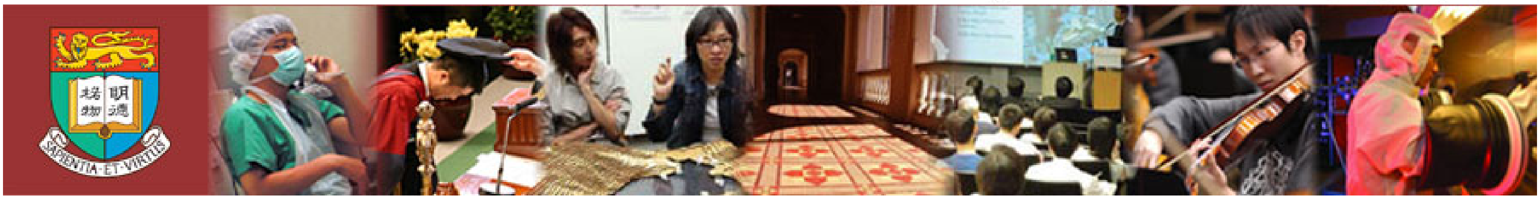

Title	Distributed Bayesian hybrid power state estimation with PMU synchronization errors
Author(s)	Du, J; Ma, S; Wu, YC; Poor, VH
Citation	The 2014 IEEE Global Communications Conference (GLOBECOM 2014), Austin, TX., 8-12 December 2014. In Conference Proceedings, 2014, p. 3174-3179
Issued Date	2014
URL	http://hdl.handle.net/10722/214822
Rights	IEEE Conference and Exhibition of Global Telecommunications (GLOBECOM). Copyright © IEEE.

Distributed Bayesian Hybrid Power State Estimation with PMU Synchronization Errors

Jian Du*, Shaodan Ma*, Yik-Chung Wu† and H. Vincent Poor ‡

* Department of Electrical and Computer Engineering, University of Macau, Macau
{jiandu,shaodanma}@umac.mo
†Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong
ycwu@eee.hku.hk
‡Department of Electrical Engineering, Princeton University, Princeton, New Jersey
poor@princeton.edu

Abstract—**This paper presents a distributed hybrid power state estimator, with measurements from both the traditional supervisory control and data acquisition (SCADA) system and the newly invented phasor measurement units (PMUs). The proposed distributed algorithm, which jointly estimates the power states and PMU phase errors, only involves local computations and limited information exchange between neighboring areas, thus alleviating the heavy communication burden compared to the centralized approach. Simulation results show that the performance of the proposed algorithm is very close to that of centralized optimal hybrid state estimates without sampling phase error.**

I. INTRODUCTION

Due to the time-varying nature of power generation and consumption, state estimation in the power grid has always been a fundamental function for real-time monitoring of electric power networks [1]. The knowledge of the state vector at each bus, i.e., voltage magnitude and phase angle, enables the energy management system (EMS) to perform various crucial tasks, such as bad data detection, optimizing power flows, maintaining system stability and reliability, etc.

In the past several decades, the supervisory control and data acquisition (SCADA) system has been universally established in the electric power industry, and installed virtually in all EMSs around the world to manage large and complex power systems. In SCADA systems, the voltage magnitude, power injection at each bus and current flow between neighbouring buses are measured and then sent to a master terminal unit to perform state estimation. As these measurements are nonlinear functions of the power states, the state estimation programs are formulated as iterative reweighted least-squares solution [2].

The invention of phasor measurement units (PMUs) [3] has made it possible to measure power states directly, which is infeasible with SCADA systems. Despite the advantage of PMUs over SCADA, the traditional SCADA system cannot be replaced by a PMU-based system overnight. The reason is two fold. Firstly, in practice there are only sporadic PMUs deployed in the power grid due to expensive installation costs. Besides, the SCADA system involves long-term significant investment, and is currently working smoothly in existing power systems. Consequently, hybrid state estimation with both SCADA and PMU measurements is appealing. One straightforward methodology is to simultaneously process both SCADA and PMU raw measurements [4]. However, this simultaneous data processing, which leads to a totally different set of estimation equations, requires significant changes to existing EMS/SCADA systems [5], and is not preferable in practice. In fact, incorporating PMU measurements with minimal change to the SCADA system is an important research problem [5].

In addition to the challenge of integrating PMU with S-CADA data, there are also other practical concerns that need to be considered. Firstly, it is usually assumed that PMUs provide synchronized sampling of voltage and current signals [6] due to the Global Positioning System (GPS) receiver included in the PMU. However, tests [7] provided by a joint effort between the U.S. Department of Energy and the North American Electric Reliability Corporation show that PMUs from multiple vendors can yield up to $\pm 277.8\mu s$ sampling phase errors (or $\pm 6^\circ$ phase error in a 60Hz power system) due to different delays in the instrument transformers used by different vendors. Sampling phase mismatch in PMUs will make the state estimation problem nonlinear, which offsets the original motivation for introducing PMUs. It is important to develop state estimation algorithms that are robust to sampling phase errors.

Secondly, with fast sampling rates of PMU devices, a centralized approach, which requires gathering of measurements through propagating a significant computational large amounts of data from peripheral nodes to a central processing unit, imposes heavy communication burden across the whole network and imposes a significant computation burden at the control center. Decentralizing the computations across different control areas and fusing information in a hierarchical structure have thus been investigated in [8], [9]. However, these approaches need to meet the requirement of local observability of all the control areas. Consequently, fully distributed

This work was supported in part by the Macao Science and Technology Development Fund under grant 067/2013/A, in part by the Research Committee of University of Macau under grants MYRG078 and MYRG101, and in part by the U.S. Air Force Office of Scientific Research under MURI Grant FA9550-09-1-0643.

978-1-4799-3512-3/14/$31.00 ©2014 IEEE

state estimation scalable with network size is preferred [4], [10]–[12].

In view of above problems, this paper proposes a distributed power state estimation algorithm, which only involves local computations and information exchanges with immediate neighborhoods, and is suitable for implementation in large-scale power grids. In contrast to [4], [10] and [12], the proposed distributed algorithm integrates the data from both the SCADA system and PMUs while keeping the existing SCADA system intact, and the observability problem is bypassed. The challenging problem of sampling phase errors in PMUs is also considered. Simulation results show that after convergence the proposed algorithm performs very close to that of the ideal case which assumes perfect synchronization among PMUs, and centralized information processing.

The following notations are used throughout this paper. Boldface uppercase and lowercase letters will be used for matrices and vectors, respectively. $\mathbb{E}\{\cdot\}$ denotes the expectation of its argument and $\jmath \triangleq \sqrt{-1}$. Superscript T denotes transpose. The symbol $\boldsymbol{I}_N$ represents the $N \times N$ identity matrix. $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{R})$ stands for the probability density function (pdf) of a Gaussian random vector $\boldsymbol{x}$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{R}$. The symbol $\propto$ represents a linear scalar relationship between two real-valued functions. The cardinality of a set $\mathcal{V}$ is denoted by $|\mathcal{V}|$ and the difference between two sets $\mathcal{V}$ and $\mathcal{A}$ is denoted by $\mathcal{V} \setminus \mathcal{A}$.

II. Hybrid Estimation System Model

A. PMU Measurements with Sampling Errors

For a power grid, the continuous voltage on bus i is denoted as $A_i \cos(2\pi f_c t + \phi_i)$, with A_i being the amplitude and ϕ_i being the phase angle in radians. Ideally, a PMU provides measurements in rectangular coordinates: $A_i \cos(\phi_i)$ and $A_i \sin(\phi_i)$. However, for reasons of sampling phase error [6], [7] and measurement error, the measured voltage at bus i would be [7]

$$x_i^r = A_i \cos(\theta_i + \phi_i) + w_{i,E}^r, \tag{1}$$

$$x_i^\jmath = A_i \sin(\theta_i + \phi_i) + w_{i,E}^\jmath, \tag{2}$$

where θ_i is the phase error induced by an unknown and random sampling delay, and $w_{i,E}^r$ and $w_{i,E}^\jmath$ are the Gaussian measurement noises. On the other hand, a PMU also measures the current between neighboring buses. Let the admittance at the branch $\{i, j\}$ be $g_{ij} + \jmath \cdot b_{ij}$, the shunt admittance at bus i be $\jmath B_i$, and the transformer turn ratio from bus i to j be $\rho_{ij} = |\rho_{ij}| \exp\{\jmath\varphi_{ij}\}$. Under sampling phase error, the real and imaginary parts of the measured current at bus i are given by [13]

$$\begin{aligned} y_{ij}^r =& \kappa_{ij}^1 A_i \cos(\theta_i + \phi_i) - \kappa_{ij}^2 A_i \sin(\theta_i + \phi_i) \\ & -\kappa_{ij}^3 A_j \cos(\theta_i + \phi_j) + \kappa_{ij}^4 A_j \cos(\theta_i + \phi_j) + w_{i,I}^r, \end{aligned} \tag{3}$$

$$\begin{aligned} y_{ij}^\jmath =& \kappa_{ij}^2 A_i \cos(\theta_i + \phi_i) + \kappa_{ij}^1 A_i \sin(\theta_i + \phi_i) \\ & -\kappa_{ij}^4 A_j \cos(\theta_i + \phi_j) - \kappa_{ij}^3 A_j \sin(\theta_i + \phi_j) + w_{i,I}^\jmath, \end{aligned} \tag{4}$$

where $\kappa_{ij}^1 \triangleq |\rho_{ij}|^2 g_{ij}$, $\kappa_{ij}^2 \triangleq |\rho_{ij}|^2 (b_{ij} + B_i)$, $\kappa_{ij}^3 \triangleq |\rho_{ij}\rho_{ji}|(\cos\varphi_i^\jmath g_{ij} - \sin\varphi_i^\jmath b_{ij})$, $\kappa_{ij}^4 \triangleq |\rho_{ij}\rho_{ji}|(\cos\varphi_i^\jmath b_{ij} + \sin\varphi_i^\jmath g_{ij})$, and $w_{i,I}^r$ and $w_{i,I}^\jmath$ are the corresponding Gaussian measurement errors.

In general, since the phase error θ_i is small (e.g., the maximum sampling phase error measured by the North American SynchroPhasor Initiative is $6°$ [7]), the standard approximations $\sin\theta_i \approx \theta_i$ and $\cos\theta_i \approx 1$ can be applied to (1) and (2), leading to [14]

$$x_i^r \approx E_i^r - E_i^\jmath \theta_i + w_{i,E}^r, \quad x_i^\jmath \approx E_i^\jmath + E_i^r \theta_i + w_{i,E}^\jmath, \tag{5}$$

where $E_i^r \triangleq A_i \cos(\phi_i)$ and $E_i^\jmath \triangleq A_i \sin(\phi_i)$ denote the true power state. Applying the same approximations to (3) and (4) yields

$$\begin{aligned} y_{ij}^r \approx& \kappa_{ij}^1 E_i^r - \kappa_{ij}^2 E_i^\jmath - \kappa_{ij}^3 E_j^r + \kappa_{ij}^4 E_j^\jmath \\ & + \theta_i \{ -\kappa_{ij}^2 E_i^r - \kappa_{ij}^1 E_i^\jmath + \kappa_{ij}^4 E_j^r + \kappa_{ij}^3 E_j^\jmath \} + w_{i,I}^r, \end{aligned} \tag{6}$$

$$\begin{aligned} y_{ij}^\jmath \approx& \kappa_{ij}^2 E_i^r + \kappa_{ij}^1 E_i^\jmath - \kappa_{ij}^4 E_j^r - \kappa_{ij}^3 E_j^\jmath \\ & + \theta_i \{ \kappa_{ij}^1 E_i^r - \kappa_{ij}^2 E_i^\jmath - \kappa_{ij}^3 E_j^r + \kappa_{ij}^4 E_j^\jmath \} + w_{i,I}^\jmath. \end{aligned} \tag{7}$$

We gather all the PMU measurements related to bus i as $\boldsymbol{z}_i = [x_i^r, x_i^\jmath, y_{ij_1}^r, y_{ij_1}^\jmath, \ldots, y_{ij_n}^r, y_{ij_n}^\jmath]^T$ where j_k is the index of bus connected to bus i, and arranged in ascending order. Using (5), (6) and (7), $\boldsymbol{z}_i$ can be expressed as [14]

$$\boldsymbol{z}_i = \sum_{j \in \mathcal{M}(i)} \boldsymbol{H}_{ij} \boldsymbol{s}_j + \theta_i \sum_{j \in \mathcal{M}(i)} \boldsymbol{G}_{ij} \boldsymbol{s}_j + \boldsymbol{w}_i, \tag{8}$$

where $\boldsymbol{s}_i \triangleq [E_i^r, E_i^\jmath]^T$; $\mathcal{M}(i)$ is the set of all immediate neighboring buses of bus i and also includes bus i; $\boldsymbol{H}_{ij}$ and $\boldsymbol{G}_{ij}$ are known matrices containing elements 0, 1 κ_{ij}^1, κ_{ij}^2, κ_{ij}^3 and κ_{ij}^4; and the measurement error vector $\boldsymbol{w}_i$ is assumed to be Gaussian $\boldsymbol{w}_i \sim \mathcal{N}(\boldsymbol{w}_i|\boldsymbol{0}, \sigma_i^2 \boldsymbol{I})$, with σ_i^2 being the i^{th} PMU's measurement error variance.

B. Mixed Measurement from SCADA and PMUs

For the existing SCADA system, the RTUs measure active and reactive power flows in network branches, bus injections and voltage magnitudes at buses. The measurements of the whole network by the SCADA system can be described as [5] $\boldsymbol{\zeta} = \boldsymbol{g}(\boldsymbol{\xi}) + \boldsymbol{n}$, where $\boldsymbol{\zeta}$ is the vector of the measurements from RTUs in the SCADA system, $\boldsymbol{\xi} \triangleq [A_1, \phi_1, A_2, \phi_2, \ldots, A_{|\mathcal{B}|}, \phi_{|\mathcal{B}|}]^T$ with $\mathcal{B}$ denoting the set of buses, and $\boldsymbol{n} \sim \mathcal{N}(\boldsymbol{n}|\boldsymbol{0}, \boldsymbol{W})$ is the measurement noise from RTUs. Due to the nonlinear function $\boldsymbol{g}(\cdot)$, $\boldsymbol{\xi}$ can be determined by the iterative reweighted least-squares algorithm [15], and it was shown in [15] that with proper initialization, such a SCADA-based state estimate $\hat{\boldsymbol{\xi}}$ converges to the maximum likelihood (ML) solution with covariance matrix $\boldsymbol{\Upsilon} = [\nabla \boldsymbol{g}(\boldsymbol{\xi})^T \boldsymbol{W}^{-1} \nabla \boldsymbol{g}(\boldsymbol{\xi})]^{-1}|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}}$, where $\nabla \boldsymbol{g}(\boldsymbol{\xi})$ is the partial derivative of $\boldsymbol{g}$ with respect to $\boldsymbol{\xi}$.

While there are many possible ways of integrating measurements from SCADA and PMUs, in this paper, we adopt the approach that keeps the SCADA system intact, as the SCADA system involves long-term investment and is running smoothing in current power networks. In order to incorporate the polar coordinate state estimate $\hat{\boldsymbol{\xi}}$ with the PMU measurements, the work [5] advocates transforming $\hat{\boldsymbol{\xi}}$ into rectangular coordinates, denoted as $\hat{\boldsymbol{s}}_{\text{SCADA}} \triangleq \mathcal{T}(\hat{\boldsymbol{\xi}})$. Due to the invariant property of the ML estimator [16], $\hat{\boldsymbol{s}}_{\text{SCADA}}$ is also the ML estimator in rectangular coordinates. Furthermore, the mean and covariance of $\hat{s}_{\text{SCADA}}$ can be approximately computed using the linearization method [17]. It can be shown that the mean and covariance matrix of $\hat{s}_{\text{SCADA}}$ are s and $\mathbf{\Gamma}_{\text{SCADA}} = \nabla\mathcal{T}(\boldsymbol{\xi})\mathbf{\Upsilon}\nabla[\mathcal{T}(\boldsymbol{\xi})]^T|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}}$, respectively. Since the goal is to derive a distributed algorithm, it is also assumed that each bus has access only to the mean and variance of its own state from SCADA estimates, i.e., $p(\boldsymbol{s}) \approx \prod_{i\in\mathcal{B}} p(\boldsymbol{s}_i) = \prod_{i\in\mathcal{B}} \mathcal{N}(\boldsymbol{s}_i|\boldsymbol{\gamma}_i, \mathbf{\Gamma}_i)$, with $\boldsymbol{\gamma}_i = [\hat{\boldsymbol{s}}_{\text{SCADA}}]_{2i-1:2i}$ and $\mathbf{\Gamma}_i = [\boldsymbol{P}_{\text{SCADA}}]_{2i-1:2i,2i-1:2i}$. Thus, we have

$$p(\boldsymbol{s}_i) = \mathcal{N}(\boldsymbol{s}_i|\boldsymbol{\gamma}_i, \mathbf{\Gamma}_i). \tag{9}$$

For $p(\boldsymbol{\theta})$, we adopt the truncated Gaussian model:

$$\begin{aligned} p(\theta_i) &= \mathcal{TN}(\theta_i|\underline{\theta_i}, \bar{\theta}_i, \tilde{v}_i, \tilde{C}_i) \\ &\triangleq \frac{[U(\theta_i - \underline{\theta_i}) - U(\theta_i - \bar{\theta}_i)]\mathcal{N}(\theta_i|\tilde{v}_i, \tilde{C}_i)}{\text{erf}(\frac{\bar{\theta}_i - \tilde{v}_i}{\tilde{C}_i^{1/2}}) - \text{erf}(\frac{\underline{\theta_i} - \tilde{v}_i}{\tilde{C}_i^{1/2}})}, \end{aligned} \tag{10}$$

where $\underline{\theta_i}$ and $\bar{\theta}_i$ are lower and upper bounds of the truncated Gaussian distribution, respectively; $U(x)$ is the unit step function, whose value is zero for negative x and one for non-negative x; $\tilde{v}_i$ and $\tilde{C}_i$ are the mean and covariance of the original, non-truncated Gaussian distribution; and $\text{erf}(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_0^x \exp\{-\frac{y^2}{2}\}dy$. Moreover, the first order moment of (10) is

$$\begin{aligned} \tilde{\varpi}_i = \mathbb{E}\{\theta_i\} &= \tilde{v}_i - \tilde{C}_i^{1/2} \frac{\mathcal{N}(\bar{\theta}_i|\tilde{v}_i, \tilde{C}_i) - \mathcal{N}(\underline{\theta_i}|\tilde{v}_i, \tilde{C}_i)}{\text{erf}(\frac{\bar{\theta}_i - \tilde{v}_i}{\tilde{C}_i^{1/2}}) - \text{erf}(\frac{\underline{\theta_i} - \tilde{v}_i}{\tilde{C}_i^{1/2}})} \\ &\triangleq \Xi_1[\underline{\theta_i}, \bar{\theta}_i, \tilde{v}_i, \tilde{C}_i] \end{aligned} \tag{11}$$

and the second order moment is

$$\begin{aligned} \tilde{\tau}_i &= \mathbb{E}\{\theta_i^2\} \\ &= \tilde{v}_i^2 - 2\tilde{v}_i\tilde{C}_i^{1/2} \frac{\mathcal{N}(\bar{\theta}_i|\tilde{v}_i, \tilde{C}_i) - \mathcal{N}(\underline{\theta_i}|\tilde{v}_i, \tilde{C}_i)}{\text{erf}(\frac{\bar{\theta}_i - \tilde{v}_i}{\tilde{C}_i^{1/2}}) - \text{erf}(\frac{\underline{\theta_i} - \tilde{v}_i}{\tilde{C}_i^{1/2}})} + \tilde{C}_i \\ &\times \left\{1 - \frac{(\bar{\theta}_i - \tilde{v}_i)\mathcal{N}(\bar{\theta}_i|\tilde{v}_i, \tilde{C}_i) - (\underline{\theta_i} - \tilde{v}_i)\mathcal{N}(\underline{\theta_i}|\tilde{v}_i, \tilde{C}_i)}{\tilde{C}_i^{1/2}\left[\text{erf}(\frac{\bar{\theta}_i - \tilde{v}_i}{\tilde{C}_i^{1/2}}) - \text{erf}(\frac{\underline{\theta_i} - \tilde{v}_i}{\tilde{C}_i^{1/2}})\right]}\right\} \\ &\triangleq \Xi_2[\underline{\theta_i}, \bar{\theta}_i, \tilde{v}_i, \tilde{C}_i]. \end{aligned} \tag{12}$$

In general, the parameters of $p(\theta_i)$ can be obtained through pre-deployment measurements. For example, the truncated range $[\underline{\theta_i}, \bar{\theta}_i]$ is founded to be $[-6\pi/180, 6\pi/180]$ according to the test results [7]. $\tilde{v}_i$ and $\tilde{C}_i$ can also be obtained from a histogram generated during PMU testing [18]. On the other extreme, (10) also incorporates the case when we have no statistical information about the unknown phase error: setting $[\underline{\theta_i}, \bar{\theta}_i] = [-\pi, \pi]$, $\tilde{v}_i = 0$, and $\tilde{C}_i = \infty$, giving an uniform distributed θ_i in one sampling period of the PMU.

III. Distributed State Estimation under Sampling Phase Error

Given all the prior distributions (9), (10) and the likelihood function (8), the MMSE estimator can be written as

$$\begin{aligned} \hat{\boldsymbol{s}}_i = &\frac{1}{c} \int \dots \int \boldsymbol{s}_i \prod_{i\in\mathcal{B}} p(\boldsymbol{s}_i) \prod_{i\in\mathcal{P}} p(\theta_i) \\ &\times \prod_{i\in\mathcal{P}} p(\boldsymbol{z}_i|\theta_i, \{\boldsymbol{s}_j\}_{j\in\mathcal{M}(i)}) d\{\theta_i\}_{i\in\mathcal{P}} d\{\boldsymbol{s}\}_{i\in\mathcal{B}}, \end{aligned} \tag{13}$$

where $c = \int \dots \int \prod_{i\in\mathcal{B}} p(\boldsymbol{s}_i) \prod_{i\in\mathcal{P}} p(\theta_i) p(\boldsymbol{z}_i|\theta_i, \{\boldsymbol{s}_j\}_{j\in\mathcal{M}(i)})$ $d\{\theta_i\}_{i\in\mathcal{P}} d\{\boldsymbol{s}\}_{i\in\mathcal{B}}$. The integration is complicated as θ_i is coupled with $\{\boldsymbol{s}_j\}_{j\in\mathcal{M}(i)}$, and its expression is not analytically tractable. Furthermore, the dimensionality of the state space of the integrand (of the order of number of buses in a power grid, which is typically more than a thousand) prohibits direct numerical integration. In this case, approximate schemes need to be resorted to. One example is the Markov Chain Monte Carlo (MCMC) method, which approximates the distributions and integration operations using a large number of random samples. However, sampling methods can be computationally demanding, often limiting their use to small-scale problems. Even if it can be successfully applied, the solution is centralized, meaning that the network still suffers from heavy communication overhead.

To facilitate the estimation, variational inference is used to approximate the complicated posterior distribution. The goal of variational inference (VI) is to find a tractable variational distribution $q(\boldsymbol{\theta}, \boldsymbol{s})$ that closely approximates the true posterior distribution $p(\boldsymbol{\theta}, \boldsymbol{s}|\boldsymbol{z}) \propto p(\boldsymbol{\theta})p(\boldsymbol{s})p(\boldsymbol{z}|\boldsymbol{\theta}, \boldsymbol{s})$. The criterion for finding the approximating $q(\boldsymbol{\theta}, \boldsymbol{s})$ is to minimize the Kullback-Leibler (KL) divergence between $q(\boldsymbol{\theta}, \boldsymbol{s})$ and $p(\boldsymbol{\theta}, \boldsymbol{s}|\boldsymbol{z})$ [19]:

$$\text{KL}\left[q(\boldsymbol{\theta}, \boldsymbol{s})||p(\boldsymbol{\theta}, \boldsymbol{s}|\boldsymbol{z})\right] \triangleq -\mathbb{E}_{q(\boldsymbol{\theta}, \boldsymbol{s})}\left\{\ln \frac{p(\boldsymbol{\theta}, \boldsymbol{s}|\boldsymbol{z})}{q(\boldsymbol{\theta}, \boldsymbol{s})}\right\}. \tag{14}$$

If there is no constraint on $q(\boldsymbol{\theta}, \boldsymbol{s})$, then the KL divergence vanishes when $q(\boldsymbol{\theta}, \boldsymbol{s}) = p(\boldsymbol{\theta}, \boldsymbol{s}|\boldsymbol{z})$. However, in this case, we still face the intractable integration in (13). In the VI framework, a common practice is to apply the mean-field approximation to $q(\boldsymbol{\theta}, \boldsymbol{s})$. For a large-scale power grid, to achieve distributed computation for the power state $\boldsymbol{s}_i$, a mean-field approximation is applied to $q(\boldsymbol{\theta}, \boldsymbol{s})$, and the variational distribution is in the form $q(\boldsymbol{\theta}, \boldsymbol{s}) = \prod_{i\in\mathcal{P}} b(\theta_i) \prod_{i\in\mathcal{B}} b(\boldsymbol{s}_i)$. Under this mean-field approximation, the optimal $b(\theta_i)$ and $b(\boldsymbol{s}_i)$ that minimize (14) are given by (15) and (16) at the top of the next page. Next, we will evaluate the expressions for $b(\theta_i)$ and $b(\boldsymbol{s}_i)$ in (15) and (16), respectively.

- **Computation of $b(\theta_i)$:**

Assume $b(\boldsymbol{s}_i)$ is known for all $i \in \mathcal{B}$ with mean and covariance denoted by $\boldsymbol{\mu}_i$ and $\boldsymbol{P}_{i,i}$, respectively. By substituting the prior distributions $p(\boldsymbol{s}_i)$ from (9), $p(\theta_i)$ from (10) and the likelihood function from (8) into (15), the variational distribution $b(\theta_i)$ can be shown to be

$$b(\theta_i) \propto \mathcal{TN}(\theta_i|\underline{\theta_i}, \bar{\theta}_i, v_i, C_i) \tag{17}$$

with

$$C_i = \frac{\tilde{C}_i}{\sigma_i^{-2}\text{Tr}\{\boldsymbol{B}_{i,2}\}\tilde{C}_i + 1} \tag{18}$$

$$b(\theta_i) \propto \exp\Big\{\mathbb{E}_{\prod_{j\in\mathcal{P}\setminus i} b(\theta_j)\prod_{i\in\mathcal{B}} p(\boldsymbol{s}_i)}\big\{\ln \prod_{i\in\mathcal{P}} p(\boldsymbol{z}_i|\theta_i,\{\boldsymbol{s}_j\}_{j\in\mathcal{M}(i)})p(\theta_i)\prod_{i\in\mathcal{B}} p(\boldsymbol{s}_i)\big\}\Big\} \quad i\in\mathcal{P}, \tag{15}$$

$$b(\boldsymbol{s}_i) \propto \exp\Big\{\mathbb{E}_{\prod_{i\in\mathcal{P}} b(\theta_i)\prod_{j\in\mathcal{B}\setminus i} b(\boldsymbol{s}_j)}\big\{\prod_{i\in\mathcal{P}} p(\boldsymbol{z}_i|\theta_i,\{\boldsymbol{s}_j\}_{j\in\mathcal{M}(i)})p(\theta_i)\prod_{i\in\mathcal{B}} p(\boldsymbol{s}_i)\big\}\Big\} \quad i\in\mathcal{B}. \tag{16}$$

$$v_i = C_i\big[\tilde{v}_i/\tilde{C}_i + \sigma_i^{-2}\mathrm{Tr}\big\{\boldsymbol{z}_i\sum_{j\in\mathcal{M}(i)}(\boldsymbol{G}_{ij}\boldsymbol{\mu}_j)^T - \boldsymbol{B}_{i,1}\big\}\big] \tag{19}$$

with $\boldsymbol{B}_{i,1} = \sum_{j\in\mathcal{M}(i)}\boldsymbol{H}_{ij}(\boldsymbol{P}_{j,j} + \boldsymbol{\mu}_j\boldsymbol{\mu}_j^T)\boldsymbol{G}_{ij}^T + \sum_{j,k\in\mathcal{M}(i),j\neq k}\boldsymbol{H}_{ij}\boldsymbol{\mu}_j\boldsymbol{\mu}_k^T\boldsymbol{G}_{ik}^T$ and $\boldsymbol{B}_{i,2} = \sum_{j\in\mathcal{M}(i)}\boldsymbol{G}_{ij}(\boldsymbol{P}_{j,j} + \boldsymbol{\mu}_j\boldsymbol{\mu}_j^T)\boldsymbol{G}_{ij}^T + \sum_{j,k\in\mathcal{M}(i),j\neq k}\boldsymbol{G}_{ij}\boldsymbol{\mu}_j\boldsymbol{\mu}_k^T\boldsymbol{G}_{ik}^T$. With C_i and v_i in (18) and (19), to facilitate the computation of $b(\boldsymbol{s}_i)$ in the next step, the first and second order moments of $b(\theta_i)$ are computed through (11) and (12) as

$$\varpi_i = \Xi_1[\underline{\theta_i},\bar{\theta}_i,v_i,C_i], \quad \tau_i = \Xi_2[\underline{\theta_i},\bar{\theta}_i,v_i,C_i]. \tag{20}$$

- **Computation of $b(\boldsymbol{s}_i)$:**

Assume $b(\theta_i)$ for all $i\in\mathcal{P}$ are known with first and second order moments denoted by ϖ_i and τ_i, respectively. Furthermore, it is assumed that $b(\boldsymbol{s}_j)$ for $j\in\mathcal{B}\setminus i$ are also known with their covariance matrices given by $\boldsymbol{P}_{j,j}$. Now, rewrite (16) as

$$b(\boldsymbol{s}_i) \propto p(\boldsymbol{s}_i) \times \prod_{j\in\mathcal{M}(i)}\underbrace{\exp\big\{\mathbb{E}_{b(\theta_j)\prod_{k\in\mathcal{M}(j)\setminus i} b(\boldsymbol{s}_k)}\{\ln p(\boldsymbol{z}_j|\theta_j,\{\boldsymbol{s}_{\tilde{k}}\}_{\tilde{k}\in\mathcal{M}(j)})\}\big\}}_{\triangleq m_{j\to i}(\boldsymbol{s}_i)}. \tag{21}$$

After some matrix multiplications, it can be shown that $m_{j\to i}(\boldsymbol{s}_i)$ is in Gaussian form

$$m_{j\to i}(\boldsymbol{s}_i) \propto \mathcal{N}(\boldsymbol{s}_i|\boldsymbol{v}_{j\to i},\boldsymbol{C}_{j\to i}) \tag{22}$$

with

$$\boldsymbol{C}_{j\to i} = \sigma_j^2[\boldsymbol{H}_{ji}^T\boldsymbol{H}_{ji} + \varpi_j(\boldsymbol{G}_{ji}^T\boldsymbol{H}_{ji} + \boldsymbol{H}_{ji}^T\boldsymbol{G}_{ji}) + \tau_j\boldsymbol{G}_{ji}^T\boldsymbol{G}_{ji}]^{-1}, \tag{23}$$

$$\boldsymbol{v}_{j\to i} = \sigma_j^{-2}\boldsymbol{C}_{j\to i}\Big\{(\boldsymbol{H}_{ji} + \varpi_j\boldsymbol{G}_{ji})^T\boldsymbol{z}_j - \sum_{k\in\mathcal{M}(j)\setminus i}\big[\boldsymbol{H}_{ji}^T\boldsymbol{H}_{jk} + \varpi_j(\boldsymbol{G}_{ji}^T\boldsymbol{H}_{jk} + \boldsymbol{H}_{ji}^T\boldsymbol{G}_{jk}) + \tau_j\boldsymbol{G}_{ji}^T\boldsymbol{G}_{jk}\big]^T\boldsymbol{\mu}_k\Big\}. \tag{24}$$

Then, putting $p(\boldsymbol{s}_i) = \mathcal{N}(\boldsymbol{s}_i|\gamma_i,\boldsymbol{\Gamma}_i)$ and (22) into (21), we obtain

$$\begin{aligned} b(\boldsymbol{s}_i) &\propto \mathcal{N}(\boldsymbol{s}_i|\gamma_i,\boldsymbol{\Gamma}_i)\mathcal{N}(\boldsymbol{s}_i|\boldsymbol{v}_{j\to i},\boldsymbol{C}_{j\to i}) \\ &\propto \mathcal{N}(\boldsymbol{s}_i|\boldsymbol{\mu}_i,\boldsymbol{P}_{i,i}), \end{aligned} \tag{25}$$

with

$$\boldsymbol{P}_{i,i} = (\boldsymbol{\Gamma}_i^{-1} + \sum_{j\in\mathcal{M}(i)}\boldsymbol{C}_{j\to i}^{-1})^{-1} \tag{26}$$

$$\boldsymbol{\mu}_i = \boldsymbol{P}_{i,i}(\boldsymbol{\Gamma}_i^{-1}\gamma_i + \sum_{j\in\mathcal{M}(i)}\boldsymbol{C}_{j\to i}^{-1}\boldsymbol{v}_{j\to i}). \tag{27}$$

Inspection of (23) and (24) reveals that these expressions can be readily computed at bus j and then $\boldsymbol{C}_{j\to i}$ and $\boldsymbol{v}_{j\to i}$ can be sent to its immediate neighbouring bus i for computation of $b(\boldsymbol{s}_i)$ according to (25).

- **Updating Schedule and Summary:**

From the expressions for $b(\theta_i)$ and $b(\boldsymbol{s}_i)$ in (17) and (25), it should be noticed that these functions are coupled. Consequently, $b(\theta_i)$ and $b(\boldsymbol{s}_i)$ should be iteratively updated. Since updating any $b(\theta_i)$ or $b(\boldsymbol{s}_i)$ corresponds to minimizing the KL divergence in (14), the iterative algorithm is guaranteed to converge monotonically to at least a stationary point [19] and there is no requirement that $b(\theta_i)$ or $b(\boldsymbol{s}_i)$ should be updated in any particular order. Besides, the variational distributions $b(\theta_i)$ and $b(\boldsymbol{s}_i)$ in (17) and (25) keep the form of truncated Gaussian and Gaussian distributions during the iterations, thus only their parameters are required to be updated.

However, the successive update scheduling might take too long in large-scale networks. Fortunately, from (23)-(27), it is found that updating $b(\boldsymbol{s}_i)$ only involves information within two hops from bus i. Besides, from (18) and (19), it is observed that updating $b(\theta_i)$ only involves information from direct neighbours of bus i. Hence, if buses within two hops from each other do not update their variational distributions $b(\cdot)$ at the same time, the KL divergence in (14) is guaranteed to be decreased in each iteration and the distributed algorithm keeps the monotonic convergence property. This can be achieved by grouping the buses using a distance-2 coloring scheme [20], which colors all the buses under the principle that buses within a two-hop neighborhood are assigned different colors and the number of colors used is the least (for the IEEE-300 system, only 13 different colors are needed). Then, all buses with the same color update at the same time and buses with different colors are updated in succession. Notice that the complexity order of the distance-2 coloring scheme is $\mathcal{Q}(\lambda|\mathcal{B}|)$ [20], where λ is the maximum number of branches linked to any bus. Since λ is usually small compared to the network size (e.g., $\lambda = 9$ for the IEEE 118-bus system), the complexity of distance-2 coloring depends only on the network size and it is independent of the specific topology of the power network.

In summary, all the buses are first colored by the distance-2 coloring scheme, and the iterative procedure is formally given in Algorithm 1. Notice that although the modelling and formulation of state estimation under phase error is complicated, the final result and processing are simple. During each iteration, the first and second order moments of the phase error estimate are computed via (20); while the covariance and mean of the state estimate are computed using (26) and (27). Due to the fact that computing these quantities at one bus depends

Algorithm 1 Distributed states estimation

Initialization: $\boldsymbol{\mu}_i = [\hat{\boldsymbol{s}}_{\text{SCADA}}]_{2i-1:2i}$ and $\boldsymbol{P}_{i,i} = [\boldsymbol{P}_{\text{SCADA}}]_{2i-1:2i;2i-1:2i}$.
Neighboring buses exchange $\boldsymbol{\mu}_i$ and $\boldsymbol{P}_{i,i}$.
Buses with PMUs update ϖ_i and τ_i via (20).
Every bus i computes $\boldsymbol{C}_{i\to j}$ $\boldsymbol{v}_{i\to j}$ $\boldsymbol{P}_{i,i}$, $\boldsymbol{\mu}_i$ via (23) (24) (26) (27), and sends these four entities to bus j, where $j \in \mathcal{M}(i)$.
for the l^{th} iteration **do**
  Select a group of buses with the same color.
  Buses with PMUs in the group compute ϖ_i and τ_i via (20).
  Every bus in the group updates its $\boldsymbol{C}_{i\to j}$ $\boldsymbol{v}_{i\to j}$ $\boldsymbol{P}_{i,i}$, $\boldsymbol{\mu}_i$ via (23) (24) (26) (27), and sends them out to its neighbor j.
  Bus j computes $\boldsymbol{v}_{j\to k}$ via (24) and send to its neighbor $k \in \mathcal{M}(j)$.
end for

on information from neighboring buses, these equations are computed iteratively. After convergence, the state estimate is given by $\boldsymbol{\mu}_i$ at each bus.

Although the proposed distributed algorithm advocates each bus to perform computations and message exchanges, but it is also applicable if computations of several buses are executed by a local control center. Then any two control centers only need to exchange the messages for their shared power states.

IV. Simulation Results and Discussions

This section provides results on the numerical tests of the developed centralized and distributed state estimators in Section III. The network parameters g_{ij}, b_{ij}, B_i, ρ_{ij} are loaded from the test cases in MATPOWER4.0 [21]. In each simulation, the value at each load bus is varied by adding a uniformly distributed random value within $\pm 10\%$ of the value in the test case. Then the power flow program is run to determine the true states. The RTUs measurements are composed of active/reactive power injection, active/reactive power flow, and bus voltage magnitude at each bus, which are also generated from MATPOWER4.0 and perturbed by independent zero-mean Gaussian measurement errors with standard deviation 1×10^{-2}. For the SCADA system, the estimates $\hat{\boldsymbol{\xi}}$ and $\boldsymbol{\Upsilon}$ are obtained through the classical iterative reweighted least-squares with initialization $[A_i, \phi_i]^T = [1, 0]^T$ [13]. In general, the proposed algorithms are applicable regardless of the number of PMUs and their placements. But for the simulation study, the placement of PMUs is obtained through the method proposed in [22]. As experiments in [7] show the maximum phase error is 6° in a 60Hz power system, θ_i is generated uniformly from $[-6\pi/180, 6\pi/180]$ for each Monte-Carlo simulation run. The PMU measurement errors follow a zero-mean Gaussian distribution with standard deviation $\sigma_i = 1 \times 10^{-2}$. 1000 Monte-Carlo simulation runs are averaged for each point in the figures. Furthermore, it is assumed that bad data from RTUs and PMU measurements has been successfully handled [5], [23, Chap 7].

For comparison, we consider the following three existing methods: 1) Centralized WLS [5] assuming no sampling phase errors in the PMUs. This represents the best performance one can achieve in the state estimation and is a benchmark for the proposed algorithms. 2) Centralized WLS under sampling phase errors in the PMUs. This will show how much degradation one would have if phase errors are ignored. 3) The centralized alternating minimization (AM) scheme [14] with $p(\boldsymbol{s})$ and $p(\theta_i)$ incorporated as prior information.

Fig. 1 shows the convergence behavior of the proposed algorithms with average mean square error (MSE) defined as $\frac{1}{2|\mathcal{B}|}\sum_{i\in\mathcal{B}} ||\hat{\boldsymbol{s}}_i - \boldsymbol{s}_i||^2$. It can be seen that the proposed distributed algorithm performs close to the optimal performance after convergence. The seemingly slow convergence is a result of sequential updating of buses with different colors to guarantee convergence. If one iteration is defined as one round of updating of all buses, the distributed algorithm would converge only in a few iterations. On the other hand, part of the degradation from the centralized AM solution is due to the fact that in the distributed algorithm, the covariance of states $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ in prior distributions and variational distributions cannot be taken into account. Finally, if the sampling phase error is ignored, we can see that the performance of centralized WLS shows significant degradation, illustrating the importance of simultaneous power state and phase error estimation. Fig. 2 shows the MSE of the sampling phase error estimation $\frac{1}{|\mathcal{P}|}\sum_{i\in\mathcal{P}} ||\varpi_i^* - \theta_i||^2$, where ϖ_i^* is the converged ϖ_i in (20). It can be seen from the figure that same conclusions as in Fig. 1 can be drawn.

V. Conclusions

In this paper, a distributed state estimation scheme integrating measurements from a traditional SCADA system and newly deployed PMUs has been proposed, with the aim that the existing SCADA system is kept intact, and unknown sampling phase errors among PMUs incorporated in the estimation procedure. The proposed distributed power state estimation algorithm only involves limited message exchanges between neighboring buses and is guaranteed to converge. Numerical results have shown that the converged state estimates of the distributed algorithm are very close to those of the optimal centralized estimates assuming no sampling phase error.

References

[1] A. Monticelli, "Electric power system state estimation," *Proc. IEEE*, vol. 88, no. 2, pp. 262–282, 2000.

[2] G. Giannakis, V. Kekatos, N. Gatsis, S.-J. Kim, H. Zhu, and B. Wollenberg, "Monitoring and optimization for power grids: A signal processing perspective," *IEEE Signal Process. Mag*, vol. 30, no. 5, pp. 107–128, 2013.

[3] A. Phadke, J. Thorp, and M. Adamiak, "A new measurement technique for tracking voltage phasors, local system frequency, and rate of change of frequency," *IEEE Trans. Power App. Syst.*, vol. PAS-102, no. 5, pp. 1025–1038, 1983.

[4] X. Li and A. Scaglione, "Robust decentralized state estimation and tracking for power systems via network gossiping," *IEEE J. Select. Areas Commun.*, vol. 31, no. 7, pp. 1184–1194, 2013.

[5] M. Zhou, V. Centeno, J. Thorp, and A. Phadke, "An alternative for including phasor measurements in state estimators," *IEEE Trans. Power Syst.*, vol. 21, no. 4, pp. 1930–1937, 2006.

[6] A. Phadke and J. Thorp, *Synchronized Phasor Measurements and Their Applications*. New York: Springer, 2008.

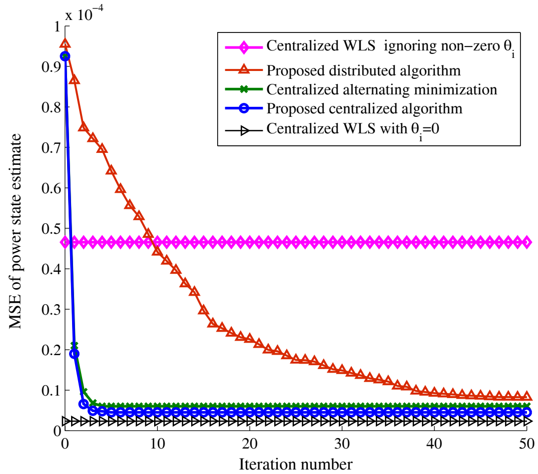

Fig. 1. MSE of the power state versus iteration number for the IEEE 118-bus system.

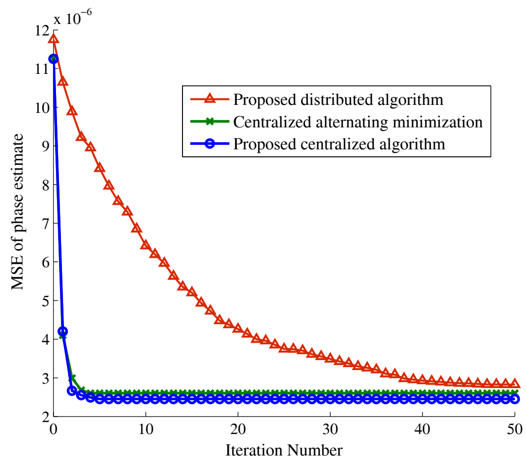

Fig. 2. MSE of the phase error versus iteration number for the IEEE 118-bus system.

[7] A. P. Meliopoulos, V. Madani, D. Novosel, G. Cokkinides, and *et al.*, "Synchrophasor measurement accuracy characterization," *North American SynchroPhasor Initiative Performance & Standards Task Team (Consortium for Electric Reliability Technology Solutions)*, 2007.

[8] A. Gómez-Expósito, A. Abur, A. de la Villa Jaén, and C. Gómez-Quiles, "A multilevel state estimation paradigm for smart grids," *Proc. IEEE*, vol. 99, no. 6, pp. 952–976, 2011.

[9] R. Ebrahimian and R. Baldick, "State estimation distributed processing for power systems," *IEEE Trans. Power Syst.*, vol. 15, no. 4, pp. 1240–1246, 2000.

[10] V. Kekatos and G. Giannakis, "Distributed robust power system state estimation," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1617–1626, 2013.

[11] J. Du and Y.-C. Wu, "Distributed clock skew and offset estimation in wireless sensor networks: Asynchronous algorithm and convergence analysis," *IEEE Trans. Wireless Commun.*, vol. 12, no. 11, pp. 5908–5917, November 2013.

[12] L. Xie, D.-H. Choi, S. Kar, and H. V. Poor, "Fully distributed state estimation for wide-area monitoring systems," *IEEE Trans. Smart Grid*, vol. 3, no. 3, pp. 1154–1169, 2012.

[13] A. G. E. Ali Abur, *Power System State Estimation: Theory and Implementation*. Artech House, Mar 2004.

[14] P. Yang, Z. Tan, A. Wiesel, and A. Nehorai, "Power system state estimation using PMUs with imperfect synchronization," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4162–4172, 2013.

[15] F. Schweppe and J. Wildes, "Power system static-state estimation, part I: Exact model," *IEEE Trans. Power App. Syst.*, vol. PAS-89, no. 1, pp. 120–125, 1970.

[16] S. M. Kay, *Fundamentals of Statistical Signal Processing Estimation Theory*. Upper Saddle River, NJ: Prentice-Hall, 1993.

[17] D. Simon, *State Estimation: Kalman, H-Infinity, and Nonlinear Approaches*. Hoboken, NJ: Wiley, 2006.

[18] S. M. Shah and M. C. Jaiswal, "Estimation of parameters of doubly truncated normal distribution from first four sample moments," *Annals of the Institute of Mathematical Statistics*, vol. 18, no. 1, pp. 107–111, 1966.

[19] C. Bishop, *Pattern Recognition and Machine Learning*. Artech House, January 2006.

[20] S. T. McCormick, "Optimal approximation of sparse hessians and its equivalence to a graph coloring problem," *Math. Programming*, 1983.

[21] R. Zimmerman, C. Murillo-Sanchez, and R. Thomas, "Matpower: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 12–19, 2011.

[22] B. Gou, "Generalized integer linear programming formulation for optimal PMU placement," *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 1099–1104, 2008.

[23] L. Xie, D.-H. Choi, S. Kar, and H. V. Poor, *Bad-data detection in smart grid: a distributed approach*. in E. Hossain, Z. Han, and H. V. Poor, editors Smart Grd Communications and Networking, Cambridge University Press, 2012.