Performance Comparison of Distributed State Estimation Algorithms for Power Systems*

SUN Yibing · FU Minyue · ZHANG Huanshui

Abstract A newly proposed distributed dynamic state estimation algorithm based on the maximum a posteriori (MAP) technique is generalised and studied for power systems. The system model involves linear time-varying load dynamics and nonlinear measurements. The main contribution of this paper is to compare the performance and feasibility of this distributed algorithm with several existing distributed state estimation algorithms in the literature. Simulations are tested on the IEEE 39-bus and 118-bus systems under various operating conditions. The results show that this distributed algorithm performs better than distributed quasi-steady state estimation algorithms which do not use the load dynamic model. The results also show that the performance of this distributed method is very close to that by the centralized state estimation method. The merits of this algorithm over the centralized method lie in its low computational complexity and low communication load. Hence, the analysis supports the efficiency and benefits of the distributed algorithm in applications to large-scale power systems.

Keywords Distributed MAP estimation, distributed state estimation, extended Kalman filter, power systems.

1 Introduction

State estimation is a fundamental and vital part of the energy management system (EMS) in power system operation and control[1], which has remained an attractive and contentious

SUN Yibing

School of Control Science and Engineering, Shandong University, Jinan 250061, China.

Email: sun_yibing@126.com.

FU Minyue

School of Electrical Engineering and Computer Science, University of Newcastle, NSW 2308, Australia; School of Automation, Guangdong University of Technology, Guangzhou 510006, China.

Email: minyue.fu@newcastle.edu.au.

ZHANG Huanshui

School of Control Science and Engineering, Shandong University, Jinan 250061, China.

Email: hszhang@sdu.edu.cn.

◇*This paper was recommended for publication by Editor HONG Yiguang.*

research field in recent years. From redundant noisy measurements, state estimators are broadly utilized to estimate state variables including voltage magnitudes and phase angles of the buses in power systems, which are not directly monitored for economical and computational reasons. The accuracy of state estimation depends on measurements gathered from the network. Then the development of fast measuring and processing devices, such as Phasor Measurement Units (PMUs) are extremely crucial for state estimation techniques[2]. With growing system sizes and wide usage of PMUs, the need for faster state estimation becomes urgent, which promotes the development of distributed estimation and control in wide-area power systems. Furthermore, distributed state estimation algorithms are also necessary for many other types of large-sized networked systems, such as sensor networks, which have been developed extensively; see [3–6].

State estimation in power systems has been classically performed by a static (or quasi-steady state) approach, based on the weighted least square (WLS) method, where a single set of measurements is used to estimate the system's quasi-steady state. For the last few decades, there has been much research activity focused on distributed methods to static state estimation[7–14]. At present, static state estimators are widely used in power systems under the reliable operation of the transmission and distribution systems. Pasqualetti, et al.[11] presented a distributed static estimation method to estimate the state of power systems, in which the proposed algorithm for each monitor returned an approximate estimate within a finite number of iterations. Xie, et al.[12] presented a fully distributed state estimation algorithm, and all control centers achieved an almost sure estimate as the centralized method. However, in [11] and [12] each control center had to transmit its entire high-dimensional state estimate to its neighbors, the communication load of which was heavy.

In fact, smart grids are typically dynamic systems due to the dynamic nature of system loads. When the requirement of real-time and accurate monitoring in power systems becomes urgent, dynamic features are considered to improve state estimation methods, but static state estimators have no capacity for capturing dynamic behaviors efficiently and accurately. Furthermore, the forecasting ability of dynamic state estimation techniques plays an important role in the improvement of the overall EMS control and operation. In this case, dynamic state estimation techniques have been the focus of attention in recent years[15–24]. These dynamic methods obtained in literature mainly related to the extended Kalman filtering (EKF) technique, in which state vectors are estimated based on a prediction-correction process; see [18–22]. Shih and Huang[21] proposed a robust algorithm for dynamic state estimation by using the exponential weighting function, and showed its immunity to polluted measurements. Actually, the above mentioned algorithm is a typically centralized method, which is not well suited for applications in wide-area power systems. At present, there are few results studying the distributed state estimation for dynamic systems. Sun, et al.[23] presented a fully distributed state estimation method for a linear dynamic system, which extended and improved some known results in literature.

In this paper, we will investigate and generalize the distributed MAP estimation algorithm proposed in [23] to a nonlinear system, which is improved by the following methods: (1) Based on the exponential weighting function in [21], we incorporate new exponential weighting functions

 Springer

to ensure the robustness of the distributed algorithm; (2) For improving the estimated precision, we also use a exponential weighted method to identify model parameters. The main contribution of this paper is to validate the effectiveness of the distributed MAP estimation algorithm by comparing its performance with a centralized method, which can be seen as a combined method of [21] and [22], and two distributed static state estimation approaches in [11] and [12]. In the simulation results, the above mentioned methods are tested under various operating conditions, including normal operation condition, sudden load change, bad measurements and topology error condition. We can see that the distributed MAP estimation algorithm together with parameter identification perform better than the algorithm in [23], when sudden load changes and bad data are included in measurements. This verifies the robustness and estimated precision of the distributed algorithm. From the simulation results, the state estimates and performance indices obtained by the distributed MAP estimation algorithm are slightly worse than the centralized method, especially for large-scale systems, and the performance of this method is more accurate than the distributed static methods. Moreover, in this method each control center transmits much lower dimensional data than the whole system state to its neighbors. Hence, the communication load of the distributed MAP estimation algorithm is much less than the centralized approach and distributed static methods shown in [11] and [12], and based on the analysis of computational complexity, this algorithm is more scalable to wide-area power systems.

The rest of the paper is organized as follows. Section 2 describes some preliminaries and problem setup. The centralized algorithm for multi-area state estimation and identification of model parameters are presented in Section 3. In Section 4, we introduce two existing distributed algorithms for static systems in literature and the distributed MAP estimation algorithm for dynamic and nonlinear systems. In Section 5, illustrative studies are performed on the IEEE 39-bus and 118-bus systems. The concluding remarks are discussed in Section 6.

2 System Modeling and Problem Setup

In this section, we will introduce the system model in power systems, an example of which is shown here to depict the partition of the system model. Then the state and measurement models of each subsystem are described. Before giving the system model for state estimation, we first present some notations and preliminaries, which will be used in this 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. $\mathbb{N}_0$ is the set of non-negative integers, while $\mathbb{N}$ is the set of positive integers. We denote the transpose of M by M^{T}, in which M is a column vector or a matrix. The shorthand $\mathrm{diag}[A_1, A_2, \cdots, A_n]$ denotes a block diagonal matrix with diagonal blocks being matrices $A_1, A_2, \cdots, A_n$.

We use a communication graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ to represent a multi-area power system, where $\mathcal{V} = \{1, 2, \cdots, N\}$ is the set of nodes and $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ is the set of edges with an unordered pair $(i, j) \in \mathcal{E}$ expressing that there exists an edge between nodes i and j. We also assume that the graph $\mathcal{G}$ is connected and undirected, and the graph is void of self-loops and multiple

edges. For node i, we denote the set of its neighbors as $\mathcal{N}_i = \{j \in \mathcal{V} : (i, j) \in \mathcal{E}\}$. The notion $\overline{\mathcal{N}}_i = \mathcal{N}_i \bigcup \{i\}$ denotes the set of node i and its neighbors, and $\mathcal{N}_i / \{j\}$ denotes the set which does not include node j.

Definition 2.1 For a graph without loops, a path is known as a concatenation of adjacent edges and its length is the number of edges forming it. The diameter Γ of the graph is defined as the maximum length of a path between any two nodes.

2.1 Model Building in Power Systems

The operating state of a power system can be uniquely defined by a multidimensional vector including voltage magnitudes and phase angles. The state space representation of discrete time-varying systems is usually shown as

$$x(k + 1) = A(k)x(k) + G(k) + \omega(k), \tag{1}$$

$$z(k) = f(x(k)) + \nu(k), \tag{2}$$

where $k \in \mathbb{N}_0$ is the time sample, $x(k)$ is the state vector, $z(k)$ is the measurement vector; $A(k)$ is the state transition matrix which is usually assumed diagonal (see [20] and [21]), $G(k)$ describes the trend behavior of the state trajectory, f represents the load-flow function for the current network configuration; $\omega(k) \sim \mathcal{N}(0, R(k))$ is the process noise, and $\nu(k) \sim \mathcal{N}(0, R_*(k))$ is the measurement noise. Voltage magnitudes, phase angles as well as active/reactive power flows form the measurement vector.

2.2 Partition of Power Systems

Throughout this paper, we require the following assumption:

Assumption 2.1 The considered graph $\mathcal{G}$ is acyclic.
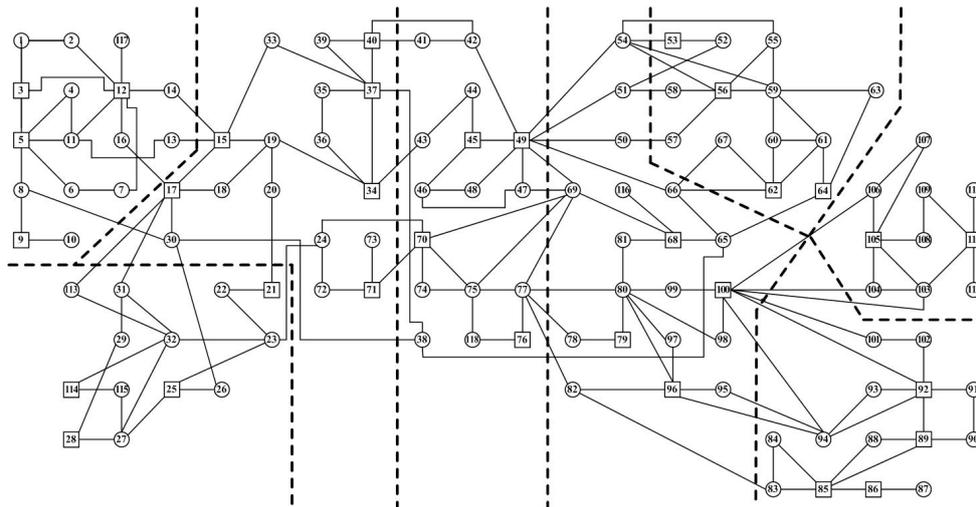
Figure 1 Topological structure of the IEEE 118-bus system

The graph $\mathcal{G}$ of Figure 2 is generally used to describe the partition of the IEEE 118-bus system as shown in Figure 1, where each node corresponds to a subsystem in Figure 1. That is to say, all buses within a subsystem are treated as a whole to measure its local state and exchange information with its neighboring nodes. Obviously, Figure 2 is acyclic, which corresponds to Assumption 2.1. The state and measurement vectors of each node are stacked by that of all buses within the corresponding subsystem. Since there exist power flows between neighboring buses, it is natural to use edge measurements to describe physical interactions in power systems.
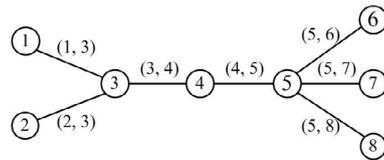
Figure 2 The graph $\mathcal{G}$ depicting the partition of the 118-bus system

From the local viewpoint of node i, which contains s_i buses, we consider the linear dynamic system as follows

$$x_i(k+1) = A_i(k)x_i(k) + G_i(k) + \omega_i(k), \tag{3}$$

where $x_i(k) \in \mathbb{R}^{2s_i}$ is the local state of node i, and $\omega_i(k)$ is assumed to be independent white Gaussian with zero mean and covariance $R_i(k)$. It is also assumed that the initial state $x_i(0)$ is an independent Gaussian variable with mean $\overline{x}_i(0)$ and covariance $\Sigma_i(0)$.

The measurements of each node can be classified into two types, i.e., local measurements which are only functions of the state of every node, and edge measurements representing the tie-line measurements related to the neighboring nodes. Therefore, the two types of measurement equations of node i can be represented respectively as follows

$$z_{i,i}(k) = f_i(x_i(k)) + \nu_{i,i}(k), \tag{4}$$

$$z_{i,j}(k) = h_{i,j}(x_i(k), x_j(k)) + \nu_{i,j}(k), \tag{5}$$

where $z_{i,i}(k)$ is the measurement vector of node i, $z_{i,j}(k)$ describes the interaction between nodes i and j, $\nu_{i,i}(k)$ and $\nu_{i,j}(k)$ are the associated measurement noises, assumed to be independent white Gaussian with zero mean and covariances $S_i(k)$ and $T_{i,j}(k)$. $f_i(\cdot)$ represents a nonlinear vector, and $h_{i,j}(\cdot)$ is the stacked power flow equations.

Remark 2.2 Notice that the edge measurement (5) is shared by both neighboring nodes i and j. Then we assume that $z_{i,j}(k)$ and $z_{j,i}(k)$ represent the same measurement and the same goes to $\nu_{i,j}(k)$ and $\nu_{j,i}(k)$.

Linearizing around the operating points $x_i^0(k)$ and $x_j^0(k)$, the Jacobian matrices of $f_i(\cdot)$ and $h_{i,j}(\cdot)$ are derived as

$$H_i(x_i^0) = \left.\frac{\partial f_i(x_i)}{\partial x_i}\right|_{x_i = x_i^0},$$

$$B_{i,j}(x_i^0) = \left.\frac{\partial h_{i,j}(x_i, x_j)}{\partial x_i}\right|_{x_i = x_i^0}, \quad B_{j,i}(x_j^0) = \left.\frac{\partial h_{i,j}(x_i, x_j)}{\partial x_j}\right|_{x_j = x_j^0}.$$

Dynamic state estimation of power systems is mainly built on the EKF theory, which utilize the measurements and the predicted values of state variables for the filtering operation. We will propose the centralized state estimation method in the next section.

3 Centralized MAP Estimation Method

The dynamic state estimation process can be divided into the following stages: State estimation, state prediction and parameter identification. In order to propose the centralized MAP estimator, we firstly describe the state and measurement models by aggregating the state and measurement equations of the nodes included in the whole system. This also explains the derivation of (1) and (2) in this paper. For (3), (4) and (5), combining states and measurements, the stacked state and measurement equations can be written as (1) and (2), where

$$x(k) = (x_1^{\mathrm{T}}(k), x_2^{\mathrm{T}}(k), \cdots, \ x_N^{\mathrm{T}}(k))^{\mathrm{T}},$$
$$z(k) = (\cdots, \ z_{i,i}^{\mathrm{T}}(k), \ \cdots, \ z_{i,j}^{\mathrm{T}}(k), \ \cdots)^{\mathrm{T}},$$
$$R(k) = \mathrm{diag}[R_1(k), R_2(k), \cdots, \ R_N(k)],$$
$$R_*(k) = \mathrm{diag}[\cdots, \ S_i(k), \ \cdots, \ T_{i,j}(k), \ \cdots].$$

The Jacobian matrix of $f(x(k))$ is $H(x^0(k))$, which is an aggregation of Jacobian matrices of local and edge functions. In the below, we choose the state prediction $\widetilde{x}(k)$ as $x^0(k)$ at each time k, and we express $H(\widetilde{x}(k))$ as $H(k)$ for simplicity. $H_i(k)$, $B_{i,j}(k)$ and $B_{j,i}(k)$ can be similarly described. $x(0)$ is the initial state with mean $\overline{x}(0) = (\overline{x}_1^{\mathrm{T}}(0), \overline{x}_2^{\mathrm{T}}(0), \cdots, \ \overline{x}_N^{\mathrm{T}}(0))^{\mathrm{T}}$ and covariance $\Sigma(0) = \mathrm{diag}[\Sigma_1(0), \Sigma_2(0), \cdots, \ \Sigma_N(0)]$. Also, we make the following assumption:

Assumption 3.1 The matrix $H(k)$ has full column rank and covariances of noises $R(k) \geq 0$, $R_*(k)$ and the initial state $\Sigma(0)$ are positive definite.

3.1 Centralized MAP Estimation Process

When the measurements from time 0 to k are known, the state estimate $\widehat{x}(k)$ and the prediction $\widetilde{x}(k)$ can be obtained by the centralized MAP estimation process.

1) State estimation.

In the filtering process, the predicted state vector $\widetilde{x}(k)$ together with its covariance matrix $M(k)$ can be filtered to obtain $\widehat{x}(k)$, if the measurement $z(k)$ is available. At time instant k, an objective function can be formulated as

$$J(x) = [z - f(x)]^{\mathrm{T}} R_*^{-1}[z - f(x)] + [x - \widetilde{x}]M^{-1}[x - \widetilde{x}],$$

where the time index k is omitted from all variables in the above equation. Minimizing $J(x)$ with respect to the state vector as shown in [21], the filtering state vector and the error covariance matrix can be expressed as

$$\widehat{x}(k) = \widetilde{x}(k) + K(k)e(k), \tag{6}$$
$$P(k) = M(k) - K(k)H(k)M(k), \tag{7}$$

where the innovation vector $e(k)$ and the gain matrix $K(k)$ are formulated as follows:

$$e(k) = z(k) - f(\widetilde{x}(k)),$$
$$K(k) = M(k)H^{\mathrm{T}}(k)\Omega^{-1}(k),$$
$$\Omega(k) = H(k)M(k)H^{\mathrm{T}}(k) + R_*(k).$$

In fact, the centralized state estimator used in this paper is the above mentioned EKF. Rewriting (6) and (7), we have

$$\widehat{x}(k) = Q^{-1}(k)\alpha(k), \tag{8}$$
$$P(k) = Q^{-1}(k), \tag{9}$$

where

$$\alpha(k) = H^{\mathrm{T}}(k)R_*^{-1}(k)\big(z(k) + H(k)\widetilde{x}(k) - f(\widetilde{x}(k))\big) + M^{-1}(k)\widetilde{x}(k),$$
$$Q(k) = H^{\mathrm{T}}(k)R_*^{-1}(k)H(k) + M^{-1}(k).$$

2) State prediction.

Based on the past data observed until time k, the system state can be forecasted for time instant $k + 1$. Executing the conditional expectation on (1), we obtain

$$\widetilde{x}(k + 1) = A(k)\widehat{x}(k) + G(k), \tag{10}$$
$$M(k + 1) = A(k)P(k)A^{\mathrm{T}}(k) + R(k), \tag{11}$$

which will become the prior statistical information to compute the state estimate $\widehat{x}(k + 1)$, while the measurement at time $k + 1$ is available.

3.2 Parameter Identification

Parameters $A(k)$ and $G(k)$ can be identified on-line by using the Holt's 2-parameter linear exponential smoothing method of forecasting[25]. Let $\widetilde{x}_m(k + 1)$ and $\widehat{x}_m(k)$ be the m-th components of the centralized state prediction $\widetilde{x}(k + 1)$ and state estimate $\widehat{x}(k)$, respectively. The following equations are given by Holt's method

$$\widetilde{x}_m(k + 1) = a_m(k) + b_m(k),$$
$$a_m(k) = \alpha_m\widehat{x}_m(k) + (1 - \alpha_m)\widetilde{x}_m(k),$$
$$b_m(k) = \beta_m(a_m(k) - a_m(k - 1)) + (1 - \beta_m)b_m(k - 1),$$

where α_m and β_m are constant parameters with the arbitrary values between 0 and 1. $A(k)$ is a diagonal matrix and $G(k)$ is a vector, the elements of which are written as

$$A_m(k) = \alpha_m(1 + \beta_m),$$
$$G_m(k) = (1 + \beta_m)(1 - \alpha_m)\widetilde{x}_m(k) - \beta_m a_m(k - 1) + (1 - \beta_m)b_m(k - 1).$$

In the centralized state prediction step, $A(k)$ and $G(k)$ are composed by $A_m(k)$ and $G_m(k)$. Furthermore, $A_i(k)$ and $G_i(k)$ of each node i are also composed by them in Subsection 4.4.

In the state estimation step, the measurement error covariance $R_*(k)$ is defined as[21]

$$R_*(k) = W_*^{-1}(k),$$
$$W_*(k) = W(k)\exp(-|e(k)|),$$

where $W(k)$ is a weighting function and $\exp(-|e(k)|)$ is the absolute residual vector. The elements of $W(k)$ are composed by the standard deviation of each measurement. When some measured values are considerably distorted at time instant k, the measurement $z(k)$ will significantly change. The variable $R_*(k)$ is used to restrain inaccurate measurements. Since the state prediction $\widetilde{x}(k)$ does not detect the abnormal measurement, $f(\widetilde{x}(k))$ is still in a normal state, which can add the absolute value of $e(k)$. Yet the value of the exponential function will then decrease. Therefore, it can reduce the weighting function and mitigate the measurement error. Since this method can adjust the weighting function at each time step according to the change in measurements, it is more efficient than the standard EKF, due to all the unchanged weighting functions throughout the estimation process.

The parameter $R(k)$ can be determined by either historical records of state variable changes or a series of off-line simulations. Actually, the noise statistics are not known accurately beforehand, which would deprive the optimality of the state estimation algorithm, even leading to filtering divergence. In this paper, we use the time-varying noise estimator to identify $R(k)$. In this method, different weights are assigned to the new data and the old data, which make the new data more important and the bad data will be forgotten gradually. This is the so-called exponential weighted method. In order to maintain the stability of the state estimator and suppress the filtering divergence, we use the biased estimator[22] as follows:

$$\widehat{R}(k) = (1 - d(k-1))\widehat{R}(k-1) + d(k-1)\big(P(k) + K(k)e(k)e^{\mathrm{T}}(k)K^{\mathrm{T}}(k)\big),$$

where $d(k-1) = \frac{1-b}{1-b^k}$, and b is the forgetting factor in the range of 0.95 and 0.99, when the process noise is changed slowly. Since $R(k)$ is a diagonal matrix, we use the diagonal elements of $\widehat{R}(k)$ to compose the new $\widehat{R}^{\#}(k)$, which will be used in the state prediction step.

Remark 3.1 We can see that the centralized state estimation method presented in this section is a combination of algorithms in [21] and [22]. When the power system encounters anomaly conditions, the robustness of the centralized method is better than the classical EKF.

It is obvious that the centralized state estimator for computing $\widehat{x}(k)$ requires the complete knowledge of the Jacobian matrix $H(k)$, the covariance matrix $R_*(k)$ and measurement vector $z(k)$, and it also needs to compute the inversion of the matrix $Q(k)$ at the single control center. For a large-scale power system, such estimator imposes a limitation on processing the real-time state estimate efficiently. In the next section, we will introduce some distributed state estimation algorithms, which can solve this complex problem.

4   Distributed State Estimation Methods

In this section, we will introduce some distributed state estimation algorithms as follows. Subsections 4.1 and 4.2 propose two distributed algorithms for static state estimation problems

in [11] and [12]. In Subsection 4.3, we present a distributed MAP estimation algorithm obtained from the algorithm in [23], and analyze the computational complexity of this algorithm. In Subsection 4.4, we will identify parameters for the distributed MAP estimation algorithm.

4.1 The Distributed Static State Estimator in [11]

Pasqualetti, et al.[11] considered a linear measurement model as follows

$$z = Cx + \nu,$$

where $z = (z_1^{\mathrm{T}}, z_2^{\mathrm{T}}, \cdots, z_N^{\mathrm{T}})^{\mathrm{T}}$ and $C = (C_1^{\mathrm{T}}, C_2^{\mathrm{T}}, \cdots, C_N^{\mathrm{T}})^{\mathrm{T}}$. Then the measurement vector of node i can be written as

$$z_i = C_i x + \nu_i, \tag{12}$$

which is obtained by combining the local measurement and edge measurements of node i into a single vector, i.e.,

$$z_i = \left(H_i + \sum_{j \in \mathcal{N}_i} B_{i,j} \right) x_i + \sum_{j \in \mathcal{N}_i} B_{j,i} x_j + \nu_i.$$

Denote $\dagger$ as the pseudo-inverse operation. Given a subspace V, $\mathrm{Basis}(V)$ is denoted as any full rank matrix, whose columns span V. $\mathrm{Im}(H)$ denotes the range space spanned by the matrix H, and $\mathrm{Ker}(H)$ denotes the null space of H.

Since the measurement noises of all nodes are independent, the covariance matrix R_* of ν is diagonal. Then we choose $B = R_*^{\frac{1}{2}}$ satisfying the condition in [11], which proposes the diffusive state estimation algorithm. Denote $\widehat{x}_i(0) = [C_i \ \varepsilon B_i]^{\dagger} z_i$, $K_i(0) = \mathrm{Basis}(\mathrm{Ker}([C_i \ \varepsilon B_i]))$, where ε is a positive integer and $B = (B_1^{\mathrm{T}}, B_2^{\mathrm{T}}, \cdots, B_N^{\mathrm{T}})^{\mathrm{T}}$. At the h-th iteration, using the exchanged data from node $j \in \mathcal{N}_i$, node i updates its local state estimate:

$$\widehat{x}_i(h+1) = \widehat{x}_i(h) + [K_i(h) \ 0][-K_i(h) \ K_j(h)]^{\dagger}(\widehat{x}_i(h) - \widehat{x}_j(h)),$$
$$K_i(h+1) = \mathrm{Basis}(\mathrm{Im}(K_i(h)) \cap \mathrm{Im}(K_j(h))),$$

and then transmits $\widehat{x}_i(h+1)$ and $K_i(h+1)$ to its neighbors.

Based on this algorithm, the estimate $\widehat{x}_i(\varepsilon)$ of system state x is computed within a finite number of iterations. But [11] has proven that $\widehat{x}_i(\varepsilon)$ is an approximate estimate, i.e., when the parameter ε is fixed, $\widehat{x}_i(\varepsilon)$ differs from the minimum variance estimate $\widehat{x}_{\mathrm{wls}}$ and only in the limit for $\varepsilon \to 0^+$, $\widehat{x}_i(\varepsilon)$ coincides with $\widehat{x}_{\mathrm{wls}}$.

4.2 The Distributed Static State Estimator in [12]

Xie, et al.[12] also studied the linear measurement model (12) of each node i. Under the assumption that the matrix $G = \sum_{i=1}^{N} C_i^{\mathrm{T}} C_i$ is full-rank, and based on the current state estimate $\widehat{x}_i(h)$ at the h-th iteration, the exchanged data $\{\widehat{x}_j(h)\}_{j \in \mathcal{N}_i}$ and the measurement vector z_i, this paper proposes the following distributed iterative algorithm to update the state estimate of node i:

$$\widehat{x}_i(h+1) = \widehat{x}_i(h) - \left[\beta(h) \sum_{j \in \mathcal{N}_i} (\widehat{x}_i(h) - \widehat{x}_j(h)) - \alpha(h) C_i^{\mathrm{T}}(z_i - C_i \widehat{x}_i(h)) \right],$$

where $\{\alpha(h)\}$ and $\{\beta(h)\}$ are appropriately chosen time-varying weight sequences and time implies iteration here.

Under the assumption that the inter-area communication network is connected, the estimate sequence $\{\widehat{x}_i(h)\}$ converges almost surely to the centralized least square estimate as $h \to \infty$. Specially, when the considered network is acyclic, $\{\widehat{x}_i(h)\}$ converges within a finite number of iterations. However, each node i has to estimate the state vector of entire system, which contains some elements unrelated to node i. Hence, this will increase the computational burden at each node, and transmitting such estimate to neighboring nodes will make the communication load very heavy.

4.3 Distributed MAP Estimation Algorithm

In this section, we will introduce a distributed state estimation algorithm, which generalizes the algorithm in [23] for a linear system to the nonlinear case. Based on the idea of [23], we firstly treat the edge measurement $z_{i,j}(k)$ as the local measurement of node i to estimate its local state. Then following by the local MAP estimator as (8) and (9), it is clear to get that

$$\breve{\alpha}_i(k) = H_i^{\mathrm{T}}(k)S_i^{-1}(k)Z_{i,i}(k) + \sum_{j \in \mathcal{N}_i} B_{i,j}^{\mathrm{T}}(k)T_{i,j}^{-1}(k)Z_{i,j}(k) + M_i^{-1}(k)\widetilde{x}_i(k), \qquad (13)$$

$$\breve{Q}_i(k) = H_i^{\mathrm{T}}(k)S_i^{-1}(k)H_i(k) + \sum_{j \in \mathcal{N}_i} B_{i,j}^{\mathrm{T}}(k)T_{i,j}^{-1}(k)B_{i,j}(k) + M_i^{-1}(k), \qquad (14)$$

where

$$Z_{i,i}(k) = z_{i,i}(k) + H_i(k)\widetilde{x}_i(k) - f_i(\widetilde{x}_i(k)),$$
$$Z_{i,j}(k) = z_{i,j}(k) + B_{i,j}(k)\widetilde{x}_i(k) + B_{j,i}(k)\widetilde{x}_j(k) - h_{i,j}(\widetilde{x}_i(k), \widetilde{x}_j(k)).$$

If $k = 0$, $\widetilde{x}_i(k)$ and $M_i(k)$ are replaced by $\overline{x}_i(0)$ and $\Sigma_i(0)$, respectively. So the initial state estimate of node i at each time instant k can be expressed as

$$\breve{x}_i(k, 0) = \breve{Q}_i^{-1}(k)\breve{\alpha}_i(k), \quad \breve{P}_i(k, 0) = \breve{Q}_i^{-1}(k).$$

Secondly, we update the state estimate of node i via the following distributed algorithm.

Algorithm 1 Distributed MAP estimation method

Initialization At time step $k \in \mathbb{N}_0$, each node $i \in \mathcal{V}$ computes $\breve{x}_i(k, 0)$ and $\breve{P}_i(k, 0)$, and transmits the following data to node $j \in \mathcal{N}_i$:

$$\theta_{j,i}(k, 0) = B_{i,j}(k)\breve{x}_i(k, 0),$$
$$\Theta_{j,i}(k, 0) = B_{i,j}(k)\breve{P}_i(k, 0)B_{i,j}^{\mathrm{T}}(k).$$

Main loop $h \in \mathbb{N}$ is the step number of iteration.

1) Let $\theta_{i,j}(k, h-1)$ and $\Theta_{i,j}(k, h-1)$ be the data transmitted from node j. Node i updates the local estimation and the associated covariance as follows:

$$\widehat{x}_i(k, h) = P_i(k, h)\left(\breve{\alpha}_i(k) - \sum_{j \in \mathcal{N}_i} \beta_{i,j}(k, h-1)\right),$$

$$P_i(k, h) = \left(\breve{Q}_i(k) - \sum_{j \in \mathcal{N}_i} \Phi_{i,j}(k, h-1)\right)^{-1},$$

where

$$\beta_{i,j}(k,h-1) = B_{i,j}^{\mathrm{T}}(k)T_{i,j}^{-1}(k)\theta_{i,j}(k,h-1),$$
$$\Phi_{i,j}(k,h-1) = B_{i,j}^{\mathrm{T}}(k)T_{i,j}^{-1}(k)\Theta_{i,j}(k,h-1)T_{i,j}^{-1}(k)B_{i,j}(k).$$

2) Node i also computes

$$\breve{x}_{j,i}(k,h) = \breve{P}_{j,i}(k,h)\left(\breve{\alpha}_i(k) - \sum_{n_i \in \mathcal{N}_i/\{j\}} \beta_{i,n_i}(k,h-1)\right),$$
$$\breve{P}_{j,i}(k,h) = \left(\breve{Q}_i(k) - \sum_{n_i \in \mathcal{N}_i/\{j\}} \Phi_{i,n_i}(k,h-1)\right)^{-1},$$

and then transmits the following information to node j:

$$\theta_{j,i}(k,h) = B_{i,j}(k)\breve{x}_{j,i}(k,h),$$
$$\Theta_{j,i}(k,h) = B_{i,j}(k)\breve{P}_{j,i}(k,h)B_{i,j}^{\mathrm{T}}(k).$$

It is important to note that the information transmitted from node i to node j do not include the information that node i previously receives from node j. Since the inter-area communication graph considered in this paper is acyclic, Theorem 2 in [23] has proven that the local state estimates on all nodes converge after Γ steps at each time instant.

Remark 4.1 The local state of node i can also be predicted, based on the past information. With employment of the conditional expectation operator on (3), the local state prediction $\widetilde{x}_i(k+1)$ and its covariance matrix $M_i(k+1)$ can be obtained, which are similar to (10) and (11).

Computational complexity The centralized state estimator requires $O((\sum_{i=1}^N 2s_i)^3)$ computation, whereas the computational complexity of the distributed MAP estimator of node i is $\widetilde{n}_i O((2s_i)^3)$ at each time stamp, where $\widetilde{n}_i$ 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 power systems, each node only has a few neighbors, i.e., $\widetilde{n}_i \ll N$. Thus, the computational complexity of Algorithm 1 is much smaller than that of the centralized method in Section 3.

4.4 Parameter Identification

In this section, we also use exponential weighting functions in the distributed form to strengthen the robustness of the distributed state estimation algorithm. Since $R_*(k)$ and $W_*(k)$ are diagonal matrices, we have

$$S_i(k) = (W_{i,i}^*(k))^{-1}, \ T_{i,j}(k) = (W_{i,j}^*(k))^{-1},$$
$$W_{i,i}^*(k) = W_{i,i}(k)\exp\left(-|e_i(k)|\right),$$
$$W_{i,j}^*(k) = W_{i,j}(k)\exp\left(-|z_{i,j}(k) - h_{i,j}(\widetilde{x}_i(k),\widetilde{x}_j(k))|\right),$$

where $e_i(k) = z_{i,i}(k) - f_i(\widetilde{x}_i(k))$, $W_{i,i}(k)$ and $W_{i,j}(k)$ are diagonal elements of $W(k)$.

The formula deduction of parameters $A_i(k)$ and $G_i(k)$ has been given in Section 3, but the state estimate and prediction used in these equations are obtained by the distributed method in Section 4.3.

Using the exponential weighted method in each node i, the estimating equation of $R_i(k)$ is as follows:

$$\widehat{R}_i(k) = (1 - d_i(k-1))\widehat{R}_i(k-1) + d_i(k-1)\big(P_i(k) + K_i(k)e_i(k)e_i^{\mathrm{T}}(k)K_i^{\mathrm{T}}(k)\big),$$

where

$$d_i(k-1) = \frac{1 - b_i}{1 - b_i^k}, \quad K_i(k) = M_i(k)H_i^{\mathrm{T}}(k)(H_i(k)M_i(k)H_i^{\mathrm{T}}(k) + S_i(k))^{-1},$$

and every b_i has the arbitrary value between 0.95 and 0.99. We also use the diagonal elements of $\widehat{R}_i(k)$ to compose the new $\widehat{R}_i^{\#}(k)$.

The main purpose of this paper is to verify the effectiveness and robustness of Algorithm 1 together with the identification of model parameters. In the next section, numerical experiments are used to illustrate our conclusions.

5 Simulations

In this section, we will show the effectiveness of Algorithm 1 with the identified parameters (DSE), and compare its performance with the following algorithms: The centralized state estimator (CSE) containing the methods in [21] and [22], the diffusive state estimator (DSSE1) in Subsection 4.1 and the distributed state estimator (DSSE2) in Subsection 4.2. Test results are made on the IEEE 39-bus and 118-bus systems. In Subsection 5.1, we describe the test systems for simulation results. In Subsection 5.2, we introduce performance indices, which are used for assessing the performance of these algorithms. In Subsection 5.3, detailed numerical simulations are investigated under various operating conditions, which are constructed to explore the effectiveness of DSE over other methods.

5.1 Simulation Description

The IEEE 39-bus and 118-bus systems are used to test the above mentioned methods. As shown in Figure 3, the IEEE 39-bus system has four nonoverlapping control areas, which contain 10, 10, 10 and 9 buses, respectively. Combining all buses of each control area containing into a single node, one can see that the obtained network is acyclic. It needs to be pointed out that the IEEE 39-bus system is only monitored by SCADA measurements. Furthermore, the IEEE 118-bus system is split into eight control areas as shown in Figure 1, and the resulting network in Figure 2 is also acyclic. Hence, Assumption 2.1 is satisfied. In the IEEE 118-bus system, the buses installed with PMU measurements are marked as blocks in Figure 1, and these algorithms are based on the mixed SCADA and PMU measurements[24]. The values of parameters in power flows are taken from [26].

Figure 3 Topological structure of the IEEE 39-bus system

In the simulations, we consider the slow dynamics of power systems, i.e., the smooth load changes over a period of time are obtained by running successful load flows. According to [21], a linear trend between 1% and 3% along with a randomly distributed fluctuation of ±4% are added to the load curve, where different values are used for different buses. For each time sample, a load flow is carried out using the load and generation values obtained through load curve and generation participation factor. The outcome of the load flow on each node i at each time instant k is served as true state $x_i^+(k)$, true values of the local measurement $z_{i,i}^+(k)$ and the edge measurement $z_{i,j}^+(k)$ including voltage magnitudes and phase angles, and power flows. The actual measurements $z_{i,i}(k)$ and $z_{i,j}(k)$ of node i are obtained by adding $z_{i,i}^+(k)$ with random normally distributed noise of 1% for SCADA voltage magnitude measurements, 0.2% for PMU voltage magnitude and phase angle measurements, and adding $z_{i,j}^+(k)$ with 2% standard deviation for power flows, respectively. Additionally, the diagonal elements of $R_i(0)$ are set to 10^{-6}.

5.2 Performance Indices

The performance of DSE and its comparison with CSE, DSSE1 and DSSE2 at time instant k are assessed by the following indices.

1) Prediction Step: The forecasting performance is evaluated by using the relative error between the predicted state vector $\widetilde{x}_i(k)$ obtained respectively by CSE and DSE, and the true state vector $x_i^+(k)$, which is the index

$$\varepsilon^p(k) = \frac{\Sigma_{i=1}^N |\widetilde{x}_i(k) - x_i^+(k)|}{\Sigma_{i=1}^N |x_i^+(k)|} \times 100\%.$$

Since the static state estimator fails to predict the future state, the $\varepsilon^p(k)$ values of DSSE1 and DSSE2 are null.

2) Filtering Step: The filtering performance is evaluated by using the relative error between the state estimation $\widehat{x}_i(k)$ obtained respectively by CSE, DSE, DSSE1 and DSSE2, and the true state vector $x_i^+(k)$, which is

$$\varepsilon^f(k) = \frac{\Sigma_{i=1}^N |\widehat{x}_i(k) - x_i^+(k)|}{\Sigma_{i=1}^N |x_i^+(k)|} \times 100\%.$$

3) Performance Index: The performance index is

$$J(k) = \frac{\Sigma|\widehat{z}_{i,m}(k) - z_{i,m}^+(k)|}{\Sigma|z_{i,m}(k) - z_{i,m}^+(k)|}, \quad m \in \overline{\mathcal{N}}_i,$$

where $\widehat{z}_{i,m}(k)$ is the filtered measurement vector, and $J(k)$ is the mean absolute ratio of estimated and true error of measurements, computed to assess the entire measurement estimation achievement.

We can see from [22] that, the smaller the indices $\varepsilon^p(k)$ and $\varepsilon^f(k)$ are, the better the effects of state prediction and estimation are. The same conclusion is drawn with $J(k)$. The uncertainty level existing in the measurements is reduced and the effectiveness of state filtering is confirmed, when $J(k)$ is smaller than 1.

5.3 Test Results

In the simulations, the time step of tests is chosen as one second. DSE can be divided into two types. The one is the DSE1, which means that we run one iteration of DSE per time step and then predict the local state, and the other one is the DSE2, which denotes that at each time instant DSE iterates two steps for the IEEE 39-bus system and four steps for the IEEE 118-bus system. The diameters of networks partitioned from the IEEE 39-bus and 118-bus systems are 2 and 4, respectively, and Theorem 2 in [23] has proven that per instant of time the local state estimate on each node converges after a finite number of iterations, which equal to the diameters of the considered networks. Meanwhile, DSSE1 and DSSE2 iterate the same steps as DSE2. We use Monte Carlo simulations to compute performance indices, and 1000 Monte Carlo runs are taken in these methods for comparison.

In order to verify the validity of Algorithm 1 with parameter identification in this paper, four test scenarios including normal operating condition, sudden load change, presence of bad data and topology error condition are investigated. Results are presented and discussed below.

Case 1 Normal Operation Condition: Table 1 shows the results of using CSE, DSSE1, DSSE2 and DSE, when the IEEE 39-bus and 118-bus systems are operated at the normal operating condition. The simulation results are carried out through 20 time samples. Figure 4 describes the time evolution of the corresponding performance indices of these methods, which contain four parts corresponding to the four test conditions and simulations of each part are also carried out through 20 time samples. The first parts of Figure 4 correspond to Table 1. One can see from Table 1 and Figure 4 that the performance index of CSE is optimal, followed by the two types of DSE, and the average prediction and filtering indices of DSE are marginally worse than that of CSE. For example, when CSE, DSE1 and DSE2 are applied in the IEEE 118-bus system, the average values of $\varepsilon^f(k)$ are 0.329%, 0.547% and 0.486%, respectively. Furthermore, among the considered methods, DSSE1 and DSSE2 have higher $\varepsilon^f(k)$ and $J(k)$ values than that of DSE, because state estimation in static case does not base on the prediction-correction process. From the index values in Table 1, the state estimate and prediction of DSE2 are more accurate than that of DSE1, which means that the choice of iterations can affect the filter behavior, and they also verify the benefits of DSE over the distributed static estimation methods.

$\underline{\textcircled{2}}$ Springer

Table 1 Performance indices for normal operation condition

system	Indices	$\varepsilon^p(k)(\%)$		$\varepsilon^f(k)(\%)$		$J(k)$	
		Max.	Ave.	Max.	Ave.	Max.	Ave.
39 bus	CSE	1.126	0.645	0.752	0.513	0.535	0.467
	DSE1	2.463	1.102	1.341	0.865	0.806	0.723
	DSE2	1.871	0.894	1.132	0.740	0.721	0.647
	DSSE1	/	/	1.896	1.216	1.013	0.918
	DSSE2	/	/	2.303	1.322	1.052	0.983
118 bus	CSE	0.822	0.416	0.533	0.329	0.325	0.304
	DSE1	1.587	0.709	0.873	0.547	0.483	0.445
	DSE2	0.986	0.608	0.655	0.486	0.451	0.411
	DSSE1	/	/	1.146	0.694	0.621	0.552
	DSSE2	/	/	1.465	0.776	0.652	0.583
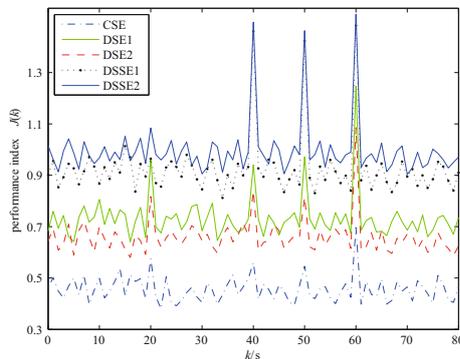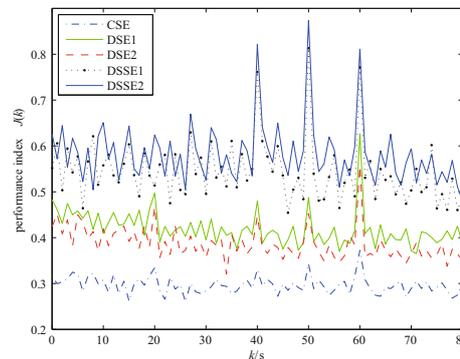
(a) IEEE 39-bus system (b) IEEE 118-bus system

Figure 4 Variations of $J(k)$ for the presence of anomalies of different methods at the IEEE 39-bus and 118-bus systems

Figure 5 depicts the time evolution of the real values and the corresponding estimates of bus 23 in the IEEE 118-bus system filtered by CSE and DSE, when the system is operated at the normal operating condition. We can see that the values obtained by DSE are probably correct at each time instant, especially DSE2, the estimated values of which are close to that of CSE, when we estimate voltage magnitudes and voltage phases. Furthermore, by using the MATLAB (MathWorks, Inc., Natick, MA, USA), running times of CSE, DSE1 and DSE2 for processing Figure 5 (a) are 0.02637s, 0.01592s and 0.01965s, respectively. This offers the computational advantage of DSE over CSE for large-scale systems.

(a) voltage magnitude   (b) phase angle
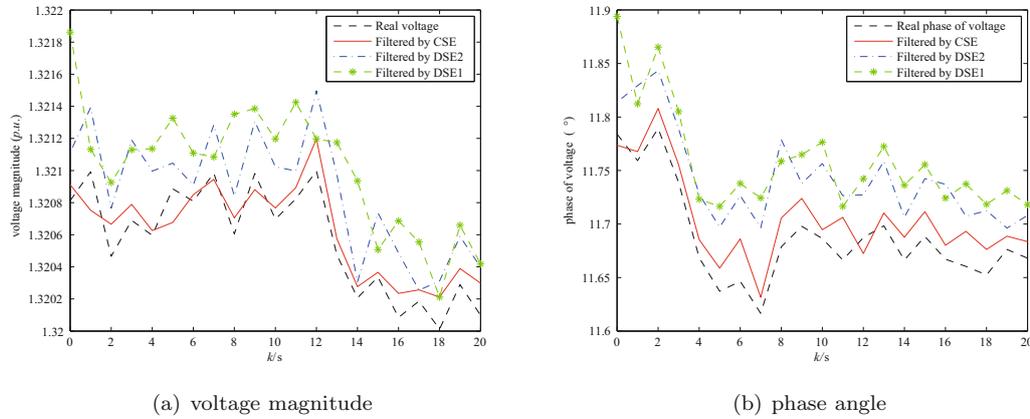
Figure 5   Variations of real and estimated values of bus 23 in the IEEE 118-bus system

Case 2 Sudden Load Change Condition: In this case, all the state estimators are assessed under sudden load change conditions. For the IEEE 39-bus system, the load change is assumed to occur as follows: 50% loads are cut on buses 2, 5, 18, 23, 29 at the 20th time sample. While for the IEEE 118-bus system, these methods are applied to the following scenario: 30% loads are increased on buses 8, 11, 12, 15, 16, 21, 24, 30, 31, 34, 37, 40, 49, 50, 51, 54, 56, 65, 69, 70, 77, 85, 89, 92, 94, 95, 97, 100, 103, 106 at $k = 20$. The performance indices of these approaches are shown in Table 2 and the second parts of Figure 4. We know that static state estimation never needs to consider the previous state prediction, so DSSE1 and DSSE2 are not affected by the sudden load change. From the simulation results, CSE and DSE are little impacted at time $k = 20$, which means that the adjustment of parameters in our algorithm is effective. The $J(k)$ values of the two types of DSE reduce quickly after the sudden load change, which means that DSE can come back to the normal level adaptively. From the index values in Table 2, the performance comparisons of DSE1 and DSE2 with CSE, DSSE1 and DSSE2 are similar to Case 1.

In order to demonstrate the effect of the parameter identification more clearly, we compare the simulation results of DSE2 with that of the DSE3, which also iterates 4 steps per each time k for the IEEE 118-bus system, but we do not revise parameters $R_i(k)$, $S_i(k)$ and $T_{i,j}(k)$ in DSE3. Figure 6 shows the variations of $\varepsilon^p(k)$ values obtained by DSE2 and DSE3. Since the state prediction before the sudden load change is very different to the actual state value, the index $\varepsilon^p(k)$ of DSE3 becomes larger and the forecasting precision is depressed as soon as the sudden load change occurs. In this figure, the maximal value of $\varepsilon^p(k)$ obtained by DSE3 decreases from 2.3658% down to 0.9778% by DSE2, which still supports the feasibility of the proposed algorithm with parameter identification. We can also see from Figure 6 that the $\varepsilon^p(k)$ values of DSE2 are more accurate than that of DSE3 at all time samples, even when the sudden load is not changed.

Table 2 Performance indices for sudden load change condition

system	Indices	$\varepsilon^p(k)(\%)$		$\varepsilon^f(k)(\%)$		$J(k)$	
		Max.	Ave.	Max.	Ave.	Max.	Ave.
39 bus	CSE	1.582	0.658	0.867	0.518	0.574	0.449
	DSE1	3.125	1.144	1.685	0.879	0.966	0.731
	DSE2	2.209	0.917	1.363	0.752	0.817	0.668
	DSSE1	/	/	1.845	1.213	0.980	0.907
	DSSE2	/	/	2.152	1.314	1.084	0.981
118 bus	CSE	0.913	0.418	0.591	0.331	0.334	0.292
	DSE1	1.837	0.720	0.864	0.544	0.497	0.416
	DSE2	1.124	0.606	0.743	0.491	0.467	0.384
	DSSE1	/	/	1.137	0.681	0.629	0.547
	DSSE2	/	/	1.448	0.762	0.670	0.575
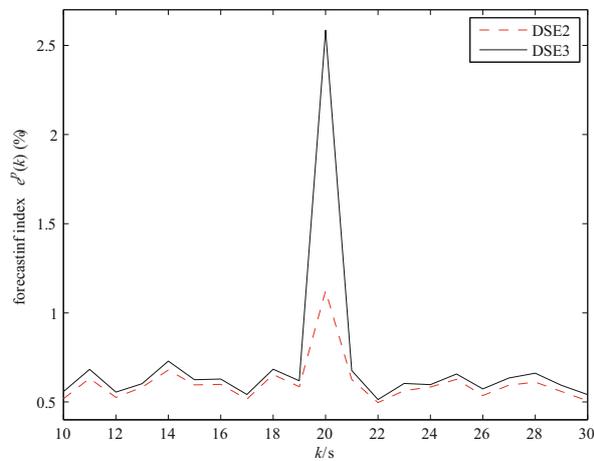
Figure 6 Variations of forecasting indices for the sudden load change condition at the IEEE 118-bus system

Table 3 Performance indices for bad data condition

system	Indices	$\varepsilon^p(k)(\%)$		$\varepsilon^f(k)(\%)$		$J(k)$	
		Max.	Ave.	Max.	Ave.	Max.	Ave.
39 bus	CSE	1.826	0.714	0.948	0.537	0.560	0.451
	DSE1	3.597	1.235	1.844	0.923	0.972	0.742
	DSE2	2.488	0.978	1.659	0.796	0.839	0.670
	DSSE1	/	/	4.921	1.537	1.461	0.951
	DSSE2	/	/	5.360	1.661	1.496	1.030
118 bus	CSE	0.806	0.425	0.547	0.338	0.343	0.298
	DSE1	1.671	0.734	0.826	0.553	0.488	0.406
	DSE2	1.054	0.617	0.726	0.478	0.453	0.380
	DSSE1	/	/	3.540	0.924	0.813	0.552
	DSSE2	/	/	4.183	1.056	0.873	0.608

Case 3 Bad Data Condition: In this part, these methods are applied to the condition, where bad data are included in power system measurements. The following scenarios are investigated at the IEEE 39-bus and 118-bus systems:

(i) One raw measurement error of 30% at the 40th time sample;

(ii) One raw measurement is mistaken as zero at the 50th time sample.

Suppose that no measures are taken to identify these bad data. Results of performance indices of these methods are shown in Table 3 and the third parts of Figure 4. From the simulation results, the considered indices of all the methods are affected by bad data, the impact of which on DSSE1 and DSSE2 is very serious. But the degree of variations of CSE and the two types of DSE is not drastic, which means that our method restrains the influence of polluted measurements. In Table 3, the average values of $\varepsilon^p(k)$, $\varepsilon^f(k)$ and $J(k)$ for DSE1 and DSE2 are smaller than that of two distributed static methods, but larger than the ones of CSE. The two types of DSE take less time to run than CSE, the running times of which are similar to that in Case 1. According to Figure 4, $J(k)$ values of DSE2 are smaller than DSE1 and the robustness of DSE is confirmed. When bad data comes, all indices of DSE are not drastically changed and they can also reduce to the normal level at the next time sample, which reveals that DSE owns excellent performance to the influence of polluted measurements.

Figure 7 plots the variations of the index $\varepsilon^f(k)$ obtained by DSE2 and DSE3 for all time samples considered. We can see that DSE3 is heavily affected by bad data, and the maximal value obtained by DSE3 is enormously larger than that of DSE2. Meanwhile, the $\varepsilon^f(k)$ values obtained from DSE2 at each time instant are smaller than that of DSE3, due to the exponential weighting functions. Figures 6 and 7 mean that the robustness of the proposed algorithm in

this paper is strengthened better than the method in [23], and the state estimate obtained by DSE at each time sample is also more accurate than Algorithm 1 in [23].
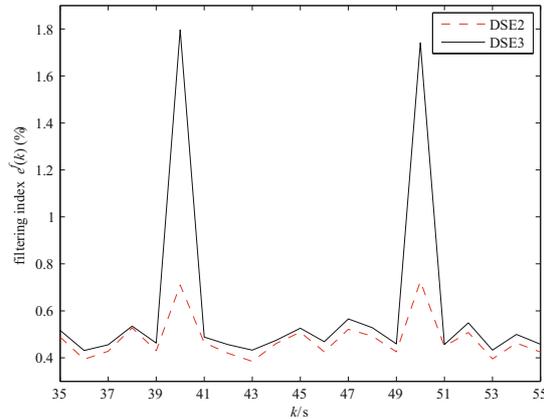
Figure 7 Variations of filtering indices for bad data conditions at the IEEE 118-bus system

Case 4 Topology Error Condition: In this case, the performance of these methods are evaluated when the topology is reported incorrectly. For the IEEE 39-bus system, the lines 2–3, 5–8, 13–14, 17–18 and 23–24 are assumed wrongly reported at the 60th time sample. While for the IEEE 118-bus system, the following operation conditions are simulated: The branches 14–15, 17–31, 37–38, 49–51, 54–56, 69–77, 70–71, 94–95, 92–100 and 100–104 are disconnected at the 60th time sample but they are reported connected. Table 4 and the last parts of Figure 4 describe the performance evaluation results of this case. We can see that all the methods are seriously affected by the topology error condition, but the performance of DSE1 and DSE2 is still more accurate than that of DSSE1 and DSSE2. Besides, Figure 4 shows that the performance indices of DSE can also reduce quickly after topological structures of the considered systems return to normal. In Table 4, the maximal and average values of the considered indices obtained by DSE in the IEEE 118-bus system are smaller than that in the IEEE 39-bus system. This affirms that when the system size increases, the influence caused by the topology error condition may be better restrained[21].

Summary of simulation results From the above mentioned analysis of simulations, the performance of DSE is somewhat worse than CSE. As shown in Figure 5, DSE2 offers marginally less accurate estimates than that of CSE, and the maximal and average values of the prediction and filtering indices obtained by the two types of DSE are quite close to that of CSE shown in these tables. Under different scenarios in the IEEE 118-bus system, running times of CSE, DSE1, DSE2, DSSE1 and DSSE2 for processing $J(k)$ are as follows: 0.11470s, 0.07265s, 0.08236s, 0.05463s and 0.05175s, which together with the analysis of computational complexity in Subsection 4.3 indicate that DSE is more applicable than CSE to large-scale power systems. Although running times of DSE are a few longer than that of DSSE1 and DSSE2, the effectiveness of DSE is clearly better than the distributed static estimation methods, and more

importantly, DSE has the forecasting ability which is essential for real-time state estimation. Parameter identification is also effective to deal with abnormal conditions. Thus, DSE can be a potential candidate for applications of dynamic state estimation.

Table 4 Performance indices for topology error condition

system	Indices	$\varepsilon^p(k)(\%)$		$\varepsilon^f(k)(\%)$		$J(k)$	
		Max.	Ave.	Max.	Ave.	Max.	Ave.
39 bus	CSE	2.478	0.711	1.207	0.533	0.697	0.459
	DSE1	6.632	1.286	3.526	0.954	1.247	0.733
	DSE2	5.246	1.047	2.914	0.829	1.086	0.654
	DSSE1	/	/	5.467	1.394	1.483	0.922
	DSSE2	/	/	6.025	1.518	1.527	0.988
118 bus	CSE	1.174	0.434	0.862	0.344	0.373	0.294
	DSE1	3.106	0.785	2.387	0.603	0.627	0.411
	DSE2	2.193	0.645	1.534	0.507	0.556	0.374
	DSSE1	/	/	4.122	0.843	0.771	0.525
	DSSE2	/	/	4.615	0.937	0.813	0.562

6 Conclusions

In this paper, we use a linear exponential smoothing technique and exponential weighting functions to identify model parameters, which increase the robustness and enhance the precision of the distributed MAP estimation algorithm. In order to validate the effectiveness of DSE, two example power systems under different conditions are tested, where measurements are composed by the conventional SCADA measurements and the limited PMU measurements. Low computational complexity gives DES the merit over CSE, and flexibility and efficiency in applications also support DSE, since in the simulation results, state estimates and values of performance indices obtained by DSE are marginally worse than that of CSE. Compared with DSSE1 and DSSE2, DSE is verified by the forecasting ability, accurate performance and light communication load. All this leads up to the application value of the distributed MAP estimation algorithm in wide-area power systems.

References

[1] Wu F F, Moslehi K, and Bose A, Power system control centers: Past, present, and future, *Proceedings of the IEEE*, 2005, **93**(11): 1890–1908.
[2] Ree J D L, Centeno V, Thorp J S, et al., Synchronized phasor measurement applications in power systems, *IEEE Trans. Smart Grid*, 2010, **1**(1): 20–27.

[3] Kar S, Moura J M F, and Ramanan K, Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication, *IEEE Trans. Information Theory*, 2012, **58**(6): 3575–3605.

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

[5] Yu W, Chen G, Wang Z, et al., Distributed consensus filtering in sensor networks, *IEEE Trans. Systems, Man and Cybernetics*, 2009, **39**(6): 1568–1577.

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

[7] Jiang W, Vittal V, and Heydt G T, A distributed state estimator utilizing synchronized phasor measurements, *IEEE Trans. Power Systems*, 2007, **22**(2): 563–571.

[8] Falcao D M, Wu F F, and Murphy L, Parallel and distributed state estimation, *IEEE Trans. Power Systems*, 1995, **10**(2): 724–730.

[9] Korres G N, A distributed multiarea state estimation, *IEEE Trans. Power Systems*, 2011, **26**(1): 73–84.

[10] Kar S, Moura J M F, and Ramanan K, Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication, *IEEE Trans. Information Theory*, 2012, **58**(6): 3575–3605.

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

[12] Xie L, Choi D H, Kar S, et al., Fully distributed state estimation for wide-area monitoring systems, *IEEE Trans. Smart Grid*, 2012, **3**(3): 1154–1169.

[13] Karimipour H and Dinavahi V, Parallel domain decomposition based distributed state estimation for large-scale power systems, *Proceedings of the Industrial and Commercial Power Systems Technical Conference*, Calgary, Canada, 2015, 1–5.

[14] Tai X, Lin Z Y, Fu M Y, et al., A new distributed state estimation technique for power networks, *Proceedings of the 2013 American Control Conference*, Washington, USA, 2013, 3338–3343.

[15] Debs A and Larson R, A dynamic estimator for tracking the state of a power system, *IEEE Trans. Power Apparatus and Systems*, 1970, **89**(7): 1670–1678.

[16] Beides H M and Heydt G T, Dynamic state estimation of power system harmonics using Kalman filter methodology, *IEEE Trans. Power Delivery*, 1991, **6**(4): 1663–1670.

[17] Durgaprasad G and Thakur S S, Robust dynamic state estimation of power systems based on M-estimation and realistic modeling of system dynamics, *IEEE Trans. Power Systems*, 1998, **13**(4): 1331–1336.

[18] Lin J M, Huang S J, and Shih K R, Application of sliding surface-enhanced control for dynamic state estimation of a power system, *IEEE Trans. Power Systems*, 2003, **18**(2): 570–577.

[19] Do Coutto Filho M B and de Souza J C S, Forecasting-aided state estimation-part I: Panorama, *IEEE Trans. Power Systems*, 2009, **24**(4): 1667–1677.

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

[21] Shih K R and Huang S J, Application of a robust algorithm for dynamic state estimation of a power system, *IEEE Trans. Power Systems*, 2002, **17**(1): 141–147.

[22] Li H and Li W G, Estimation and forecasting of dynamic state estimation in power systems, *Procssdings of the 2009 International Conference on Sustainable Power Generation and Supply*, Nanjing, China, 2009, 1–6.

[23] Sun Y B, Fu M Y, Wang B C, et al., Dynamic state estimation in power systems using a distributed MAP method, *Proceedings of the 34th Chinese Control Conference*, Hangzhou, China, 2015, 47–52.

[24] Sun Y B, Fu M Y, Wang B C, et al., Dynamic state estimation for power networks using distributed MAP technique, *Automatica*, 2016, **73**: 27–37.

[25] Makridakis S and Wheelwright S C, *Forecasting Methods and Applications*, Wiley, New York, 1978.

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