Automatica 73 (2016) 27–37

Contents lists available at ScienceDirect

Automatica

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

Dynamic state estimation for power networks using distributed MAP technique☆

Yibing Sun [a], Minyue Fu [c,b], Bingchang Wang [a], Huanshui Zhang [a], Damián Marelli [c,d]

[a] School of Control Science and Engineering, Shandong University, Jinan 250061, China
[b] School of Electrical Engineering and Computer Science, University of Newcastle, NSW 2308, Australia
[c] School of Automation, Guangdong University of Technology, Guangzhou, China
[d] CIFASIS/CONICET, 2000 Rosario, Argentina

ARTICLE INFO

Article history:
Received 15 May 2015
Received in revised form
21 May 2016
Accepted 31 May 2016
Available online 3 September 2016

Keywords:
Distributed state estimation
Distributed MAP estimation
Kalman filter
Power systems

ABSTRACT

This paper studies a distributed state estimation problem for a network of linear dynamic systems (called nodes), which evolve autonomously, but their measurements are coupled through neighborhood interactions. Power networks are typical networked systems obeying such features, with other examples including traffic networks, sensor networks and many multi-agent systems. We develop a new distributed state estimation approach, for each node to update its local state. The core of this distributed approach is a distributed maximum *a posteriori* (MAP) estimation technique, which delivers a globally optimal estimate under certain assumptions. We apply the distributed approach to an IEEE 118-bus system, and compare it with a centralized approach, which provides the optimal state estimate using all the measurements, and with a local state estimation approach, which uses only local measurements to estimate local states. Simulation results show that under different scenarios including normal operation, bad measurements and sudden load change, the distributed approach is clearly more accurate than the local state estimation approach and distributed static state estimation approach. Although the result is a bit less accurate than that by a centralized algorithm, the distributed algorithm enjoys low computational complexity and communication load, and is scalable to large power networks.

1. Introduction

As our society emphasizes the importance of smarter electricity networks to support sustainable energy utilization, power networks are undergoing tremendous changes. The main goal is to maintain the smart grid in an efficient, secure and reliable operating environment, for the production and distribution of electricity. In practice, a wide range of uncertainties in measurements and communication give rise to inaccuracies, which may affect the performance of optimization and control algorithms, and finally affect the stability of the power plant. State estimation is a necessary tool to deal with these uncertainties. Ever since the introduction of state estimation, by Schweppe and Wildes (1970), this area has received increasing attention from researchers in different fields, and particularly in recent years, owing to the multidisciplinary nature of smart grids (Wu, 1990). State estimators are broadly utilized to obtain, from redundant noisy measurements, an estimation of the state of a subnetwork, which is not directly monitored, for computational or economical reasons. State estimation is a key module in the energy management system (EMS) and plays a vital role in security analysis, power dispatch, voltage stability analysis, economic optimization, optimal power flow, diagnosis and recovery (Wu, Moslehi, & Bose, 2005).

State estimation algorithms for power networks can be roughly divided into three classes: centralized, hierarchical and distributed ones. In centralized state estimation, there is a control center which aggregates all measurements over the whole power network and provides the optimal state estimate of the entire network. In the 1980s, some researchers put forward hierarchical

☆ This work is supported by the Taishan Scholar Construction Engineering by Shandong Government, the National Natural Science Foundation of China under Grants 61120106011, 61573221, 61403233. The material in this paper was partially presented at the 14th annual European Control Conference, July 15–17, 2015, Linz, Austria. This paper was recommended for publication in revised form by Associate Editor Wei Xing Zheng under the direction of Editor Torsten Söderström.

E-mail addresses: sun_yibing@126.com (Y. Sun), minyue.fu@newcastle.edu.au (M. Fu), bcwang@sdu.edu.cn (B. Wang), hszhang@sdu.edu.cn (H. Zhang), damian.marelli@newcastle.edu.au (D. Marelli).

http://dx.doi.org/10.1016/j.automatica.2016.06.015
0005-1098/

estimation techniques (Bose, Abur, Poon, & Emami, 2010; Cutsem & Ribbens-Pavella, 1983; Gomez-Exposito & de la Villa Jaen, 2009; Jiang, Vittal, & Heydt, 2008). Cutsem and Ribbens-Pavella (1983) proposed a hierarchical estimation method for multi-area power systems, in which the local state estimation, obtained at the first hierarchical level, is coordinated at a higher level. In Bose et al. (2010), a *star-like* hierarchical state estimation scheme was proposed. However, hierarchical estimation methods still require a centralized coordinator to perform global estimation. Thus, centralized and hierarchical state estimation approaches may suffer from communication bottlenecks and reliability issues, due to the following reasons.

There are at least two major aspects that make it difficult to support the development of the above two state estimation methods. Firstly, the measurements used in the state estimation methods are traditionally captured by a Supervisory Control and Data Acquisition (SCADA) system (Abur & Exposito, 2004). However, this system has intrinsic limitations, i.e., low sampling rate and relatively low accuracy of measurements which limit the reliability of state estimation. Modern power systems are likely to involve many fast measuring and processing devices, such as Phasor Measurement Units (PMUs), featured with synchronous sampling and high data updating rates (Huang, Werner, Huang, Kashyap, & Gupta, 2012; Tai, Marelli, Rohr, & Fu, 2013). More importantly, PMUs can directly measure both the voltage magnitudes and phase angles of the bus, which make simple linear state estimation possible, giving rise to higher precision and faster calculation than conventional nonlinear methods (Tai et al., 2013). However, it is impossible to install a PMU on each local area, due to the expensive cost of PMUs. Thus, it is a challenge to combine PMU measurements with conventional measurements to obtain an optimal state estimation. Secondly, policy and market pricing competition require utility companies to share more information and monitor the power network over wide-scale areas. Furthermore, the deregulation of electricity industry has led to the creation of regional transmission organizations (RTOs), within a large-scale complex system (Wu et al., 2005). Therefore, there is an urgent need for distributed estimation algorithms. Distributed state estimation algorithms are necessary for many other types of large-sized networked systems, such as traffic networks and sensor networks; see Feng and Zeng (2012), Liang, Wang, and Liu (2011), Olfati-Saber (2007) and Yu, Chen, Wang, and Yang (2009).

Distributed state estimation algorithms can also be divided into static and dynamic ones. Under stationary operational conditions, the power system is usually treated as a quasi-static system, whose operating condition is fully characterized by variables such as bus loads, line flows, generation, and bus voltages (magnitudes and phase angles) at a given point of time. Among these interdependent variables, one can choose the bus voltages as the system state, which is referred to as the static-state of the system to avoid misunderstanding on what type of variable is being considered (Do Coutto Filho & Stacchini de Souza, 2009). Many studies have been made on distributed static estimation problems; see Conejo, de la Torre, and Canas (2007), Falcao, Wu, and Murphy (1995), Lin (1992), Pasqualetti, Carli, and Bullo (2012), Tai, Lin, Fu, and Sun (2013) and Xie, Choi, and Kar (2011). Tai et al. (2013) proposed a distributed weighted least-squares (WLS) estimation approach for static state estimation with the property that the local estimates converge to the same estimates obtained via a centralized estimator. In this scheme, each local estimator only needed to know its local measurements and low dimensional boundary information exchanged from neighboring nodes, which results in a much lighter communication load than that of Pasqualetti et al. (2012) and Xie et al. (2011). However, when transient dynamics are considered, power networks are typically modeled as dynamic systems. The dynamic change of loads gives rise to the adjustment of generators, which in turn leads to a change in flows and injections at all buses. These dynamic changes cannot be captured by the static state estimation methods, thus prompting for the development of dynamic state estimation methods (Cattivelli, Lopes, & Sayed, 2008; Do Coutto Filho & Stacchini de Souza, 2009; Khan, Ilic, & Moura, 2008; Shih & Huang, 2002; Valverde & Terzija, 2011; Wang, Gao, & Meliopoulos, 2012). These methods are mostly based on the Kalman filtering technique; see Do Coutto Filho and Stacchini de Souza (2009), Huang et al. (2012), Shih and Huang (2002), Valverde and Terzija (2011) and Wang et al. (2012). Compared with traditional static estimation schemes, dynamic state estimation methods have better accuracy and the ability to predict the future state, which is valuable for performing security analysis and real-time control.

The objective of this paper is to develop a fully distributed dynamic estimation method for large-scale interconnected networks. Our method requires that the communication graph of the power network is acyclic, which is a valid assumption for many practical systems. The major advantage of the distributed MAP estimation algorithm is that only local computation and communication are needed. We also show that the distributed algorithm at steady state converges in a finite number of iterations, which is equal to the maximum path length of the acyclic graph. The main contribution of this paper is that it generalizes known results on distributed static state estimation (Tai et al., 2013) to the dynamic case, resulting in a distributed dynamic state estimator which delivers better estimation accuracy than that by Tai et al. (2013). Furthermore, we apply our distributed approach to an IEEE 118-bus system, and compare it with a centralized approach, which provides the optimal state estimate using all the measurements, a local state estimation approach, which uses only local measurements to estimate local states, and the distributed static state estimation algorithm obtained in Tai et al. (2013). Simulation results using the IEEE 118-bus system in different scenarios show that the distributed approach clearly offers more accurate estimates than the local estimation and static estimation approaches do. Although the simulation experiments show that the accuracy of distributed MAP estimator is somehow worse than that of the centralized state estimator, the distributed algorithm enjoys low computational complexity and communication load, making the method scalable to large-sized power networks.

The rest of the paper is organized as follows. Section 2 introduces the system model and problem formulation. Section 3 details the centralized state estimation approach. Section 4 describes the local state estimation scheme. Section 5 gives the distributed MAP estimation algorithm. Simulation examples are presented in Section 6, and Section 7 concludes the paper. Some proofs are contained in the Appendix.

2. Problem description

For completeness, we first introduce some notation and preliminaries on algebraic graph theory and matrices, which will be used in the rest of the paper.

Notation. $\mathbb{R}^l$ denotes the set of l-dimensional real column vectors and $\mathbb{R}^{l\times q}$ denotes the set of $l \times q$ real matrices. The superscript ^{T} denotes the transpose of a vector or matrix. $\mathbb{E}\{x\}$ denotes the expectation of the random variable x. Also, $\text{diag}\{A_1, A_2, \ldots, A_n\}$ denotes a block diagonal matrix with the diagonal blocks being the matrices $A_1, \ldots, A_n$.

Throughout the paper, we consider a multi-area interconnected power network, to which we associate a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. The set $\mathcal{V} = \{1, \ldots, n\}$ contains the nodes, each of which corresponds to a control area, and $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ denotes the set of edges

$(i,\ j)$, connecting nodes i and j. Each edge in $\mathcal{E}$ corresponds to a pair of nodes indicating that there is an edge measurement $z_{i,j}$ depending on x_i and x_j, which are the states of nodes i and j. We use $\mathcal{N}_i = \{j : (i, j) \in \mathcal{E}\}$ to denote the set of neighbors of node i. We assume that the graph is connected, undirected (i.e., there exists a two-way path between each pair of nodes) and acyclic (i.e., it does not contain loops). For a connected graph $\mathcal{G}$, without loops, the length of a path is the number of edges forming it. The radius ε_i of node i is defined as the maximum length of a path between node i and any other node in the graph. The diameter of the graph is $\Gamma = \max\{\varepsilon_i : i \in \mathcal{V}\}$.

This paper proposes a distributed state estimation algorithm using mixed SCADA and PMU measurements. The available SCADA measurements are limited to slow varying quantities, such as voltage and power flows, whereas the PMU measurements include instantaneous voltages and their phases. With the high sampling rate, we can obtain several PMU measurements per second in the linear form, while the SCADA system captures only one sample per several seconds in the nonlinear form. To deal with this problem, we first average the PMU measurements within each SCADA sampling period to give a single measurement so that the averaged measurement is synchronized with the SCADA measurements and that the amount of data to deal with is reduced. Secondly, we apply the standard linearization technique to the SCADA measurements so that they can be regarded as linear measurements around the operating points, available at each SCADA sampling time (see more details later). These two types of measurements are then used together to estimate the state of the power system.[1]

Consider the state vector of each bus i expressed as the following dynamic model:

$$x_i(k+1) = f_i(x_i(k), k) + \omega_i(k), \tag{1}$$

where $x_i(k) = (|V_i(k)|,\ \theta_i(k))^T$ is the state of bus i, $V_i(k)$ and $\theta_i(k)$ are voltage magnitude and angle at time instant k, the nonlinear function f_i represents the dynamic drift of the state due to load changes, and $\omega_i(k)$ is the Gaussian process noise. The sampling time k takes values of $0, 1, 2, 3, \ldots$. The initial state $x_i(0)$ is assumed to be Gaussian.

Using the standard linearization approach around an operating point $x_i^0(k)$, we can convert the state space model (1) into the following model (Do Coutto Filho & Stacchini de Souza, 2009; Valverde & Terzija, 2011):

$$x_i(k+1) = A_i(x_i^0(k), k)x_i(k) + G_i(x_i^0(k), k) + \omega_i(k), \tag{2}$$

where $A_i(x_i^0(k), k)$ is an diagonal matrix and $G_i(x_i^0(k), k)$ represents the trend behavior of the state trajectory, namely,

$$A_i(x_i^0(k), k) = \frac{\partial f_i(x_i, k)}{\partial x_i}\Big|_{x_i = x_i^0};$$

$$G_i(x_i^0(k), k) = f_i(x_i^0(k), k) - A_i(x_i^0(k), k)x_i^0(k).$$

In practice, we can replace $x_i^0(k)$ with the state prediction. Alternatively, the matrices $A_i(x_i^0(k), k)$ and $G_i(x_i^0(k), k)$ can be identified on-line by using the Holt's 2-parameter linear exponential smoothing method of forecasting (Makridakis & Wheelwright, 1978).

We assume that all the buses are installed with SCADA measurements, which can measure voltage magnitudes of each bus and power flows between two connected buses. So the measurement equations of each bus i can be divided into two types: *local measurements* and *edge measurements* which are given by

$$\begin{aligned} z_{i,i}(k) &= \begin{bmatrix} 1 & 0 \end{bmatrix} x_i(k) + \nu_{i,i}(k), \\ z_{i,j}(k) &= h_{i,j}(x_i(k), x_j(k)) + \nu_{i,j}(k), \quad j \in \mathcal{N}_i, \end{aligned} \tag{3}$$

where $z_{i,i}(k)$ is the local measurement of node i, $z_{i,j}(k)$ is the edge measurement describing the interaction between buses i and j, $\nu_{i,i}(k)$ and $\nu_{i,j}(k)$ are the associated measurement noises, and $h_{i,j}(x_i(k), x_j(k)) = (P_{i,j}(k), Q_{i,j}(k))^T$ is the power flow equations in transmission lines (Monticelli, 1999), with

$$\begin{aligned} P_{i,j}(k) = {} & g_{ij}|V_i(k)|^2 - g_{ij}|V_i(k)|\,|V_j(k)|\cos(\theta_{ij}(k)) \\ & - b_{ij}|V_i(k)|\,|V_j(k)|\sin(\theta_{ij}(k)), \end{aligned} \tag{4}$$

$$\begin{aligned} Q_{i,j}(k) = {} & -(b_{ij} + b_{ij}^{sh})|V_i(k)|^2 + b_{ij}|V_i(k)|\,|V_j(k)| \\ & \times \cos(\theta_{ij}(k)) - g_{ij}|V_i(k)|\,|V_j(k)|\sin(\theta_{ij}(k)), \end{aligned} \tag{5}$$

where the formulas g_{ij} and b_{ij} are the series conductance and susceptance, b_{ij}^{sh} is admittance between circuits and ground, and $\theta_{i,j}(k) = \theta_i(k) - \theta_j(k)$ is the difference of voltage angles between buses i and j.

If bus i is installed with a PMU, then there exists an additional local measurement equation, which is

$$z'_{i,i}(k) = x_i(k) + \nu'_{i,i}(k).$$

Since PMU measurements are much more accurate than SCADA measurements, the covariance of noise $\nu'_{i,i}(k)$ is much smaller than that of $\nu_{i,i}(k)$. Moreover, PMU can directly measure the voltage angle $\theta_i(k)$, so we use $z'_{i,i}(k)$ rather than $z_{i,i}(k)$ as the local measurement of bus i, and for uniformity, we write $z_{i,i}(k)$ as follows:

$$z_{i,i}(k) = C_i x_i(k) + \nu_{i,i}(k). \tag{6}$$

Similarly in Leite da Silva, Do Coutto Filho, and de Queiroz (1983), linearizing around the operating points $x_i^0(k)$ and $x_j^0(k)$ of (3), we have

$$\begin{aligned} z_{i,j}(k) = {} & h_{i,j}(x_i^0(k), x_j^0(k)) + B_{ij}(k)(x_i(k) - x_i^0(k)) \\ & + B_{ji}(k)(x_j(k) - x_j^0(k)) + \nu_{i,j}(k), \end{aligned} \tag{7}$$

where

$$B_{ij}(k) = \frac{\partial h_{i,j}(x_i, x_j)}{\partial x_i}\Big|_{x_i = x_i^0, x_j = x_j^0},$$

$$B_{ji}(k) = \frac{\partial h_{i,j}(x_i, x_j)}{\partial x_j}\Big|_{x_i = x_i^0, x_j = x_j^0},$$

are the Jacobian matrix of $h_{i,j}(\cdot)$ depending on $x_i^0(k)$ and $x_j^0(k)$. For notational simplicity, we will drop the time, $x_i^0(k)$ and $x_j^0(k)$ dependence in A_i, G_i, B_{ij} and B_{ji} in the sequel. Furthermore, the vector $z_{i,j}(k)$ will replace $z_{i,j}(k) - h_{i,j}(x_i^0(k), x_j^0(k)) + B_{ij}x_i^0(k) + B_{ji}x_j^0(k)$; i.e., (2) and (7) can be written as

$$x_i(k+1) = A_i x_i(k) + G_i + \omega_i(k), \tag{8}$$

$$z_{i,j}(k) = B_{ij}x_i(k) + B_{ji}x_j(k) + \nu_{i,j}(k). \tag{9}$$

As mentioned earlier, A_i and G_i can be evaluated at the estimated operating point of $x_i^0(k)$, or by an online identification method. The computation of B_{ij} and B_{ji} can also be done using the estimates of $x_i^0(k)$ and $x_j^0(k)$. We will ignore the resulting approximation errors.

Remark 1. Actually, the parameters A_i, G_i, B_{ij} and B_{ji} are time-variant, and we just simplify these notations here.

[1] Strictly speaking, the use of a Gaussian noise model $w_i(k)$ results in an inaccurate dynamic model because the voltage magnitudes should always be non-negative. However, voltage magnitudes are in the range of 100s–100,000s volts and are typically well regulated, which means that the variances of $w_i(k)$ should be small, which in turn means that the probability of voltage magnitudes becoming negative is negligible.

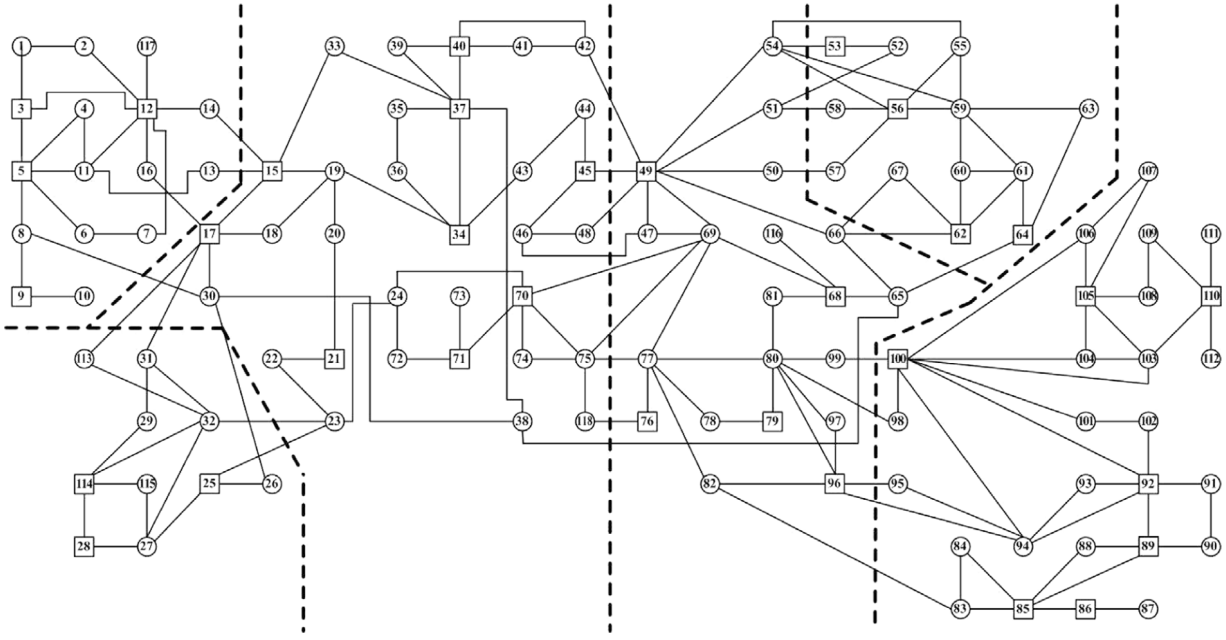

Fig. 1. Topological structure of the IEEE 118-bus system.

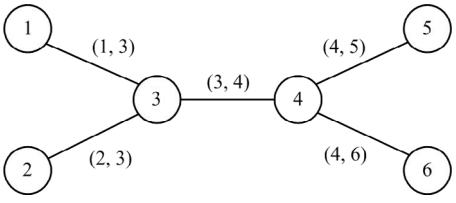

Fig. 2. The graph $\mathcal{G}$ depicting the partition of the 118-bus system.

Remark 2. We use the term 'node' to refer to a cluster of buses, as depicted in Fig. 1. The state of each node is formed by stacking the states of all its buses. Fig. 2 shows the resulting network, where a link between two nodes indicates the existence of joint measurements involving the corresponding two states. In order to avoid introducing more notation, in the rest of the paper, we use i to index the different nodes (rather than buses) of a given system, and we use Eqs. (8), (6) and (9) to describe the model of node (rather than bus) i. We use s_i to denote the dimension of the state x_i.

The local measurements include those involving the buses within node i only (e.g., voltage magnitudes and angles at the buses), whereas the edge measurements are the so-called tie-line measurements, including those involving both nodes i and j (e.g., power flow measurements across two buses, one in node i and one in node j). It is assumed that the measurement vector $z_{i,j}(k)$ is shared by both nodes i and j. We also assume that the noises $\omega_i(k)$, $\nu_i(k)$ and $\nu_{i,j}(k)$ are independent white Gaussian with zero mean and covariances R_i, S_i and $T_{i,j}$, respectively. Furthermore, the initial state $x_i(0)$ is a Gaussian variable, independent from $\omega_i(k)$, $\nu_i(k)$ and $\nu_{i,j}(k)$, with mean $\bar{x}_i(0)$ and covariance $\Sigma_i(0)$.

Fig. 2 shows the graph $\mathcal{G}$ resulting from the partition made of the 118-bus system in Fig. 1. Here, edge (1, 3), for example, means that there exists an edge measurement associated with nodes 1 and 3, and they can exchange data with each other.

Let us denote the aggregate state and measurement by

$$x(k) = (x_1^T(k),\ \ldots,\ x_n^T(k))^T;$$
$$z(k) = (\ldots,\ z_{i,i}^T(k),\ \ldots,\ z_{i,j}^T(k),\ \ldots)^T,$$

respectively. Then, the state and measurement equations of the whole power network, take the following forms:

$$x(k+1) = Ax(k) + G + \omega(k), \tag{10}$$
$$z(k) = Hx(k) + \nu(k), \tag{11}$$

where $A = \text{diag}\{A_1,\ \ldots,\ A_n\}$ is a diagonal matrix according with the assumption in Shih and Huang (2002),

$$G = (G_1^T,\ \ldots,\ G_n^T)^T,$$

$$H = \begin{bmatrix} & & & \vdots & & & \\ \cdots & 0 & C_i & 0 & \cdots & & \\ & & & \vdots & & & \\ \cdots & 0 & B_{ij} & 0 & B_{ji} & 0 & \cdots \\ & & & \vdots & & & \end{bmatrix},$$

and the noises $\omega(k) = (\omega_1^T(k),\ \ldots,\ \omega_n^T(k))^T$ and $\nu(k) = (\ldots,\ \nu_{i,i}^T(k),\ \ldots,\ \nu_{i,j}^T(k),\ \ldots)^T$, have covariance matrices $R = \text{cov}(\omega(k)) = \text{diag}\{R_1,\ \ldots,\ R_n\}$ and $R_* = \text{cov}(\nu(k)) = \text{diag}\{\ldots,\ S_i,\ \ldots,\ T_{i,j},\ \ldots\}$, respectively. Due to the interconnection structure of the power system, we note that the measurement matrix H is usually sparse. Also, the initial state $x(0)$ has mean $\bar{x}(0) = (\bar{x}_1^T(0),\ \ldots,\ \bar{x}_n^T(0))^T$ and covariance $\Sigma(0) = \text{diag}\{\Sigma_1(0),\ \ldots,\ \Sigma_n(0)\}$.

We make the following two assumptions:

Assumption 1. The graph $\mathcal{G}$ is acyclic.

Assumption 2. The matrix H has full column rank and the covariances of noises R, R_* and the initial state $\Sigma(0)$ are positive definite.

Notice that Assumption 1 means that the graph resulting from a partition of the power network does not have cycles. There is a large number of practical power systems which can be abstracted as acyclic graphs, such as the power systems constructed along the coastline. Also, many microgrids are designed specifically to have a radial, and therefore acyclic, network. Fig. 2 shows how the IEEE 118-bus system of Fig. 1 can be partitioned into an acyclic network. More precisely, each node in Fig. 2 corresponds to a cluster of nodes from the IEEE 118-bus system, as indicated in Fig. 1.

Assumption 2 implies that, when the measurements for all the nodes are collectively available, the network satisfies the so-called *topological observability* (TO) condition (Xie et al., 2011). TO is an essential requirement for power networks. This guarantees that state estimation error covariance will be bounded when the measurements of all the nodes are jointly used. Due to operational safety reasons, EMS for every power network is designed to ensure that TO is always satisfied (Wu, 1990). It is also a common requirement for EMS that TO is guaranteed with sufficient redundancy in the sense that even in the event of partial measurement failure (loss of certain measurements or existence of certain outliers), TO is still preserved using the remaining measurements. For EMSs equipped with SCADA measurements, not only TO is guaranteed, the so-called *local topological observability* is usually ensured as well, which means that the local measurements in one area are sufficient to give a local state estimate with a bounded estimation error covariance. The IEEE 118-bus system is an example of such. For EMSs equipped with PMUs only, numerous algorithms exist for the placement of PMUs to ensure the TO; see, e.g., Tai et al. (2013) and the references thereof. With this background, it is practical and realistic to assume that Assumption 2 holds.

3. Centralized state estimation

This section describes the (standard) centralized state estimation approach. Let $\hat{x}(k|k)$ denote the estimate of $x(k)$, conditioned on the measurements from time 0 to k, and $\Sigma(k|k)$ be the associated estimation error covariance matrix. Also, let $\hat{x}(k|k-1)$ denote the one-step-ahead prediction of $x(k)$, conditioned on the measurements from time 0 to $k-1$, and $\Sigma(k|k-1)$ denote its covariance matrix. We have (Anderson & Moore, 1979)

$$\hat{x}(k|k) = \mathbb{E}\{x(k)|Z(k)\},$$
$$\Sigma(k|k) = \mathbb{E}\{(x(k)-\hat{x}(k|k))(x(k)-\hat{x}(k|k))^T|Z(k)\},$$

where $Z(k) = \{z(0),\ z(1),\ \ldots,\ z(k)\}$, and $\hat{x}(k|k-1)$ and $\Sigma(k|k-1)$ are expressed in a similar way.

Let $p(x(k)|Z(k))$ denote the probability density function of $x(k)$ conditioned on the measurements $Z(k)$, and $p(x(k)|Z(k-1))$ can be defined similarly. Then, they are given by the following Gaussian distributions:

$$p(x(k)|Z(k-1)) = \mathcal{N}(x(k);\ \hat{x}(k|k-1),\ \Sigma(k|k-1)),$$
$$p(x(k)|Z(k)) = \mathcal{N}(x(k);\ \hat{x}(k|k),\ \Sigma(k|k)). \tag{12}$$

The centralized state estimator is obtained using a standard Kalman filter. It comprises two steps, namely, update and prediction. The update step is done by using MAP estimator and the prediction step uses a one-step-ahead predictor. We detail each step below.

3.1. Centralized MAP estimation

In this section, our main focus is on the update step. Its purpose is to compute $\hat{x}(k|k)$ from the posterior density $p(x(k)|Z(k))$, using a centralized MAP estimator. Different from the weighted least square (WLS) estimator, the MAP estimator also incorporates the prior density $p(x(k)|Z(k-1))$ in computing $p(x(k)|Z(k))$, thus offering a much more accurate estimate. As described in Kay (1993), the MAP estimator $\hat{x}^{MAP}(k|k)$ is the value that maximizes the posterior density, i.e.,

$$\hat{x}^{MAP}(k|k) = \arg\max_{x(k)} p(x(k)|Z(k)).$$

Using the Bayesian rule (Arulampalam, Maskell, Gordon, & Clapp, 2002) and from (12), we get

$$\begin{aligned} p(x(k)|Z(k)) &\propto p(z(k)|x(k))p(x(k)|Z(k-1)) \\ &\propto \exp\Big[-\frac{1}{2}(e^T(k)R_*^{-1}e(k) \\ &\quad + \Delta^T(k)\Sigma^{-1}(k|k-1)\Delta(k))\Big], \end{aligned}$$

where $e(k) = z(k) - Hx(k)$ and $\Delta(k) = x(k) - \hat{x}(k|k-1)$. We can see that maximizing $p(x(k)|Z(k))$ is equivalent to minimizing its negative logarithm, which leads to

$$\begin{aligned} \hat{x}^{MAP}(k|k) = \arg\min_{x(k)}\big[&(e^T(k)R_*^{-1}e(k) \\ &+\Delta^T(k)\Sigma^{-1}(k|k-1)\Delta(k))\big]. \end{aligned}$$

Then, it is easy to obtain that the optimal estimate and the estimation error covariance are

$$\begin{aligned} \hat{x}^{MAP}(k|k) = &\big(H^TR_*^{-1}H + \Sigma^{-1}(k|k-1)\big)^{-1} \\ &\times \big(H^TR_*^{-1}z(k) + \Sigma^{-1}(k|k-1)\hat{x}(k|k-1)\big), \end{aligned} \tag{13}$$

$$\Sigma^{MAP}(k|k) = \big(H^TR_*^{-1}H + \Sigma^{-1}(k|k-1)\big)^{-1}, \tag{14}$$

initialized by $\hat{x}(0|-1) = \bar{x}(0)$ and $\Sigma(0|-1) = \Sigma(0)$.

3.2. Centralized prediction

This step computes the optimal prediction of the state vector $x(k+1)$ based on the available measurements $Z(k)$, i.e., the conditional mean $\hat{x}(k+1|k)$. By taking conditional expectation on both sides of (10), and using the previously obtained state estimation $\hat{x}^{MAP}(k|k)$, with its associated error covariance $\Sigma^{MAP}(k|k)$, we have

$$\hat{x}(k+1|k) = A\hat{x}^{MAP}(k|k) + G,$$
$$\Sigma(k+1|k) = A\Sigma^{MAP}(k|k)A^T + R.$$

Once $z(k+1)$ becomes available, the centralized state estimation $\hat{x}^{MAP}(k+1|k+1)$, and its error covariance $\Sigma^{MAP}(k+1|k+1)$, are computed using the priors $\hat{x}(k+1|k)$ and $\Sigma(k+1|k)$, as described in the previous subsection. In this way, the optimal estimation and its associated covariance are obtained in a recursive manner.

As mentioned earlier, the state estimate obtained from the centralized MAP estimation scheme is optimal. However, the centralized state estimator creates a heavy computational burden and a communication bottleneck for a large-scale network. To avoid this, we will describe a local state estimation method and a distributed one in the next two sections.

4. Local state estimation

We describe a local state estimation method, which only uses local measurements. This method is essentially the same as the centralized one, with the difference in that edge measurements are not employed.

Using (6) and (8), the local Kalman filtering algorithm for every node $i \in \mathcal{V}$, is given as follows:

Measurement-update equations:

$$\hat{x}_i(k|k) = \hat{x}_i(k|k-1) + K_i(k)\big(z_{i,i}(k) - C_i\hat{x}_i(k|k-1)\big),$$
$$\Sigma_i(k|k) = \Sigma_i(k|k-1) - K_i(k)C_i\Sigma_i(k|k-1),$$

initialized by $\hat{x}_i(0|-1) = \bar{x}_i(0)$, $\Sigma_i(0|-1) = \Sigma_i(0)$, and

$$\Xi_i(k) = C_i\Sigma_i(k|k-1)C_i^T + S_i,$$
$$K_i(k) = \Sigma_i(k|k-1)C_i^T\Xi_i^{-1}(k)$$

are the covariance of the innovation term $z_{i,i}(k) - C_i\hat{x}_i(k|k-1)$ and the Kalman gain, respectively.

Time-update equations:

$$\hat{x}_i(k+1|k) = A_i\hat{x}_i(k|k) + G_i,$$
$$\Sigma_i(k+1|k) = A_i\Sigma_i(k|k)A_i^T + R_i.$$

Obviously, this method leads to a suboptimal estimator. However, it will play a role in the performance comparison of Section 6.

5. Distributed state estimation

This section bears the main contribution of this paper. We will describe the proposed distributed state estimation method. It also involves two steps: a distributed MAP estimator, which replaces the centralized MAP estimator in the centralized control center, and a local state predictor, which is the same as the local state estimator described in Section 4.

5.1. Distributed MAP estimation

In the distributed MAP estimation algorithm, each node $i \in \mathcal{V}$ obtains, at time instant k, an estimate $\hat{x}_i(k|k)$ of the local state $x_i(k)$. This is done by using (6) and (9), and the exchanged information from its neighboring nodes $j \in \mathcal{N}_i$, along an iterative procedure. The purpose is to minimize the global objective function

$$J(x(k)) = e^T(k)R_*^{-1}e(k) + \Delta^T(k)\tilde{\Sigma}^{-1}(k|k-1)\Delta(k),$$

where $\tilde{\Sigma}(k|k-1) = \text{diag}\{\Sigma_1(k|k-1), \ldots, \Sigma_n(k|k-1)\}$ with $\Sigma_1(k|k-1), \ldots, \Sigma_n(k|k-1)$ being the diagonal elements of $\Sigma(k|k-1)$.

Remark 3. Although we assume that $\Sigma(0)$ is a block diagonal matrix, the centralized estimation method leads to $\Sigma(k-1|k-1)$, for $k > 1$, being no longer diagonal. Hence, $\Sigma(k|k-1)$ is not diagonal either. As explained above, in our distributed MAP estimation method we replace $\Sigma(k|k-1)$ with $\tilde{\Sigma}(k|k-1)$. According to the Kalman filter theory (Anderson & Moore, 1979), such modification means that optimality is no longer maintained, i.e., the resulting state estimation becomes suboptimal. However, this algorithm will eliminate many of the numerical difficulties present when a full covariance matrix is used, and in practice this suboptimal estimator may perform better than the optimal one (Larson, Tinney, & Peschon, 1970). We denote this suboptimal estimate by $\hat{x}^*(k|k) = \arg\min_{x(k)} J(x(k))$, and its associated estimation error covariance by $\Sigma^*(k|k)$. Also, for each $i = 1, \ldots, n$, we use $\Sigma_i^*(k|k)$ to denote the block diagonal sub-matrix in $\Sigma^*(k|k)$ corresponding to the state $x_i(k)$.

Our next step is to derive the proposed distributed MAP estimator, for carrying out the update step. This requires computing the state estimate $\hat{x}_i(k|k)$, and its error covariance $\Sigma_i(k|k)$, for each $i \in \mathcal{V}$. We do this derivation in three steps. In Lemma 1 we solve the distributed MAP estimation problem for a network formed only by two nodes. Then, Lemma 2 generalizes this result to networks forming a *radial graph*, i.e., in which there exists a single node (called the central node), whose radius is one, and the radii of all remaining nodes (called leaf nodes) are two (an example is shown in Fig. 3). Finally, using Lemma 2, we derive our proposed distributed MAP estimation algorithm, and we show in Theorem 1 that it is applicable to networks with arbitrary acyclic topology. The proofs of these results appear in the Appendix.

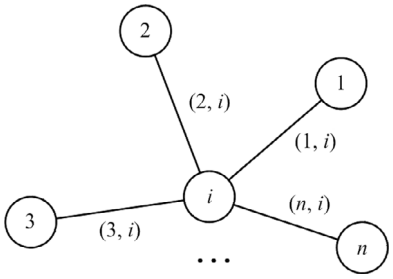

Fig. 3. Topological structure of a radial graph.

Lemma 1. *Consider the system (8)–(9) with $\mathcal{V} = \{1, 2\}$ (i.e., an connected graph with two nodes). Then, for each $i \in \mathcal{V}$, and $j \in \mathcal{N}_i$, we have*

$$\hat{x}_i^*(k|k) = \Sigma_i^*(k|k)(\bar{\alpha}_i(k) + B_{ij}^T S_{ji}^{-1}(k)y_{ji}(k)),$$
$$\Sigma_i^*(k|k) = (\bar{Q}_i(k) + B_{ij}^T S_{ji}^{-1}(k)B_{ij})^{-1},$$

where

$$\begin{aligned}
\bar{\alpha}_i(k) &= C_i^T S_i^{-1} z_{i,i}(k) + \Sigma_i^{-1}(k|k-1)\hat{x}_i(k|k-1),\\
\bar{Q}_i(k) &= C_i^T S_i^{-1} C_i + \Sigma_i^{-1}(k|k-1),\\
y_{ji}(k) &= z_{i,j}(k) - \beta_j^i(k),\\
S_{ji}(k) &= T_{i,j} + \Phi_j^i(k),\\
\beta_j^i(k) &= B_{ji}\bar{Q}_j^{-1}(k)\bar{\alpha}_j(k),\ \Phi_j^i(k) = B_{ji}\bar{Q}_j^{-1}(k)B_{ji}^T,
\end{aligned} \tag{15}$$

initialized by $\hat{x}_i(0|-1) = \bar{x}_i(0)$ and $\Sigma_i(0|-1) = \Sigma_i(0)$.

Lemma 2. *Suppose that Assumptions 1 and 2 hold. Consider an interconnected system represented by a radial graph $\mathcal{G}$. At time instant k and node $i \in \mathcal{V}$, the distributed MAP estimation and its error covariance matrix are*

$$\hat{x}_i^*(k|k) = \Sigma_i^*(k|k)\left(\bar{\alpha}_i(k) + \sum_{j\in\mathcal{N}_i} B_{ij}^T S_{ji}^{-1}(k)y_{ji}(k)\right), \tag{16}$$

$$\Sigma_i^*(k|k) = \left(\bar{Q}_i(k) + \sum_{j\in\mathcal{N}_i} B_{ij}^T S_{ji}^{-1}(k)B_{ij}\right)^{-1}. \tag{17}$$

Our proposed distributed MAP estimation algorithm follows from Lemma 2. Its basic idea is as follows. Initially, each node i computes a local estimation $\breve{x}_i(k|k, 0)$ and its associated estimation error covariance $\breve{\Sigma}_i(k|k, 0)$, using only its prior information about $x_i(k)$ and its local measurements $z_{i,i}(k)$. Then it builds $\beta_i^j(k, 0)$ and $\Phi_i^j(k, 0)$, and sends them to node $j \in \mathcal{N}_i$. Then, the method follows an iterative procedure. At iteration h, using the values of $\beta_j^i(k, h-1)$ and $\Phi_j^i(k, h-1)$, received from its neighbors $j \in \mathcal{N}_i$, node i updates the edge information $y_{ji}(k, h)$ and $S_{ji}(k, h)$, and uses them to compute its current estimation $\breve{x}_i(k|k, h)$ and covariance $\breve{\Sigma}_i(k|k, h)$. Also, for each neighbor $j \in \mathcal{N}_i$, node i computes $\beta_i^j(k, h)$ and $\Phi_i^j(k, h)$ and sends them to node j. The details of these steps are given in Algorithm 1. A key observation is that the information $\beta_i^j(k, h)$ and $\Phi_i^j(k, h)$ transmitted from node i to node j, do not include the information that node i previously receives from node j.

Distributed MAP estimation algorithm

Initialization:

(1) At time instant $k = 1, 2, \ldots$, each node $i \in \mathcal{V}$ computes the local state estimation and its error covariance

$$\breve{x}_i(k|k, 0) = \bar{Q}_i^{-1}(k)\bar{\alpha}_i(k),$$
$$\breve{\Sigma}_i(k|k, 0) = \bar{Q}_i^{-1}(k),$$

with $\bar{\alpha}_i(k)$ and $\bar{Q}_i(k)$ defined as in Lemma 1. If $k = 0, \hat{x}_i(0|-1)$ and $\Sigma_i(0|-1)$ are replaced by $\bar{x}_i(0)$ and $\Sigma_i(0)$, respectively.
(2) For each node $j \in \mathcal{N}_i$, node i computes

$$\beta_i^j(k,\ 0) = B_{ij}\breve{x}_i(k|k,\ 0),$$
$$\Phi_i^j(k,\ 0) = B_{ij}\breve{\Sigma}_i(k|k,\ 0)B_{ij}^T,$$

and sends them to node j.
Main loop: At iteration $h = 1, 2, \ldots$, and for each i:
(1) Using $\beta_j^i(k,\ h-1)$ and $\Phi_j^i(k,\ h-1)$ received from its neighbors $j \in \mathcal{N}_i$, node i updates the edge information

$$y_{ji}(k,\ h) = z_{i,j}(k) - \beta_j^i(k,\ h-1),$$
$$S_{ji}(k,\ h) = T_{i,j} + \Phi_j^i(k,\ h-1).$$

(2) Node i calculates the current state estimation and its covariance:

$$\breve{x}_i(k|k,\ h) = [Q_i(k,h)]^{-1}\alpha_i(k,h),$$
$$\breve{\Sigma}_i(k|k,\ h) = [Q_i(k,h)]^{-1},$$

where

$$\alpha_i(k,\ h) = \bar{\alpha}_i(k) + \sum_{j\in\mathcal{N}_i} B_{ij}^T S_{ji}^{-1}(k,\ h)y_{ji}(k,\ h),$$
$$Q_i(k,\ h) = \bar{Q}_i(k) + \sum_{j\in\mathcal{N}_i} B_{ij}^T S_{ji}^{-1}(k,\ h)B_{ij}.$$

(3) For each $j \in \mathcal{N}_i$, node i computes

$$\beta_i^j(k,\ h) = B_{ij}[Q_i^j(k,\ h)]^{-1}\alpha_i^j(k,\ h),$$
$$\Phi_i^j(k,\ h) = B_{ij}[Q_i^j(k,\ h)]^{-1}B_{ij}^T,$$

where

$$\alpha_i^j(k,\ h) = \bar{\alpha}_i(k) + \sum_{m\in\mathcal{N}_i/\{j\}} B_{im}^T S_{mi}^{-1}(k,\ h)y_{mi}(k,\ h),$$
$$Q_i^j(k,\ h) = \bar{Q}_i(k) + \sum_{m\in\mathcal{N}_i/\{j\}} B_{im}^T S_{mi}^{-1}(k,\ h)B_{im},$$

and transmits $\beta_i^j(k,\ h)$ and $\Phi_i^j(k,\ h)$ to node j.

Theorem 1, shows that, by doing so, the algorithm converges in a finite number of steps to the suboptimal estimate $\hat{x}_i^*(k|k)$ and $\Sigma_i^*(k|k)$. See Appendix for proof.

Theorem 1. *Suppose that Algorithm* 1 *is used under Assumptions* 1 *and* 2. *Then, for each* $k = 0, 1, \ldots$ *and each node* $i \in \mathcal{V}$, *we have*

$$\hat{x}_i(k|k,\ \varepsilon_i + l) = \hat{x}_i^*(k|k),$$
$$\Sigma_i(k|k,\ \varepsilon_i + l) = \Sigma_i^*(k|k), \quad \text{for all } l \geq 0,$$

where $\varepsilon_i + l$ *is the step number of iteration.*

Remark 4. Theorem 1 states that the local state estimates on all nodes converge to the suboptimal estimates after $\Gamma = \max\{\varepsilon_i,\ i \in \mathcal{V}\}$ steps.

Remark 5. Notice that the edge measurement $z_{i,j}(k)$ is usually of low dimension. Hence, every control center only needs to exchange the low-dimensional information $B_{ji}\hat{x}_j(k|k)$ with its neighbors, instead of the entire local estimate $\hat{x}_j(k|k)$. This leads to a very light communication requirement; see Tai et al. (2013).

5.2. Local prediction

In this section we describe the prediction step of our distributed state estimation method. From (8), and using the MAP estimation $\hat{x}_i^*(k|k)$ and its error covariance matrix $\Sigma_i^*(k|k)$, obtained from Theorem 1, we get

$$\hat{x}_i(k+1|k) = A_i\hat{x}_i^*(k|k) + G_i,$$
$$\Sigma_i(k+1|k) = A_i\Sigma_i^*(k|k)A_i^T + R_i.$$

The state prediction is used as the prior information to initialize Algorithm 1 once the new measurements at time $k+1$ become available.

6. Simulation

In this section, we compare the performances of the distributed state estimator (or distributed Kalman filter (DKF)), with the centralized state estimator (or centralized Kalman filter (CKF)), the distributed static state estimator (DSSE) in Tai et al. (2013) and the local Kalman filter (LKF). To this end, we use the IEEE 118-bus system (Christie, 1993). We split this system into six areas, as shown in Fig. 1. Each area (considered as a single node) is monitored and operated by a control center. The resulting network has the radial topology as shown in Fig. 2, and therefore Assumption 1 is satisfied. We assume that PMUs (which produce linear measurements) are located at some buses, which are marked as blocks in Fig. 1. The placement of PMUs is designed using the method proposed in Tai et al. (2013), which guarantees the topological observability. Measurements for the system state estimation consist of voltage magnitudes at all the buses, voltage angles at 32 buses on which PMUs are installed, and power flows measured at all the line terminals. The parameters g_{ij}, b_{ij} and b_{ij}^{sh} used in (4) and (5) are taken from Christie (1993).

Following Shih and Huang (2002), the linear trend A having a standard deviation of 2% along with the fluctuation G is added to the load curve and the fluctuation in G is represented by a normally distributed random number $\pm 3\%$ of the value of the trend component. For the node i which is installed with SCADA measurements, the covariance of the local measurement noise is $S_i = \sigma_1^2$ with $\sigma_1 = 0.1$. For the node j which is installed with a PMU, the covariance of the local measurement noise is $S_j = \mathrm{diag}(\sigma_2^2, \sigma_2^2)$ with $\sigma_2 = 0.05$. The covariance of the edge measurement noise is $T_{i,j} = \mathrm{diag}(\sigma_1^2, \sigma_1^2)$.

The centralized MAP and the local state estimators require $O((\sum_{i=1}^n s_i)^3)$ and $O(s_i^3)$ computations, whereas the computational complexity of the distributed MAP estimator of node i is $\tilde{n}_i O(s_i^3)$ at each time stamp, where $\tilde{n}_i$ denotes the cardinality of $\mathcal{N}_i$. We can see that the computational complexity of each distributed estimator relates to the number of its neighbors. Due to the interconnection structure of the power system, each node only has a few neighbors, i.e., $\tilde{n}_i \ll n$. Thus, the computational complexity of the proposed distributed algorithm is smaller than that of the centralized method, but it is larger than that of the local method.

The time step of simulations is selected to be one second, in the simulation. In Figs. 4 and 5, we run one iteration of the DKF and the DSSE between time steps t and $t+1$. Fig. 4 shows the evolution of the state estimation error of each method, which is made over a period of 45 time-sample intervals. To quantify this error we use $\sum_{i=1}^n \mathrm{Tr}\{\Sigma_i(k|k)\}$ as the sum of the trace of the estimation error covariance of each subsystem. The definition of each method can be similarly described by their own estimation error covariance. We use 100 Monte Carlo runs to compute each estimation error covariance. The tests presented in Fig. 4 are discussed below.

Case 1. *Normal Operation Condition*: These simulations are carried out through 15 time samples. The first 15 time samples in Fig. 4 show the results of using the four methods when IEEE 118-bus system is operated at normal operating condition. We see how the DKF largely outperforms the LKF and the DSSE, though its values are somehow worse than that of the CKF. The state estimation error covariance of the DKF is bounded.

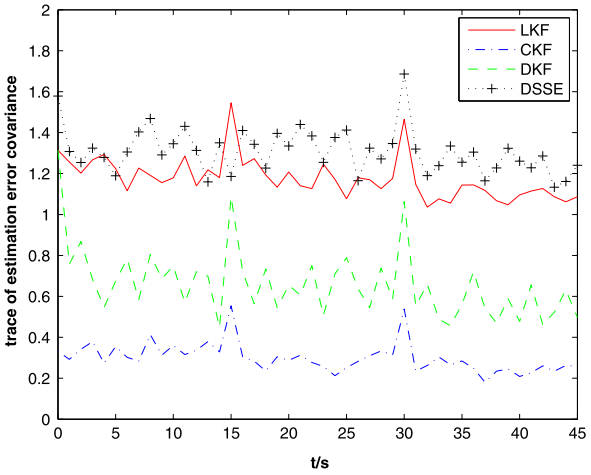

Fig. 4. The traces of the estimation error covariances of the CKF, LKF, DKF and DSSE estimates.

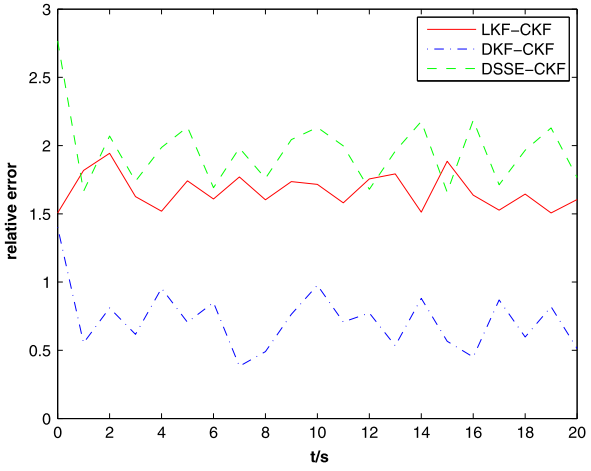

Fig. 5. Relative errors by different estimation algorithms.

Case 2. *Sudden Load Change Condition*: In this case, the proposed method is applied to the scenario in which sudden load is changed in power system. For the IEEE 118-bus system, the following scenario is simulated: 20% load increase occurs on the buses 12, 13, 14, 16, 17, 19, 26, 32, 37, 42, 45, 56, 62, 63, 66, 69, 70, 75, 77, 80, 85, 89, 92, 96, 100, 105, 114 at the 15th time sample, achieved by adding appropriate input noise $\omega(k)$ at $k = 15$. Fig. 4 plots the evaluation results of aforementioned scenarios in different methods between 15 and 29. Here all methods have low estimation performance at $k = 15$, since the dynamic state estimators depend on the previous state prediction, which is different to the actual state condition. After the sudden load occurs, the estimation error covariance of the DKF reduces quickly, which means that the proposed method can come back adaptively.

Case 3. *Bad Data Condition*: In this case, the proposed method under bad data scenario is investigated. The following condition is simulated: one additional measurement error of 20% is added at the 30th time sample. It can be seen that the estimation results of all methods are affected when polluted measurements are not detected. However, the four estimation methods still return to the normal values rapidly as the first case. This also reveals that the proposed method owns higher immunity to the influence of bad data, and this approach can be a potential candidate of dynamic state estimation in addition to conventional methods.

The run times of the CKF, LKF, DKF and DSSE are 0.06532 s, 0.05386 s, 0.05974 s and 0.05052 s, respectively.

Fig. 5 shows the evolution of the relative errors between the state estimates of the DKF and CKF, i.e.,

$$\frac{\sum_{i=1}^{n} \|\hat{x}_i^*(k|k) - \hat{x}_i^{MAP}(k|k)\|}{\sum_{i=1}^{n} \|\hat{x}_i^{MAP}(k|k)\|},$$

as well as the relative errors between the LKF and CKF estimates, i.e.,

$$\frac{\sum_{i=1}^{n} \|\hat{x}_i(k|k) - \hat{x}_i^{MAP}(k|k)\|}{\sum_{i=1}^{n} \|\hat{x}_i^{MAP}(k|k)\|}.$$

The relative errors between the DSSE and CKF estimates can be similarly defined. Again, we see the clear advantage of the DKF over the LKF and DSSE, and its close accuracy to that of the CKF.

7. Conclusions

We proposed a distributed algorithm for local state estimation in large-scale interconnected networks, which could be applied in multi-area networked power systems. The key component of this method is a distributed MAP estimation step, which only requires local measurements and low dimensional boundary state estimates from neighbors. It is shown that, under the assumption that the network topology is acyclic, at each time instant, the distributed MAP estimator converges after a finite number of iteration, which equals the diameter of the graph. Although simulation experiments using the IEEE 118-bus system in different scenarios show that the accuracy of distributed MAP estimator is somehow worse than that of the centralized state estimator, the former has low requirements in terms of computational complexity and communication load at each node. This makes it scalable to large-sized networked systems.

Appendix. Proofs of Section 5

A.1. Proof of Lemma 1

At time k, consider the centralized MAP estimator, in which $\Sigma(k|k-1)$ is replaced by $\tilde{\Sigma}(k|k-1)$. The measurement equation of the two-node graph can be written as (11), where

$$x(k) = (x_1^T(k),\ x_2^T(k))^T,$$
$$z(k) = (z_{1,1}^T(k),\ z_{2,2}^T(k),\ z_{1,2}^T(k))^T,$$
$$\nu(k) = (\nu_{1,1}^T(k),\ \nu_{2,2}^T(k),\ \nu_{1,2}^T(k))^T,$$
$$H = \begin{bmatrix} C_1 & 0 \\ 0 & C_2 \\ B_{12} & B_{21} \end{bmatrix},$$
$$R_* = \mathrm{diag}\{S_1,\ S_2,\ T_{1,2}\}.$$

According to (14), we obtain

$$\Sigma^*(k|k) = \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{12}^T & \Gamma_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \Sigma_1^*(k|k) & -\Sigma_1^*(k|k)\Gamma_{12}\Gamma_{22}^{-1} \\ -\Sigma_2^*(k|k)\Gamma_{21}\Gamma_{11}^{-1} & \Sigma_2^*(k|k) \end{bmatrix}, \tag{A.1}$$

where

$$\Gamma_{11} = \bar{Q}_1(k) + B_{12}^T T_{1,2}^{-1} B_{12},$$
$$\Gamma_{12} = B_{12}^T T_{1,2}^{-1} B_{21},$$
$$\Gamma_{22} = \bar{Q}_2(k) + B_{21}^T T_{1,2}^{-1} B_{21}.$$

From the block matrix inversion formula (Kailath, Sayed, & Hassibi, 2000, A.1(v)), we obtain that the first diagonal block of $\Sigma^*(k|k)$ is given by

$$\begin{aligned}\Sigma_1^*(k|k) &= (\Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{12}^T)^{-1} \\ &= \big[\bar{Q}_1(k) + B_{12}^T T_{1,2}^{-1} B_{12} - B_{12}^T T_{1,2}^{-1} B_{21} \\ &\quad \times \left(\bar{Q}_2(k) + B_{21}^T T_{1,2}^{-1} B_{21}\right)^{-1} B_{21}^T T_{1,2}^{-1} B_{12}\big]^{-1}.\end{aligned}$$

Using the matrix inversion lemma (Kailath et al., 2000, A.1(vii)), we get

$$\begin{aligned}&T_{1,2}^{-1} - T_{1,2}^{-1} B_{21}\left(\bar{Q}_2(k) + B_{21}^T T_{1,2}^{-1} B_{21}\right)^{-1} B_{21}^T T_{1,2}^{-1} \\ &\quad = (T_{1,2} + B_{21}\bar{Q}_2^{-1}(k) B_{21}^T)^{-1} = S_{21}^{-1}(k). \end{aligned} \tag{A.2}$$

Therefore, we obtain

$$\Sigma_1^*(k|k) = [\bar{Q}_1(k) + B_{12}^T S_{21}^{-1}(k) B_{12}]^{-1}.$$

Following the same argument, we have that

$$\Sigma_2^*(k|k) = [\bar{Q}_2(k) + B_{21}^T S_{12}^{-1}(k) B_{21}]^{-1}.$$

On the other hand, from (13), we have

$$\hat{x}^*(k|k) = \begin{bmatrix}\Gamma_{11} & \Gamma_{12} \\ \Gamma_{12}^T & \Gamma_{22}\end{bmatrix}^{-1} \begin{bmatrix}\Upsilon_1(k) \\ \Upsilon_2(k)\end{bmatrix},$$

where

$$\Upsilon_i(k) = \bar{\alpha}_i(k) + B_{ij}^T T_{i,j}^{-1} z_{i,j}(k), \quad i = 1,\ 2.$$

From (A.1), we get

$$\hat{x}_1^*(k|k) = \Sigma_1^*(k|k)\Upsilon_1(k) - \Sigma_1^*(k|k)\Gamma_{12}\Gamma_{22}^{-1}\Upsilon_2(k). \tag{A.3}$$

Similar to (A.2), and using (15), we obtain

$$\Gamma_{22}^{-1} = \bar{Q}_2^{-1}(k) - \bar{Q}_2^{-1}(k) B_{21}^T S_{21}^{-1}(k) B_{21} \bar{Q}_2^{-1}(k),$$

and

$$\begin{aligned}B_{21}\Gamma_{22}^{-1} &= B_{21}(\bar{Q}_2^{-1}(k) - \bar{Q}_2^{-1}(k) B_{21}^T S_{21}^{-1}(k) B_{21}\bar{Q}_2^{-1}(k)) \\ &= (S_{21}(k) - B_{21}\bar{Q}_2^{-1}(k) B_{21}^T) S_{21}^{-1}(k) B_{21}\bar{Q}_2^{-1}(k) \\ &= T_{1,2} S_{21}^{-1}(k) B_{21}\bar{Q}_2^{-1}(k).\end{aligned}$$

Using $z_{1,2}(k) = z_{2,1}(k)$ and $T_{1,2} = T_{2,1}$, it follows that

$$\begin{aligned}&\Upsilon_1(k) - \Gamma_{12}\Gamma_{22}^{-1}\Upsilon_2(k) \\ &= \bar{\alpha}_1(k) + B_{12}^T T_{1,2}^{-1} z_{1,2}(k) \\ &\quad - B_{12}^T T_{1,2}^{-1} B_{21}\Gamma_{22}^{-1}(\bar{\alpha}_2(k) + B_{21}^T T_{2,1}^{-1} z_{2,1}(k)) \\ &= \bar{\alpha}_1(k) + B_{12}^T T_{1,2}^{-1} z_{1,2}(k) \\ &\quad - B_{12}^T S_{21}^{-1}(k) B_{21}\bar{Q}_2^{-1}(k)(\bar{\alpha}_2(k) + B_{21}^T T_{2,1}^{-1} z_{2,1}(k)) \\ &= \bar{\alpha}_1(k) - B_{12}^T S_{21}^{-1}(k) B_{21}\bar{Q}_2^{-1}(k)\bar{\alpha}_2(k) \\ &\quad + B_{12}^T S_{21}^{-1}(k)(S_{21}(k) - B_{21}\bar{Q}_2^{-1}(k) B_{21}^T) T_{1,2}^{-1} z_{1,2}(k) \\ &= \bar{\alpha}_1(k) + B_{12}^T S_{21}^{-1}(k)(z_{1,2}(k) - B_{21}\bar{Q}_2^{-1}(k)\bar{\alpha}_2(k)) \\ &= \bar{\alpha}_1(k) + B_{12}^T S_{21}^{-1}(k) y_{21}(k).\end{aligned}$$

The last step above used (A.2). Returning to (A.3) yields

$$\hat{x}_1^*(k|k) = \Sigma_1^*(k|k)(\bar{\alpha}_1(k) + B_{12}^T S_{21}^{-1}(k) y_{21}(k)).$$

Using the same proof, we also get

$$\hat{x}_2^*(k|k) = \Sigma_2^*(k|k)(\bar{\alpha}_2(k) + B_{21}^T S_{12}^{-1}(k) y_{12}(k)).$$

This completes the proof. □

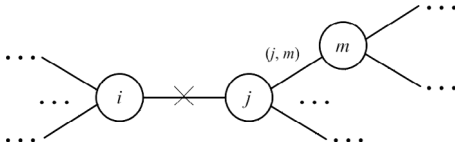

Fig. A.1. Illustration for the proof of Theorem 1.

A.2. Proof of Lemma 2

A radial graph, as shown in Fig. 3, can be considered as a two-node graph, by combining nodes 1, ..., n into a single node. Doing so, we can use Lemma 1 to obtain the local update on node i at time instant k, which is

$$\hat{x}_i^*(k|k) = \Sigma_i^*(k|k)\alpha_i(k),$$
$$\Sigma_i^*(k|k) = Q_i^{-1}(k),$$

where

$$\alpha_i(k) = \bar{\alpha}_i(k) + \begin{bmatrix}B_{i1} \\ B_{i2} \\ \vdots \\ B_{in}\end{bmatrix}^T \tilde{S}^{-1}(k) \begin{bmatrix}y_{1i}(k) \\ y_{2i}(k) \\ \vdots \\ y_{ni}(k)\end{bmatrix}, \tag{A.4}$$

$$Q_i(k) = \bar{Q}_i(k) + \begin{bmatrix}B_{i1} \\ B_{i2} \\ \vdots \\ B_{in}\end{bmatrix}^T \tilde{S}^{-1}(k) \begin{bmatrix}B_{i1} \\ B_{i2} \\ \vdots \\ B_{in}\end{bmatrix}, \tag{A.5}$$

and $\tilde{S}(k) = \text{diag}\{S_{1i}(k),\ S_{2i}(k),\ \ldots,\ S_{ni}(k)\}$. Then, it is easy to see that (16)–(17) follow from (A.4)–(A.5). □

A.3. Proof of Theorem 1

The proof is based on the estimation at node i. Consider the acyclic graph $\mathcal{G}$ depicted in Fig. A.1. Consider also the radial sub-graph of $\mathcal{G}$, having i as its central node, as well as all its neighbors as leaf nodes. At the first iteration $h = 1$, Algorithm 1 consists in applying the result of Lemma 2 to this radial sub-graph. Hence, after the first iteration, node i can compute the desired suboptimal estimate corresponding to this sub-graph.

For the second iteration, consider the radial sub-graph having i as central node. Also, each leaf node of this sub-graph, is formed by grouping together, as a single node, each neighbor $j \in \mathcal{N}_i$ of i, with the neighbors of j different from i. Now, according to Algorithm 1, each node $j \in \mathcal{N}_i$ builds the quantities $\beta_j^i(k, 2)$ and $\Phi_j^i(k, 2)$ to be transmitted to node i, without using the information $\beta_i^j(k, 1)$ and $\Phi_i^j(k, 1)$ that it previously received from node i. Hence, the second iteration is equivalent to applying the result of Lemma 2 to the aforementioned radial graph, and therefore node i is able to compute the suboptimal estimate of the sub-graph formed by all nodes which are two hops away from it.

The above argument applies at each iteration. More precisely, at iteration h, we consider the radial sub-graph having i as central node. Also, each leaf node of this sub-graph, is formed by grouping together, as a single node, each neighbor $j \in \mathcal{N}_i$ of i, with all the neighbors of j which are $h - 1$ hops away from it, but not having node i as intermediate node. Then, following the above argument, we obtain that node i is able to compute the suboptimal estimate corresponding to this radial sub-graph, i.e., to the sub-graph formed by all nodes which are h hops away from i. Hence, node i will achieve the suboptimal estimate corresponding to the whole graph in a number of steps equal to its radius. Moreover, this estimate will remain unchanged in all the subsequent iterations. □

References

Abur, A., & Exposito, A. G. (2004). *Power system state estimation: theory and implementation*. New York: Marcel Dekker.

Anderson, B. D. O., & Moore, J. B. (1979). *Optimal filtering*. Englewood Cliffs, NJ: Prentice-Hall.

Arulampalam, S., Maskell, S., Gordon, N., & Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, *50*(2), 174–188.

Bose, A., Abur, A., Poon, K. Y. K., & Emami, R. (2010). *Implementation issues for hierarchical state estimators. Final Project Report*.

Cattivelli, F. S., Lopes, C. G., & Sayed, A. H. (2008). Diffusion recursive least-squares for distributed estimation over adaptive networks. *IEEE Transactions on Signal Processing*, *56*(5), 1865–1877.

Christie, R. (1993). 118 bus power flow test case. http://www.ee.washington.edu/research/pstca/pf118/pg_tca118bus.htm.

Conejo, A. J., de la Torre, S., & Canas, M. (2007). An optimization approach to multiarea state estimation. *IEEE Transactions on Power Systems*, *22*(1), 213–221.

Cutsem, T. V., & Ribbens-Pavella, M. (1983). Critical survey of hierarchical methods for state estimation of electric power systems. *IEEE Transactions on Power Apparatus and Systems*, *PAS-102*(10), 3415–3424.

Do Coutto Filho, M. B., & Stacchini de Souza, J. C. (2009). Forecasting-aided state estimation–part I: panorama. *IEEE Transactions on Power Systems*, *24*(4), 1667–1677.

Falcao, D. M., Wu, F. F., & Murphy, L. (1995). Parallel and distributed state estimation. *IEEE Transactions on Power Systems*, *10*(2), 724–730.

Feng, J., & Zeng, M. (2012). Optimal distributed Kalman filtering fusion for a linear dynamic system with cross-correlated noises. *International Journal of Systems Science*, *43*(2), 385–398.

Gomez-Exposito, A., & de la Villa Jaen, A. (2009). Two-level state estimation with local measurement pre-processing. *IEEE Transactions on Power Systems*, *24*(2), 676–684.

Huang, Y. F., Werner, S., Huang, J., Kashyap, N., & Gupta, V. (2012). State estimation in electric power grids: meeting new challenges presented by the requirements of the future grid. *IEEE Signal Processing Magazine*, *29*(5), 33–43.

Jiang, W., Vittal, V., & Heydt, G. T. (2008). Diakoptic state estimation using phasor measurement units. *IEEE Transactions on Power Systems*, *23*(4), 1580–1589.

Kailath, T., Sayed, A. H., & Hassibi, B. (2000). *Linear estimation*. Upper Saddle River, NJ: Prentice-Hall.

Kay, S. M. (1993). *Fundamentals of statistical signal processing*. Upper Saddle River, NJ: Prentice-Hall.

Khan, U.A., Ilic, M.D., & Moura, J.M.F. (2008). Cooperation for aggregating complex electric power networks to ensure system observability. In *International conference on infrastructure systems and services: developing 21st century infrastructure networks*, (pp. 1–6).

Larson, R. E., Tinney, W. F., & Peschon, J. (1970). State estimation in power systems part I: theory and feasibility. *IEEE Transactions on Power Apparatus and Systems*, *PAS-89*(3), 345–352.

Leite da Silva, A. M., Do Coutto Filho, M. B., & de Queiroz, J. F. (1983). State forecasting in electric power systems. *IEE Proceedings C (Generation, Transmission and Distribution)*, *130*(5), 237–244.

Liang, J., Wang, Z., & Liu, X. (2011). Distributed state estimation for discrete-time sensor networks with randomly varying nonlinearities and missing measurements. *IEEE Transactions on Neural Networks*, *22*(3), 66–86.

Lin, S. Y. (1992). A distributed state estimator for electric power systems. *IEEE Transactions on Power Systems*, *7*(2), 551–557.

Makridakis, S., & Wheelwright, S. C. (1978). *Forecasting methods and applications*. New York: Wiley.

Monticelli, A. (1999). *State estimation in electric power systems: a generalized approach*. Massachusetts: Norwell: Kluwer Academic Publishers.

Olfati-Saber, R. (2007). Distributed Kalman filtering for sensor networks. In *Proceedings of the 46th IEEE conference on decision and control*, (pp. 5492–5498).

Pasqualetti, F., Carli, R., & Bullo, F. (2012). Distributed estimation via iterative projections with application to power network monitoring. *Automatica*, *48*(5), 747–758.

Schweppe, F. C., & Wildes, J. (1970). Power system static-state estimation, part I, II, III. *IEEE Transactions on Power Apparatus and Systems*, *89*, 120–135.

Shih, K. R., & Huang, S. J. (2002). Application of a robust algorithm for dynamic state estimation of a power system. *IEEE Transactions on Power Systems*, *17*(1), 141–147.

Tai, X., Lin, Z., Fu, M., & Sun, Y. (2013). A new distributed state estimation technique for power networks. In *Proceedings of the 2013 American control conference*, (pp. 3338–3343).

Tai, X., Marelli, D., Rohr, E., & Fu, M. (2013). Optimal PMU placement for power system state estimation with random component outages. *Journal of Electrical Power and Energy Systems*, *51*, 35–42.

Valverde, G., & Terzija, V. (2011). Unscented Kalman filter for power system dynamic state estimation. *IET Generation, Transmission & Distribution*, *5*(1), 29–37.

Wang, S., Gao, W., & Meliopoulos, A. P. S. (2012). An alternative method for power system dynamic state estimation based on unscented transform. *IEEE Transactions on Power Systems*, *27*(2), 942–950.

Wu, F. F. (1990). Power system state estimation: a survey. *International Journal of Electrical Power and Energy Systems*, *12*(1), 80–87.

Wu, F. F., Moslehi, K., & Bose, A. (2005). Power system control centers: past, present, and future. *Proceedings of the IEEE*, *93*(11), 1890–1908.

Xie, L., Choi, D.H., & Kar, S. (2011). Cooperative distributed state estimation: local observability relaxed. In *Proceedings of IEEE power and energy society general meeting*, (pp. 1–11).

Yu, W., Chen, G., Wang, Z., & Yang, W. (2009). Distributed consensus filtering in sensor networks. *IEEE Transactions on Systems, Man and Cybernetics*, *39*(6), 1568–1577.

Yibing Sun received his B.S. and M.S. degrees in pure mathematics from School of Mathematical Sciences, University of Jinan, Jinan, China, in 2009 and 2012, respectively. He is now working toward his Ph.D. degree in School of Control Science and Engineering, Shandong University, Jinan, China. His research interests include distributed state estimation in power systems and Kalman filter.

Minyue Fu received his Bachelor's Degree in Electrical Engineering from the University of Science and Technology of China, Hefei, China, in 1982, and M.S. and Ph.D. degrees in Electrical Engineering from the University of Wisconsin-Madison in 1983 and 1987, respectively. From 1983 to 1987, he held a teaching assistantship and a research assistantship at the University of Wisconsin-Madison. He worked as a Computer Engineering Consultant at Nicolet Instruments, Inc., Madison, Wisconsin, during 1987. From 1987 to 1989, he served as an Assistant Professor in the Department of Electrical and Computer Engineering, Wayne State University, Detroit, Michigan. He joined the Department of Electrical and Computer Engineering, the University of Newcastle, Australia, in 1989. Currently, he is a Chair Professor in Electrical Engineering and Head of School of Electrical Engineering and Computer Science. He is a Fellow of IEEE. His main research interests include control systems, signal processing and communications. He has been an Associate Editor for the IEEE Transactions on Automatic Control, Automatica and Journal of Optimization and Engineering. He was a Visiting Associate Professor at University of Iowa in 1995–1996, a Senior Fellow/Visiting Professor at Nanyang Technological University, Singapore, 2002, a Qian-ren Professorship at Zhejiang University, China, in 2010–2015, and a Bai-ren Professorship at Guangdong University of Technology, China, in 2016.

Bingchang Wang received the Ph.D. degree in System Theory from Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China, in 2011. From September 2011 to August 2012, he was with Department of Electrical and Computer Engineering, University of Alberta, Canada, as a postdoctoral fellow. From September 2012 to September 2013, he was with School of Electrical Engineering and Computer Science, University of Newcastle, Australia, as a research academic. From October 2013, he has been with School of Control Science and Engineering, Shandong University, China, as an associate professor. During November 2014–May 2015, he visited School of Mathematics, Carleton University, Canada as a research associate. His current research interests are stochastic control and games, distributed estimation and event based control.

Huanshui Zhang graduated in mathematics from the Qufu Normal University in 1986 and received his M.Sc. and Ph.D. degrees in control theory from Heilongjiang University, China, and Northeastern University, China, in 1991 and 1997, respectively. He worked as a postdoctoral fellow at Nanyang Technological University from 1998 to 2001 and Research Fellow at Hong Kong Polytechnic University from 2001 to 2003. He is currently a Changjiang Professorship at Shandong University, China. He held Professor in Harbin Institute of Technology from 2003 to 2006. He also held visiting appointments as Research Scientist and Fellow with Nanyang Technological University, Curtin University of Technology and Hong Kong City University from 2003 to 2006. His interests include optimal estimation and control, time-delay systems, stochastic systems, signal processing and wireless sensor networked systems.

Damián Marelli received his Bachelors Degree in Electronics Engineering from the Universidad Nacional de Rosario, Argentina in 1995. He also received his Bachelor (Honours) degree in Mathematics and Ph.D. degree in Electrical Engineering, both from the University of Newcastle, Australia in 2003. From 2004 to 2005 he held a postdoctoral position at the Laboratoire d'Analyse Topologie et Probabilités, CNRS/Université de Provence, France. From 2005 to 2015 he was Research Academic at the Centre for Complex Dynamic Systems and Control, the University of Newcastle, Australia. In 2007 he received a Marie Curie

Postdoctoral Fellowship, hosted at the Faculty of Mathematics, University of Vienna, Austria, and in 2011 he received a Lise Meitner Senior Fellowship, hosted at the Acoustics Research Institute of the Austrian Academy of Sciences. Since 2016, he holds an Independent Researcher appointment at the French-Argentinean International Center for Information and Systems Sciences, National Scientific and Technical Research Council, Argentina. His main research interests include time–frequency analysis, system identification, statistical signal processing and sensor networks.