A Distributed Pareto-Optimal Dynamic Estimation Method

Francesca Boem, Yuzhe Xu, Carlo Fischione and Thomas Parisini

Abstract— **In this paper, a novel distributed model-based prediction method is proposed using sensor networks. Each sensor communicates with the neighboring nodes for state estimation based on a consensus protocol without centralized coordination. The proposed distributed estimator consists of a consensus-filtering scheme, which uses a weighted combination of sensors information, and a model-based predictor. Both the consensus-filtering weights and the model-based prediction parameter for all the state components are jointly optimized to minimize the variance and bias of the prediction error in a Pareto framework. It is assumed that the weights of the consensus-filtering phase are unequal for the different state components, unlike consensus-based approaches from literature. The state, the measurements, and the noise components are assumed to be individually correlated, but no probability distribution knowledge is assumed for the noise variables. The optimal weights are derived and it is established that the consensus-filtering weights and the model-based prediction parameters cannot be designed separately in an optimal way. The asymptotic convergence of the mean of the prediction error is demonstrated. Simulation results show the performance of the proposed method, obtaining better results than distributed Kalman filtering.**

I. Introduction

In this paper, we consider the problem of distributed prediction of the state of a dynamical system from noisy measurements taken by a wireless sensor network. This approach, which is thought for the networked control field, can be useful in many applications, such as wireless localization, target tracking, environment and agriculture monitoring, where the state of the processes has to be estimated in a distributed manner at each sensor node. We assume that each sensor locally estimates the state of a common dynamic system with an uncertain model. This is done by letting each node communicate measurements and estimates only to neighboring nodes in the sensor network. Each sensor filters the measurements by taking a linear combination of its own and neighboring measurements and predictions. Subsequently, the sensor uses the current filtered measurements to implement a classic model-based predictor, smoothing the previous prediction error. The consensus-filtering weights and the model-based prediction parameters are time-varying and are computed locally by each sensor at each step in an optimal way so as to minimize the mean and variance of the current prediction error. In this paper, for the sake of simplicity, we assume that the entire system state is measured by each sensor. The non-completely measurable state case is left for further works.

Considering decentralized or distributed scenarios, the estimates are affected by bias due to the measurement and modeling uncertainty and the possibly incomplete communication between the sensors of the network. This motivates the need of simultaneously minimizing the bias and the variance of the estimation error. The proposed estimation method minimizes at each discrete-time step both the mean and the variance of the prediction error by posing a multi-objective optimization problem based on a Pareto optimization framework and this is the first main contributions of the paper. Compared to Kalman-based approaches, beyond the asymptotic behavior of the estimator – we show that asymptotic convergence is achieved – we aim at guaranteeing a good transient behavior of the estimator. Moreover, we jointly optimize the design of both the consensus-filtering weights and the model-based prediction parameters, and we establish that it is not possible to separate the two designs. This is the second main contribution of this work.

As far as our previous results are concerned, in [1] we developed a distributed estimation method based on Pareto optimization, but we considered only the measurement filtering task of a scalar signal, whereas in the present paper, we investigate the much more challenging task of prediction with a multi-variable scenario having possibly correlated components. The distributed state prediction goal is also motivated by the purpose of applying the proposed method to model-based fault detection in future works. Besides, in [2] we proposed a distributed prediction approach for fault detection purposes, but we did not consider the correlations between different state components, i.e., we considered each component as independent, and we did not optimize the model-based prediction parameter (we assumed it was given). On the other hand, in this paper, we minimize the bias and the variance of the prediction error with reference to both the consensus-filtering weights and the model-based prediction parameter at the same time in a jointly optimization problem. In this paper we obtain an important result: filtering and prediction weights cannot be designed in an optimal way separately. Therefore the approach proposed in this paper gives a substantial innovative contribution to the distributed estimation problem and considers a more challenging scenario. To the best of the authors' knowledge, this is the first time that a Pareto optimization approach has been designed for distributed prediction using sensor networks.

Proposed Approach vs. State of the Art

In the state of the art, roughly two different approaches have been proposed to the problem of distributed state

This work has 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's Energy Programme (contract no: EP/L014343/1).

F. Boem is with the Dept. of Electrical and Electronic Engineering at Imperial College London, UK. (f.boem@imperial.ac.uk)

Y. Xu and C. Fischione are with ACCESS Linnaeus Center, KTH Royal Institute of Technology, Stockholm, Sweden. (yuzhe@kth.se, carlofi@kth.se)

T. Parisini is with the Dept. of Electrical and Electronic Engineering at Imperial College London, UK, and with Dept. of Engineering and Architecture at University of Trieste, Italy (t.parisini@gmail.com)

estimation and prediction. First, the approaches based on diffusion strategies, such as the ones proposed in [3], [4] and the references therein of the same authors, where the diffusion of the local estimations in neighbors is applied after incremental update. These are in contrast with the second approach: consensus strategies, used, e.g., in [5], where consensus approaches are applied to obtain average observations or estimations at each iteration. Moreover, in [6] an optimal decentralized Kalman-Consensus Filtering approach is designed with the objectives of estimating the state of the system and reaching a consensus with neighboring estimator agents on the estimate.

Unlike Kalman filtering approaches (such as the optimal decentralized Kalman-consensus filter proposed in [6] that assumes a Gaussian distribution), in the proposed method no assumptions on the probability distribution of the measurement and modeling noises are made: we do not need to assume Gaussian distribution, but we assume knowledge of the mean and covariance matrix of the noise components. When our estimation problem 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 [7]. However, the scenario we are facing is more challenging, since we consider a distributed case, i.e., the prediction is computed locally without the coordination of a central agent. Moreover, we consider the multi variable case, unlike [8] which was for scalar cases. This is done by managing correlations between the different components of the state and of the noises, which is a major analysis challenge. In this paper, we minimize in a multi-objective optimization problem both the mean and the variance of the prediction error at each iteration. This is in contrast to [8], where Kalman-Consensus based approaches are investigated for the scalar case to obtain optimal Kalman gains and consensus weights such that the variance of the estimate error is minimized. Compared to [6], [5], [9], [10], we do not assume that the distribution of the disturbance is known. We only assume to know the mean and variance of such disturbance. Compared to [11], we do not select the most accurate estimate to propagate through the network. By using local estimation and measurements fusion, our method updates the filtering weights at each iteration. Furthermore, our work is in contrast to approaches that are focused on average initial samples [12], or where the estimation variables are static parameters [13], or where the only measurements are diffused over the network [14], [15]. Compared to [5], [6], we do not use the consensus algorithms in the estimator.

The paper is organized as follows. In Section II, we introduce the problem formulation. We describe the distributed prediction method in Section III, where the convergence of the predictor is demonstrated in Subsection III-C. In Section IV, the optimization problem of the time-varying weights is designed and the optimal values are derived in Section IV-A. Finally, some simulation results are provided in Section V and concluding remarks in Section VI.

II. Problem formulation

We introduce some notations that will be used throughout the paper. By $|\cdot|$ we denote the cardinality of the argument and by $\|\cdot\|$ the spectral norm of a matrix. Given a stochastic variable x, we represent as $\mathbb{E}x$ its expected value. By $\mathbf{1}$ and I we denote the vector $(1,\dots,1)^\top$ and the identity matrix with appropriate size, respectively. Finally, $\otimes$ denotes the Kronecker product and the operator $\circ$ represents the component-by-component product: as example, applied to a matrix M as $M \circ I$, it is equivalent to M with off-diagonal elements set to zero.

The process the state of which we want to predict is modeled as

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

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

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

A *sensor network*, made of n nodes, monitors system (1) by taking measurements, so that, for each sensor node i, with $i = 1,\dots,n$, we have:

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

where $y_i \in \mathbb{R}^m$ denotes the measurements vector taken by sensor i and $v_i \in \mathbb{R}^m$ is the measurement noise. The non-completely measurable state case is left for further works.

Assumption 2: We assume that $v_i(t)$ is a zero-mean measurement noise, with Σ_{v_i} its covariance matrix.

The extension to the non-zero mean case is trivial.

The original contribution of this paper concerns the state filtering and one-step-ahead prediction. We wish to answer the questions: how to optimally choose the weights of the two phases? Can the two phases be designed separately? The main underlying assumption is that each node of the network observes the entire state vector and exchanges measurements and predictions with the neighboring nodes to implement a consensus mechanism. The communication network is modeled by an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{N}_i = \{j \in \mathcal{V} : (j,i) \in \mathcal{E}\} \cup \{i\}$ is the set of neighbors of node $i \in \mathcal{V}$ plus the node itself.

III. Distributed state prediction

We propose that each node i of the sensor network implements a two phases estimator: a *filtering* phase and a *prediction* phase. First, by communicating with neighboring nodes, the estimator filters the measurements noise by computing a linear combination of its own and neighboring measurements and predictions. This is similar to a consensus-phase already used for distributed filtering:

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

where $k_{i,j}(t)$ and $h_{i,j}(t)$ are time-varying filter weights that we design, differently from consensus filtering approaches, where they are given. The objective of this first phase is for each node to reduce its own measurement uncertainty and to reach a consensus, without centralized coordination.

After the consensus-filtering phase, each node implements a classical model-based prediction, using filtered measurements (obtained from the previous step), instead of the

observed measurements. The one-step-ahead prediction is

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

where $\lambda_i'(t) \in \mathbb{R}^{m\times m}$ is a matrix collecting the time-varying filter parameters that will be designed in an optimal form in the following sections. The term $\lambda_i'(t)(\hat{x}_i(t) - \bar{x}_i(t))$ represents a correction of the previous prediction errors.

To facilitate the analysis, the distributed estimator-predictor formulation can be expressed in a vector form that includes all the nodes in the sensor network:

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

where $\bar{x}$, $\hat{x}$ and $y \in \mathbb{R}^{mn\times 1}$ 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 w_E is a column vector of appropriate length, where the process disturbance vector w is repeated n times. $K(t)$ and $H(t)$ are block matrices, where each (i,j)-th block, with $i,j = 1,\ldots,n$, collects on the diagonal the weights with which the i-th sensor weights measurements or estimates components developed by the j-th node, if they are neighbors; the block is a null matrix if i and j are not neighboring nodes. λ' is a diagonal block matrix, collecting matrices λ_i', $i = 1,\ldots,n$, on the diagonal.

The main goal of the paper is to devise a design principle of the time-varying weights $H(t)$, $K(t)$ and $\lambda'(t)$ in order to minimize bias and variance of the prediction error at each time-instant t. In particular, we aim at establishing whether the consensus 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 subsection.

A. Analysis of the estimation error

Let us define the following estimation errors: 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 dimension, where the process state vector x is repeated. They can be computed as follows:

$$\begin{aligned} \bar{e}(t) =& K(t)\hat{x}(t) + H(t)y(t) - x_E(t) \\ =& K(t)\hat{E}(t) + H(t)v(t), \end{aligned} \quad (6)$$

where we use the following assumption, needed for the tractability of the problem:

Assumption 3: Suppose that $(K(t) + H(t))\mathbf{1} = \mathbf{1}$.

For simplicity, we express the prediction parameter as $\lambda'(t) = A_E\lambda(t)$, being $\lambda(t)$ a diagonal matrix. Accordingly, the prediction error is

$$\begin{aligned} \hat{E}(t+1) =& A_E\bar{x}(t) + \bar{w}_E(t) + \lambda'(t)(\hat{x}(t) - \bar{x}(t)) - A_E x_E(t) \\ & - \bar{w}_E(t) - \tilde{w}_E(t) \\ =& A_E\left[(I - \lambda(t))K(t) + \lambda(t)\right]\hat{E}(t) \\ & + A_E(I - \lambda(t))H(t)v(t) - \tilde{w}_E(t). \end{aligned} \quad (7)$$

Let us compute mean and variance of the global estimation and prediction errors. The expected value expressions are

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

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

We introduce the following further assumption:

Assumption 4: The measurement noise v and process disturbances w are not correlated.

The covariance matrices can then be computed as

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

where $\Gamma_{\hat{E}}(t) = \mathbb{E}\left[(\hat{E}(t) - \mathbb{E}\hat{E}(t))(\hat{E}(t) - \mathbb{E}\hat{E}(t))^\top\right]$, $\Sigma_v(t) = \mathbb{E}\left[(v(t) - \mathbb{E}v(t))(v(t) - \mathbb{E}v(t))^\top\right]$, $\mathrm{Cov}(\hat{E}(t), v(t)) = 0$, and $\Gamma_{\hat{E}}(t+1)$ is

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

with

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

$$W_2(t) = A_E(I - \lambda(t))H(t)$$

and $\Sigma_{w_E}(t) = \mathbb{E}\left[(w_E(t) - \mathbb{E}w_E(t))(w_E(t) - \mathbb{E}w_E(t))^\top\right]$, being $\mathrm{Cov}(\hat{E}(t), v(t)) = 0$, $\mathrm{Cov}(\hat{E}(t), \tilde{w}_E(t)) = 0$ and $\mathrm{Cov}(v(t), \tilde{w}_E(t)) = 0$.

B. Local estimation and prediction errors

In this section, we derive the expressions of the local estimation and prediction errors. Each node computes locally a consensus estimation and a model-based prediction as

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

where $\hat{x}_i^{\mathrm{reg}}$ and y_i^{reg} are two column vectors collecting the prediction vectors and the measurements vectors respectively available at node i, ordered according to their indexes: $\hat{x}_i^{\mathrm{reg}} = (\hat{x}_{i_1}^\top, \ldots, \hat{x}_{i_{N_i}}^\top)^\top$, $i_1 < \ldots < i_{N_i}$, and $y_i^{\mathrm{reg}} = (y_{i_1}^\top, \ldots, y_{i_{N_i}}^\top)^\top$, $i_1 < \ldots < i_{N_i}$, where N_i is the cardinality of set $\mathcal{N}_i$. Moreover, $\kappa_i(t)$ and $\eta_i(t)$ are the time-varying matrices corresponding to the non–zero elements of the m rows related to the i-th node of matrices $K(t)$ and $H(t)$, respectively. Let us now define the local filtering error $\bar{e}_i(t) = \bar{x}_i(t) - x(t)$ and the local prediction error $\hat{E}_i(t) = \hat{x}_i(t) - x(t)$. Similarly to the global quantities, they can be computed as

$$\bar{e}_i(t) = \kappa_i(t)\hat{\epsilon}_i(t) + \eta_i(t)v_{\epsilon_i}(t), \quad (13)$$

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

where $\hat{\epsilon}_i$ collects the prediction error vectors available at node i and v_{ϵ_i} collects the measurement noise vectors related to the measurements available at node i, both ordered following their indexes.

Now, we can derive the expressions of the bias and the variance for the local estimation and prediction errors. The mean values can be computed as

$$\mathbb{E}\bar{e}_i(t) = \kappa_i(t)\mathbb{E}\hat{\epsilon}_i(t),$$

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

whereas the variance can be computed as

$$\mathbb{E}[(\hat{E}_i(t+1) - \mathbb{E}\hat{E}_i(t+1))(\hat{E}_i(t+1) - \mathbb{E}\hat{E}_i(t+1))^\top]$$
$$= W_{1i}(t)\Gamma_{\hat{\epsilon}_i}(t)W_{1i}(t)^\top + W_{2i}(t)\Sigma_{v_{\hat{\epsilon}_i}}W_{2i}(t)^\top + \Sigma_w(t), \tag{15}$$

where

$$W_{1i}(t) = A[(I-\lambda_i(t))\kappa_i(t) + \lambda_i^0(t)],$$
$$W_{2i}(t) = A(I-\lambda_i(t))\eta_i(t),$$
$$\Gamma_{\hat{\epsilon}_i}(t) = \mathbb{E}(\hat{\epsilon}_i(t) - \mathbb{E}[\hat{\epsilon}_i(t))(\hat{\epsilon}_i(t) - \mathbb{E}\hat{\epsilon}_i(t))^\top]$$

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

C. Convergence analysis

In this subsection, we show that the mean of the global prediction error (Eq. (9)) of our proposed estimation method converges to zero. In this subsection, for the sake of notational simplicity, we drop the dependence on t of the matrices.

Proposition 3.1: Consider the global prediction error mean of Eq. (9). The following local conditions are sufficient to guarantee its convergence to zero: if $\frac{1}{\|A_E\|_\infty} > 1$, then

$$\sum_{j=1}^{n} \left|k_{i,j}^l\right| < \frac{1}{\|A_E\|_\infty}$$
$$\frac{-\frac{1}{\|A_E\|_\infty} + \sum_{j=1}^{n}\left|k_{i,j}^l\right|}{1+\sum_{j=1}^{n}\left|k_{i,j}^l\right|} < \lambda_i^l < \frac{\frac{1}{\|A_E\|_\infty} + \sum_{j=1}^{n}\left|k_{i,j}^l\right|}{1+\sum_{j=1}^{n}\left|k_{i,j}^l\right|}; \tag{16}$$

if $\frac{1}{\|A_E\|_\infty} = 1$, then

$$\sum_{j=1}^{n} \left|k_{i,j}^l\right| < \frac{1}{\|A_E\|_\infty}$$
$$\frac{-\frac{1}{\|A_E\|_\infty} + \sum_{j=1}^{n}\left|k_{i,j}^l\right|}{1+\sum_{j=1}^{n}\left|k_{i,j}^l\right|} \leq \lambda_i^l < 1; \tag{17}$$

if $\frac{1}{\|A_E\|_\infty} < 1$, then

$$\sum_{j=1}^{n} \left|k_{i,j}^l\right| < \frac{1}{\|A_E\|_\infty}$$
$$\frac{-\frac{1}{\|A_E\|_\infty} + \sum_{j=1}^{n}\left|k_{i,j}^l\right|}{1+\sum_{j=1}^{n}\left|k_{i,j}^l\right|} < \lambda_i^l < 1 - \frac{1-\frac{1}{\|A_E\|_\infty}}{1-\sum_{j=1}^{n}\left|k_{i,j}^l\right|}. \tag{18}$$

Proof: Eq. (9) represents the dynamics of a linear time-varying system. We use the results in [16] to derive convergence conditions on the time-varying matrices $K(t)$ and $\lambda(t)$. A necessary and sufficient condition for global asymptotic stability is given in [16], by which there exists a finite $k > 0$ such that

$$\left\|\prod_{t=1}^{k} A_E\left[(I-\lambda(t))K(t) + \lambda(t)\right]\right\|_\infty < 1. \tag{19}$$

This condition is sufficiently satisfied if, for each t,

$$\|A_E\left[(I-\lambda(t))K(t) + \lambda(t)\right]\|_\infty < 1.$$

Remember that $\lambda(t)$ and $I-\lambda(t)$ are two diagonals 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. Thanks to the submultiplicative property of the norm, the convergence condition is implied by

$$\|(I-\lambda)K + \lambda\|_\infty < \frac{1}{\|A_E\|_\infty}.$$

This is satisfied if, for each l-th row of each i-th block row of matrix $(I-\lambda(t))K(t) + \lambda(t)$, with $i = 1,\ldots,n$ and $l = 1,\ldots,m$, we have

$$\left|1-\lambda_i^l\right|\sum_{j=1}^{n}\left|k_{i,j}^l\right| + \left|\lambda_i^l\right| < \frac{1}{\|A_E\|_\infty}. \tag{20}$$

We analyzed this last condition for all the alternative cases of values of λ_i^l to extract it from the absolute value and we obtained the convergence conditions in the statement. □

Note that the previous proposition gives local convergence conditions, meaning that they can be computed at each node. Therefore, in this section, we have proved that the proposed estimator converges under the derived convergence conditions and we can therefore formulate an optimization problem to build the distributed estimator.

IV. The optimization problem

The proposed distributed estimation methodology is based on minimizing the bias and variance of the prediction error at every time step. To do that, each sensor at each step, computes the optimal time-varying weights by solving a Pareto optimization problem:

$$\min_{\kappa_i,\eta_i,\lambda_i} \ \text{tr}\ [\rho_i B_i^2 + (1-\rho_i)V_i] \tag{21a}$$
$$\text{s.t. } (\kappa_i(t) + \eta_i(t))\mathbf{1}_{m_{N_i}} = \mathbf{1}_m, \tag{21b}$$
$$\text{convergence conditions Eq.(16)} \vee (17) \vee (18), \tag{21c}$$

where $m_{N_i} = mN_i$, $0 \leq \rho_i \leq 1$ is the Pareto parameter, $B_i = \mathbb{E}\hat{E}_i(t+1)$ is the prediction error bias, $V_i = \mathbb{E}(\hat{E}_i(t+1) - \mathbb{E}\hat{E}_i(t+1))^2$ is the variance of the prediction error. According to the derived expressions for the bias and the variance of the prediction error, the cost function of problem (21) is

$$\begin{aligned}
&\text{tr}\ [\rho_i B_i^2 + (1-\rho_i)V_i] \\
=&\text{tr}\ [\rho_i W_{1i}(t)\mathbb{E}\hat{\epsilon}_i(t)\mathbb{E}\hat{\epsilon}_i^\top(t)W_{1i}(t)^\top + (1-\rho_i)[\Sigma_w(t) \\
&+ W_{1i}(t)\Gamma_{\hat{\epsilon}_i}(t)W_{1i}(t)^\top + W_{2i}(t)\Sigma_{v_{\hat{\epsilon}_i}}W_{2i}(t)^\top]] \\
=&\text{tr}\ [A(I-\lambda_i(t))\kappa_i(t)[\rho\ \mathbb{E}\hat{\epsilon}_i(t)\mathbb{E}\hat{\epsilon}_i^\top(t) \\
&+ (1-\rho_i)\Gamma_{\hat{\epsilon}_i}(t)]\kappa_i(t)^\top(I-\lambda_i(t))^\top A^\top + (1-\rho_i)\Sigma_w(t) \\
&+ A\lambda_i(t)[\rho\mathbb{E}\hat{E}_i(t)\mathbb{E}\hat{E}_i^\top(t) + (1-\rho_i)\Gamma_{\hat{E}_i}(t)]\lambda_i(t)^\top A^\top \\
&+ A(I-\lambda_i(t))\eta_i(t)[(1-\rho_i)\Sigma_{v_{\hat{\epsilon}_i}}]\eta_i(t)^\top(I-\lambda_i(t))^\top A^\top],
\end{aligned} \tag{22}$$

where the last equality comes by inserting the definitions of W_{1i} and W_{2i}, with $\Gamma_{\hat{E}_i}(t) := \mathbb{E}(\hat{E}_i(t) - \mathbb{E}\hat{E}_i(t))^2$. The

first constraint in problem (21) is the local condition given by Assumption 3; the other constraints are the convergence conditions derived in Section III-C, depending on matrix A norm. We briefly analyze the convergence conditions (16), (17), (18). The first condition is common to all the three cases. In the optimization problem, for tractability reasons, we use the more restrictive conditions

$$\begin{cases} \kappa_i^\top(t)\mathbf{1}_m \geq \mathbf{0}_{m_{N_i}} \\ \kappa_i(t)\mathbf{1}_{m_{N_i}} < \dfrac{1}{\|A\|_\infty}\mathbf{1}_m \end{cases} \tag{23}$$

instead of $\sum_{j=1}^n \left|k_{i,j}^l\right| < 1/\|A_E\|_\infty$, $\forall l$. Moreover, as regards the quantities in Eqs.(16),(17),(18), we note that

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

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

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

In this paper, due to space restrictions, we consider only the case for $1/\|A_E\|_\infty = 1/\|A\|_\infty > 1$ (Eq.(16)). The other two cases, for different norms of system matrix A, can be investigated in a similar way. Therefore, for tractability reasons, we simplify the second condition in (16) with

$$0 \leq \lambda_i \leq I, \tag{24}$$

using the element-wise inequality, which implies the more complex condition in (16). To highlight the dependencies on the the decision variables $\kappa_i(t)$, $\eta_i(t)$ and $\lambda_i(t)$, we rewrite the Pareto optimization problem (21) by substituting the following expressions which are data of the problem or can be computed using samples:

$$M_i(t,\rho) = \rho\mathbb{E}\hat{e}_i(t)\mathbb{E}\hat{e}_i^\top(t) + (1-\rho_i)\Gamma_{\hat{e}_i}(t),$$
$$M_i(t,\rho)^{\text{loc}} = \rho\mathbb{E}\hat{E}_i(t)\mathbb{E}\hat{E}_i^\top(t) + (1-\rho_i)\Gamma_{\hat{E}_i}(t),$$
$$S_i(\rho) = (1-\rho_i)\Sigma_{v_{\hat{e}_i}}.$$

We obtain

$$\begin{aligned} \min_{\kappa_i(t),\eta_i(t),\lambda_i(t)} \ & \text{tr}\,[A(I-\lambda_i(t))\kappa_i(t)M_i(t,\rho)\kappa_i(t)^\top\cdot \\ & (I-\lambda_i(t))^\top A^\top + A\lambda_i(t)M_i(t,\rho)^{\text{loc}}\lambda_i(t)^\top A^\top \\ & + A(I-\lambda_i(t))\eta_i(t)S_i(\rho)\eta_i(t)^\top(I-\lambda_i(t))^\top A^\top \\ & + (1-\rho_i)\Sigma_w(t)] \end{aligned} \tag{25a}$$

$$\text{s. t. } (\kappa_i(t)+\eta_i(t))\mathbf{1}_{m_{N_i}} = \mathbf{1}_m, \tag{25b}$$
$$\kappa_i^\top(t)\mathbf{1}_m \geq \mathbf{0}_{m_{N_i}} \tag{25c}$$
$$\kappa_i(t)\mathbf{1}_{m_{N_i}} < \frac{1}{\|A\|_\infty}\mathbf{1}_m \tag{25d}$$
$$\lambda_i(t)\mathbf{1}_m < \mathbf{1}_m, \tag{25e}$$
$$\lambda_i(t)\mathbf{1}_m > \mathbf{0}_m, \tag{25f}$$

where the last two constraints are equivalent to the original ones since $\lambda_i(t)$ is a diagonal matrix. It is worth noting that problem (21) and problem (25) have the same objective function, but, for tractability reasons, problem (25) is constrained by more restrictive conditions. The analytical solution of this approximated problem is given in Proposition 4.2. To derive it we analyze the optimization problem. Note that problem (25) is convex since the objective function has a quadratic form and M_i, M_i^{loc} and S_i are positive definite matrices. Coherently, we can use Lagrangian duality to solve the problem. Let introduce the dual variables $\xi_{i1}, \xi_{i2}, \xi_{i4}$ and ν_i, which are $m \times 1$ vectors, and the $m_{N_i} \times 1$ vector ξ_{i3}.

Proposition 4.1: Consider optimization problem (25). 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 equations:

$$(\kappa_i(t)+\eta_i(t))\mathbf{1}_{m_{N_i}} - \mathbf{1}_m = 0_m, \tag{26}$$
$$\lambda_i(t)\mathbf{1}_m - \mathbf{1}_m \leq 0_m, \tag{27}$$
$$-\lambda_i(t)\mathbf{1}_m \leq 0_m, \tag{28}$$
$$\xi_{i1}^\top(t)(\lambda_i(t)\mathbf{1}_m - \mathbf{1}_m) = 0 \qquad \xi_{i1}(t) \geq 0, \tag{29}$$
$$\xi_{i2}^\top(t)(-\lambda_i(t)\mathbf{1}_m) = 0 \qquad \xi_{i2}(t) \geq 0, \tag{30}$$
$$-\kappa_i(t)\mathbf{1}_m \leq 0_m, \tag{31}$$
$$\kappa_i(t)\mathbf{1}_{m_{N_i}} - \frac{1}{\|A\|_\infty}\mathbf{1}_m + \epsilon_1 \leq 0_m, \tag{32}$$
$$\xi_{i3}^\top(t)(-\kappa_i(t)\mathbf{1}_m) = 0 \qquad \xi_{i3}(t) \geq 0, \tag{33}$$
$$\xi_{i4}^\top(t)(\kappa_i(t)\mathbf{1}_{m_{N_i}} - \frac{1}{\|A\|_\infty}\mathbf{1}_m + \epsilon_1) = 0 \qquad \xi_{i4}(t) \geq 0, \tag{34}$$

$$\begin{aligned} & 2[M_i(t,\rho)\kappa_i(t)^\top(I-\lambda_i(t))^\top A^\top A(I-\lambda_i(t))] \circ (\mathbf{1}\otimes I) \\ & + [\mathbf{1}_{m_{N_i}}\nu_i^\top] \circ (\mathbf{1}\otimes I) + [\mathbf{1}_{m_{N_i}}\xi_{i4}^\top] \circ (\mathbf{1}\otimes I) \\ & - [\xi_{i3}(t)\mathbf{1}_{m_{N_i}}^\top] \circ (\mathbf{1}\otimes I) = 0, \end{aligned} \tag{35}$$

$$\begin{aligned} & 2[S_i(t,\rho)\eta_i(t)^\top(I-\lambda_i(t))^\top A^\top A(I-\lambda_i(t))] \circ (\mathbf{1}\otimes I) \\ & + [\mathbf{1}_{m_{N_i}}\nu_i^\top] \circ (\mathbf{1}\otimes I) = 0, \end{aligned} \tag{36}$$

$$\begin{aligned} & 2[[\kappa_i(t)M_i(t,\rho)\kappa_i(t)^\top + \eta_i(t)S_i(t,\rho)\eta_i(t)^\top \\ & + M_i(t,\rho)^{\text{loc}}]\lambda_i(t)A^\top A \\ & - [\kappa_i(t)M_i(t,\rho)\kappa_i(t)^\top + \eta_i(t)S_i(t,\rho)\eta_i(t)^\top]A^\top A] \circ I \\ & + [\mathbf{1}_m\xi_{i1}^\top(t)] \circ I - [\mathbf{1}_m\xi_{i2}^\top(t)] \circ I = 0, \end{aligned} \tag{37}$$

Proof: Problem (25) is convex. Therefore the Karush Kuhn Tucker conditions are both necessary and sufficient for optimality. In Eqs. (26)–(28) we derive the canonical form of the constraints. Since $\kappa_i(t)$, $\eta_i(t)$ and $\lambda_i(t)$ are composed by diagonal blocks, the Eqs. (35)–(37) are obtained by using matrix derivatives for diagonal and symmetric matrices [17], [18]: given matrices A, $M \succeq \mathbf{0}$ and diagonal matrix X, then

$$\frac{d\,\text{tr}(AXMX^TA^T)}{d\,X} = 2(MXA^TA) \circ I\,. \tag{38}$$

As an extension to this result, given block-diagonal matrix κ, positive semidefinite matrix M and matrix A, then

$$\frac{d\,\mathrm{tr}(A\kappa M\kappa^T A^T)}{d\,\kappa} = 2(M\kappa^T A^T A)\circ(\mathbf{1}\otimes I). \tag{39}$$

□

In the following we analyze the derived KKT conditions to obtain the optimal weight values.

A. The optimal weights

We formulated problem (21) to minimize mean and variance of the local prediction errors and we obtained some optimality conditions for the equivalent problem (25) with more restrictive constraints. Here 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 1: Suppose square matrices M, A and X with same size, where A is symmetric and X is diagonal. Then

$$\mathrm{diag}(MXA) = (M\circ A)\mathrm{diag}(X).$$

Proof: Suppose all matrices have size $n\times n$. We have

$$MX = \begin{bmatrix} M_{11}X_1 & M_{12}X_2 & \cdots & M_{1n}X_n \\ M_{21}X_1 & M_{22}X_2 & \cdots & M_{2n}X_n \\ \vdots & \vdots & \ddots & \vdots \\ M_{n1}X_1 & M_{n2}X_2 & \cdots & M_{nn}X_n \end{bmatrix},$$

then we obtain

$$\mathrm{diag}(MXA) = \begin{bmatrix} \sum_{i=1}^n M_{1i}A_{i1}X_i \\ \sum_{i=1}^n M_{2i}A_{i2}X_i \\ \vdots \\ \sum_{i=1}^n M_{ni}A_{in}X_i \end{bmatrix}.$$

On the other side, we have $(M\circ A)\mathrm{diag}(X) =$

$$= \begin{bmatrix} M_{11}A_{11} & M_{12}A_{12} & \cdots & M_{1n}A_{1n} \\ M_{21}A_{21} & M_{22}A_{22} & \cdots & M_{2n}A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ M_{n1}A_{n1} & M_{n2}A_{n2} & \cdots & M_{nn}A_{nn} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix},$$

Thus for $j = 1,\dots,n$ we have

$$\begin{aligned}[\mathrm{diag}(MXA)]_j &= \sum_{i=1}^n M_{ji}A_{ij}X_i = \sum_{i=1}^n M_{ji}A_{ji}X_i \\ &= (M\circ A)\mathrm{diag}(X).\end{aligned} \tag{40}$$

□

We use this lemma to solve Equations (35) - (37). We now define the $m_{N_i}\times 1$ vector $\kappa_i^{\mathrm{vec}}(t) = \kappa_i^\top(t)\mathbf{1}_m$, collecting all the decision variables of the diagonals of $\kappa_i(t)$ on a column vector. Similarly, we define also $\eta_i^{\mathrm{vec}}(t) = \eta_i^\top(t)\mathbf{1}_m$, $\lambda_i^{\mathrm{vec}}(t) = \lambda_i^\top(t)\mathbf{1}_m$ and $\nu_i^E(t) = [\mathbf{1}_{m_{N_i}}\nu_i^\top]\circ(\mathbf{1}\otimes I)\mathbf{1}_m$, collecting N_i times the vector ν_i. For the sake of notation simplicity, in the following we will omit the dependence on the time when it is not strictly necessary. We now provide an analytical solution for problem (25). It is worth noting that only these formula have to be implemented in each node to compute at each step the optimal weights.

Proposition 4.2: A solution for problem (25) is given by $\xi_{i_1} = 0$, $\xi_{i_2}\neq 0$, $\xi_{i_3} = 0$, $\xi_{i_4} = 0$, $\lambda_i = 0$,

$$\kappa_i^{\mathrm{vec}} = -\left(2M_i\circ(\mathbf{1}_{N_i}\mathbf{1}_{N_i}^T\otimes D_i)\right)^{-1}\nu_i^E, \tag{41}$$

$$\eta_i^{\mathrm{vec}} = -\left(2S_i\circ(\mathbf{1}_{N_i}\mathbf{1}_{N_i}^T\otimes D_i)\right)^{-1}\nu_i^E, \tag{42}$$

$$\begin{aligned}\nu_i = -2(I_{m\times m_{N_i}}((M_i\circ\mathbf{D}_i)^{-1} \\ + (S_i\circ\mathbf{D}_i)^{-1})I_{m\times m_{N_i}}^T)^{-1}\mathbf{1}_m,\end{aligned} \tag{43}$$

where $I_{m\times m_{N_i}} := \mathbf{1}_{N_i}^T\otimes I_m$, $D_i := A^TA$ and $\mathbf{D}_i = \mathbf{1}_{N_i}\mathbf{1}_{N_i}^T\otimes D_i$.

Proof: We observe that the Eq. (37) is equivalent to

$$\begin{aligned}&\mathrm{diag}\left(2(M_i^{\mathrm{loc}} + \kappa_i M_i\kappa_i^T + \eta_i S_i\eta_i^T)\lambda_i A^TA\right) \\ &= \mathrm{diag}\left(2(\kappa_i M_i\kappa_i^T + \eta_i S_i\eta_i^T)A^TA\right) + \xi_{i_2} - \xi_{i_1}.\end{aligned} \tag{44}$$

By using the result from (1), we have

$$\begin{aligned}\lambda_i^{\mathrm{vec}} &= \left(2(M_i^{\mathrm{loc}} + \kappa_i M_i\kappa_i^T + \eta_i S_i\eta_i^T)\circ(A^TA)\right)^{-1} \\ &\left(\mathrm{diag}\left(2(\kappa_i M_i\kappa_i^T + \eta_i S_i\eta_i^T)A^TA\right) + \xi_{i_2} - \xi_{i_1}\right).\end{aligned} \tag{45}$$

Let us now analyze KKT conditions in Eqs. (35) and (36). We find a solution with $\xi_{i_3} = 0$ and $\xi_{i_4} = 0$. Let note that Eq. (35) holds if and only if it holds

$$\begin{aligned}&2(M_i\kappa_i^T(I-\lambda_i)^TA^TA(I-\lambda_i))\circ(\mathbf{1}\otimes I)\,\mathbf{1}_m \\ &\qquad = -\mathbf{1}_{m_{N_i}}v_i^T\circ(\mathbf{1}\otimes I)\,\mathbf{1}_m = -\nu_i^E\end{aligned} \tag{46}$$

Let denote $\kappa_i = [\kappa_{i_1},\cdots,\kappa_{i_{N_i}}]$, where κ_{i_j} is the (j)-th $m\times m$ diagonal block of κ_i, with $j = 1,\dots,N_i$. Similarly we consider $M_{i_{pq}}$, which is the $m\times m$ block of matrix M_i:

$$M_i = \begin{bmatrix} M_{i_{11}} & \cdots & M_{i_{1N_i}} \\ \vdots & \ddots & \vdots \\ M_{i_{N_i1}} & \cdots & M_{i_{N_iN_i}} \end{bmatrix}.$$

Then, by using block operations, we have

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

Thus by using block by block the result in (1):

$$2M_i\circ(\mathbf{1}_{N_i}\mathbf{1}_{N_i}^T\otimes D_i)\kappa_i^{\mathrm{vec}} = -\mathbf{1}_{N_i}\otimes v_i = -\nu_i^E. \tag{47}$$

Remembering that by definition κ_i^{vec} is the vector that collects all the elements from κ_i, from Eq. (47), we obtain

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

and similarly

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

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

$$(\kappa_i+\eta_i)\mathbf{1}_{m_{N_i}} = \mathbf{1}_{N_i}^T\otimes I_m(\kappa_i^{\mathrm{vec}} + \eta_i^{\mathrm{vec}}) = \mathbf{1}_m, \tag{48}$$

which implies that

$$\nu_i = -2(I_{m\times m_{N_i}}((M_i \circ \mathbf{D}_i)^{-1} + (S_i \circ \mathbf{D}_i)^{-1})I^T_{m\times m_{N_i}})^{-1}\mathbf{1}_m \,,$$

where $I_{m\times m_{N_i}}$ denotes $\mathbf{1}^T_{N_i} \otimes I_m$, and $\mathbf{D}_i = \mathbf{1}_{N_i}\mathbf{1}^T_{N_i} \otimes D_i$. In this way we have obtained the optimal value of the dual variable ν_i and, by substituting the result in Eq. (41) and (42), we obtain the optimal values of κ_i and η_i depending on λ_i. We can observe that the KKT conditions hold by imposing $\xi_{i_2} \neq 0$: λ_i has to be null because of Eq.(30) and $\xi_{i_1} = 0$ because of Eq.(29). Therefore, we can find optimal solutions in closed form for κ_i and η_i by substituting $\lambda_i = 0$ in Eq. (41), (42) and (43). Remember that the weights are adaptive and can be computed locally by node i at each step. This is an optimal solution for the Pareto optimization problem (25), if, once computed the solutions on-line, in addition they satisfy the constraints in (31) and (32). □

In the analytical solution given by the previous proposition, one of the convergence conditions constraints is active. Since in problem (25) we have imposed more restrictive conditions than the ones obtained in Section III-C for tractability reasons, and so the solution could be sub–optimal, we propose here an alternative solution of the problem to use the more complex less restrictive computed constraints (16).

Proposition 4.3: A solution for problem (21) is given by $\xi_{i_1} = 0$, $\xi_{i_2} = 0$, $\xi_{i_3} = 0$, $\xi_{i_4} = 0$, Eqs. (45), (41), (42) and (43) with $D_i := (I - \lambda_i)^T A^T A(I - \lambda_i)$.

Proof: "Sketch": we follow the same procedure as in proof of Prop.4.2, but imposing the dual variables ξ_{i_1}, ξ_{i_2}, ξ_{i_3} and ξ_{i_4} all null, so that the constraints are all non active, and letting the time-varying variable λ_i be non-null. In fact, with these conditions, in Eq. (45) we have found an optimal solution of λ_i depending on κ_i and η_i, and in Eqs. (41) and (42) we have found an optimal solution of κ_i and η_i depending on λ_i. □

To use the result in Prop. 4.3, we adopt a computational method to find optimal values κ_i^*, η_i^* and λ_i^*. At each step, we initialize $\lambda_i = I_m/2$, and define a small positive constant ϵ. We first compute κ_i and η_i with the initial given value of λ_i. Then we update λ_i with the computed values of κ_i and η_i. We repeat this procedure until the difference between current value of λ_i and the updated one is smaller than ϵ as shown in Algorithm 1. At each step, we verify that the obtained values satisfy the convergence conditions obtained in Section III-C. Such a second solution method can be used with any matrix A norm, since the convergence conditions constraints are not active.

In this section we have proposed two methods to obtain the values of the optimal time-varying weights for the proposed distributed estimator. Moreover, we showed that in general

Theorem 4.4: Consider optimization problem (21). The consensus-filtering weights κ_i and η_i, and the model-based prediction parameters λ_i cannot be computed by two separate optimization problems.

Proof: "Sketch": from Prop. 4.3, it follows that the optimal solution of λ_i depends on the optimal solutions of κ_i and η_i, and vice versa. □

Algorithm 1 Algorithm

Set $\lambda_i^+ = I_m/2$, ϵ
$I_{m\times m_{N_i}} = \mathbf{1}^T_{N_i} \otimes I_m$
repeat
$\lambda_i = \lambda_i^+$
$\mathbf{D}_i = \mathbf{1}_{N_i}\mathbf{1}^T_{N_i} \otimes (I_m - \lambda_i)^T A^T A(I_m - \lambda_i)$
$\nu_i = -(I_{m\times m_{N_i}}((2M_i \circ \mathbf{D}_i)^{-1} + (2S_i \circ \mathbf{D}_i)^{-1})I^T_{m\times m_{N_i}})^{-1}\mathbf{1}_m$
$\kappa_i^{\text{vec}} = \mathbf{P}_K - (2M_i \circ \mathbf{D}_i)^{-1}\,\mathbf{1}_{N_i} \otimes I_m\,\nu_i$
$\eta_i^{\text{vec}} = -\,(2S_i \circ \mathbf{D}_i)^{-1}\,\mathbf{1}_{N_i} \otimes I_m\,\nu_i$
$\mathbf{R}_i = \kappa_i M_i \kappa_i^T + \eta_i S_i \eta_i^T$
$\lambda_i^{\text{vec}} = \mathbf{P}_\Lambda\left((M_i^{\text{loc}} + \mathbf{R}_i) \circ (A^T A)\right)^{-1} \text{diag}\left(\mathbf{R}_i A^T A\right)$
$\lambda_i^+ = \text{diag}(\lambda_i^{\text{vec}})$
until $|\lambda_i - \lambda_i^+| \leq \epsilon$
return $\kappa_i(t) = \text{diag}(\kappa_i^{\text{vec}})$, $\eta_i(t) = \text{diag}(\eta_i^{\text{vec}})$, and $\lambda_i(t) = \lambda_i^+$

This is an important result, since in many works of the state of the art (see [10], as example), the two phases (consensus-filtering and prediction) are designed separately.

V. Simulation results

In this section, simulation results are given with the purpose of illustrating the analysis performed in this paper. In [1], it is explained how to derive the estimates of the needed quantities to implement the proposed method.

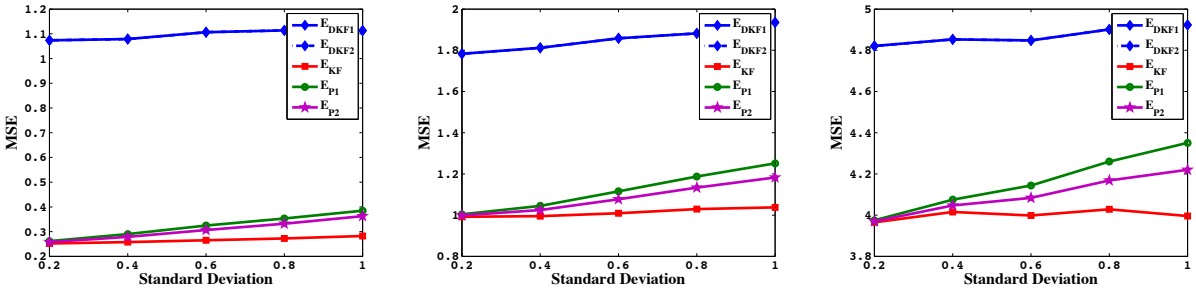
To quantify the performance of the distributed estimator proposed in this paper, we consider a linear system, whose dynamics are described by the matrix:

$$A = \begin{pmatrix} 0.9 & 0.05 \\ 0.25 & 0.6 \end{pmatrix}.$$

We compare the following prediction methods:

E_{DKF1}: Decentralized Kalman filter as in [5].
E_{DKF2}: Decentralized Kalman filter as in [6].
E_{KF}: Centralized Kalman filter. The estimator collects all the measurements and implements the classical Kalman filter.
E_{p1}: The proposed estimator using Algorithm 1.
E_{p2}: The proposed estimator using Prop. 4.2.

Note that the Kalman Filter method can not be implemented in the distributed scenario we are considering due to communication and computation limits. 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 $1.7\sqrt{N}$. The evaluated performance metric is the Mean Square Error (MSE) of the predictions at each node, that is then averaged over all the nodes of the network. We consider different noise scenarios. For each scenario, we repeat the experiment 50 times with a random network topology. We denote MSE the average value over the 50 experiments. We investigate the performances of the considered methods with different values of modeling uncertainty/disturbances and measurement noise. Figs. 1(a), 1(b) and 1(c) show the MSE values for all the estimators with respect to the measurement noise standard deviation, with different values of disturbances standard deviation for each figure. In this first case, we consider Gaussian colored



(a) (b) (c)

Fig. 1. MSE obtained by the three estimators with different values of measurement noise standard deviation. The standard deviation of the disturbance noise is 0.5 in 1(a), 1 in 1(b) and 2 in 1(c).

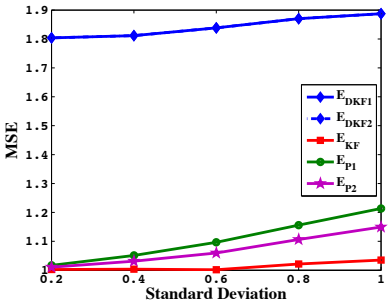


Fig. 2. MSE obtained by the five estimators with different values of measurement noise standard deviation. The measurement noise is a white noise. The standard deviation of the disturbance noise is 1.

noises for the measurement noise and white noises for the disturbances. We can see that in these conditions, the proposed estimator obtains always better results than the distributed Kalman filters and it performs equal or slightly worse than the centralized Kalman filter. Even in the ideal case with Gaussian white noises (Fig. 2), where the Kalman filter is optimal, the proposed predictor continues to behave similarly.

VI. Concluding Remarks

In this paper, we propose a novel distributed prediction method for dynamic systems using sensor networks, able to simultaneously minimize both the bias and the variance of the prediction error. The optimal consensus-filtering and prediction weights are time-varying and can be computed locally at each step by each sensor node. We consider a multi-variable scenario where the components of the state and therefore measurement and modeling noises can be correlated. No knowledge of the noise distribution is assumed.

As a future work, we aim at extending the proposed method by considering the case the state is not observable by each node, and a scenario with the topology changing along the time. Moreover, future efforts will be devoted to a more extensive simulative comparison with distributed prediction methods at the state of the art.

References

[1] F. Boem, Y. Xu, C. Fischione, and T. Parisini, "A distributed estimation method for sensor networks based on pareto optimization," in *Decision and Control, Proc. of 51th Conf. on*, 2012, pp. 775–781.

[2] ——, "Distributed fault detection using sensor networks and pareto estimation," in *European Control Conf.*, July 2013, pp. 932–937.

[3] A. Speranzon, C. Fischione, K. Johansson, and A. Sangiovanni-Vincentelli, "A distributed minimum variance estimator for sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 4, pp. 609–621, 2008.

[4] F. S. Cattivelli and A. H. Sayed, "Diffusion strategies for distributed kalman filtering and smoothing," *Automatic Control, IEEE Trans. on*, vol. 55, no. 9, pp. 2069–2084, 2010.

[5] R. Olfati-Saber, "Distributed kalman filtering for sensor networks," in *Decision and Control, IEEE Conf. on*, 2007, pp. 5492–5498.

[6] ——, "Kalman-consensus filter: Optimality, stability, and performance," in *Decision and Control, IEEE Conf. on, held jointly with 28th Chinese Control Conf.* IEEE, 2009, pp. 7036–7042.

[7] M. H. Davis, *Linear estimation and stochastic control*. London, 1977.

[8] R. Carli, A. Chiuso, L. Schenato, and S. Zampieri, "Distributed kalman filtering based on consensus strategies," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 4, pp. 622 –633, 2008.

[9] ——, "Distributed kalman filtering using consensus strategies," in *Decision and Control, IEEE Conf. on*, Dec 2007, pp. 5486–5491.

[10] P. Alriksson and A. Rantzer, "Distributed kalman filtering using weighted averaging," in *Proc. of the 17th Int. Symposium on Mathematical Theory of Networks and Systems*, 2006, pp. 1–60.

[11] D. Di Paola, A. Petitti, and A. Rizzo, "Distributed kalman filtering via node selection in heterogeneous sensor networks," *International Journal of Systems Science*, no. ahead-of-print, pp. 1–12, 2014.

[12] R. Carli, G. Como, P. Frasca, and F. Garin, "Distributed averaging on digital erasure networks," *Automatica*, vol. 47, no. 1, pp. 115–121, 2011.

[13] V. Delouille, R. Neelamani, and R. G. Baraniuk, "Robust distributed estimation using the embedded subgraphs algorithm," *IEEE Trans. on Signal Processing*, vol. 54, pp. 2998–3010, August 2006.

[14] C. G. Lopes and A. H. Sayed, "Diffusion lest-mean squares over adaptive networks: Formulation and performance analysis," *IEEE Trans. on Signal Processing*, vol. 56, no. 7, pp. 3122–3135, 2008.

[15] F. S. Cattivelli, C. G. Lopes, and A. H. Sayed, "Diffusion recursive least-squares for distributed estimation over adaptive networks," *IEEE Trans. on Signal Processing*, vol. 56, pp. 1865–1877, 2008.

[16] P. H. Bauer, K. Premaratne, and J. Duran, "A necessary and sufficient condition for robust asymptotic stability of time-variant discrete systems," *IEEE Trans. on Automatic Control*, vol. 38, no. 9, pp. 1427–1430, 1993.

[17] K. B. Petersen and M. S. Pedersen, "The matrix cookbook," 2006.

[18] T. P. Minka, "Old and new matrix algebra useful for statistics," 2001.