

State Estimation of Three-Phase Four-Conductor Distribution Systems with Real-Time Data from Selective Smart Meters

Yikui Liu, *Student Member, IEEE*, Jie Li, *Member, IEEE*, Lei Wu, *Senior Member, IEEE*

Abstract— **Distribution system state estimation (DSSE) has recently been tested and experimentally deployed in some practical distribution networks.** Distinct features of distribution systems, such as diverse and unsymmetrical configurations as well as limited real-time measurements, prohibit the direct application of mature state estimation methods for transmission systems. Targeting at three-phase four-conductor configured unsymmetrical medium-voltage distribution systems (MDS) with neutral conductors and ground resistances, this paper proposes a weighted least square (WLS) based DSSE approach, in which voltages are chosen as state variables and load pseudo measurements of low-voltage distribution systems (LDS) are considered to compensate insufficient real-time measurements in MDS. Both rectangular and polar coordinates are studied, and voltage variables of neutrals and zero-injection phases are eliminated to reduce the scale of the DSSE problem. Moreover, in order to enhance load pseudo measurement accuracy of LDSs, a clustering and partial least square (PLS) regression based load estimation model is proposed to leverage real-time communication ability of smart meters. Case studies on a modified IEEE 123-bus distribution system with actual smart meter data illustrate effectiveness of the proposed approaches.

Index Terms—Distribution system state estimation, load estimation, partial least square regression, three-phase four-conductor.

I. INTRODUCTION

DISTRIBUTION system state estimation (DSSE) processes raw measurements from the supervisory control and data acquisition (SCADA) system and supplementary pseudo measurements of loads to provide real-time monitoring. DSSE is considered as the foundation of a variety of key applications, such as voltage control, system reconfiguration, and demand side management [1], that are under development and testing to manage emerging distribution systems with an increasing penetration of distributed energy resources and flexible demand assets. Moreover, DSSE could also facilitate the development of new applications, aiming at the next-generation distribution systems [2]. For instance, utilizing DSSE results, recent works [3]–[4] determined electricity price signals that could be potentially used in a deregulation paradigm of future distribution systems.

Compared with transmission system state estimation (TSSE) that has been developed and deployed for over half a century, DSSE is relatively young since it was first studied in the mid-1990s [5]. Indeed, most DSSE implementations are originated from TSSE models with moderate modifications. One notable implementation is the bus voltage based DSSE models, which use bus voltages in polar or rectangular coordinates as state variables [6]–[7]. In addition, in

recognizing that majority of distribution systems are operated in a radial topology, branch current based models are also customized for DSSE studies by taking the advantage of their computational benefits [8]–[9]. However, DSSE of medium-voltage distribution systems (MDS), as studied in this paper, is more challenging than TSSE because of the following two facts: (i) unsymmetrical and heterogeneous system configurations and (ii) limited real-time measurements.

Indeed, MDSs are essentially unsymmetrical because of single-/two-phase laterals, untransposed lines, and unbalanced loads. Consequently, a three-phase DSSE model is a must to accurately simulate their unsymmetrical characteristics. Reference [10] considered multi-phase distribution systems with industry-grade models of various electrical components. That is, a large number of additional state variables and associated equations are involved to model details of electrical components, and those equations are used as equality constraints in the DSSE model. A hybrid particle swarm optimization (PSO) based three-phase state estimation method was proposed in [11], while solution quality and consistency are of major concerns. An improved three-phase admittance matrix based DSSE model for MDSs was proposed in [12], by leveraging certain measurements so that the Jacobian matrix can be reasonably treated as constant. Reference [13] used discrete variables to indicate transformer tap positions in the DSSE model, which was solved via the ordinary optimization technique. Reference [14] proposed a two-step procedure for multi-area DSSE, in which the second step uses information from adjacent areas to refine local results from the first step. However, case study showed that solution accuracies among different areas were inconsistent and benefits over the integrated DSSE were not noticeable.

Moreover, heterogeneous configurations of MDSs, in terms of numbers, connection styles, and grounding modes of conductors, further limit the universal applicability of a specified DSSE model. Indeed, models developed in [10]–[14] are customized for three-phase four-conductor MDSs with solidly multi-grounded configuration, i.e., neutral voltages are equal to zero and can be naturally excluded. In turn, an original 4×4 line admittance matrix can be reduced to a 3×3 phase frame one. However, rather than solidly grounded, in practice neutral conductors are usually grounded via resistances, which render non-zero potentials for neutrals [15]. Actually, ground resistances have been widely studied in power flow problems of three-phase four-conductor distribution systems, and salient studies [16]–[17] have shown that ground resistances of reasonably large values could significantly impact branch flows and bus voltages. Thus, they should be adequately considered in practical three-phase four-conductor distribution system studies. However, to our best knowledge, ground resistances have not been considered in existing DSSE studies. Indeed, when neutrals are considered,

This work was supported in part by the U.S. National Science Foundation grants PFI-1534035, CNS-1647135, and ECCS-1653179. The authors are with Electrical and Computer Engineering Department, Clarkson University, Potsdam, NY 13699 USA. (E-mail: yikliu, jieli, lwu@clarkson.edu).

two **issues** need to be properly addressed. One is the increased number of state variables and the increased size of **Jacobian** matrix, which introduce significant computational burden. The other is the difficulty in setting initial values of neutral voltages. As phase angles of neutral voltages at different buses could vary considerably, improper initial values will deteriorate DSSE convergence performance. These two **issues** will be properly mitigated by the proposed approach.

In addition, in an MDS, except the substation bus, very few buses are equipped with real-time measurement devices to record bus voltages and/or line currents. Indeed, existing real-time measurements in an MDS are far from enough to guarantee observability. Thus, load pseudo measurements of connected low-voltage distribution systems (LDS) are usually used to assist DSSE. In old-fashioned distribution systems without advanced metering technology, load pseudo measurements are estimated based on monthly electricity bills of customers connected in LDSs [18], which are far from accurate [19]. On the other hand, zero-injection buses are considered as highly trusted virtual measurements in DSSE. Therefore, in a weighted least square (WLS) based DSSE model, large weights are assigned to virtual measurements for enhancing solution accuracy, which however may cause numerical issues. Alternatively, early work [20] formulated zero injections as constraints in DSSE models and adopted Hachtel's augmented matrix method to seek for solutions. Moreover, treating virtual measurements as fully trustworthy sources, [21] eliminated voltages of zero-injection buses via linear functions of voltages of non-zero injection buses.

Recently, proliferated smart meters [22] offer new opportunities to improve DSSE accuracy via accurate load pseudo measurements. Indeed, as smart meters can automatically upload their measurements to control centers on a daily or weekly basis [23], replacing monthly data with higher-fidelity and more accurate records is expected to achieve load pseudo measurements of higher accuracy for LDSs [24]-[26]. An artificial neural network approach was proposed in [27] to generate load pseudo measurements by utilizing load profiles as training data, and the associated load estimation error variances were obtained via a Gaussian mixture model. A closed-loop state estimation framework was proposed in [28], in which nonlinear auto-regressive exogenous load estimation models were developed to provide pseudo measurements, and their corresponding error variances were iteratively adjusted base on the most recent performance. Reference [29] proposed a Bayesian theory based DSSE model to deal with pseudo measurements with non-Gaussian estimation errors. The impact of integrating higher-fidelity smart meter data to DSSE was discussed in [30] with well-designed case studies. Reference [31] proposed a home energy management system (HEMS) to optimize energy costs and comfort levels in residential houses, and a sensitive analysis framework was further developed to study the impact of HEMS on DSSE.

Indeed, smart meters can upload measurements to the control center in real time as long as communication is not an issue [32]. Ideally, in an LDS with a full smart meter

coverage, if all smart meters can simultaneously upload their measurements in real time, together with a reasonable estimation on the network losses, the total load of the entire LDS can be directly calculated. In reality, however, simultaneously uploading massive measurement data from multiple smart meters to a control center could potentially compromise the communication network, while approaches to alleviate communication network congestion such as the random back-off method could induce significant delays [33]. Indeed, it is reported in [33] that the average communication delay increases linearly in the number of smart meters. Therefore, considering that an MDS connects dozens of LDSs and each LDS contains dozens to hundreds of customers, transmitting measurements from all smart meters to the control center in real time will foreseeably cause unacceptable delays. In order to effectively utilize accurate measurements of smart meters in a timely manner to facilitate DSSE, a load aggregation based idea is proposed in this paper. That is, by leveraging actual communication capability, real-time measurements from a limit number of selective smart meters will be uploaded to the control center to estimate loads of the entire LDS timely and accurately.

This paper targets at DSSE of three-phase four-conductor configured MDSs with grounded wye-connected loads. The proposed work targets on addressing certain industry's practical interests and needs, such as the "Data Analytics Cases" described in the EPRI's Data Mining Initiative on Distribution Systems [34]. Specifically, one of the case focuses on developing electrical load models via SCADA and AMI data, which can be used to conduct accurate power flow analysis and allow engineers to monitor, manage, and plan operations of the distribution network [35]. The contributions of this paper are summarized as follows:

- *An accurate multi-phase DSSE model:* A WLS based DSSE model for three-phase four-conductor configured MDSs is proposed, while neutral conductors and ground resistances are explicitly considered. Voltages in rectangular or polar coordinates are formulated as state variables.
 - *An effective state variable reduction approach to improve DSSE computational performance:* The number of state variables and consequently the scale of the DSSE model are reduced via two strategies: (i) by applying the KCL theorem at neutrals, voltages of neutrals can be linearly represented by those of non-neutral phases, and (ii) zero-injection phases are defined by extending the concept of zero-injection bus, and their voltages can be linearly represented by voltages of non-neutral and non-zero injection phases. Finally, only voltages of non-neutral and non-zero injection phases are state variables, which significantly reduces the scale of the DSSE.
 - *An effective approach to leverage smart meter measurements in a timely manner to improve DSSE accuracy:* A load estimation model is proposed to calculate load pseudo measurements of LDSs, which is built by a regression oriented agglomerative hierarchical clustering algorithm together with real-time measurements from selective smart meters.
- The rest of this paper is organized as follows. The DSSE model is discussed in Section II. Section III presents the load

estimation model for LDSs. Case studies are discussed in Section IV, and conclusions are drawn in Section V.

II. THE PROPOSED DSSE MODEL

A. Three-Phase Four-Conductor Distribution System Model

In this paper, matrices are shown as bold, vectors and sets are shown as *italic and bold*, and scalars are shown as *non-bold italic style*. For a distribution system with N buses, i and j are used to index buses, and distribution lines are indexed as $i-j$. $\Psi = \{a, b, c, n\}$ and $\Psi^{abc} = \{a, b, c\}$ are sets of phases including and excluding the neutral. ϕ and φ are used to index phases. Phase-to-ground and phase-to-neutral voltages of phase ϕ at bus i are respectively denoted as complex variables V_i^ϕ and U_i^ϕ . Current injection at phase ϕ of bus i is denoted as complex variable I_i^ϕ . Current on phase ϕ of line $i-j$ is denoted as complex variable I_{i-j}^ϕ .

For an equivalent π model of line $i-j$ with a 4×4 series admittance matrix $\mathbf{y}_{i-j}$, its carrying current can be calculated by two terminal voltages as in (1). With the assumption of solidly multi-grounded neutral, i.e., voltages of the two terminals for the neutral line are zeros, namely $V_i^n = V_j^n = 0$, equation (1) can be simplified as (2), where $\mathbf{y}_{i-j}^{abc}$ is a 3×3 line series admittance matrix by removing the row and column corresponding to neutral from $\mathbf{y}_{i-j}$. The 4×4 shunt admittance matrix $\mathbf{y}_{i-j}^s$ of line $i-j$ can be similarly reduced to 3×3 .

$$[I_{i-j}^a \ I_{i-j}^b \ I_{i-j}^c \ I_{i-j}^n]^T = \mathbf{y}_{i-j} \cdot ([V_i^a \ V_i^b \ V_i^c \ V_i^n]^T - [V_j^a \ V_j^b \ V_j^c \ V_j^n]^T) \quad (1)$$

$$[I_{i-j}^a \ I_{i-j}^b \ I_{i-j}^c]^T = \mathbf{y}_{i-j}^{abc} \cdot ([V_i^a \ V_i^b \ V_i^c]^T - [V_j^a \ V_j^b \ V_j^c]^T) \quad (2)$$

However, instead of solidly grounded, neutral conductors are commonly grounded via resistances at buses, and the assumption $V_i^n = V_j^n = 0$ will not hold. Thus, line admittance matrix remains 4×4 to explicitly simulate neutral conductors.

Consequently, for a three-phase four-conductor distribution system with neutral conductors and ground resistances, system nodal admittance matrix $\mathbf{Y}$ shown in (3) can be built via (4)-(5), where Ω_i is the set of buses adjacent to bus i . Equation (4) calculates self-admittance $\mathbf{Y}_{i,i}$ of bus i , and equation (5) calculates mutual-admittance $\mathbf{Y}_{i,j}$ between buses i and j , where $\mathbf{0}$ represents a zero matrix of proper dimensions and $\mathbf{0}$ represents a zero column vector of proper size. In the last term of (4), y_i^g represents ground conductance, i.e., reciprocal of ground resistance, at bus i . System current injection equations can be written in a compact form as in (6), where $\mathbf{I} = [I_1^a \ I_1^b \ I_1^c \ I_1^n \ \dots \ I_N^a \ I_N^b \ I_N^c \ I_N^n]^T$ and $\mathbf{V} = [V_1^a \ V_1^b \ V_1^c \ V_1^n \ \dots \ V_N^a \ V_N^b \ V_N^c \ V_N^n]^T$ are respectively vectors of current injections and phase-to-ground voltages.

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_{1,1} & \dots & \mathbf{Y}_{1,N} \\ \vdots & \ddots & \vdots \\ \mathbf{Y}_{N,1} & \dots & \mathbf{Y}_{N,N} \end{bmatrix} \quad (3)$$

$$\mathbf{Y}_{i,i} = \sum_{j \in \Omega_i} \mathbf{y}_{i-j} + \sum_{j \in \Omega_i} \mathbf{y}_{i-j}^s + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0}^T & y_i^g \end{bmatrix} \quad (4)$$

$$\mathbf{Y}_{i,j} = \mathbf{Y}_{j,i} = -\mathbf{y}_{i-j} \quad (5)$$

$$\mathbf{I} = \mathbf{Y} \cdot \mathbf{V} \quad (6)$$

B. Model the Effect of Neutral Phases

Vectors $\mathbf{I}$ and $\mathbf{V}$ can be reordered as $\mathbf{I}^{abcn} = [\mathbf{I}^{aT} \ \mathbf{I}^{bT} \ \mathbf{I}^{cT} \ \mathbf{I}^{nT}]^T$ and $\mathbf{V}^{abcn} = [\mathbf{V}^{aT} \ \mathbf{V}^{bT} \ \mathbf{V}^{cT} \ \mathbf{V}^{nT}]^T$ via a row permutation matrix $\hat{\mathbf{A}}$, where $\mathbf{I}^\phi = [I_1^\phi \ \dots \ I_N^\phi]^T$ and $\mathbf{V}^\phi = [V_1^\phi \ \dots \ V_N^\phi]^T$. Thus, the system current injection equation (6) can be reformulated as in (7), where $\mathbf{Y}^{abcn}$ is defined in a block form as (8).

$$\mathbf{I}^{abcn} = \mathbf{Y}^{abcn} \cdot \mathbf{V}^{abcn} \quad (7)$$

$$\mathbf{Y}^{abcn} = \hat{\mathbf{A}} \cdot \mathbf{Y} \cdot \hat{\mathbf{A}}^{-1} = \begin{bmatrix} \mathbf{Y}^{a,a} & \mathbf{Y}^{a,b} & \mathbf{Y}^{a,c} & \mathbf{Y}^{a,n} \\ \mathbf{Y}^{b,a} & \mathbf{Y}^{b,b} & \mathbf{Y}^{b,c} & \mathbf{Y}^{b,n} \\ \mathbf{Y}^{c,a} & \mathbf{Y}^{c,b} & \mathbf{Y}^{c,c} & \mathbf{Y}^{c,n} \\ \mathbf{Y}^{n,a} & \mathbf{Y}^{n,b} & \mathbf{Y}^{n,c} & \mathbf{Y}^{n,n} \end{bmatrix} \quad (8)$$

Equation (9) describes that the summation of current injections at a neutral is zero. Thus, by substituting current injections in (9) with voltages, equation (10) is derived.

$$\mathbf{I}^n + \mathbf{I}^a + \mathbf{I}^b + \mathbf{I}^c = \mathbf{0} \quad (9)$$

$$\sum_{\phi \in \Psi^{abc}} (\sum_{\psi \in \Psi} \mathbf{Y}^{\phi,\psi}) \cdot \mathbf{V}^\psi + (\sum_{\phi \in \Psi} \mathbf{Y}^{\phi,n}) \cdot \mathbf{V}^n = \mathbf{0} \quad (10)$$

Because of the presence of ground resistances, $\sum_{\phi \in \Psi} \mathbf{Y}^{\phi,n}$ is nonsingular. $\mathbf{V}^n$ can be represented via $\mathbf{V}^{abc} = [\mathbf{V}^{aT} \ \mathbf{V}^{bT} \ \mathbf{V}^{cT}]^T$ as in (11), where $\mathbf{T}_{NN}$ is a representation matrix from $\mathbf{V}^{abc}$ to $\mathbf{V}^n$. By further defining $\mathbf{I}^{abc} = [\mathbf{I}^{aT} \ \mathbf{I}^{bT} \ \mathbf{I}^{cT}]^T$ and $\mathbf{Y}^{abc}$ as in (12), we have equation $\mathbf{I}^{abc} = \mathbf{Y}^{abc} \cdot \mathbf{V}^{abc}$. Furthermore, equation $\mathbf{V} = \hat{\mathbf{A}}^{-1} \cdot \mathbf{T}_N \cdot \mathbf{V}^{abc}$ is obtained, where $\mathbf{T}_N$ is extended from $\mathbf{T}_{NN}$ as in (13) and $\mathbf{E}$ is an identity matrix of proper dimensions.

$$\mathbf{V}^n = -(\sum_{\phi \in \Psi} \mathbf{Y}^{\phi,n})^{-1} \cdot [\sum_{\phi \in \Psi} \mathbf{Y}^{\phi,a} \ \sum_{\phi \in \Psi} \mathbf{Y}^{\phi,b} \ \sum_{\phi \in \Psi} \mathbf{Y}^{\phi,c}] \cdot \mathbf{V}^{abc} = \mathbf{T}_{NN} \cdot \mathbf{V}^{abc} \quad (11)$$

$$\mathbf{Y}^{abc} = \begin{bmatrix} \mathbf{Y}^{a,a} & \mathbf{Y}^{a,b} & \mathbf{Y}^{a,c} \\ \mathbf{Y}^{b,a} & \mathbf{Y}^{b,b} & \mathbf{Y}^{b,c} \\ \mathbf{Y}^{c,a} & \mathbf{Y}^{c,b} & \mathbf{Y}^{c,c} \end{bmatrix} + \begin{bmatrix} \mathbf{Y}^{a,n} \\ \mathbf{Y}^{b,n} \\ \mathbf{Y}^{c,n} \end{bmatrix} \cdot \mathbf{T}_{NN} \quad (12)$$

$$\mathbf{T}_N = \begin{bmatrix} \mathbf{E} \\ \mathbf{T}_{NN} \end{bmatrix} \quad (13)$$

Equations (11)-(13) can also properly handle the solidly-grounded situation by adopting the following strategy: if bus i is solidly grounded, $y_i^g = \infty$ and $V_i^n = 0$; I_i^ϕ in $\mathbf{I}^{abcn}$ for $\phi \in \Psi$ and V_i^n in $\mathbf{V}^{abcn}$ are removed; rows of $\mathbf{Y}^{abcn}$ corresponding to I_i^ϕ and columns of $\mathbf{Y}^{abcn}$ corresponding to V_i^n are removed.

C. Model the Effect of Zero-Injection Phases

A zero-injection phase refers to a physical phase of a non-substation bus that is not connected with generators and loads. In this paper, KCL is applied on each zero-injection phase to derive the relationship of voltages for this phase and its adjacent non-zero injection phases. That is, the voltage variable of a zero-injection phase can be represented as a linear function of those of adjacent non-zero injection phases, and in turn be eliminated.

First, we define a row permutation matrix $\hat{\mathbf{B}}$ to reorder $\mathbf{I}^{abc}$ and $\mathbf{V}^{abc}$ as in (14), with which $\mathbf{I}^{abc} = \mathbf{Y}^{abc} \cdot \mathbf{V}^{abc}$ can be represented in a block matrix form (15). In (14)-(15), $\mathbf{I}_{NZ}^{abc}$ is the vector of current injections corresponding to non-zero injection phases, the zero vector $\mathbf{0}$ corresponds to zero-injection phases, and $\mathbf{V}_Z^{abc}$ and $\mathbf{V}_{NZ}^{abc}$ are two reordered sub-

vectors of $\mathbf{V}^{abc}$ corresponding to zero and non-zero injection phases. According to the first set of equations in (15), $\mathbf{V}_Z^{abc}$ can be represented via $\mathbf{V}_{NZ}^{abc}$ as in (16), with which equation (17) can be derived. We define $\mathbf{T}_Z$ as the representation matrix from $\mathbf{V}_{NZ}^{abc}$ to $\begin{bmatrix} \mathbf{V}_Z^{abc} \\ \mathbf{V}_{NZ}^{abc} \end{bmatrix}$. Furthermore, equation $\mathbf{V}^{abc} = \hat{\mathbf{B}}^{-1} \cdot \mathbf{T}_Z \cdot \mathbf{V}_{NZ}^{abc}$ can be obtained.

$$\begin{bmatrix} \mathbf{0} \\ \mathbf{I}_{NZ}^{abc} \end{bmatrix} = \hat{\mathbf{B}} \cdot \mathbf{I}^{abc}, \quad \begin{bmatrix} \mathbf{V}_Z^{abc} \\ \mathbf{V}_{NZ}^{abc} \end{bmatrix} = \hat{\mathbf{B}} \cdot \mathbf{V}^{abc}; \quad (14)$$

$$\begin{bmatrix} \mathbf{Y}_{Z,Z}^{abc} & \mathbf{Y}_{Z,NZ}^{abc} \\ \mathbf{Y}_{NZ,Z}^{abc} & \mathbf{Y}_{NZ,NZ}^{abc} \end{bmatrix} = \hat{\mathbf{B}} \cdot \mathbf{Y}^{abc} \cdot \hat{\mathbf{B}}^{-1} \quad (15)$$

$$\begin{bmatrix} \mathbf{0} \\ \mathbf{I}_{NZ}^{abc} \end{bmatrix} = \begin{bmatrix} \mathbf{Y}_{Z,Z}^{abc} & \mathbf{Y}_{Z,NZ}^{abc} \\ \mathbf{Y}_{NZ,Z}^{abc} & \mathbf{Y}_{NZ,NZ}^{abc} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{V}_Z^{abc} \\ \mathbf{V}_{NZ}^{abc} \end{bmatrix} \quad (15)$$

$$\mathbf{V}_Z^{abc} = -\mathbf{Y}_{Z,Z}^{abc^{-1}} \cdot \mathbf{Y}_{Z,NZ}^{abc} \cdot \mathbf{V}_{NZ}^{abc} \quad (16)$$

$$\begin{bmatrix} \mathbf{V}_Z^{abc} \\ \mathbf{V}_{NZ}^{abc} \end{bmatrix} = \begin{bmatrix} -\mathbf{Y}_{Z,Z}^{abc^{-1}} \cdot \mathbf{Y}_{Z,NZ}^{abc} \\ \mathbf{E} \end{bmatrix} \cdot \mathbf{V}_{NZ}^{abc} = \mathbf{T}_Z \cdot \mathbf{V}_{NZ}^{abc} \quad (17)$$

By combining equations $\mathbf{V} = \hat{\mathbf{A}}^{-1} \cdot \mathbf{T}_N \cdot \mathbf{V}^{abc}$ and $\mathbf{V}^{abc} = \hat{\mathbf{B}}^{-1} \cdot \mathbf{T}_Z \cdot \mathbf{V}_{NZ}^{abc}$, phase-to-ground voltage vector $\mathbf{V}$ can be represented as in (18) and further rewritten under rectangular coordinates as in (19), where $\mathbf{V}_e = \text{Re}(\mathbf{V})$, $\mathbf{V}_f = \text{Im}(\mathbf{V})$, $\mathbf{T}_e = \text{Re}(\mathbf{T})$, $\mathbf{T}_f = \text{Im}(\mathbf{T})$, $\mathbf{V}_{e,NZ}^{abc} = \text{Re}(\mathbf{V}_{NZ}^{abc})$, and $\mathbf{V}_{f,NZ}^{abc} = \text{Im}(\mathbf{V}_{NZ}^{abc})$. $\text{Re}(\cdot)$ and $\text{Im}(\cdot)$ respectively extract the real and imaginary parts of a complex matrix/vector. $\mathbf{T}$ and $\mathbf{T}_E$ are respectively representation matrices from $\mathbf{V}_{NZ}^{abc}$ to $\mathbf{V}$ in complex and real domains.

$$\mathbf{V} = (\mathbf{A}^{-1} \cdot \mathbf{T}_N \cdot \mathbf{B}^{-1}) \cdot \mathbf{T}_Z \cdot \mathbf{V}_{NZ}^{abc} = \mathbf{T} \cdot \mathbf{V}_{NZ}^{abc} \quad (18)$$

$$\begin{bmatrix} \mathbf{V}_e \\ \mathbf{V}_f \end{bmatrix} = \begin{bmatrix} \mathbf{T}_e & -\mathbf{T}_f \\ \mathbf{T}_f & \mathbf{T}_e \end{bmatrix} \cdot \begin{bmatrix} \mathbf{V}_{e,NZ}^{abc} \\ \mathbf{V}_{f,NZ}^{abc} \end{bmatrix} = \mathbf{T}_E \cdot \begin{bmatrix} \mathbf{V}_{e,NZ}^{abc} \\ \mathbf{V}_{f,NZ}^{abc} \end{bmatrix} \quad (19)$$

D. Model Various Measurements

Various types of measurements are considered in this paper, including actual measurements on voltage and current magnitudes from SCADA, voltage phasor measurements from μPMUs , and pseudo measurements on active and reactive demands from the proposed load estimation models. Specifically, both phase-to-ground and phase-to-neutral voltage magnitude measurements are included, while current and voltage measurements of both neutrals and non-neutral phases are also considered [36]-[37]. Squared current and voltage magnitude measurements as well as active and reactive load pseudo measurements are represented as functions of state variable vectors as follows [38].

- Measurements of Voltage Magnitudes

By defining $\Phi_i^{V,\phi}$ as in (20a), a squared phase-to-ground voltage magnitude measurement is represented in a compact matrix form as in (20b). In (20a), $\mathbf{e}_i^\phi$ is a standard basis vector with element corresponding to V_i^ϕ being 1. In (20b), $\bar{V}_i^\phi$ represents the magnitude of V_i^ϕ . With known variance of the phase-to-ground voltage magnitude measurement error σ_V^2 , the measurement error variance of $(\bar{V}_i^\phi)^2$ is $\sigma_{V,sq}^2 = 4 \cdot (\bar{V}_i^\phi)^2 \cdot \sigma_V^2$ through error propagation.

Similarly, with $\Phi_i^{U,\phi}$ defined as in (21a), a squared phase-to-neutral voltage magnitude measurement is represented as in (21b), where $\bar{U}_i^\phi$ represents the magnitude of U_i^ϕ . The measurement error variance of $(\bar{U}_i^\phi)^2$ is $\sigma_{U,sq}^2 = 4 \cdot (\bar{U}_i^\phi)^2 \cdot \sigma_U^2$, where σ_U^2 is the measurement error variance of the phase-to-neutral voltage magnitude.

$$\Phi_i^{V,\phi} = \begin{bmatrix} \mathbf{e}_i^\phi \cdot (\mathbf{e}_i^\phi)^T & [\mathbf{0}] \\ [\mathbf{0}] & \mathbf{e}_i^\phi \cdot (\mathbf{e}_i^\phi)^T \end{bmatrix} \quad (20a)$$

$$\begin{aligned} (\bar{V}_i^\phi)^2 &= \begin{bmatrix} \mathbf{V}_e \\ \mathbf{V}_f \end{bmatrix}^T \cdot \Phi_i^{V,\phi} \cdot \begin{bmatrix} \mathbf{V}_e \\ \mathbf{V}_f \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{V}_{e,NZ}^{abc} \\ \mathbf{V}_{f,NZ}^{abc} \end{bmatrix}^T \cdot \mathbf{T}_E^T \cdot \Phi_i^{V,\phi} \cdot \mathbf{T}_E \cdot \begin{bmatrix} \mathbf{V}_{e,NZ}^{abc} \\ \mathbf{V}_{f,NZ}^{abc} \end{bmatrix} \end{aligned} \quad (20b)$$

$$\Phi_i^{U,\phi} = \begin{bmatrix} \mathbf{e}_i^{\phi-n} \cdot (\mathbf{e}_i^{\phi-n})^T & [\mathbf{0}] \\ [\mathbf{0}] & \mathbf{e}_i^{\phi-n} \cdot (\mathbf{e}_i^{\phi-n})^T \end{bmatrix}; \quad \mathbf{e}_i^{\phi-n} = \mathbf{e}_i^\phi - \mathbf{e}_i^n \quad (21a)$$

$$\begin{aligned} (\bar{U}_i^\phi)^2 &= \begin{bmatrix} \mathbf{V}_e \\ \mathbf{V}_f \end{bmatrix}^T \cdot \Phi_i^{U,\phi} \cdot \begin{bmatrix} \mathbf{V}_e \\ \mathbf{V}_f \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{V}_{e,NZ}^{abc} \\ \mathbf{V}_{f,NZ}^{abc} \end{bmatrix}^T \cdot \mathbf{T}_E^T \cdot \Phi_i^{U,\phi} \cdot \mathbf{T}_E \cdot \begin{bmatrix} \mathbf{V}_{e,NZ}^{abc} \\ \mathbf{V}_{f,NZ}^{abc} \end{bmatrix} \end{aligned} \quad (21b)$$

- Measurements of Current Magnitudes

With $\mathbf{S}_{i-j}^\phi$ and $\Phi_{i-j}^{I,\phi}$ defined in (22a)-(22b), a squared line current magnitude measurement is represented as in (22c), where $\bar{I}_{i-j}^\phi$ represents the magnitude of I_{i-j}^ϕ . In (22a), $\mathbf{e}^a = [1 \ 0 \ 0 \ 0]$, $\mathbf{e}^b = [0 \ 1 \ 0 \ 0]$, $\mathbf{e}^c = [0 \ 0 \ 1 \ 0]$, and $\mathbf{e}^n = [0 \ 0 \ 0 \ 1]$, corresponding to superscript ϕ being a, b, c , and n ; Subscript of a zero vector explicitly indicates its size. The measurement error variance of $(\bar{I}_{i-j}^\phi)^2$ is $\sigma_{I,sq}^2 = 4 \cdot (\bar{I}_{i-j}^\phi)^2 \cdot \sigma_I^2$, where σ_I^2 is the measurement error variance of the line current magnitude.

$$\mathbf{S}_{i-j}^\phi = [\mathbf{0}_{i-1}^T \quad -\mathbf{Y}_{i,j}^\phi \quad \mathbf{0}_{j-i-1}^T \quad \mathbf{Y}_{i,j}^\phi \quad \mathbf{0}_{N-1}^T]; \quad \mathbf{Y}_{i,j}^\phi = \mathbf{e}^\phi \cdot \mathbf{Y}_{i,j} \quad (22a)$$

$$\Phi_{i-j}^{I,\phi} = \begin{bmatrix} \text{Re}((\mathbf{S}_{i-j}^\phi)^H \cdot \mathbf{S}_{i-j}^\phi) & -\text{Im}((\mathbf{S}_{i-j}^\phi)^H \cdot \mathbf{S}_{i-j}^\phi) \\ \text{Im}((\mathbf{S}_{i-j}^\phi)^H \cdot \mathbf{S}_{i-j}^\phi) & \text{Re}((\mathbf{S}_{i-j}^\phi)^H \cdot \mathbf{S}_{i-j}^\phi) \end{bmatrix} \quad (22b)$$

$$\begin{aligned} (\bar{I}_{i-j}^\phi)^2 &= \begin{bmatrix} \mathbf{V}_e \\ \mathbf{V}_f \end{bmatrix}^T \cdot \Phi_{i-j}^{I,\phi} \cdot \begin{bmatrix} \mathbf{V}_e \\ \mathbf{V}_f \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{V}_{e,NZ}^{abc} \\ \mathbf{V}_{f,NZ}^{abc} \end{bmatrix}^T \cdot \mathbf{T}_E^T \cdot \Phi_{i-j}^{I,\phi} \cdot \mathbf{T}_E \cdot \begin{bmatrix} \mathbf{V}_{e,NZ}^{abc} \\ \mathbf{V}_{f,NZ}^{abc} \end{bmatrix} \end{aligned} \quad (22c)$$

- Pseudo Measurements of Active and Reactive Loads

By defining $\mathbf{Y}_i^\phi$, $\Phi_i^{P,\phi}$, and $\Phi_i^{Q,\phi}$ as in (23a)-(23c), pseudo measurements of active and reactive loads can be represented as in (23d) and (23e).

$$\mathbf{Y}_i^\phi = \mathbf{e}_i^\phi \cdot (\mathbf{e}_i^\phi)^T \cdot \mathbf{Y}; \quad (23a)$$

$$\Phi_i^{P,\phi} = \frac{1}{2} \cdot \begin{bmatrix} \text{Re}(\mathbf{Y}_i^\phi + (\mathbf{Y}_i^\phi)^T) & \text{Im}((\mathbf{Y}_i^\phi)^T - \mathbf{Y}_i^\phi) \\ \text{Im}(\mathbf{Y}_i^\phi - (\mathbf{Y}_i^\phi)^T) & \text{Re}(\mathbf{Y}_i^\phi + (\mathbf{Y}_i^\phi)^T) \end{bmatrix} \quad (23b)$$

$$\Phi_i^{Q,\phi} = \frac{-1}{2} \cdot \begin{bmatrix} \text{Im}(\mathbf{Y}_i^\phi + (\mathbf{Y}_i^\phi)^T) & \text{Re}(\mathbf{Y}_i^\phi - (\mathbf{Y}_i^\phi)^T) \\ \text{Re}((\mathbf{Y}_i^\phi)^T - \mathbf{Y}_i^\phi) & \text{Im}(\mathbf{Y}_i^\phi + (\mathbf{Y}_i^\phi)^T) \end{bmatrix} \quad (23c)$$

$$\begin{aligned} P_i^\phi &= \begin{bmatrix} \mathbf{V}_e \\ \mathbf{V}_f \end{bmatrix}^T \cdot \Phi_i^{P,\phi} \cdot \begin{bmatrix} \mathbf{V}_e \\ \mathbf{V}_f \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{V}_{e,NZ}^{abc} \\ \mathbf{V}_{f,NZ}^{abc} \end{bmatrix}^T \cdot \mathbf{T}_E^T \cdot \Phi_i^{P,\phi} \cdot \mathbf{T}_E \cdot \begin{bmatrix} \mathbf{V}_{e,NZ}^{abc} \\ \mathbf{V}_{f,NZ}^{abc} \end{bmatrix} \end{aligned} \quad (23d)$$

$$\begin{aligned} Q_i^\phi &= \begin{bmatrix} \mathbf{V}_e \\ \mathbf{V}_f \end{bmatrix}^T \cdot \Phi_i^{Q,\phi} \cdot \begin{bmatrix} \mathbf{V}_e \\ \mathbf{V}_f \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{V}_{e,NZ}^{abc} \\ \mathbf{V}_{f,NZ}^{abc} \end{bmatrix}^T \cdot \mathbf{T}_E^T \cdot \Phi_i^{Q,\phi} \cdot \mathbf{T}_E \cdot \begin{bmatrix} \mathbf{V}_{e,NZ}^{abc} \\ \mathbf{V}_{f,NZ}^{abc} \end{bmatrix} \end{aligned} \quad (23e)$$

- Measurements of Voltage Phasors

Measurements on voltage real and imaginary parts of phase ϕ at bus i can be respectively represented as in (24a) and (24b). As measurement error variances for magnitude and phase angle are usually given, measurement variances of voltage real and imaginary parts can be calculated as in (25) [39], where σ_V^2 and σ_δ^2 are respectively measurement error variances of voltage magnitude and angle.

$$\text{Re}(V_i^\phi) = \begin{bmatrix} (\mathbf{e}_i^\phi)^T & \mathbf{0}^T \end{bmatrix} \cdot \mathbf{T}_E \cdot \begin{bmatrix} \mathbf{V}_{e,NZ}^{abc} \\ \mathbf{V}_{f,NZ}^{abc} \end{bmatrix} \quad (24a)$$

$$\text{Im}(V_i^\phi) = \begin{bmatrix} \mathbf{0}^T & (\mathbf{e}_i^\phi)^T \end{bmatrix} \cdot \mathbf{T}_E \cdot \begin{bmatrix} \mathbf{V}_{e,NZ}^{abc} \\ \mathbf{V}_{f,NZ}^{abc} \end{bmatrix} \quad (24b)$$

$$\begin{aligned} \mathbf{\Theta}_i^\phi &= \begin{bmatrix} \cos(\delta_i^\phi) & -\bar{V}_i^\phi \cdot \sin(\delta_i^\phi) \\ \sin(\delta_i^\phi) & \bar{V}_i^\phi \cdot \cos(\delta_i^\phi) \end{bmatrix} \cdot \begin{bmatrix} \sigma_V^2 & 0 \\ 0 & \sigma_\delta^2 \end{bmatrix} \\ &= \begin{bmatrix} \cos(\delta_i^\phi) & -\bar{V}_i^\phi \cdot \sin(\delta_i^\phi) \\ \sin(\delta_i^\phi) & \bar{V}_i^\phi \cdot \cos(\delta_i^\phi) \end{bmatrix}^T \end{aligned} \quad (25)$$

In summary, equations (20)-(24) describe that all measurements can be represented via state variables $\mathbf{V}_{e,NZ}^{abc}$ and $\mathbf{V}_{f,NZ}^{abc}$ in the form of $\begin{bmatrix} \mathbf{V}_{e,NZ}^{abc} \\ \mathbf{V}_{f,NZ}^{abc} \end{bmatrix}^T \cdot \Phi \cdot \begin{bmatrix} \mathbf{V}_{e,NZ}^{abc} \\ \mathbf{V}_{f,NZ}^{abc} \end{bmatrix}$, whose partial derivative over $\begin{bmatrix} \mathbf{V}_{e,NZ}^{abc} \\ \mathbf{V}_{f,NZ}^{abc} \end{bmatrix}^T$ is $\begin{bmatrix} \mathbf{V}_{e,NZ}^{abc} \\ \mathbf{V}_{f,NZ}^{abc} \end{bmatrix}^T \cdot (\Phi + \Phi^T)$. It is also noteworthy that (20) and (22) are also applicable for measurements on neutrals. Moreover, if maximum measurement errors instead of error variances are known, with a reasonable assumption that measurement errors follow the normal distribution, by setting the triple standard deviation of the normal distribution equal to the maximum measurement error, the measurement error variance can be calculated [40].

E. Extension to Polar Coordinates

The DSSE model under polar coordinates is further discussed in this subsection, in which the aforementioned measurements are represented via $\mathbf{V}_{e,NZ}^{abc}$ in polar coordinates. $\mathbf{V}_{e,NZ}^{abc}$ and $\mathbf{V}_{f,NZ}^{abc}$ can be represented by voltage magnitudes and phase angles as new state variables. That is, $\text{Re}(V_i^\phi) = \bar{V}_i^\phi \cdot \cos(\delta_i^\phi)$ and $\text{Im}(V_i^\phi) = \bar{V}_i^\phi \cdot \sin(\delta_i^\phi)$, where δ_i^ϕ represents

voltage phase angles. Thus, $\begin{bmatrix} \mathbf{V}_{e,NZ}^{abc} \\ \mathbf{V}_{f,NZ}^{abc} \end{bmatrix}$ can be represented via the function vector $\mathbf{f}(\bar{\mathbf{V}}_{NZ}^{abc}, \boldsymbol{\delta}_{NZ}^{abc})$, where $\bar{\mathbf{V}}_{NZ}^{abc}$ and $\boldsymbol{\delta}_{NZ}^{abc}$ are vectors of voltage magnitudes and phase angles of $\mathbf{V}_{e,NZ}^{abc}$.

By substituting $\mathbf{V}_{e,NZ}^{abc}$ and $\mathbf{V}_{f,NZ}^{abc}$ with $\mathbf{f}(\bar{\mathbf{V}}_{NZ}^{abc}, \boldsymbol{\delta}_{NZ}^{abc})$, equations (20)-(23) can be written in the form of $\begin{bmatrix} \mathbf{f}(\bar{\mathbf{V}}_{NZ}^{abc}, \boldsymbol{\delta}_{NZ}^{abc}) \end{bmatrix}^T \cdot \Phi \cdot \mathbf{f}(\bar{\mathbf{V}}_{NZ}^{abc}, \boldsymbol{\delta}_{NZ}^{abc})$, whose partial derivative over $\begin{bmatrix} (\bar{\mathbf{V}}_{NZ}^{abc})^T & (\boldsymbol{\delta}_{NZ}^{abc})^T \end{bmatrix}$ can be calculated via (26). It is noteworthy that (26) represents partial derivatives of real-valued functions with respect to real-valued variables. In addition, in (24)-(25) under rectangular coordinates, actual measurements on voltage magnitudes and angles from μPMU are converted to real and imaginary parts. However, under polar coordinates, they are directly represented via $\bar{\mathbf{V}}_{NZ}^{abc}$ and $\boldsymbol{\delta}_{NZ}^{abc}$. Thus, the corresponding measurement error variances can be directly obtained without conversion.

$$\begin{aligned} &\frac{\partial \left[\begin{bmatrix} \mathbf{f}(\bar{\mathbf{V}}_{NZ}^{abc}, \boldsymbol{\delta}_{NZ}^{abc}) \end{bmatrix}^T \cdot \Phi \cdot \mathbf{f}(\bar{\mathbf{V}}_{NZ}^{abc}, \boldsymbol{\delta}_{NZ}^{abc}) \right]}{\partial \left[(\bar{\mathbf{V}}_{NZ}^{abc})^T & (\boldsymbol{\delta}_{NZ}^{abc})^T \right]} \\ &= \begin{bmatrix} \mathbf{f}(\bar{\mathbf{V}}_{NZ}^{abc}, \boldsymbol{\delta}_{NZ}^{abc}) \end{bmatrix}^T \cdot (\Phi + \Phi^T) \cdot \frac{\partial \mathbf{f}(\bar{\mathbf{V}}_{NZ}^{abc}, \boldsymbol{\delta}_{NZ}^{abc})}{\partial \left[(\bar{\mathbf{V}}_{NZ}^{abc})^T & (\boldsymbol{\delta}_{NZ}^{abc})^T \right]} \end{aligned} \quad (26)$$

F. WLS Based State Estimation

Using vectors $\mathbf{Z}$ and $\mathbf{X}$ to represent measurement values and state variables, $\mathbf{Z}$ can be represented as $\mathbf{h}(\mathbf{X})$, where $\mathbf{h}(\cdot)$ is the vector of measurement functions. State variable vector $\mathbf{X}$ can be estimated via WLS as in (27), where $\mathbf{W}$ is a diagonal block weight matrix. In the matrix $\mathbf{W}$, diagonal blocks corresponding to measurements of phase-to-ground voltage magnitudes, phase-to-neutral voltage magnitudes, line current magnitudes, active loads, and reactive loads are respectively set as reciprocals of $\sigma_{V,sq}^2$, $\sigma_{U,sq}^2$, $\sigma_{I,sq}^2$, σ_P^2 , and σ_Q^2 , while diagonal blocks corresponding to measurements of voltage phasors are set as inverse of $\mathbf{\Theta}_i^\phi$. Estimation on $\mathbf{X}$ is obtained by iteratively calculating $\Delta \mathbf{X}^k$ and updating $\mathbf{X}$ via (28a)-(28b), until a certain threshold on $\Delta \mathbf{X}^k$ is met. $\mathbf{H}(\mathbf{X})$ is the Jacobian matrix of $\mathbf{h}(\mathbf{X})$ over $\mathbf{X}$.

$$\min_{\mathbf{X}} J(\mathbf{X}) = \frac{1}{2} \cdot [\mathbf{Z} - \mathbf{h}(\mathbf{X})]^T \cdot \mathbf{W} \cdot [\mathbf{Z} - \mathbf{h}(\mathbf{X})] \quad (27)$$

$$\Delta \mathbf{X}^k = \left[(\mathbf{H}(\mathbf{X}^k))^T \cdot \mathbf{W} \cdot \mathbf{H}(\mathbf{X}^k) \right]^{-1} \cdot (\mathbf{H}(\mathbf{X}^k))^T \cdot \mathbf{W} \cdot [\mathbf{Z} - \mathbf{h}(\mathbf{X}^k)] \quad (28a)$$

$$\mathbf{X}^{k+1} = \mathbf{X}^k + \Delta \mathbf{X}^k \quad (28b)$$

III. LOAD ESTIMATION MODEL TO CALCULATE LOAD PSEUDO MEASUREMENTS OF LDSS

Individual customers in an LDS may present distinct load patterns. In order to accurately characterize load profiles of the entire LDS, we propose to divide customers into multiple clusters, identify representative customers of each cluster to build a load estimation model for the cluster, and finally calculate the total load of the entire LDS by aggregating

estimated loads of individual clusters. In addition, only active load models are built, and reactive loads are calculated via the estimated active loads and pre-set power factors according to historical data.

A. Establish Clusters and Their Load Estimation Models

Clusters of an LDS and their load estimation models are established via a regression oriented agglomerative hierarchical clustering algorithm. That is, given an initial cluster scheme with a set of clusters, the algorithm merges two clusters at a time according to a certain criterion to generate a new cluster scheme, until only one cluster remains. Finally, the cluster scheme with the best performance, together with its load estimation models, is chosen as the final cluster scheme. In this paper, quality of regression QR defined in (29) is used as a metric to evaluate performance of each cluster scheme $\mathcal{C}$, where c is index of clusters; d is index of customers; $\mathcal{D}$ denotes the set of customers in an LDS on a certain phase; $\mathcal{D}_c$ is the set of customers in cluster c of scheme $\mathcal{C}$; S_c^{ps} is predicted residual error sum of squares (PRESS); S_c^{ss} is residual sum of squares for cluster c [41]; $card(\cdot)$ is cardinality of a set.

$$QR = \sum_{c \in \mathcal{C}} (card(\mathcal{D}_c) \cdot S_c^{ps} / S_c^{ss}) / card(\mathcal{D}) \quad (29)$$

The proposed algorithm is described as follows:

Step 1: Set $k = 1$ and create an initial cluster scheme $\mathcal{C}^1$. As communication bottleneck will restrict the number of smart meters that can simultaneously upload measurements to the control center in real time, $\mathcal{M}$ is used as the real-time communication smart meter quota. A bisection K -means based hierarchical division algorithm [42] is adopted to create the initial cluster scheme $\mathcal{C}^1$ with no more than $\mathcal{M}$ clusters, by measuring the Euler distance of historical load profiles of individual customers. Build PLS models for individual clusters in $\mathcal{C}^1$ and calculate QR^1 .

Step 2: Any two clusters in $\mathcal{C}^k$ could potentially merge, which presents a total of $card(\mathcal{C}^k) \cdot [card(\mathcal{C}^k) - 1] / 2$ options. Build a PLS regression model and calculate QR for each option, and record the one with the smallest QR as $\mathcal{C}^{k+1}$, together with the corresponding PLS models and QR^{k+1} .

Step 3: If $card(\mathcal{C}^{k+1}) = 1$, go to *Step 4*; Otherwise, set $k = k + 1$ and go to *Step 2*.

Step 4: Among all $\mathcal{C}^k$ for $k = 1, 2, 3, \dots, card(\mathcal{C}^1)$, choose the one with the smallest QR as the final cluster scheme. This identified cluster scheme and its corresponding PLS models, together with real-time smart meter measurements of selective customers, are used to calculate load pseudo measurements P_i^ϕ and Q_i^ϕ of an LDS at phase ϕ of bus i in real time.

B. PLS Regression Based Load Estimation Model

This subsection describes a procedure to build regression models of individual clusters in a given cluster scheme $\mathcal{C}$ and to calculate the corresponding QR , as discussed in above Step 2. PLS regression [41] is used, given its advantages of effectively combining ordinary least square regression and principle analysis to handle issues of data collinearity and observation insufficiency.

Specifically, the PLS regression based load estimation model (30) is built for each cluster c in cluster scheme $\mathcal{C}$, describing the relationship between independent variables L_c and dependent variable L_d for $d \in \mathcal{D}_c^s$. L_c represents the normalized total active load of all customers in cluster c , and L_d represents the normalized active load of customer d . $\mathcal{D}_c^s$ denotes the set of selective customers in cluster c . Regression coefficients α_d are calculated via the normalized matrix of historical loads of selective customers $\mathbf{H}_c^s$ and the normalized vector of historical loads $\mathbf{Z}_c$. For a weekday/weekend, only history weekday/weekend data are used.

$$L_c = \sum_{d \in \mathcal{D}_c^s} \alpha_d \cdot L_d \quad (30)$$

The detailed procedure for building PLS models and calculating QR values is described as follows:

(i) *Calculate VIP of Individual Customers:* The variable importance projection (VIP) of an independent variable is used as a metric to evaluate its importance in representing dependent variables through (30) [43]. That is, the bigger the VIP, the more representative an independent variable. VIP of each customer in a cluster is calculated via (31), where values of parameters $\mathbf{J}_m$ and $\beta_{m,d}$ are calculated via Algorithm 1 with inputs $\mathbf{A}_0 = \mathbf{H}_c$, $\mathbf{F}_0 = \mathbf{Z}_c$, and $\mathcal{S} = \mathcal{D}_c$. $cor(\cdot, \cdot)$ denotes correlation coefficient, and $\mathbf{H}_c$ is the normalized matrix of historical loads of all customers in cluster c . M is a pre-specified parameter representing the number of principle components, which can be determined by the cross validation method [41]. Parameter $\mathbf{e}_d$ in Algorithm 1 is a standard basis vector with the element corresponding to customer d being 1.

$$VIP_d = \sqrt{\frac{card(\mathcal{D}_c) \cdot \sum_{m=1}^M (cor(\mathbf{F}_0 \mathbf{J}_m)^2 \cdot \beta_{m,d})}{\sum_{m=1}^M cor(\mathbf{F}_0 \mathbf{J}_m)^2}}; \quad d \in \mathcal{D}_c \quad (31)$$

Algorithm 1: Regression Algorithm

Input: $\mathbf{A}_0$, $\mathbf{F}_0$, and $\mathcal{S}$

For $m = 1$ to M , calculate

$$\mathbf{B}_m = \frac{\mathbf{A}_{m-1}^T \cdot \mathbf{F}_{m-1}}{\|\mathbf{A}_{m-1}^T \cdot \mathbf{F}_{m-1}\|}; \quad \mathbf{P}_m = \frac{\mathbf{A}_{m-1}^T \cdot \mathbf{J}_m}{\|\mathbf{J}_m\|^2}; \quad \mathbf{R}_m = \frac{\mathbf{F}_{m-1}^T \cdot \mathbf{J}_m}{\|\mathbf{J}_m\|^2};$$

$$\mathbf{A}_m = \mathbf{A}_{m-1} - \mathbf{J}_m \cdot \mathbf{P}_m^T; \quad \mathbf{J}_m = \mathbf{A}_{m-1} \cdot \mathbf{B}_m;$$

$$\mathbf{F}_m = \mathbf{F}_{m-1} - \mathbf{J}_m \cdot \mathbf{R}_m$$

For each $d \in \mathcal{S}$, calculate

$$\beta_{m,d} = \begin{cases} \mathbf{e}_d^T \cdot \prod_{h=1}^{m-1} (\mathbf{E} - \mathbf{B}_h \cdot \mathbf{P}_h^T) \cdot \mathbf{B}_m; & \text{if } m > 1 \\ \mathbf{e}_d^T \cdot \mathbf{B}_m; & \text{if } m = 1 \end{cases}$$

end

end

Output: $\mathbf{J}_m$, $\mathbf{R}_m$, and $\beta_{m,d}$ for $m = 1, \dots, M$

(ii) *Determine Selective Customers:* Based on VIP values calculated above, a selective customer identification problem (32) is solved to determine $\mathcal{D}_c^s$. The objective (32a) is to maximize the total VIP value of selective customers, i.e., choosing the most representative customers in clusters in terms of regression. I_d is a binary variable with 1 indicating customer d is selected; otherwise 0. Constraint (32b) limits the number of selected customers to be no larger than the smart meter quota $\mathcal{M}$. Constraint (32c) forces that each cluster includes at least one selective customer.

$$\max_{I_d \in \{0,1\}} \sum_{d \in \mathcal{D}} VIP_d \cdot I_d \quad (32a)$$

$$\sum_{d \in \mathcal{D}} I_d \leq \mathcal{M}; \quad d \in \mathcal{D} \quad (32b)$$

$$1 \leq \sum_{d \in \mathcal{D}_c} I_d; \quad c \in \mathcal{C} \quad (32c)$$

Problem (32) can be effectively solved via the following three-step process: (a) In each cluster, select the customer with the largest VIP value; (b) Sort remaining customers from all clusters in a descending order of VIP; (c) Combine the top $(\mathcal{M} - \text{card}(\mathcal{C}))$ customers with those identified in (a) to constitute the final $\mathcal{D}_c^s$. Considering that $\sum_{d \in \mathcal{D}_c} \text{VIP}_d = \text{card}(\mathcal{D}_c)$ [43] and VIP values of top customers in a large cluster are much higher than those in small clusters, VIP values of non-top customers in large clusters are smaller than top ones in small clusters. Thus, over-concentration or over-dispersion in selecting $\mathcal{D}_c^s$ among clusters can be avoided.

(iii) *Build PLS Regression Models of Individual Clusters:* With $\mathcal{D}_c^s$ determined from (32), equation (33) is used to compute regression coefficient α_d , where R_m and $\beta_{m,d}$ are calculated via Algorithm 1 with inputs $\mathbf{A}_0 = \mathbf{H}_c^s$, $\mathbf{F}_0 = \mathbf{Z}_c$, and $\mathcal{S} = \mathcal{D}_c^s$. Finally, S_c^{ss} and S_c^{ps} can be calculated with $\mathcal{D}_c^s$ and (30) [41], and QR can be obtained via (29).

$$\alpha_d = \sum_{m=1}^M R_m \cdot \beta_{m,d} \quad (33)$$

C. Variances of Load Estimates

When using load estimations as load pseudo measurements of LDSs in DSSE, variances of estimation errors are needed to determine their weights in $\mathbf{W}$ of (20). In this paper, we use the ordinary least square type expression [44] to approximate estimation error variances. With a reasonable assumption that loads of different clusters are independent, approximate variances of active load estimation errors can be calculated via (34), where $\sigma_{P,c}^2$ represents approximate variance of active load estimation error for cluster c . It is noteworthy that as reactive power of an LDS is calculated through a pre-set power factor, active power and reactive power are fully correlated and their covariance matrix is singular. Thus, in order to derive effective weights, we opt to use equation (35) to calculate approximate variances of reactive load estimation errors by neglecting off-diagonal elements in the covariance matrix, where pf is a pre-determined power factor of an LDS.

$$\sigma_P^2 = \sum_{c \in \mathcal{C}} \sigma_{P,c}^2 \quad (34)$$

$$\sigma_Q^2 = [(1 - pf^2)/pf^2] \cdot \sigma_P^2 \quad (35)$$

IV. CASE STUDIES

Numerical case studies are conducted to evaluate the proposed load estimation model and relevant factors that could affect its performance, and to assess the DSSE model with a quantitative analysis on computational performance and solution accuracy. Load estimation models are implemented via C# in Visual Studio, and state estimation models are implemented in MATLAB. All case studies are conducted on a personal computer with i7 2.90GHz CPU and 16GB RAM.

A. Load Estimation Model

Smart meter data in CER [45] are used, which contain half-hour granularity electricity load measurements for over 4000 residential customers from 08/14/09 to 12/31/10. The first 301 customers with complete data, which are contained in a same

phase, are chosen to form an LDS. 48 half-hour load estimation models of this LDS, indexed as 1 to 48, are established to perform load estimation studies on 11/09/09.

Different parameter settings are applied in this study. Specifically, the size of history data varies from 10 to 40 with a step size of 5, and the smart meter quota ranges from 20 to 70 (i.e., 6.64% to 23.26% of total smart meters) with a step size of 10. Load estimation error is defined as the relative residual between the actual and estimated values for each measurement time point, and daily average load estimation error is the average of load estimation errors throughout a day.

Daily average load estimation errors against different parameter settings are shown in Table I. A general trend is that load estimation is more accurate with a larger value of $\mathcal{M}$. On the other hand, better performance is observed with historical data sizes of 20, 25, and 35, indicating that unlike the smart meter quota, a larger historical data set may not necessarily improve estimation accuracy. This can be understood as that an unnecessarily larger historical data set which covers a longer time period may contain exceptional load pattern variations triggered by unforeseeable events.

TABLE I AVERAGE LOAD ESTIMATION ERRORS WITH DIFFERENT SETTINGS

Smart meter quota	Size of history data						
	10	15	20	25	30	35	40
20 (6.64%)	8.06%	5.29%	5.05%	4.57%	4.91%	5.57%	6.19%
30 (9.96%)	7.95%	4.68%	4.20%	4.02%	5.75%	4.28%	5.86%
40 (13.29%)	7.56%	5.12%	4.20%	4.47%	4.61%	4.80%	5.38%
50 (16.61%)	6.04%	5.15%	4.49%	3.51%	4.23%	4.03%	4.59%
60 (19.93%)	5.85%	4.45%	3.92%	3.75%	4.91%	3.70%	4.53%
70 (23.26%)	5.44%	4.42%	3.60%	3.89%	3.64%	3.78%	4.45%
	≥6%, [5%, 6%), [4%, 5%), [3%, 4%)						

Next, load estimation results of this LDS for the week of 11/09/09 (Mon.) to 11/15/09 (Sun.) are studied. In this study, $\mathcal{M}$ is set as 60 (i.e., 19.93% of total smart meters), and the size of historical data for weekdays/weekends is set as 35/25. The proposed load estimation model is compared with several prevailing forecasting models, including the persistence method, the autoregressive moving average (ARMA) method, and the support vector machine (SVM) method. The persistence method simply uses actual load values of the previous day as estimated load values [46]. ARMA is widely applied for time series analysis. One ARMA load forecasting model is built for each measurement time point based on historical load data series of an LDS [47], in which the Akaike information criterion (AIC) is adopted to identify the best orders of the ARMA model. The reason for adopting ARMA, instead of other time series methods in the same family, is because historical load data series of individual measurement time points are stationary while their autocorrelations and partial autocorrelations tail off to zero, which better fits the properties of ARMA. SVM is a supervised learning based method that has been widely used in data regression analysis. Two types of SVM based load forecasting models, SVM-I and SVM-II, are respectively built with and without real-time smart meter measurements [46]. Specifically, in SVM-I, when limited real-time measurements are available, similar to the proposed load estimation model, loads of an LDS are considered as dependent variables while loads of selective

customers are considered as independent variables. Selective customers are identified by measuring the similarity via Euler distance between historical load profiles of the LDS and each customer. In comparison, SVM-II is built based only on historical load series of the LDS while neglecting real-time measurements. In the following studies, suitable lengths of training data sets for ARMA and SVM models are tuned via experiments. The number of selective real-time measurements used in SVM-I is equal to 60.

ARMA and SVM models are developed in MATLAB. LIBSVM [48] package is used to implement SVM models. Daily average load estimation errors of all models for the entire week are compared in Table II. It is observed that the proposed model presents the smallest weekly average load estimation error among all five models, as well as the smallest daily average load estimation errors for six days of the week except Sunday. The two SVM models show close performance and are generally better than the ARMA and the persistence models, but present the highest estimation errors in Saturday. In summary, the proposed load estimation model is consistently more accurate than the other models.

TABLE II AVERAGE ESTIMATION ERRORS OF THE WHOLE WEEK

	11/09	11/10	11/11	11/12	11/13	11/14	11/15	Weekly
	Mon.	Tue.	Wen.	Thu.	Fri.	Sat.	Sun.	Average
Proposed model	3.70%	3.79%	4.54%	3.57%	3.7%	4.86%	4.75%	4.13%
Persistence model	5.07%	6.60%	6.69%	4.50%	5.40%	6.65%	5.82%	5.82%
ARMA model	6.14%	7.61%	6.71%	6.40%	6.11%	6.93%	6.48%	6.63%
SVM-I	4.89%	5.36%	5.19%	4.64%	5.34%	7.10%	3.96%	5.21%
SVM-II	4.57%	5.62%	5.20%	4.06%	5.35%	7.50%	4.24%	5.22%

As the cluster scheme and the set of selective customers in each cluster could impact load estimation accuracy, they should be updated periodically to avoid degradation in estimation accuracy. Updates could be conducted monthly, weekly, or even daily as needed, triggered when the load estimation error is larger than a threshold. It is noteworthy that since load estimation models of individual LDSs are independent, they can be separately updated in parallel. In our case studies, it takes about 50 seconds to build a load estimation model for the LDS with 301 customers, which indicates that the computational burden is tolerable.

B. Computational Performance of The Proposed DSSE

A modified IEEE 123-bus system is used in this section to evaluate computational performance of the proposed DSSE model. Detailed system data can be found in [49]. A 10hm ground resistance is added at each bus of the original system, and per unit line impedances are recalculated according to the given line type codes and spacing IDs. In the original 123-bus system, 119 buses remain after removing isolated, switch, and secondary buses. In addition, considering single-/two-phase laterals, there is a total of 244 non-neutral phases (including 98 non-zero injection phases) and 119 neutrals. Thus, initially, there are 363×2 (i.e., $(244+119) \times 2$) variables, in which 363 is the total number of phases and 2 represents two variables for each phase voltage (i.e., real and imaginary terms under rectangular coordinates, or magnitude and angle under polar coordinates). On the other hand, if neutrals and zero-injection

phases are eliminated, only 98 non-neutral and non-zero injection phases remain, corresponding to 98×2 variables. That is, by applying the conversion via (19), the number of state variables is reduced by 73%. Locations of measurement devices are shown in Fig. 1. μ PMUs are only installed at the substation bus with the maximum measurement error of 1% for magnitudes and 10^{-2} rad for angles [39]. The maximum measurement error of all load pseudo measurements is set as 50%. Measurement settings of SCADA are summarized in Table III.

DSSE models with and without the proposed variable elimination approach as well as under rectangular and polar coordinates are compared. Comparisons are conducted under both Setting 1 and Setting 2 shown in Table III. The DSSE model with variable elimination is referred to as “reduced DSSE model”, while the one without variable elimination is called “unreduced DSSE model”.

For each model and under each measurement setting, 5000 independent DSSE runs with randomly generated measurement errors are executed. Initial voltage values of phases a , b , and c are set as $1 \angle 0^\circ$ p.u., $1 \angle -120^\circ$ p.u., and $1 \angle 120^\circ$ p.u.. In addition, in unreduced DSSE models, initial voltage values of neutrals are all set as $0 \angle 0^\circ$ p.u.. The DSSE converges when the maximum value of $|\Delta X^k|$ is no larger than 10^{-10} . MATLAB function “sparse” is used to handle sparse matrices, and backslash operator “\” is used to solve the linear equation (28a).

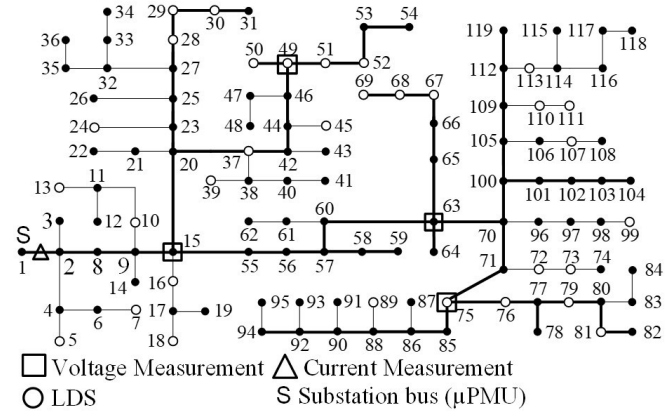


Fig. 1 A modified IEEE 123-bus test system and its measurement deployment

TABLE III SETTINGS ON MAXIMUM MEASUREMENT ERRORS

Measurement	Phase-to-ground voltage		Phase-to-neutral voltage	Current	
	neutral	non-neutral		neutral	non-neutral
Setting 1	-	1%	-	-	3%
Setting 2	1%	1%	-	3%	3%
Setting 3	-	-	1%	-	3%

The four model combinations under both Setting 1 and Setting 2 are compared in Table IV. As calculating Jacobian matrix $\mathbf{H}$ and solving equation (28a) are two major tasks of the DSSE computation, besides the entire DSSE calculation time, average computational performance for calculating $\mathbf{H}$ and solving (28a) per iteration is also recorded. It is noted that in unreduced DSSE models, the dimension of $(\mathbf{H}(\mathbf{X}^k))^T \cdot \mathbf{W} \cdot \mathbf{H}(\mathbf{X}^k)$ is 726×726 with 29,202 non-zeros, while it becomes

196×196 with 38,416 non-zeros in the reduced DSSE models.

Comparison between the reduced and unreduced DSSE models under rectangular coordinates in Setting 1 shows that, average DSSE time is significantly reduced when the proposed variable elimination approach is adopted. Similar levels of computational benefits can also be observed when applying the proposed variable elimination approach under Setting 2. In addition, comparison on reduced DSSE models under polar and rectangular coordinates shows that, computational times for solving (28a) are close while those for calculating $\mathbf{H}$ are significantly different, which is caused by heavier computational burden in calculating $\mathbf{H}$ through (26) under polar coordinates. Moreover, the unreduced DSSE model under polar coordinates fails to converge in most runs. The reason is that in this model, initial setting of neutral voltages has to be provided, while the flat initial setting (i.e., 0 $\angle 0^\circ$ p.u.) could be far from their actual values, especially the neutral voltage angles. Indeed, it is difficult to determine proper initial values due to their strong variabilities caused by volatile loads on individual buses. On the other hand, the proposed variable elimination approach can effectively avoid this issue, because neutrals are eliminated and their initial settings are no longer needed. This clearly shows the advantage of eliminating variables related to neutrals for enhancing DSSE computational performance. In addition, comparison of the same model under Setting 1 and Setting 2 further shows that including neutral measurements in Setting 2 would need more iterations and longer computational time.

TABLE IV COMPUTATIONAL PERFORMANCE

Setting	DSSE model	Ave. DSSE time	Ave. # of iterations	Ave. time to calculate $\mathbf{H}$ per iteration	Ave. time to solve (28a) per iteration
1	Reduced/Rect.	13.1 ms	6	1.4 ms	0.7 ms
	Unreduced/Rect.	93.2 ms	7	8.7 ms	4.1 ms
	Reduced/Polar	298.5 ms	6	47.7 ms	0.8 ms
	Unreduced/Polar	-	-	-	-
2	Reduced/Rect.	78.3 ms	30	1.6 ms	0.8 ms
	Unreduced/Rect.	461.3 ms	34	8.8 ms	4.3 ms
	Reduced/Polar	1215.4 ms	23	53.0 ms	0.8 ms
	Unreduced/Polar	-	-	-	-

C. Solution Accuracy of the Proposed DSSE Models

In this section, solution accuracies of DSSE models as well as impacts of load estimation models are studied. Reduced DSSE models are used in this section. 11/09/09 (Mon.) is chosen for the study. 3 types of LDSs with different numbers of customers and smart meter quotas, as summarized in Table V, are used to replace 39 (out of 95) loads in the original test system. Load estimation models of the 39 LDSs are built via historical data with the length of 35.

TABLE V INFORMATION OF THE 39 REPLACED LDSS

	LDS Type 1	LDS Type 2	LDS Type 3
Number of LDSs in the test system	18	11	10
Number of included customers	50	80	100
Smart meter quota	10 (20%)	16 (20%)	20 (20%)

Pseudo measurements of the 39 LDSs are estimated via the proposed load estimation models, which, together with approximated load estimation error variances, are used as inputs to the DSSE model. Maximum load pseudo

measurement errors of the remaining 56 LDSs are set as 50%. Thus, assuming estimation errors of these 56 loads follow normal distributions, their average relative errors are about 13.33%.

5000 DSSE runs are executed for each measurement time point. Measurement values for each DSSE run are generated via actual power flow solutions plus random measurement errors sampled from the error distributions. **Measurement errors are considered to follow normal distributions, whose means are zero and variances are calculated via the approach discussed in Section II.D using the maximum measurement errors. Relative root-mean-square errors (RRMSE) of voltage magnitudes and absolute errors of voltage phase angles** are used as metrics to quantify DSSE solution accuracy. Three settings shown in Table III are studied to illustrate solution accuracy of the proposed DSSE model.

The DSSE models under rectangular and polar coordinates with Setting 1 are compared first. Maximum RRMSEs over all non-neutral phase voltage magnitudes for each measurement time point are compared in Fig. 2. The three curves interlace closely, indicating their similarities in solution accuracy. We further include the result of DSSE under rectangular coordinates with Setting 3 in Fig. 2 for comparison, which replaces phase-to-ground voltage magnitude measurements with phase-to-neutral magnitude measurements. It shows that using phase-to-ground and phase-to-neutral voltage measurements would deliver DSSE solutions of similar accuracy. Maximum and minimum RRMSEs of non-neutral phase voltage magnitudes and angles over all measurement time points are shown in Table VI, which further illustrates that using rectangular and polar coordinates as well as phase-to-ground and phase-to-neutral voltage magnitude measurements would derive DSSE results of similar accuracy.

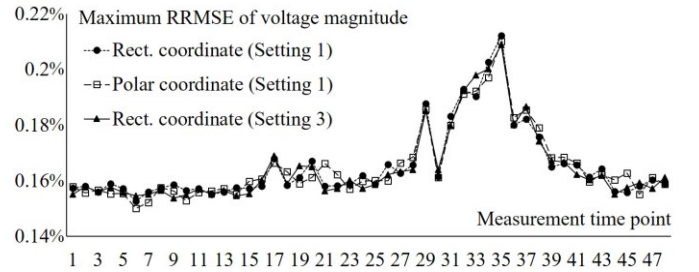


Fig. 2 Maximum RRMSE of non-neutral phase voltage magnitudes using rectangular and polar coordinates and under Settings 1&3

TABLE VI DSSE RESULTS OF THE TWO MODELS WITH TWO SETTINGS

DSSE model & Setting		Minimum RRMSE	Maximum RRMSE
Voltage magnitude	Rect. coordinates (Setting 1)	0.153%	0.212%
	Polar coordinates (Setting 1)	0.150%	0.210%
	Rect. coordinates (Setting 3)	0.153%	0.209%
DSSE model & Setting		Minimum absolute error	Maximum absolute error
Voltage angle	Rect. coordinates (Setting 1)	0.332 crad	0.359 crad
	Polar coordinates (Setting 1)	0.336 crad	0.360 crad
	Rect. coordinates (Setting 3)	0.333 crad	0.360 crad

Maximum RRMSEs over all non-neutral phase voltage magnitudes with Setting 1 and Setting 2 under rectangular coordinates are further compared in Fig. 3. It shows that extra

measurements on neutrals in Setting 2, although do not reduce the maximum RRMSE at every measurement time point, do help eliminate several extreme high RRMSE situations, such as time points 34-36. Indeed, with neutral measurements, RRMSEs of non-neutral phase voltage magnitudes and absolute errors of non-neutral phase voltage angles are respectively in the ranges of [0.152%, 0.194%] and [0.332 crad, 0.357 crad], which are better than those without neutral measurements as shown in the first row of Table VI.

Fig. 4 further shows average RRMSEs over all neutral voltage magnitudes, which are much larger than those of non-neutral phases in Fig. 3. This is because neutral voltages are close to zero and small deviations could cause large RRMSEs. Fig. 4 also indicates that improvements in DSSE accuracy are of significance when measurements on neutrals are included.

We further execute DSSE runs using various load forecasts from SVM-I and SVM-II to evaluate their impacts. SVM-I and SVM-II are chosen because of their better performance as shown in Section IV.A. However, as load estimation error variances of SVM-I and SVM-II are not readily available, maximum relative regression errors of history load series are used as maximum load estimation errors to determine weights of load pseudo measurements.

DSSE results with load pseudo measurements from different models are compared in Fig. 5. It shows that at most measurement time points (i.e., 1-15, 17-24, 26-31, and 40-47), DSSE results derived from the proposed load estimation model as well as SVM-I and SVM-II are close. Although DSSE result using the proposed load estimation model does not necessarily outperform those using SVM-I and SVM-II at every measurement time point, it shows better performance at measurement time points with more intense load volatilities, such as time points 31-40. Specifically, the maximum RRMSE of voltage magnitudes is effectively reduced when the proposed load estimation model is used.

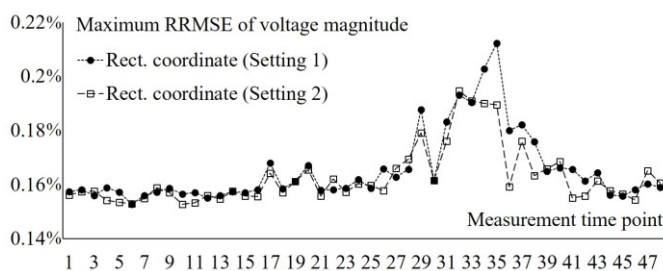


Fig. 3 Maximum RRMSE of non-neutral phase voltages with and without neutral measurements

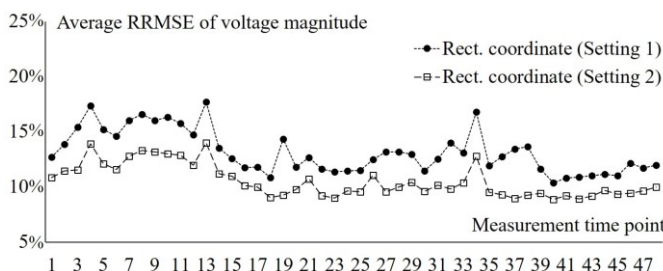


Fig. 4 Average RRMSE of neutral voltage magnitudes with and without neutral measurements

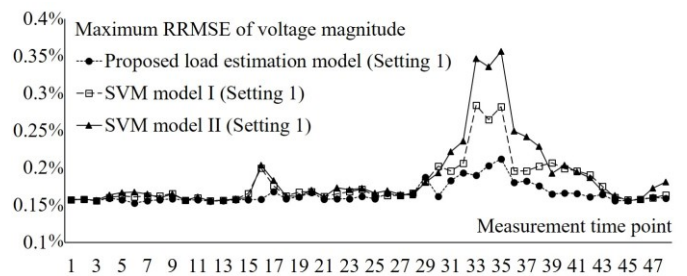


Fig. 5 Maximum RRMSE of non-neutral phase voltage magnitudes with different load pseudo measurement models

V. CONCLUSION

This paper discusses a WLS based DSSE model which accurately considers non-zero potential neutral conductors and effectively leverages voltages of non-neutral and non-zero injection phases in both rectangular and polar coordinates as state variables. Voltage variables of neutrals and zero-injection phases are linearly represented by state variables and eliminated from the DSSE model. With real-time load measurements from selective smart meters, load estimation models generate pseudo measurements of LDS loads to compensate real-time measurement insufficiency in MDS.

Comparisons with other load forecasting models show that the proposed load estimation model can effectively provide load pseudo measurements of LDSs to DSSE with consistently higher accuracy. Numerical studies on the modified IEEE 123-bus system further illustrate advantages of the proposed DSSE model in terms of computational performance and solution accuracy. Specifically, the proposed neutral and zero-injection phase elimination approach can reduce the number of iterations and total computational time. Including neutral measurements could improve DSSE solution accuracy, especially at neutrals, at the cost of increased computational time. In addition, it is observed that measuring phase-to-ground voltages is just as important as measuring phase-to-neutral voltages, while the DSSE model under rectangular coordinates presents similar accuracy as, but shorter computational time than, the one under polar coordinates.

REFERENCES

- [1] A. Primadianto and C.N. Lu, "A review on distribution system state estimation," *IEEE Trans. Power Syst.*, vol. 32, no. 5, pp. 3875-3883, Sept. 2017.
- [2] G.T. Heydt, "The next generation of power distribution systems," *IEEE Trans. on Smart Grid*, vol. 1, no. 3, pp. 225-235, Dec. 2010.
- [3] G.T. Heydt, B.H. Chowdhury, M.L. Crow, D. Haughton, B. Kiefer, F. Meng, and B.R. Sathyanarayana, "Pricing and control in the next generation power distribution system," *IEEE Trans. on Smart Grid*, vol. 3, no. 2, pp. 907-914, Jun. 2012.
- [4] Y. Liu, J. Li, L. Wu, and Q. Liu, "Ex-post real-time distribution LMP based on state estimation," in *IEEE PES General Meeting*, Boston, MA, USA, 2016.
- [5] C. Lu, J. Teng, and W. Liu, "Distribution system state estimation," *IEEE Trans. Power Syst.*, vol. 10, no. 1, pp. 229-240, Feb. 1995.
- [6] K. Li, "State estimation for power distribution system and measurement impacts," *IEEE Trans. Power Syst.*, vol. 11, no. 2, pp. 911-916, May 1996.
- [7] J. Wu, Y. He, and N. Jenkins, "A robust state estimator for medium voltage distribution networks," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1008-1016, May 2012.
- [8] M.E. Baran and A.W. Kelley, "A branch current based state estimation

- method for distribution systems," *IEEE Trans. Power Syst.*, vol. 10, no. 1, pp. 483-491, Feb. 1995.
- [9] H. Wang and N.N. Schulz, "A revised branch current-based distribution system state estimation algorithm and meter placement impact," *IEEE Trans. Power Syst.*, vol. 19, no. 1, pp. 207-213, Feb. 2004.
 - [10] I. Dzaifc, R.A. Jabr, I. Huseinagic, and B.C. Pal, "Multi-phase state estimation featuring industrial-grade distribution network models," *IEEE Trans. on Smart Grid*, vol. 8, no. 2, pp. 609-618, Mar. 2017.
 - [11] S. Nanchian, A. Majumdar, and B.C. Pal, "Three-phase state estimation using hybrid particle swarm optimization," *IEEE Trans. on Smart Grid*, vol. 8, no. 3, pp. 60-68, May 2017.
 - [12] M.C. Almeida and L.F. Ochoa, "An improved three-phase AMB distribution system state estimator," *IEEE Trans. on Power Syst.*, vol. 32, no. 2, pp. 1463-1473, Mar. 2017.
 - [13] S. Nanchian, A. Majumdar, and B.C. Pal, "Ordinal optimization technique for three-phase distribution network state estimation including discrete variables," *IEEE Trans. Sustainable Energy*, vol. 8, no. 4, pp. 1528-1535, Oct. 2017.
 - [14] M. Pau, F. Ponci, A. Monti, S. Sulis, C. Muscas, and P.A. Pegoraro, "An efficient and accurate solution for distribution system state estimation with multiarea architecture" *IEEE Trans. Instrum. Meas.*, vol. 66, no. 5, pp. 910-919, May 2017.
 - [15] W. Kersting, *Distribution System Modeling and Analysis*. CRC Press LLC, Boca Raton, FL, USA, 2002.
 - [16] D.R.R. Penido, L.R. Araujo, S. Carneiro, J.L.R. Pereira, and P.A.N. Garcia, "Three-phase power flow based on four-conductor current injection method for unbalanced distribution networks," *IEEE Trans. on Power Syst.*, vol. 23, no. 2, pp. 494-503, May 2008.
 - [17] R.M. Ciric, A.P. Feltrin, and L.F. Ochoa, "Power flow in four-wire distribution networks-general approach," *IEEE Trans. on Power Syst.*, vol. 18, no. 4, pp. 1283-1290, Nov. 2003.
 - [18] A.K. Ghosh, D.L. Lubkeman, and R.H. Jones, "Load modeling for distribution circuit state estimation," *IEEE Trans. Power Delivery*, vol. 12, no. 2, pp. 999-1005, Apr. 1997.
 - [19] J. Liu, J. Tang, F. Ponci, A. Monti, C. Muscas, and P.A. Pegoraro, "Trade-offs in PMU deployment for state estimation in active distribution grids," *IEEE Trans. on Smart Grid*, vol. 3, no. 2, pp. 915-924, Jun. 2012.
 - [20] A. Gjelsvik, S. Aam, and L. Holten, "Hachtel's augmented matrix method--A rapid method improving numerical stability in power system static state estimation," *IEEE Trans. Power Appar. Syst.*, vol. 104, no. 11, pp. 2987-2993, Nov. 1985.
 - [21] Y. Guo, W. Wu, B. Zhang, and H. Sun, "An efficient state estimation algorithm considering zero injection constraints," *IEEE Trans. on Power Syst.*, vol. 28, no. 3, pp. 2651-2659, Aug. 2013.
 - [22] N. Uribe-Perez, L. Hernández, D. Vega, and I. Angulo, "State of the art and trends review of smart metering in electricity grids," *Appl. Sci.*, vol. 6, no. 68, pp. 1-24, Feb. 2016.
 - [23] K. Samarakoon, J. Wu, J. Ekanayake, and N. Jenkins, "Use of delayed smart meter measurements for distribution state estimation," in *IEEE PES General Meeting*, Detroit, MI, USA, 2011.
 - [24] X. Feng, F. Yang, and W. Peterson, "A practical multi-phase distribution state estimation solution incorporating smart meter and sensor data," in *IEEE PES General Meeting*, San Diego, CA, USA, 2012.
 - [25] S. Huang, C. Lu, and Y. Lo, "Evaluation of AMI and SCADA data synergy for distribution feeder modeling," *IEEE Trans. on Smart Grid*, vol. 6, no. 4, pp. 1639-1647, Jul. 2015.
 - [26] E. Manitsas, R. Singh, B. Pal, and G. Strbac, "Distribution system state estimation using an artificial neural network approach for pseudo measurement modeling," *IEEE Trans. Power Syst.*, vol. 27, no. 4, Nov. 2012.
 - [27] A. Angioni, T. Schlösser, F. Ponci, and A. Monti, "Impact of pseudo-measurements from new power profiles on state estimation in low-voltage grids," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 1, pp. 70-77, Jan. 2016.
 - [28] B.P. Hayes, J.K. Gruber, and M. Prodanovic, "A closed-loop state estimation tool for MV network monitoring and operation," *IEEE Trans. on Smart Grid*, vol. 6, no. 4, pp. 2116-2125, Jul. 2015.
 - [29] A. Angioni, M. Pau, F. Ponci, A. Monti, C. Muscas, S. Sulis, and P.A. Pegoraro, "Bayesian distribution system state estimation in presence of non-gaussian pseudo-measurements," in *IEEE International Workshop on Applied Measurements for Power Systems (AMPS)*, Aachen, Germany, Sept. 2016.
 - [30] Q. Chen, D. Kaleshi, Z. Fan, and S. Armour, "Impact of smart metering data aggregation on distribution system state estimation," *IEEE Trans. on Ind. Informat.*, vol. 12, no. 4, Aug. 2016.
 - [31] J. Kang, L. Xie, and D. Choi, "Impact of data quality in home energy management system on distribution system state estimation," *IEEE Access*, Digital Object Identifier 10.1109/ACCESS.2018.2804380, Mar. 2018.
 - [32] Q. Chen, D. Kaleshi, S. Armour, and Z. Fan, "Reconsidering the smart metering data collection frequency for distribution state estimation," in *IEEE International Conf. on SmartGridComm*, Venice, Italy, 2014.
 - [33] H. Li, R. Mao, L. Lai, and R. Qiu, "Compressed meter reading for delay-sensitive and secure load report in smart grid," in *Proc. 1st IEEE SmartGridComm*, pp. 114-119, Oct. 2010.
 - [34] EPRI Distribution Modernization Demonstration (DMD) Data Mining Initiative, [Online]. Available: <http://smartgrid.epri.com/DMD-DMI.aspx>, accessed: Aug. 10, 2018.
 - [35] EPRI Data Analytics Case: Development of Electrical Load Model Utilizing SCADA and AMI Data, [Online]. Available: <http://smartgrid.epri.com/doc/DMD-Use-Cases/Development%20of%20Electrical%20Load%20Model%20Utilizing%20SCADA%20and%20AMI%20Data%20-%20Short%20Version.pdf>, accessed: Aug. 10, 2018.
 - [36] J.C. Balda, A.R. Oliva, D.W. McNabb, and R.D. Richardson, "Measurements of neutral currents and voltages on a distribution feeder," *IEEE Trans. Power Delivery*, vol. 12, no. 4, pp. 1799-1804, Oct. 1997.
 - [37] T. Chen and W. Yang, "Analysis of multi-grounded four-wire distribution systems considering the neutral grounding," *IEEE Trans. Power Delivery*, vol. 16, no. 4, pp. 710-717, Oct. 2001.
 - [38] S. Bose, D. Gayme, and S. Low, "Optimal power flow over tree networks," *49th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, USA, Sep. 2011.
 - [39] M. Pau, P.A. Pegoraro, and S. Sulis, "Efficient branch-current-based distribution system state estimation including synchronized measurements," *IEEE Trans. Instrum. Meas.*, vol. 62, no. 9, pp. 2419-2429, Sept. 2013.
 - [40] H. Wang, W. Zhang, and Y. Liu, "A robust measurement placement method for active distribution system state estimation considering network reconfiguration," *IEEE Trans. Smart Grid*, vol. 9, no. 3, pp. 2108-2117, May 2018.
 - [41] A.S. Turkmen, "Robust partial least squares for regression and classification," Ph.D Dissertation, Dept. of Mathematics and Statistics, Auburn University, 2008.
 - [42] P. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Pearson, London, United Kingdom, 2005.
 - [43] N. Akarachantachote, S. Chadcham, and K. Saithanu, "Cutoff threshold of variable importance in projection for variable selection," *Int. J. of Pure and Appl. Mathematics*, vol. 94, no. 3, pp. 307-322, 2014.
 - [44] L. Zhang and S. Garcia-Munoz, "A comparison of different methods to estimate prediction uncertainty using Partial Least Squares (PLS): A practitioner's perspective," *Chemom. Intell. Lab. Syst.*, vol. 97, pp. 152-158, 2009.
 - [45] Smart Metering Electricity Customer Behavior Trials (CBTs), *Commission for Energy Regulation (CER)*, [Online]. Available: <http://www.ucd.ie/issda/data/commissionforenergyregulationcer/>, accessed: Aug. 10, 2018.
 - [46] Y. Wang and L. Wu, "On practical challenges of decomposition-based hybrid forecasting algorithms for wind speed and solar irradiation," *Energy*, vol. 112, 1, pp. 208-220, Oct. 2016.
 - [47] L. Wu and M. Shahidehpour, "A hybrid model for integrated day-ahead electricity price and load forecasting in smart grid," *IET Generation Transmission and Distribution*, vol. 8, no. 12, pp. 1937-1950, 2014.
 - [48] C. Chang and C. Lin, LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1-27:27, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>, accessed: Aug. 10, 2018.
 - [49] Test Data of the modified IEEE 123-bus system for DSSE, [Online]. Available: <http://people.clarkson.edu/~lwu/data/DSSE/>, accessed: Aug. 10, 2018.

BIOGRAPHIES

Yikui Liu (S'15) received the B.S. degree in electrical engineering and automation from Nanjing Institute of Technology, China, in 2012 and the M.S. degree in power system and automation from Sichuan University, China, in 2015.

He is currently pursuing the Ph.D. degree with Electrical and Computer Engineering Department, Clarkson University, USA. His current research interests are power market and OPF in distribution system.

Jie Li (M'14) received the B.S. degree in information engineering from Xi'an Jiaotong University in 2003 and the M.S. degree in system engineering from Xi'an Jiaotong University, China in 2006, and the Ph.D. degree from the Illinois Institute of Technology (IIT), Chicago, in 2012. From 2006 to 2008, she was a Research Engineer with IBM China Research Lab. From 2012 to 2013, she was a Power System Application Engineer with GE Energy Consulting.

Presently, she is an Assistant Professor in the Electrical and Computer Engineering Department at Clarkson University. Her research interests include power systems restructuring and bidding strategy.

Lei Wu (SM'13) received the B.S. degree in electrical engineering and the M.S. degree in systems engineering from Xi'an Jiaotong University, Xi'an, China, in 2001 and 2004, respectively, and the Ph.D. degree in electrical engineering from Illinois Institute of Technology (IIT), Chicago, IL, USA, in 2008. From 2008 to 2010, he was a Senior Research Associate with the Robert W. Galvin Center for Electricity Innovation, IIT. He worked as summer Visiting Faculty at NYISO in 2012.

Currently, he is a Professor with the Electrical and Computer Engineering Department, Clarkson University, Potsdam, NY, USA. His research interests include power systems operation and planning, energy economics, and community resilience microgrid.