-Doc at the University of Trieste with the Machine Learning Group from 2013 to 2014. From 2014 to 2018, she was Research Associate at the Department of Electrical and Electronic Engineering, Imperial College London, with the Control and Power Research Group. Since 2015 she has been part of the team at Imperial College working on the flagship EU H2020-WIDESPREAD-TEAMING project for the development of the EU KIOS Research and Innovation Centre of Excellence, a strategic partnership between University of Cyprus and Imperial College London. In 2018 Dr. Boem has been awarded the Imperial College Research Fellowship. From April 2018 she has been appointed as a Lecturer in the Department of Electronic and Electrical Engineering at University College London. Her current research interests include distributed fault diagnosis and fault-tolerant control methods for large-scale networked systems and distributed estimation methods for sensor networks. Dr. Boem is member of the IFAC Technical Committee 6.4 (“Fault Detection, Supervision & Safety of Technical Processes - SAFEPROCESS”) and Associate Editor for the IEEE Control System Society Conference Editorial Board and for the EUCA Conference Editorial Board.



Yilun Zhou received the B. Eng. in Electrical Engineering and Automation from Huazhong University of Science and Technology, China, in 2011.

He received the M.Sc degree in Control Systems in 2012 and the Ph.D. degree in Control Theory in 2017 from Imperial College London. He is currently a research engineer at Huawei Technologies, China. His research interests include distributed fault detection, distributed estimation, and sensor fusion.

Carlo Fischione is currently a tenured Associate Professor at KTH Royal Institute of Technology, Electrical Engineering and ACCESS Linnaeus Center, Stockholm, Sweden. He received the Ph.D. degree in Electrical and Information Engineering (3/3 years) in May 2005 from University of L’Aquila, Italy, and the Laurea degree in Electronic Engineering (Laurea, Summa cum Laude, 5/5 years) in April 2001 from the same University. He has held research positions at Massachusetts Institute of Technology, Cambridge, MA (2015, Visiting Professor); Harvard University, Cambridge, MA (2015, Associate); and University of California at Berkeley, CA (2004–2005), Visiting Scholar, and 2007–2008, Research Associate. His research interests include optimization with applications to networks, wireless and sensor networks, networked control systems, security and privacy. He has co-authored over 150 publications, including a book, book chapters, international journals and conferences, and international patents. He received or co-received a number of awards, including the best paper award from the IEEE Transactions on Industrial Informatics (2007), the best paper awards at the IEEE International Conference on Mobile Ad-hoc and Sensor System 05 and 09 (IEEE MASS 2005 and IEEE MASS 2009), the Best Paper Award of the IEEE Sweden VT-COM-IT Chapter (2014), the Best Business Idea awards from VentureCup East Sweden (2010) and from Stockholm Innovation and Growth (STING) Life Science in Sweden (2014), the “Ferdinando Filauro” award from University of L’Aquila, Italy (2003), the “Higher Education” award from Abruzzo Region Government, Italy

(2004), the Junior Research award from Swedish Research Council (2007), the Silver Ear of Wheat award in history from the Municipality of Tornimparte, Italy (2012). He is Editor of IEEE Transactions on Communications and Associated Editor of IFAC Automatica, has chaired or served as a technical member of program committees of several international conferences and is serving as referee for technical journals. Meanwhile, he also has offered his advice as a consultant to numerous technology companies such as Berkeley Wireless Sensor Network Lab, Ericsson Research, Synopsys, and United Technology Research Center. He is co-funder and Scientific Director of MIND Music Labs. He is Member of IEEE (the Institute of Electrical and Electronic Engineers), and Ordinary Member of DASP (the academy of history Deputazione Abruzzese di Storia Patria).

Thomas Parisini received the Ph.D. degree in Electronic Engineering and Computer Science in 1993 from the University of Genoa. He was with Politecnico di Milano and since 2010 he holds the Chair of Industrial Control and is Director of Research at Imperial College London. He is a Deputy Director of the KIOS Research and Innovation Centre of Excellence, University of Cyprus. Since 2001 he is also Danieli Endowed Chair of Automation Engineering with University of Trieste. In 2009–2012 he was Deputy Rector of University of Trieste. He authored or co-authored more than 280 research papers in archival journals, book chapters, and international conference proceedings. His research interests include neural-network approximations for optimal control problems, fault diagnosis for nonlinear and distributed systems, nonlinear model predictive control systems and nonlinear estimation. He is a co-recipient of the IFAC Best Application Paper Prize of the Journal of Process Control, Elsevier, for the three-year period 2011–2013 and of the 2004 Outstanding Paper Award of the IEEE Trans. on Neural Networks. He is also a recipient of the 2007 IEEE Distinguished Member Award. In 2016 he was awarded as Principal Investigator at Imperial of the H2020 European Union flagship Teaming Project KIOS Research and Innovation Center of Excellence led by University of Cyprus. In 2012 he was awarded an ABB Research Grant dealing with energy-autonomous sensor networks for self-monitoring industrial environments. Thomas Parisini currently serves as Vice-President for Publications Activities of the IEEE Control Systems Society and during 2009–2016 he was the Editor-in-Chief of the IEEE Trans. on Control Systems Technology. Since 2017, he is Editor for Control Applications of Automatica and since 2018 he is the Editor in Chief of the European Journal of Control.

He is also the Chair of the IFAC Technical Committee on Fault Detection, Supervision & Safety of Technical Processes — SAFEPROCESS.

He was the Chair of the IEEE Control Systems Society Conference Editorial Board and a Distinguished Lecturer of the IEEE Control Systems Society. He was an elected member of the Board of Governors of the IEEE Control Systems Society and of the European Control Association (EUCA) and a member of the board of evaluators of the 7th Framework ICT Research Program of the European Union. Thomas Parisini is currently serving as an Associate Editor of the Int. J. of Control and served as Associate Editor of the IEEE Trans. on Automatic Control, of the IEEE Trans. on Neural Networks, of Automatica, and of the Int. J. of Robust and Nonlinear Control.

Among other activities, he was the Program Chair of the 2008 IEEE Conference on Decision and Control and General Co-Chair of the 2013 IEEE Conference on Decision and Control. Prof. Parisini is a Fellow of the IEEE and of the IFAC.