Bayesian Learning-Based Harmonic State Estimation in Distribution Systems with Smart Meter and DPMU Data

Wei Zhou, Omid Ardakanian, *Member, IEEE*, Hai-Tao Zhang†, *Senior Member, IEEE*, Ye Yuan, *Member, IEEE*

***Abstract*—This paper studies the problem of locating harmonic sources and estimating the distribution of harmonic voltages in unbalanced three-phase power distribution systems. We develop an approach for harmonic state estimation utilizing two types of measurements from smart meters and distribution-level phasor measurement units (DPMUs). It involves regression analysis for power flow calculation, prediction of demands using recurrent neural networks, and sparse Bayesian learning for state estimation. The proposed approach requires fewer DPMUs than nodes, making it more applicable to existing distribution grids. We show the effectiveness of the proposed estimator through extensive numerical simulations on an IEEE test feeder. We also investigate how the increased penetration of distributed energy resources could affect the performance of our state estimator.**

***Index Terms*—Power system harmonics, harmonic state estimation, load forecasting, supervised learning, power distribution.**

I. Introduction

THE growing adoption of power electronic devices and distributed energy resources (DERs) has exacerbated harmonic-related power quality issues in the distribution system. Harmonics introduce distortion in the fundamental waveform of voltages and currents, which may damage electric devices and interfere with control and protection equipment [1]. Given the adverse consequences of harmonics, it is imperative to monitor harmonic generation and propagation in real-time. Thus, an advanced distribution system monitoring would rely on Harmonic State Estimation (HSE) to locate harmonic sources and estimate the harmonic voltage distribution.

There is an abundant literature addressing HSE for power transmission grids using various techniques, such as least squares estimation (LSE) [2], singular value decomposition (SVD) [3], evolutionary strategies [4], Lasso regression [5], and sparse Bayesian learning (SBL) [6]. HSE was first solved using harmonic load flow and LSE [7]. An adaptive Kalman estimator [8] is proposed later to track harmonics without the exact knowledge of the noise covariance matrix. Ref. [9] addresses the uncertainty of network parameters by solving a parametric interval linear system of equations. This approach cannot cope with the situation where DPMUs are sparse, hence it cannot be presently applied to distribution grids. Ref. [10] utilizes a Newton method based on numerical differentiation to calculate a fast periodic steady state in time domain. This is combined with Kalman filter for performing HSE. In recent work [11], HSE in time domain is formulated for three-phase unbalanced systems exploiting the half-wave symmetry property. Despite the well-established tradition of solving HSE at the transmission level, this problem was not studied at the distribution level until recently. This is primarily because measurements used to be very scarce and often nonexistent beyond the distribution substation.

Distribution-level phasor measurement units (DPMUs) have been a "game changer". The availability of high-frequency harmonic measurements from multiple locations across the distribution grid has made it possible to study HSE at the distribution level in recent years [12]–[18]. For example, the weighted least squares (WLS) method is used in [14] to identify the type of harmonic sources, leveraging the fact that different types of loads have distinct angle distributions of the fifth harmonic. In [17], the Metropolis-Hastings approach is utilized to compute the posterior distribution of harmonic sources, improving the estimation performance. There are usually limited measurements in distribution networks, making the system not fully observable. Thus, a manifold of states correspond to the same measurement [19]. To address this issue, a SVD-based method [16] is adopted to estimate harmonics in unbalanced systems. More recently, deep learning is employed to learn the mapping between the states and measurements using training data sampled from the load distribution [19], [20]. The optimal meter placement is explored in [18] using branch-and-bound and genetic algorithms. A Dirichlet process mixture model [21], [22] is proposed to capture the uncertainty of load power and wind generation. Refs. [23]–[25] deal with the problem of detecting and correcting gross errors commonly encountered in field measurements. To enhance the accuracy of the state estimator, bad data identification and correction techniques [23] must be applied before carrying out HSE.

Despite these attempts to address the HSE problem at the distribution level, related work has one fundamental shortcoming which renders it of limited practical value. It assumes

This work is partly supported by the National Natural Science Foundation of China under Grants U1713203, 51721092, 61703172, 51729501 and 61673189, and in part by the 111 Project on Computational Intelligence and Intelligent Control, China under Grant B18024. As a part of the University of Albertas Future Energy Systems research initiative, this research was made possible in part thanks to funding from the Canada First Research Excellence Fund. (Corresponding author: Hai-Tao Zhang).

W. Zhou, H.-T. Zhang and Y. Yuan are with the School of Artificial Intelligence and Automation, the Key Laboratory of Image Processing and Intelligent Control and the State Key Laboratory of Digital Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan, 430074, China. (email: zhouwei@hust.edu.cn, zht@mail.hust.edu.cn, yye@hust.edu.cn.)

O. Ardakanian is with the Department of Computing Science, University of Alberta, Canada. (email: ardakanian@ualberta.ca)

the knowledge of the *measurement matrix*, which describes the relationship between measurements and state variables. This matrix is unknown in real-world scenarios because it is difficult to track the real-time operational structure of the distribution network, and to estimate aggregated demands at load buses in the primary distribution network. This paper addresses this shortcoming by learning the measurement matrix from smart metering data, thereby improving the accuracy of HSE in unbalanced three-phase distribution networks. In particular, we utilize supervised learning techniques to (a) predict aggregate demands at nodes in the primary network, and (b) learn the relationship between power flow in the primary network and demands of downstream customers measured by smart meters. The measurement matrix is updated with these *pseudo measurements*, and the state estimation is then performed using the high-sample-rate DPMU measurements.

We postulate that harmonic sources are *sparse* in a distribution network as only a small number of devices produce significant harmonics simultaneously. For distribution networks that are not fully observable, we propose an SBL-based harmonic state estimator by drawing on the idea of sparse Bayesian learning [26], which is capable of locating harmonic sources with sufficiently high accuracy. The contribution of this paper is threefold:

- We propose a data-driven approach to HSE which copes with the unknown measurement matrix leveraging data from smart meters.
- We propose an SBL-based estimator for networks that are not fully observable to locate the harmonic sources, and to estimate the voltages using considerably fewer DPMUs than distribution nodes.
- We show through extensive simulations that a photovoltaic system (PV) connected to the primary distribution network does not negatively affect the performance of the proposed state estimator.

We assume that DPMUs installed in the network are capable of providing synchronized harmonic phasors, which are obtained from voltage or current waveforms and precisely referenced to a common time base. Harmonic synchrophasor measurement techniques have been proven useful in improving power quality monitoring and state estimation in power grids [27]–[33]. In recent years, extensive research has been devoted to estimation methods [29], [33], measuring instruments [27], [28] and applications [34] of harmonic synchrophasors. For example, a harmonic phasor measurement unit (PMU), which includes a digital signal processor (DSP) and a time synchronization unit, is designed in [27] for wide-area distribution systems. Furthermore, a prototype PMU based on PXI modular hardware [28] could provide precise harmonic voltage and current phasors.

This paper is organized as follows. Section II presents harmonic models of the main distribution components and formulates the HSE problem. Section III outlines the proposed methodology. Section IV describes the evaluation metrics and simulation scenarios. Section V summarizes the simulation results. Finally, the conclusion is drawn in Section VI.

II. Problem Formulation

In the following, we denote the set of real numbers (or matrices) by $\mathbb{R}$, the set of complex numbers (or matrices) by $\mathbb{C}$, the set of symmetric complex matrices by $\mathbb{S}$, transpose and inverse of a matrix A by $A^\top$ and A^{-1} respectively, and the element in its ith row and jth column by A_{ij}. The magnitude of a complex number u is denoted by $|u|$, and the cardinality of a set $\mathcal{S}$ is also denoted by $|\mathcal{S}|$. Function $\mathrm{diag}(\cdot)$ arranges the elements in a vector into a diagonal block matrix, $\mathrm{cond}(\cdot)$ returns the condition number of a matrix, and $\odot$ denotes the Hadamard product.

Let n_b and n_e respectively denote the number of buses and branches in the distribution network, T_{rd} and T_{rm} denote the reporting period of the DPMU and the smart meter, respectively, and T_{sm} denote the sampling period of the smart meter. Let $\{tT_{rd}|t=1,\cdots,k\}$ be the times at which DPMUs report voltage and current phasor measurements, and $\{tT_{sm}|t=1,\cdots,k\}$ and $\{tT_{rm}|t=1,\cdots,k\}$ be the times at which smart meters sample and report voltage magnitude and real power consumption of customers, respectively. Let $L(i)$ be the index of the load bus or distribution line monitored by the i^{th} DPMU, and $N(i)$ be the index of the end node monitored by the i^{th} smart meter.

A. Preliminaries

A distribution network can be described by an undirected graph $\mathcal{G}(\mathcal{V},\mathcal{E})$, where $\mathcal{V}=\{1,\cdots,n_b\}$ is the set of buses and $\mathcal{E}\subseteq\mathcal{V}\times\mathcal{V}$ ($|\mathcal{E}|=n_e$) is the set of lines, each connecting two distinct buses. Let $\mathcal{P}_i\subseteq\{a,b,c\}$ be the set of phase indices of a bus i, $\mathcal{P}_{(i,j)}\subseteq\{a,b,c\}$ be the phase indices of a line (i,j), and $y_{ij}(h)\in\mathbb{S}^{|\mathcal{P}_{ij}|\times|\mathcal{P}_{ij}|}$ be the admittance matrix of this line at harmonic order h. The line admittance matrix $Y^{\mathrm{line}}(h)\in\mathbb{S}^{\sum_{(i,j)\in\mathcal{E}}|\mathcal{P}_{(i,j)}|\times\sum_{(i,j)\in\mathcal{E}}|\mathcal{P}_{(i,j)}|}$ consists of the admittance matrices of all lines at harmonic order h:

$$Y^{\mathrm{line}}(h)=\mathrm{diag}\big(\{y_{ij}(h)\}_{(i,j)\in\mathcal{E}}\big).$$

We define two other matrices: the *incidence matrix* denoted by $A\in\mathbb{R}^{\sum_{(i,j)\in\mathcal{E}}|\mathcal{P}_{(i,j)}|\times\sum_{i\in\mathcal{V}}|\mathcal{P}_i|}$ and the branch admittance matrix denoted by $Y^{\mathrm{B}}(h)\in\mathbb{C}^{\sum_{(i,j)\in\mathcal{E}}|\mathcal{P}_{(i,j)}|\times\sum_{i\in\mathcal{V}}|\mathcal{P}_i|}$. The branch admittance matrix relates line currents to nodal voltages and satisfies the following equation $Y^{\mathrm{B}}(h)=Y^{\mathrm{line}}(h)A$. Let the *harmonic admittance matrix* of the network at order h be denoted by $Y^{\mathrm{H}}(h)\in\mathbb{S}^{\sum_{i\in\mathcal{V}}|\mathcal{P}_i|\times\sum_{i\in\mathcal{V}}|\mathcal{P}_i|}$. It can be written as

$$Y^{\mathrm{H}}_{ij}(h)=\begin{cases}-y_{ij}(h), & (i,j)\in\mathcal{E},\\ \sum_{(i,k)\in\mathcal{E}}y_{ik}(h)+y^{sh}_i(h)+y_{\ell_i}(h), & i=j,\\ 0, & \text{otherwise},\end{cases}$$

where y^{sh}_i and y_{ℓ_i} are the admittance matrices of the shunt branch and the load connected to bus i, respectively.

B. Modelling Network Components

We now describe how network components are represented for harmonic analysis. A typical distribution network can be divided into a substation, and primary and secondary distribution circuits which are comprised of multiple distribution

transformers, voltage regulators, lines, capacitor banks, loads, and DERs. We focus on HSE in the primary network and model the secondary networks as equivalent Norton circuits connected to primary nodes as in [35]. We use time-invariant harmonic models to represent all distribution components except for the loads. Specifically, lines are represented using multiphase coupled equivalent π circuits [36], [37], and transformers are represented using the constant short-circuit impedance model [37]. We adopt the short-circuit impedance model for voltage regulators, and the Thevenin model of supply sources (e.g., the substation).

Since residential PV systems that are connected to low-voltage networks have negligible harmonic impacts on a medium- or high-voltage system [38], we only investigate the impact of harmonics generated by large PV systems which are directly connected to nodes in the primary distribution network. The PV systems are represented using the Norton model with a specified spectrum for harmonic analysis.

Loads are equivalent to secondary distribution circuits, specified by the consumption of active and reactive powers. It is suggested in [39] that a parallel branch of inductive reactance and resistance can be used to model the aggregate load without accurate measurements and detailed load compositions. Predominantly passive loads are represented by a series R-L impedance, and predominantly motive loads are represented using their resistive-inductive equivalents. Capacitive effects are often neglected in utility distribution systems and industrial power grids [40]. Hence, the time-varying load model in OPENDSS [41] (shown in Fig. 1) which only includes inductive reactance and resistance, is used for harmonic analysis[1]. In this model, the current source with a specified spectrum represents harmonics produced by its nonlinear part. The c parameter represents the ratio of the consumed power of the series R-L branch to the total power demand. It is determined based on the actual load composition. The admittance of a load, denoted by $y_\ell(h)$, can be computed as described below:

$$y_\ell(h) = y_\ell^p(h) + y_\ell^s(h), \tag{1}$$

$$y_\ell^p(h) = \frac{(1-c)P}{V_n^2} - j\frac{(1-c)Q}{hV_n^2}, \tag{2}$$

$$y_\ell^s(h) = \left[\frac{V_n^2 P}{c(P^2+Q^2)} + j\frac{hV_n^2 Q}{c(P^2+Q^2)}\right]^{-1}, \tag{3}$$

where V_n is the nominal voltage, and $y^p(h)$ and $y^s(h)$ denote, respectively, the admittance of parallel and series branches at harmonic order h. These two parameters depend on c and the real and reactive power consumption denoted by P, Q (refer to [42], [43] for more details).

C. Available Measurements

Measurements in distribution networks are mainly from smart meters installed at customer premises along with a small number of DPMUs sampling voltage and current phasors at high frequency in several locations across the network [44]. We assume DPMUs are installed on specific lines in the primary distribution network. These sensors report harmonic voltage and current phasors every T_{rd} seconds[2]. We denote the harmonic phasors measured by the ith DPMU located at $L(i)$ at time t by:

$$\mathcal{D}_i^t = \{V_{L(i),\mathrm{p}}^t(h),\, I_{L(i),\mathrm{p}}^t(h)\}_{\mathrm{p}\in\mathcal{P}_{L(i)},h\in\{1,\cdots,h_m\}},$$

where h_m is the highest harmonic order that can be practically measured, and $V_{L(i),\mathrm{p}}^t(h)$ and $I_{L(i),\mathrm{p}}^t(h)$ are the harmonic measurements of voltage and current phasors, respectively.

We assume that smart meters are installed throughout the secondary distribution network, measuring the magnitude of service voltage and the average active power consumption of Ω customers at T_{sm}-minute intervals. At time t, the smart meter installed at the premise of customer i measures the voltage magnitude $V_{N(i)}^t = \{V_{N(i),\mathrm{p}}^t\}_{\mathrm{p}\in\mathcal{P}_\mathrm{i}}$, and active power consumption $P_{N(i)}^t = \{P_{N(i),\mathrm{p}}^t\}_{\mathrm{p}\in\mathcal{P}_\mathrm{i}}$. The smart metering data is buffered locally and sent to the utility's data center every T_{rm} minutes.

Suppose $T_{sm} = \beta T_{rd}$. To simplify the notation, we use the timescale of DPMUs as time reference in the following. Hence, the time series data collected by all smart meters at the end nodes would be

$$\mathcal{M}^{1\to t_m} = \{P_N^{\beta t}, V_N^{\beta t}\}_{t\in\{1,\cdots,t_m\}},$$

where $P_N^{\beta t} = [P_{N(1)}^{\beta t},\cdots,P_{N(\Omega)}^{\beta t}]$, $V_N^{\beta t} = [V_{N(1)}^{\beta t},\cdots,V_{N(\Omega)}^{\beta t}]$.

D. HSE Formulation

We formulate the *static HSE problem* for each harmonic order based on the harmonic load flow equations [6], [46] to estimate state variables, x, from DPMU measurements, z, given the measurement noise, ξ:

$$z(h) = \Phi(h)x(h) + \xi, \tag{4}$$

with

$$z(h) = \begin{bmatrix} V(h) \\ I(h) \end{bmatrix}, \quad \Phi(h) = \begin{bmatrix} S_1[Y^{\mathrm{H}}(h)]^{-1} \\ S_2 Y^{\mathrm{B}}(h)[Y^{\mathrm{H}}(h)]^{-1} \end{bmatrix}. \tag{5}$$

Here $x(h) \in \mathbb{C}^{\mathrm{n}\times 1}$ is the vector of *state variables* expressing the injected currents of order h by harmonic sources, and $z(h) \in \mathbb{C}^{2\mathrm{m}\times 1}$ ($\mathrm{m} = \sum_{i=1}^{d} |\mathcal{P}_{L(i)}|$) is a vector that collects line-to-neutral nodal voltages, denoted by $V(h) \in \mathbb{C}^{\mathrm{m}\times 1}$, and

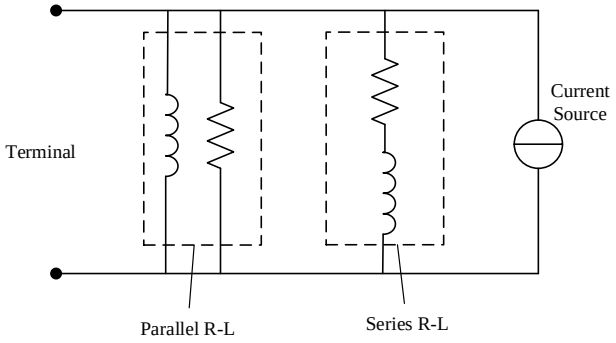

Fig. 1. The load model for harmonic analysis.

[1]We note that this model cannot satisfactorily describe feeders that are primarily capacitive.

[2]The sampling frequency of DPMUs is typically 512 samples per cycle [45].

line currents, denoted by $I(h) \in \mathbb{C}^{\mathrm{m}\times 1}$ which are measured by d distinct DPMUs. We define $S_1 \in \mathbb{R}^{\mathrm{m}\times\sum_{i\in\mathcal{V}}|\mathcal{P}_i|}$ and $S_2 \in \mathbb{R}^{\mathrm{m}\times\sum_{(i,j)\in\mathcal{E}}|\mathcal{P}_{(i,j)}|}$ matrices to encode the locations of DPMUs:

$$S_{1ij} = \begin{cases} 1, & j = L(i) \in \mathcal{V}, \\ 0, & \text{otherwise}, \end{cases} \quad S_{2ij} = \begin{cases} 1, & j = L(i) \in \mathcal{E}, \\ 0, & \text{otherwise}. \end{cases}$$

Note that Eq. (4) is underdetermined for distribution networks that are not fully observable.

Assume that measurements, $z(h)$, are cleaned through bad data identification and correction techniques. The $x(h)$ uniquely determines the state of the distribution system, and the harmonic voltage distribution can be simply computed based on the harmonic load flow given the estimated states, $\hat{x}$. The *measurement matrix*, denoted by $\Phi(h)$, depends on the inverse of the harmonic admittance matrix, $[Y^{\mathrm{H}}(h)]^{-1}$, and the branch admittance matrix, $Y^{\mathrm{B}}(h)$. The harmonic admittance matrix $Y^{\mathrm{H}}(h)$ changes over time similar to the loads and the operational structure of the distribution network (see Section II-B)[3]. Hence, the real-time measurement matrix is unknown to the distribution system operator and must be estimated using the available data.

III. Methodology

Our approach involves several steps which are described here. It starts with a preprocessing step in which we reduce the number of state variables by merging the nodes that appear on two sides of short lines and closed switches. This is because their voltages are almost identical and highly correlated. This is done using the two-port network theorem (refer to Chapter 8 of [48]). The subsequent steps are explained below.

A. Regression Analysis for Power Flow Calculation

The next step is to compute real and reactive power flow at primary nodes given measurements of smart meters installed at the end nodes. It would be simply done through power flow analysis if the model of the secondary distribution network was known. However, this model is often unavailable [49]. Thus, we use regression analysis to identify the relationship between the power flow at primary nodes and the demands of downstream customers measured by smart meters. This requires measuring the real and reactive power at each primary node for a relatively short period of time. We assume that this data is available[4]. The aggregate demand data at node i in the primary network, available for βt_m time slots, is:

$$\mathcal{L}^{1\to t_m} = \{P^{\beta t}_{i,\mathrm{p}}, Q^{\beta t}_{i,\mathrm{p}}\}_{t\in\{1,\cdots,t_m\},\mathrm{p}\in\mathcal{P}_i}.$$

Motivated by the success of linearized power flow formulations in radial distribution networks, we restrict ourselves to a linear model for secondary distribution networks. The features we use are the real power, the voltage magnitude, and the squared voltage magnitude measured by smart meters at the end nodes. We develop the following model:

$$\Upsilon = \Psi\Theta, \tag{6}$$

where

$$\Upsilon = \begin{bmatrix} P^{\beta} & Q^{\beta} \\ \vdots & \vdots \\ P^{\beta t_m} & Q^{\beta t_m} \end{bmatrix}, \Psi = \begin{bmatrix} P^{\beta}_N & V^{\beta}_N & (V^{\beta}_N)^2 \\ \vdots & \vdots & \vdots \\ P^{\beta t_m}_N & V^{\beta t_m}_N & (V^{\beta t_m}_N)^2 \end{bmatrix},$$

and Θ is the coefficient matrix.

We utilize the LS method based on QR decomposition to learn Θ given $\mathcal{L}^{1\to k_s}$ and $\mathcal{M}^{1\to k_s}$:

$$\Psi = UR, \tag{7}$$

$$\Theta = R^{-1}U^{\top}\Upsilon, \tag{8}$$

where U is a unitary matrix and R is an upper triangular matrix. We denote the estimated real and reactive powers in the primary network by $\{\widetilde{P}^t, \widetilde{Q}^t\}$. To evaluate this model, we perform three-fold cross validation.

B. Demand Prediction

Another challenge is that the reporting period of smart meters, T_{rm}, is longer than their sampling period, T_{sm}, e.g., $60\,\mathrm{min}$ versus $1\,\mathrm{min}$. This causes the harmonic admittance matrix to update every T_{rm} time slots instead of every T_{sm} time slots. To address this problem, we predict the aggregate power consumption at each node in the primary network between two successive reporting intervals using the historical smart meter data of downstream nodes. Suppose the relationship among power consumption follows Eqs. (9) and (10),

$$\widehat{P}^{\beta t}_i = f_1\Big(\widetilde{P}^{\beta t-\alpha_1}_i, \widetilde{Q}^{\beta t-\alpha_1}_i, \cdots, \widetilde{P}^{\beta t-\alpha_s}_i, \widetilde{Q}^{\beta t-\alpha_s}_i\Big), \tag{9}$$

$$\widehat{Q}^{\beta t}_i = f_2\Big(\widetilde{P}^{\beta t-\alpha_1}_i, \widetilde{Q}^{\beta t-\alpha_1}_i, \cdots, \widetilde{P}^{\beta t-\alpha_s}_i, \widetilde{Q}^{\beta t-\alpha_s}_i\Big), \tag{10}$$

where $\alpha \in \mathbb{Z}^s$ is the time-lag vector.

We learn these models using a recurrent neural network model, namely the long short-term memory (LSTM) network [50] which has superior performance for time series prediction. We build an LSTM network with four layers (an input layer, an LSTM layer, a hidden fully connected layer, and an output layer) to predict the power consumption. The time-lag term α is determined by the auto-correlation coefficients of the time series $\{\widetilde{P}^t, \widetilde{Q}^t\}$ obtained from regression analysis. The input of the LSTM network is

$$\mathcal{I}_t := [\widetilde{P}^{\beta t-\alpha_1}_i, \widetilde{Q}^{\beta t-\alpha_1}_i, \cdots, \widetilde{P}^{\beta t-\alpha_s}_i, \widetilde{Q}^{\beta t-\alpha_s}_i]^{\top},$$

and its output is, $\mathcal{O}_t = [\widehat{P}^t_i, \widehat{Q}^t_i]$, which is updated as follows:

$$i_t = \sigma(W_i\mathcal{I}_t + D_i\mathcal{O}_{t-1} + b_i), \tag{11}$$

$$f_t = \sigma(W_f\mathcal{I}_t + D_f\mathcal{O}_{t-1} + b_f), \tag{12}$$

$$o_t = \sigma(W_o\mathcal{I}_t + D_o\mathcal{O}_{t-1} + b_o), \tag{13}$$

$$g_t = \tanh(W_g\mathcal{I}_t + D_g\mathcal{O}_{t-1} + b_g), \tag{14}$$

$$\mathcal{C}_t = f_t \odot \mathcal{C}_{t-1} + i_t \odot g_t, \tag{15}$$

$$\mathcal{O}_t = o_t \odot \tanh(\mathcal{C}_t), \tag{16}$$

[3]A topology identification algorithm must be implemented to detect how the operational structure has changed before addressing the HSE problem, e.g., the online detection approach in [47].

[4]The distribution system operator may obtain this data from Distribution SCADA (D-SCADA) system or other advanced monitoring systems.

where $\sigma(\cdot)$ is the sigmoid function, and W, D and b respectively represent input weights, recurrent weights, and bias terms. We denote the layer input vector by g_t and the cell states by $\mathcal{C}_t$. Also, i_t, f_t, o_t denote the activation vectors of input, forget, and output gates, respectively.

We adopt the ADAM method [51] to train the LSTM network, obtaining the weight matrices W, D and the bias vector b. The neural units of the hidden layer are set to $n_h = 0.8 \times |\mathcal{I}|$, and the mean-squared-error of predicted values is considered as the cost function. The training data set contains many fixed length observations which are randomly sampled from the original time series.

C. Estimating the Measurement Matrix

Since the time interval between successive DPMU measurements is shorter than that of smart meter data, we use a simple interpolation [52] to obtain the missing values:

$$\widehat{P}^{t\beta+i} = \widehat{P}^{t\beta}, \qquad \forall i = 1, \cdots, \beta - 1, \tag{17}$$

$$\widehat{Q}^{t\beta+i} = \widehat{Q}^{t\beta}, \qquad \forall i = 1, \cdots, \beta - 1. \tag{18}$$

When the load compositions are determined, we can compute the admittance matrix in real-time as follows:

$$\begin{aligned} \widehat{Y}^{\mathrm{H}^t}(h) &= Y^{\mathrm{H}^0}(h) + \Delta Y^{0\to t}(h), \\ \Delta Y^{0\to t}(h) &= f(\widehat{P}^t, \widehat{Q}^t, \widetilde{P}^0, \widetilde{Q}^0), \\ \widehat{P}^t &= [\widehat{P}^t_1, \cdots, \widehat{P}^t_{n_b}]^\top, \\ \widehat{Q}^t &= [\widehat{Q}^t_1, \cdots, \widehat{Q}^t_{n_b}]^\top, \end{aligned} \tag{19}$$

where $\Delta Y^{0\to t}(h) \in \mathbb{C}^{\sum_{i\in\mathcal{V}}|\mathcal{P}_i|\times\sum_{i\in\mathcal{V}}|\mathcal{P}_i|}$ and $f(\cdot)$ is a function that returns the changes in the load admittance values according to Eqs. (1) (2) (3). The initial values of power $\widetilde{P}^0$ and $\widetilde{Q}^0$ are obtained from the regression model, and $Y^{\mathrm{H}^0}(h)$ is the initial admittance matrix.

Once the admittance matrix is updated, we utilize the kron reduction [49] to remove the nodes that are not harmonic sources. Then, the estimated measurement matrix $\widehat{\Phi}(h)$ can be established according to Eq. (5).

D. SBL-Based Harmonic State Estimation

Given the sparsity of $x(h)$ for a given order h, we develop an iterative algorithm for HSE drawing upon the SBL framework. The SBL algorithm treats unknown states as stochastic variables with certain probability distributions, and solves the HSE problem by maximizing their posterior probabilities. The solution is sparse since the derived posterior probability mass of state variables is distributed over solutions with a small number of non-zero elements (refer to [26], [53] for more details). Following [6], we run the iterative algorithm below for each harmonic order independently:

$$\hat{x}^{(j)} = \arg\min_x \frac{1}{2}\|z - \hat{\Phi}x\|_2^2 + \lambda\|u^{(j)} \odot x\|_1, \tag{20}$$

$$\gamma_i^{(j)} = \hat{x}_i^{(j)}/u_i^{(j)}, \tag{21}$$

$$u_i^{(j+1)} = [\hat{\Phi}_{\cdot i}^\top(\lambda I + \hat{\Phi}\Gamma^{(j)}\hat{\Phi}^\top)^{-1}\hat{\Phi}_{\cdot i}]^{\frac{1}{2}}. \tag{22}$$

The re-weighting parameter u_i promotes the sparsity of x, and the weight parameter λ trades off sparsity and estimation error. Note that for brevity the harmonic order h and time instant t are omitted in the above equations. Algorithm 1 summarizes different steps of the proposed approach.

Algorithm 1 Proposed framework for solving HSE

1: Learn the power flow model using LS (Section III-A);
2: Employ the LSTM network to predict the aggregate power at every node in the primary distribution network (Section III-B);
3: Compute the real-time harmonic admittance matrix based on Eq. (19) given predicted power and load compositions (Section III-C);
4: Receiving the DPMU measurements, run SBL-based harmonic state estimator to obtain $\hat{x}$ (Section III-D).

E. Meter Placement

We define the mutual coherence [54] of a matrix $\mu(\Phi)$, which measures the extent to which columns of this matrix are orthogonal, as follows:

Definition 1: Mutual coherence of a matrix $\Phi = [\phi_1, \cdots, \phi_n]$ is the largest absolute normalized inner product between any two of its columns:

$$\mu(\Phi) = \max_{1\le i\ne j\le n} \frac{|\phi_i^\top \phi_j|}{\|\phi_i\|_2 \cdot \|\phi_j\|_2}. \tag{23}$$

It is shown in [6] that an SBL-based state estimator can achieve satisfactory performance when the mutual coherence of the measurement matrix is smaller than a threshold. Given the estimation error $z - \hat{\Phi}x$ caused by the uncertainty of the measurement matrix, DPMUs should be placed such that the corresponding measurement matrix is robust to error. The condition number is typically used to quantify how sensitive the solution of the equation is to the error and how effective it acts as a diagnostic for multicollinearity in regression analysis [55]. Given d DPMUs, the optimal meter placement strategy of these sensors is the one that has a small coherence and minimizes the condition number of the corresponding measurement matrix:

$$\begin{aligned} \min_{M_i\subset\mathcal{V}} \quad & \mathrm{cond}(\Phi^{M_i}(h)), \quad \forall h \\ \text{s.t.} \quad & \mu(\Phi(h)) < \tau \end{aligned} \tag{24}$$

where $\tau \in (0, 1)$ is a positive threshold, $M_i (i = 1, \cdots, \binom{n_b}{d})$ and $\Phi^{M_i}(h)$ denote a possible meter placement strategy and the corresponding measurement matrix, respectively.

For a distribution network, the optimal meter placement can be obtained by solving Eq. (24) using a brute force search. Specifically, we search the possible meter placements starting from 2 DPMUs, and increase the number of meters until the constraint of (24) is satisfied. We subsequently compute the condition number in each case and select the smallest one among them. Note that it is computationally expensive to find the optimal placement strategy for large distribution systems in this way. We defer the full investigation of the optimal meter placement problem to future work.

IV. Case Studies

A. Test System

Since field data and real system models are scarce and not readily available, we evaluate the proposed approach on an IEEE test system using real data. Specifically, we modify the standard IEEE 13-bus test feeder [56] as shown in Fig. 2. This three-phase system contains a voltage regulator, two capacitor banks and 15 buses (a total of 36 nodes). Assume that a certain number of customers are connected via secondary networks to 12 load buses (out of the 15 buses). We treat Bus 634 as an aggregate load which is directly connected to Bus 633 and add a new bus, named Bus 635, with a single phase load connected to it. The switch 671-692 and line 692-675 in original test system are merged into a new line 671-675 in the preprocessing step using the two-port network theorem.

The ADRES data set [57] is used to generate load profiles of 180 customers over three days with one-minute resolution. The European low voltage test feeder [58] is connected to each load bus in the primary node to generate the aggregate power. Table I shows the number of customers connected to these buses. We consider four types of loads: phase-to-phase and line-to-ground single phase loads, and balanced and unbalanced three-phase loads. They are modelled as constant power factor loads.

We note that it is possible to obtain different placement strategies for different harmonic orders. In practice, the operator can place meters based on harmonic orders that are more important or concerning. We consider the optimal placement of meters for the 7^{th} order harmonics and set τ to 0.9999. Fig. 2 depicts the optimal placement of 6 DPMUs according to Eq. (24). The weight parameter λ is determined empirically. We set the reporting period of DPMUs to $T_{rd} = 1\,\text{sec}$, and the sampling and reporting intervals of smart meters to $T_{sm} = 1\,\text{min}, T_{rm} = 60\,\text{min}$. We perform state estimation every one minute over the course of the day.

TABLE I
Connection Types and Load Configurations.

Bus Name	No. Phases	Connection Type	Customers	c (%)
645	1	wye	80	60
635	1	wye	90	50
652	1	wye	90	0
611	1	wye	100	100
646	1	delta	110	60
684	1	delta	120	50
633	3	wye	180	60
670	3	wye	220	60
675	3	wye	240	70
680	3	delta	180	50
632	3	delta	200	60
671	3	delta	260	60

B. Performance Metrics and Baselines

We use three metrics, i.e., the identification error, the localization failure rate (LFR), and the normalized root-mean-square error (nRMSE), to evaluate HSE. The identification error of the h^{th} harmonic order at node i is defined as:

$$\epsilon_x(h,i) := \frac{|x_i^{es}(h) - x_i^{tr}(h)|}{\|x^{tr}(h)\|_2} \ ,$$

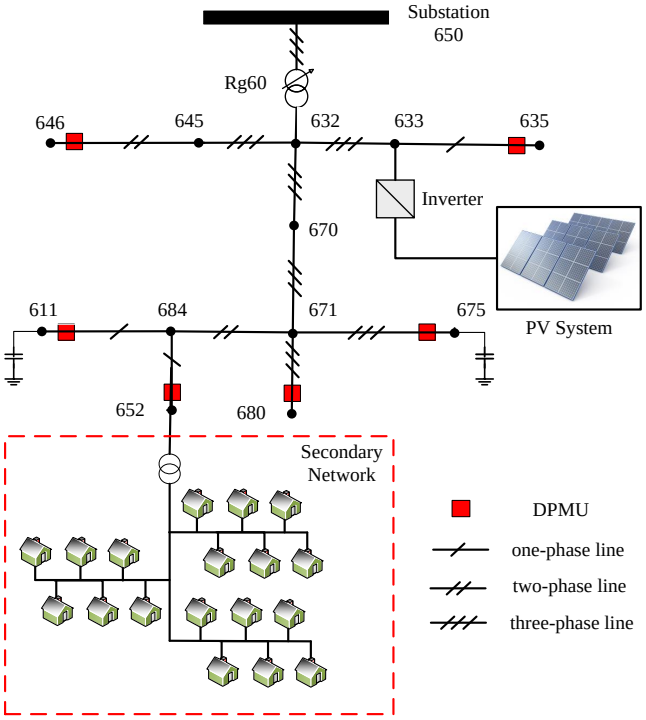

Fig. 2. A one-line diagram of the modified IEEE 13-node test feeder.

where $x_i^{tr}(h)$ and $x_i^{es}(h)$ are the true and estimated values of a state variable, respectively.

The localization failure rate is defined as the ratio of the number of experiments in which harmonic sources are not identified correctly to the total number of experiments. We say that a state estimator cannot identify harmonic sources at a given order, if the maximum identification error among all buses is larger than a threshold δ, i.e.,

$$\max_{i\in\mathcal{V}} \epsilon_x(h,i) \geq \delta \ .$$

We set δ to 0.05 in our simulations. The normalized root-mean-square error of the estimated current and voltage phasors are defined as follows:

$$\text{nRMSE}_{\text{IP}} := \|x^{tr} - x^{es}\|_2 / \|x^{tr}\|_2, \tag{25}$$

$$\text{nRMSE}_{\text{VP}} := \|V^{tr} - V^{es}\|_2 / \|V^{tr}\|_2. \tag{26}$$

Table II shows six baseline strategies that are used to benchmark the proposed harmonic state estimator, referred to as B1. Specifically, B2 is the case that the operating center collects smart meter data but does not perform prediction. It runs HSE with an outdated measurement matrix denoted by $\widetilde{\Phi}$. B3 is the case that uses the typical load profiles[5] instead of predictions to update the measurement matrix. It is the state-of-the-art approach since the typical load profiles are usually the only available information for updating the measurement matrix, denoted by $\bar{\Phi}$. B4 is the case that the model of each secondary network is known. In this case, the aggregate

[5]Load profiles describe the electricity consumption of small residential consumers on a daily basis under different operating conditions. We treat the average demand of 20 randomly selected customers over one day as the typical load profiles under light and heavy loading conditions, respectively.

TABLE II
THE PROPOSED HARMONIC STATE ESTIMATOR AND SIX BASELINES.

Cases	B1	B2	B3	B4	B5	B6	SVD
Observed Data	$z(h)$						
Measurement Matrix	$\widehat{\Phi}$	$\widetilde{\Phi}$	$\bar{\Phi}$	$\check{\Phi}$	Φ	$\widehat{\Phi}$	$\widehat{\Phi}$
Regression	Yes	Yes	No	No	No	Yes	Yes
Prediction	Yes	No	No	Yes	No	Yes	Yes
Estimator	SBL					ℓ_1	SVD

powers at all primary nodes are calculated based on the model. The corresponding measurement matrix is denoted by $\check{\Phi}$. B5 assumes the knowledge of the true measurement matrix. It is hypothetical baseline which is used to study the importance of updating measurement matrix. Since we study the HSE problem for unobservable power networks, traditional LS-based approaches are not suitable. Two classical approaches, namely SVD [16] and the ℓ_1 norm minimization [5] denoted by B6, are thus selected as the baselines. They can cope with a small number of measurement devices. For fair comparison, baselines B1, B6, and SVD utilize the proposed methods for regression analysis and demand prediction as shown in Table II.

C. Implementation

Suppose a PV system comprised of 40 solar arrays is connected to three inverters. Each array has 20 solar modules in series and 3 modules in a string. We employ the PV Powered inverter (Advanced Energy: PVP250kW-480) with $250kW$ maximum three-phase AC power output using the PV-LIB toolbox [59]. The PV system is modelled using the PVSystem object in OPENDSS.

We solve the harmonic load flow [6] in MATLAB and OPENDSS to obtain the voltages, currents, and the harmonic admittance matrix of the network denoted by V^{tr}, I^{tr} and Y^{H}, respectively. Our MATLAB script interacts with OPENDSS via the COM interface. We perform the following operations:

Step 1) Connect a certain number of customers with real load profiles to the European low voltage test feeder;
Step 2) Compute the aggregate power consumption in the primary network;
Step 3) Use the PV-LIB toolbox to compute the output power (AC) of the PV system using the temperature and solar irradiance data from NREL [60];
Step 4) Solve harmonic load flow in OPENDSS to obtain the true values of states $x^{tr}(h)$;
Step 5) Estimate the real-time measurement matrix as described in Section III-C with smart meter data;
Step 6) Run the proposed state estimator using the CVX toolbox in MATLAB.

V. MAIN TECHNICAL RESULTS

To corroborate the efficacy of the proposed state estimator, we run it under two operating conditions, namely, light and heavy loading conditions. The experiments are performed on a computer with Intel core i7-6700 CPU at 3.4 GHz, NVIDIA GTX 960 GPU and 16 GB RAM. Fig. 3 shows that the maximum nRMSE of the power flow regression is less than 1% for all loads in both operating conditions.

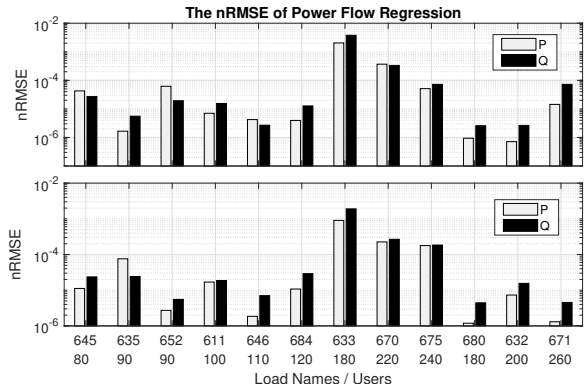

Fig. 3. The semi-logarithmic nRMSE of power flow regression between upstream nodes and the increasing numbers of customers along buses. This figure (light load (top) and heavy load (down)) presents the nRMSEs for four types of loads. The first six are the single loads of wye $(1, 2, 3, 4)$ and delta $(5, 6)$ connections, and the last six are three-phase imbalanced $(7, 8, 9)$ and balanced $(10, 11, 12)$ loads.

We select the points of the maximum 60 to 90 auto-correlation coefficients as the input units of the LSTM network for different types of loads, and randomly pick 150 observations with length of 1200 time slots from historical data to train the model that pertains to the secondary distribution network. It can be seen in Figs. 4 and 5 that the prediction error decreases as the number of customers increases (since the aggregate demand becomes smoother), and that the prediction accuracy is almost the same for loads of the same type. The maximum nRMSE values are respectively 15.28% and 14.24% under two conditions, and both are attained for Bus 645 which has the smallest number of connected customers. Observe that the average nRMSEs are 11.90% and 10.16% for the two conditions, respectively. Figs. 4 and 5 indicate that it is more difficult to predict the power drawn by unbalanced three-phase loads compared to the other three types. It took an average of 12 min and 54 sec to implement the overall offline learning for every secondary network using MATLAB's deep learning toolbox. The maximum training time of the LSTM model is 21 min and 43 sec. To predict the power consumption for the following hour, it is required to retrain (offline) the LSTM model every hour.

A. Considering Loads Only

We first study the performance of the proposed estimator with DPMUs that report data accurately. Suppose the loads connected to 645b, 670a and 670c are the significant harmonic sources with the real spectrum obtained from [61]. The weight parameter λ is 5×10^{-8} and 1×10^{-7} for the two operating conditions, respectively. Note that we solve HSE for different orders in parallel because the models for different orders are independent. Overall, it took an average of 0.85 sec to implement the SBL-based estimator. Since the harmonics are present at least for several seconds [40], the proposed estimator

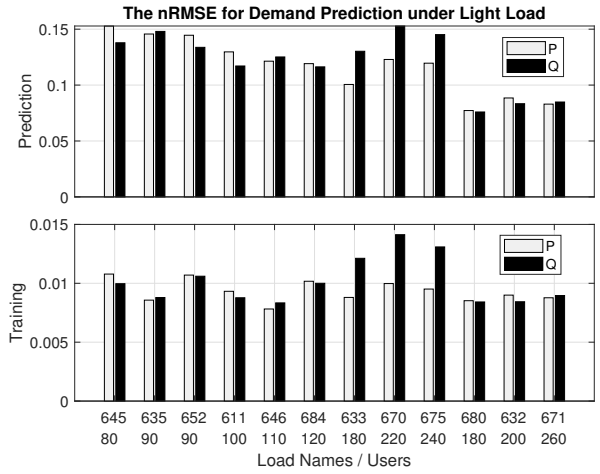

Fig. 4. The training and prediction nRMSE of an hour ahead along buses under light load condition. The first 6 loads are the single of wye $(1, 2, 3, 4)$ and delta $(5, 6)$ connections, and the last six are three-phase imbalanced $(7, 8, 9)$ and balanced $(10, 11, 12)$ loads.

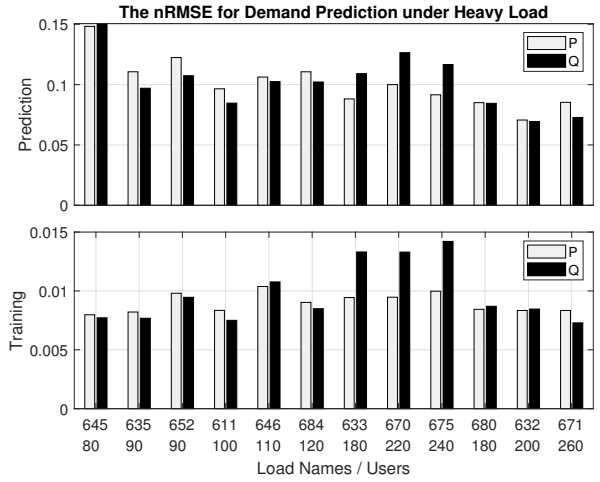

Fig. 5. The training and prediction nRMSE of an hour ahead along buses under heavy operating condition. The first 6 loads are the single of wye $(1, 2, 3, 4)$ and delta $(5, 6)$ connections, and the last six are three-phase imbalanced $(7, 8, 9)$ and balanced $(10, 11, 12)$ loads.

is capable of locating the harmonic sources in an online fashion. Fig. 6 shows the hourly maximum identification error among all orders is 2.36% which is attained for the peak hour, 12:00pm. Fig. 7 shows that the estimation error of total harmonic distortion (THD) of the voltage waveform is less than 0.0025 at three typical peak times. Table III shows that the performance of the proposed SBL-based estimator is superior to both the ℓ_1 norm minimization and SVD approaches. In particular, our estimator can locate harmonic sources with high accuracy for both loading conditions, whereas the LFR of B6 is high even under light loading condition. The SVD method cannot pinpoint the locations of the harmonic sources. For the seventh harmonic, the LFR of B3 increases by 24.64%, and the LFR of B6 increases to 99.79%. It is also evident from Fig. 8 that the nRMSE_{IP} of our approach improves by 71.27% and 73.31% compared to B3 under light and heavy operating conditions, respectively. Figs. 8 and 9 show that the proposed method has smaller nRMSE_{IP} and nRMSE_{VP} than B2, B3, B6 and SVD baselines. It is observed that the SVD method has large estimation errors for all harmonics, and B6 has large nRMSE_{IP} for the $3^{\text{rd}}, 5^{\text{th}}, 7^{\text{th}}, 13^{\text{th}}$ orders. This suggests that the proposed SBL-based estimator can accurately estimate the voltage distribution and reliably locate the harmonic sources. Moreover, it can be concluded from Figs. 8 and 9 that the nRMSE values of B1 and B4 are almost the same, implying that the proposed regression method works well for HSE. Also observe that the nRMSE values of B5 are less than 0.0001 under both operating conditions, highlighting the importance of updating the measurement matrix. The average nRMSE_{IP} of the proposed approach is 0.0065 and 0.0133 under the two conditions. This is acceptable for an advanced distribution network monitoring system.

TABLE III
COMPARING THE LFR% OF THE PROPOSED APPROACH WITH SIX BASELINES UNDER LIGHT AND HEAVY LOADING CONDITIONS.

Loading Condition	B1	B2	B3	B4	B5	B6	SVD
Light	0.07	0.80	3.53	0.25	0.00	77.32	100.00
Heavy	1.09	2.06	14.82	1.87	0.00	84.09	100.00

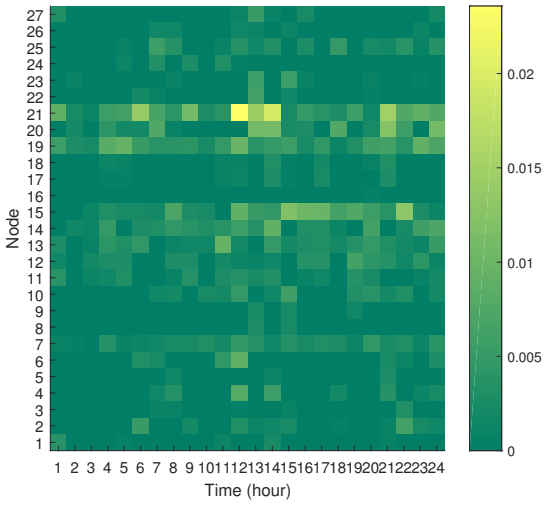

Fig. 6. The maximum identification error among all harmonic orders at different nodes for each hour of the day under the light loading condition.

We now investigate the impact of a simple interpolation method on the estimation performance. Assume that there are four kinds of smart meters with different sampling intervals, i.e., 1, 5, 10, 15 min. The measurement matrix is updated by the proposed method using these four kinds of data, respectively. The weight parameter λ is set to 6×10^{-8}. With the same DPMU data, the HSE is conducted for the light loading condition. Table IV shows that the SBL-based estimator still locates the harmonic sources accurately even with smart meter data sampled at 15-minute intervals. The estimation errors at the same level indicate that the simple interpolation method is effective for the proposed estimator.

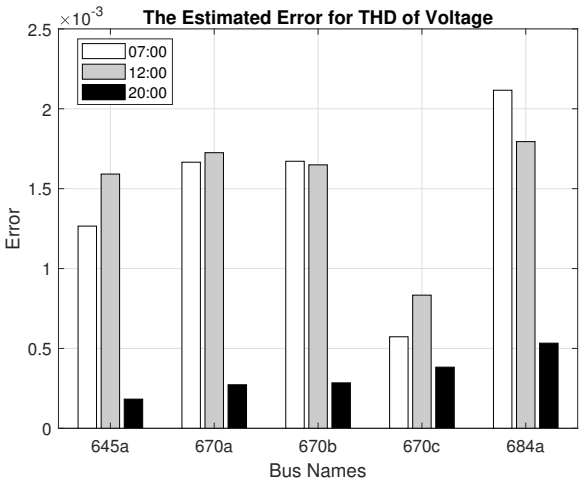

Fig. 7. Estimated error for THD of voltage waveform at three peak hours under the light loading condition.

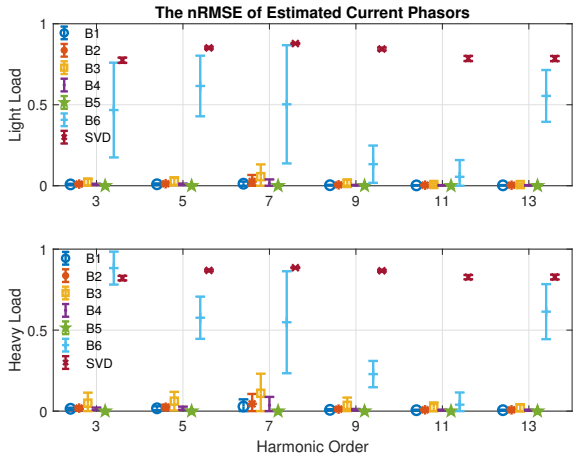

Fig. 8. Comparing the nRMSE_{IP} of the proposed approach with six baselines under light and heavy loading conditions, respectively. The error bars represent the standard deviations of nRMSE_{IP} for 1440 experiments.

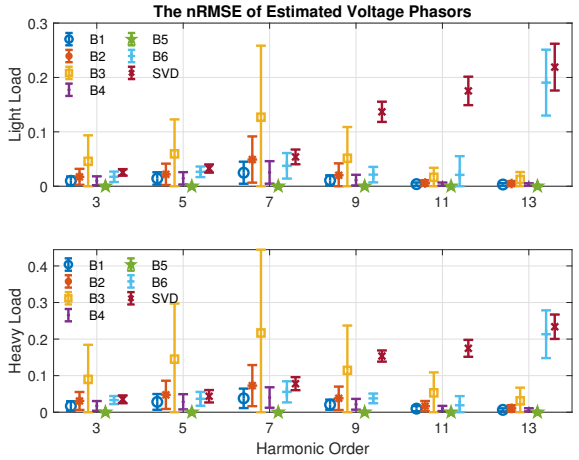

Fig. 9. Comparing the nRMSE_{VP} of the proposed approach with six baselines under light and heavy loading conditions, respectively. The error bars represent the standard deviations of nRMSE_{VP} for 1440 experiments.

We found that the prediction accuracy of consumed power is improved using smart meter data of large sampling interval, which is helpful for enhancing the performance of the state estimator. This is because the load profile is smoother for larger sampling intervals.

TABLE IV
PERFORMANCE COMPARISON OF THE ESTIMATOR EMPLOYING SMART METER DATA OF DIFFERENT SAMPLING INTERVALS UNDER LIGHT LOADING CONDITION.

T_{sm}(min)	1	5	10	15
LFR(%)	0.00	0.00	0.01	0.01
nRMSE_{IP}	0.0065	0.0062	0.0065	0.0068
nRMSE_{VP}	0.0122	0.0122	0.0124	0.0132

We next analyze the performance of the proposed harmonic state estimator with noisy measurements. Suppose DPMUs have 0.2% ratio error and $10'$ phase displacement of Class 0.2 as recommended in IEC 61869-2 [62] and 61869-3 [63]. It is observed from Table V that the average LFR of our method is 1.55%, which is still smaller than that of B2, B3, B6 and SVD. B6 locates the harmonic sources with a low accuracy of 18.33%, whereas SVD cannot pinpoint the positions of harmonics at all. Thus, even considering the measurement noise, the proposed estimator achieves a relatively high accuracy. The measurement error is neglected in the following simulations.

TABLE V
COMPARISON OF NRMSE AND LFR% WITH NOISY MEASUREMENTS.

Cases	B1	B2	B3	B4	B5	B6	SVD
LFR(%)	1.55	3.82	12.91	1.72	0.27	81.67	100
nRMSE_{IP}	0.012	0.020	0.039	0.012	0.003	0.460	0.821
nRMSE_{VP}	0.012	0.020	0.052	0.012	0.001	0.062	0.107

B. Considering PV Systems

To investigate the impact of DERs on HSE, we connect a PV system with 800kVA capacity to Bus 633. We assume that the loads at 684.a.c and 670.a.c are the significant harmonic sources with real spectra. The weight parameter λ is set to 1×10^{-7} for both operating conditions.

Table VI shows that the LFR of B1 is much smaller that that of B6 and SVD. The LFR of B6 is high and SVD approach fails to locate the harmonic sources. It is observed that the average nRMSE_{IP} values of our method are 0.0053 and 0.0095 under the two loading conditions, respectively. Figs. 10 and 11 show that the proposed method is capable of pinpointing the harmonic sources with a smaller error compared to B2, B3, B6 and SVD when DERs are connected to the primary network. The results of B5 with the smallest estimation error highlight the importance of updating the measurement matrix. Observe that nRMSE values of B1 and B4 are almost the same, indicating the efficacy of the power flow regression approach. Next, we run the proposed state estimator on four groups of harmonic sources shown in Table VII to study the impact of PV systems. It can be readily seen from Figs. 12 and 13 that both nRMSE_{IP} and nRMSE_{VP} remain unchanged for all harmonic orders when PV systems are installed in

the distribution system. According to Table VIII, the LFR values are almost the same with PV systems. This implies that the proposed estimator is not negatively affected by grid-connected DERs that do not produce significant harmonics.

TABLE VI
COMPARING THE LFR% OF THE PROPOSED APPROACH WITH SIX BASELINES CONSIDERING PV UNDER LIGHT AND HEAVY LOADING CONDITIONS.

Loading Condition	B1	B2	B3	B4	B5	B6	SVD
Light	0.00	0.00	1.26	0.00	0.00	37.97	100.0
Heavy	0.00	0.04	10.02	0.00	0.00	49.32	100.0

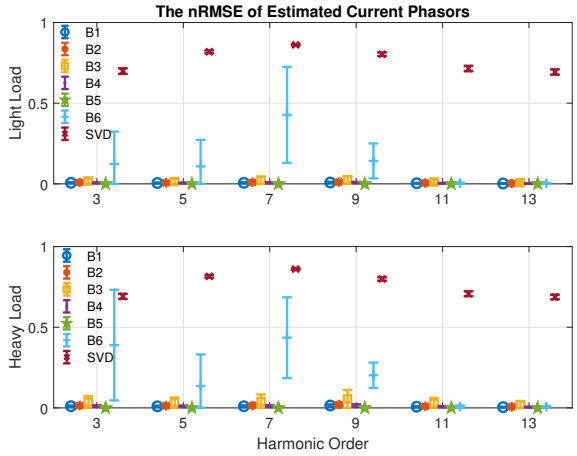

Fig. 10. Comparing the nRMSE_{IP} of the proposed estimator with other six baselines with a grid-connected PV system under light and heavy loading conditions, respectively. The error bars represent the standard deviations of nRMSE_{IP} for 1440 experiments.

C. PV Harmonic Sources

To explore the impact of PV systems that produce significant harmonics on HSE, PV systems are connected to different buses that are not equipped with DPMUs. We consider the real harmonic spectrum in [64] and the default spectrum in OPENDSS as the spectra of PV harmonic sources. The weight parameter λ is set to 2×10^{-8} and 1×10^{-8} for loads and PV harmonic sources, respectively. Fig. 14 implies that the proposed method can locate the PV harmonic sources with a small nRMSE_{IP} regardless of the type of the harmonic sources. It can be seen from Figs. 15 that the average nRMSE_{VP} values are respectively 0.0113 and 0.0118 under PV and load harmonic sources. We note that the estimation error for PV harmonic sources is almost the same as the error for load harmonic sources, which is evident from Figs. 14 and 15. The

TABLE VII
LOCATION OF LOAD HARMONIC SOURCES.

Group	1	2	3	4
Locations	684.a.c, 670.a.c	645.b, 670.a.c	645b, 684.a.c	633.a.b.c

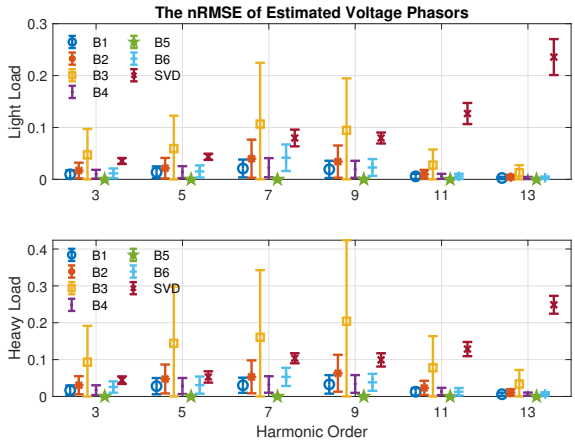

Fig. 11. Comparing the nRMSE_{VP} of the proposed estimator with other six baselines with a grid-connected PV system under light and heavy loading conditions, respectively. The error bars represent the standard deviations of nRMSE_{VP} for 1440 experiments.

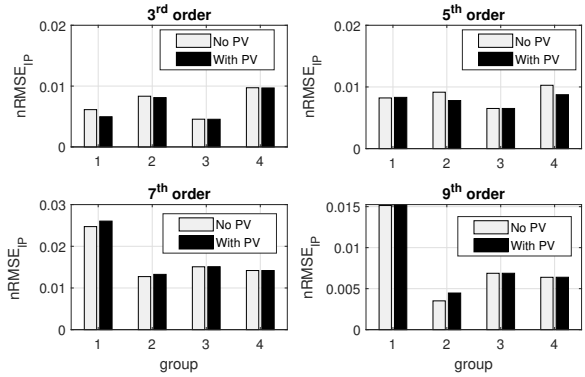

Fig. 12. Comparing the nRMSE_{IP} for harmonic orders $3, 5, 7, 9$ when a grid-connected PV system is installed (the red bar) and when there are only loads (the blue bar).

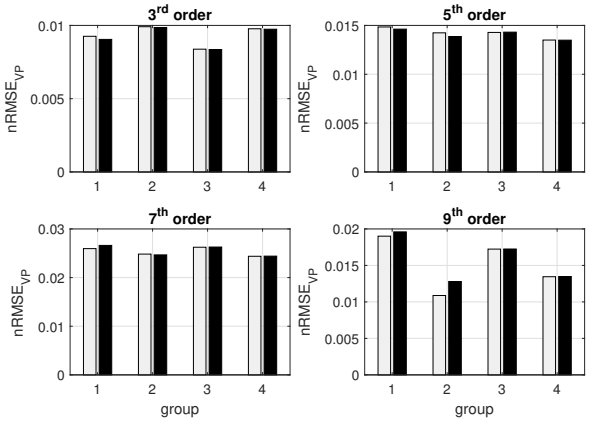

Fig. 13. Comparing the nRMSE_{VP} for harmonic orders $3, 5, 7, 9$ when a grid-connected PV system is installed (the red bar) and when there are only loads (the blue bar).

TABLE VIII
THE LFR METRIC (%) FOR EACH HARMONIC ORDER WITH GRID-CONNECTED PV SYSTEMS.

Condition	Group	3	5	7	9	11	13
No PV	1	0.00	0.00	0.49	0.00	0.00	0.00
	2	0.00	0.00	0.42	0.00	0.00	0.00
	3	0.00	0.00	0.00	0.00	0.00	0.00
	4	0.07	0.35	0.07	0.00	6.59	3.12
With PV	1	0.00	0.00	0.49	0.00	0.00	0.00
	2	0.00	0.07	0.14	0.00	0.00	0.00
	3	0.00	0.00	0.00	0.00	0.00	0.00
	4	0.07	0.14	0.07	0.00	6.66	2.98

TABLE IX
COMPARISON OF THE LFR METRIC (%) OBTAINED FOR SEVEN CASES WITH PV AND LOAD HARMONIC SOURCES UNDER LIGHT AND HEAVY LOADING CONDITIONS.

Loading Condition	B1	B2	B3	B4	B5	B6	SVD
Light	0.02	0.19	4.67	0.00	0.00	30.71	100.00
Heavy	0.64	1.30	17.73	0.54	0.00	25.24	100.00

results indicate that the proposed state estimator can pinpoint both the load harmonic sources and the PV harmonic sources.

Next, we assume that the PV system installed at Bus 670 and the load connected to 645.b are the significant harmonic sources with specified spectra. We set $\lambda = 5 \times 10^{-8}$ for both conditions. Table IX shows that, compared to B2, B3, B6 and SVD, the proposed approach is capable of locating harmonic sources with higher accuracy under both conditions in the presence of PV and load harmonic sources. As shown in Section V-A and V-B, the SVD method cannot find the harmonic sources correctly, and the LFR of B6 is much larger than that of B1. Considering the updated measurement matrix, the average LFR is almost zero under the light loading condition, and it can be reduced to 0.64% from 17.73% compared to B3 in the heavy loading situation. Fig. 16 shows that our average nRMSE_{IP} of 0.0068 is about 72% of B2 and 30% of B3 under the light loading condition. It can be seen from Fig. 17 that the nRMSE of voltage phasors is also less than that of B2, B3, B6 and SVD. These results suggest that the proposed SBL-based estimator outperforms both B6 and SVD approaches in terms of estimation error and localization accuracy. It is observed in Figs. 16 and 17 that the nRMSEs of B4 and B1 are almost the same, which shows the effectiveness of the proposed power flow regression method. As with Section V-A and V-B, the error of B5 is the smallest among all baselines. Note that nRMSE and LFR of the 7^{th} order are the largest among all orders, since it is near the resonance frequency of the system. These results indicate that the proposed approach can pinpoint harmonic sources and reliably estimate the voltage distribution in the presence of PV and load harmonic sources.

D. Sensitivity Analysis

We study how many harmonic sources can be identified reliably with 6 DPMUs using the proposed HSE. Table X shows that our approach can pinpoint 6 harmonic sources with an average nRMSE_{IP} of 0.0579 and an average LFR of

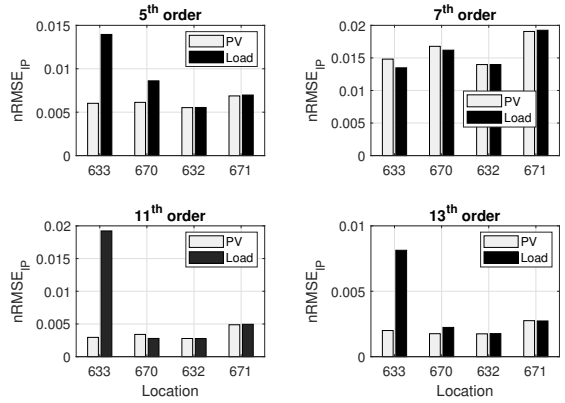

Fig. 14. Comparing the nRMSE_{IP} for harmonic orders $5, 7, 11, 13$ when harmonic sources are located at buses that are not equipped with a DPMU. The blue and red bars denote PV and load harmonic sources, respectively.

87.46%. When the number of harmonic sources increases from 6 to 7, the nRMSE_{IP} of all orders rises sharply for all baselines as shown in Fig. 18. This means that the proposed HSE cannot locate harmonic sources in a hypothetical situation where 7 sources produce significant harmonics simultaneously in a distribution network. Fig. 18 indicates that our HSE can estimate the currents with much smaller error than B3 baseline for all orders. The estimation error of B5 is much smaller than other four baselines as shown in Figs. 18 and 19. This is because B5 is performed with accurate measurement matrix and the measurement noise is ignored. It can be seen in Fig. 19 that the maximum nRMSE_{VP} is 0.0617 for 6 harmonic sources and it changes slightly for the $3^{\text{rd}}, 5^{\text{th}}, 7^{\text{th}}$ orders as we increase the number of harmonic sources. In contrast to B2 and B3 baselines, our method which attempts to update the measurement matrix reduces the estimation error and improves the location accuracy.

Finally, we analyze the performance of the proposed HSE scheme for a different number of DPMUs. Fig. 20 shows that both nRMSE_{IP} and nRMSE_{VP} decrease as we deploy more DPMUs. This curve can be used by the utility company to decide on the number of DPMUs that are necessary for a

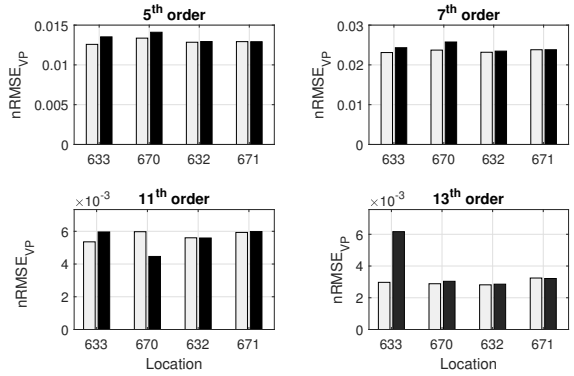

Fig. 15. Comparing the nRMSE_{VP} for harmonic orders $5, 7, 11, 13$ when harmonic sources are located at buses that are not equipped with a DPMU. The blue and red bars denote PV and load harmonic sources, respectively.

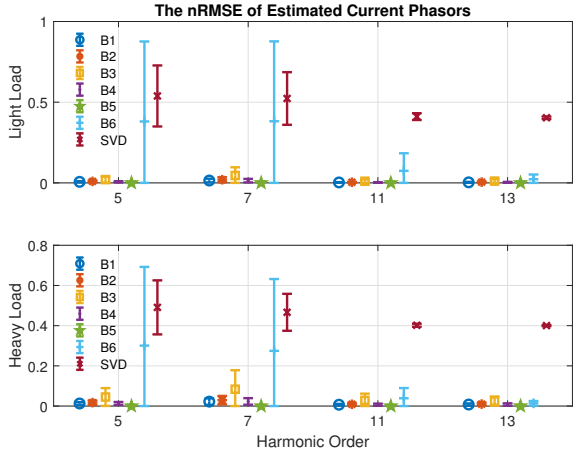

Fig. 16. Comparing the nRMSE_{IP} of the proposed estimator with six baselines with PV and load harmonic sources under light and heavy loading conditions. The error bars represent the standard deviations of nRMSE_{IP} for 1440 experiments.

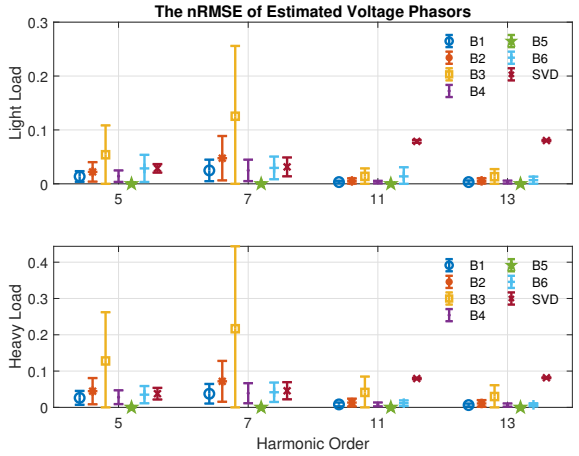

Fig. 17. Comparing the nRMSE_{VP} of the proposed estimator with six baselines with PV and load harmonic sources under light and heavy loading conditions. The error bars represent the standard deviations of nRMSE_{VP} for 1440 experiments.

desired level of accuracy, given the number of sources that produce significant harmonics in the distribution networks.

VI. Conclusion

The availability of time-synchronized phasor measurements from multiple locations in the distribution network enables a wide range of analytics applications, including state estimation, model validation, and fault detection, diagnostics, and prognostics. This paper proposed a novel approach to HSE,

TABLE X
THE NRMSE AND LFR% ALONG
THE NUMBER OF HARMONIC SOURCES (HSES)

No. HSes	2	3	4	5	6	7
LFR(%)	0.00	0.00	0.00	5.14	12.54	98.85
nRMSE_{IP}	0.0087	0.0065	0.0177	0.0068	0.0579	0.9234
nRMSE_{VP}	0.0126	0.0123	0.0123	0.0120	0.0217	0.0614

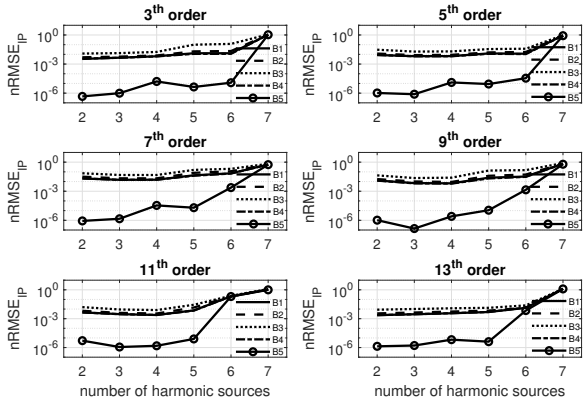

Fig. 18. Evolution of nRMSE_{IP} with increasing number of harmonic sources for harmonic orders $3, 5, 7, 9, 11, 13$ with 6 DPMUs.

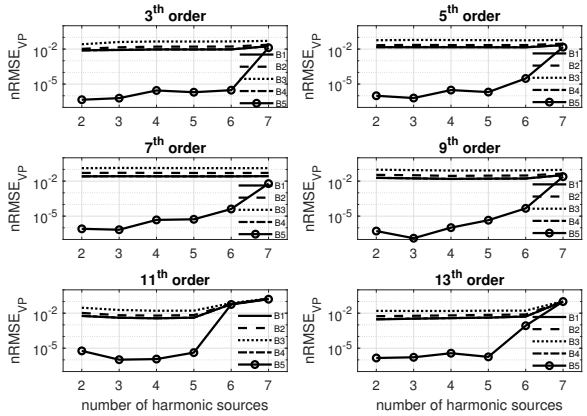

Fig. 19. Evolution of nRMSE_{VP} with increasing number of harmonic sources for harmonic orders $3, 5, 7, 9, 11, 13$ with 6 DPMUs.

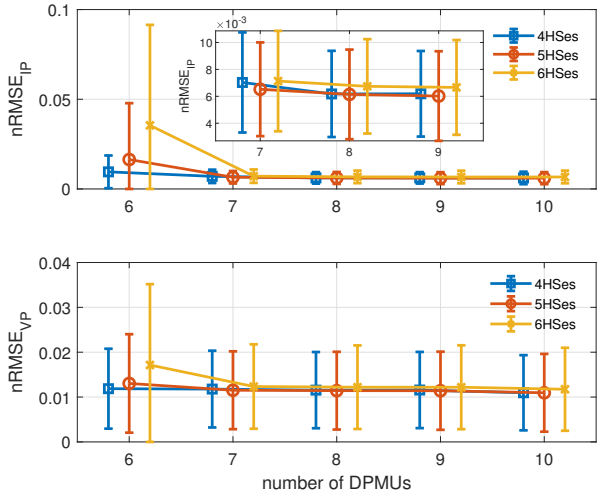

Fig. 20. Evolution of nRMSE_{IP} and nRMSE_{VP} along increasing number of DPMUs for a different number of harmonic sources (HSes). The error bar represents the standard deviation of nRMSE_{IP} for 1440 experiments.

which involves regression analysis for power flow calculation, recurrent neural network models for demand prediction, and sparse Bayesian learning. Leveraging the sparsity of harmonic sources and the availability of coarse-grained smart meter data along with fine-grained DPMU data, the proposed approach accurately identifies harmonics sources using only a small number of DPMUs compared to the number of nodes in the primary network. Extensive simulations performed on the modified IEEE 13-node test feeder with real data corroborated the efficacy of the proposed HSE. Our results showed that the proposed harmonic state estimator achieves satisfactory performance even in the presence of grid-connected PV systems. Finally, we provided a guideline to electric utilities to help them determine the number of DPMUs that must be deployed to achieve a desired level of accuracy.

In future work, we intend to investigate how other types of distributed generations, e.g., wind turbines, small hydro power units, and distributed energy storage could affect the performance of the proposed approach. We also plan to investigate accurate harmonic load models considering capacitive reactances with field measurements.

REFERENCES

[1] V. E. Wagner *et al.*, "Effects of harmonics on equipment," *IEEE Transactions on Power Delivery*, vol. 8, no. 2, pp. 672–680, 1993.

[2] N. Kanao *et al.*, "Power system harmonic analysis using state-estimation method for Japanese field data," *IEEE Transactions on Power Delivery*, vol. 20, no. 2, pp. 970–977, 2005.

[3] K. K. Yu and N. R. Watson, "Three-phase harmonic state estimation using SVD for partially observable systems," in *International Conference on Power System Technology*, vol. 1, 2004, pp. 29–34.

[4] E. F. de Arruda, N. Kagan, and P. F. Ribeiro, "Harmonic distortion state estimation using an evolutionary strategy," *IEEE Transactions on Power Delivery*, vol. 25, no. 2, pp. 831–842, 2010.

[5] H. W. Liao, "Power system harmonic state estimation and observability analysis via sparsity maximization," *IEEE Transactions on Power Systems*, vol. 22, no. 1, pp. 15–23, 2007.

[6] Y. Yuan, W. Zhou, H.-T. Zhang, Z. Ping, and O. Ardakanian, "Sparse bayesian harmonic state estimation," in *Proc. International Conference on Communications, Control, and Computing Technologies for Smart Grids*, 2018, pp. 1–6.

[7] A. P. S. Meliopoulos, F. Zhang, and S. Zelingher, "Power system harmonic state estimation," *IEEE Transactions on Power Delivery*, vol. 9, no. 3, pp. 1701–1709, 1994.

[8] K. K. Yu, N. Watson, and J. Arrillaga, "An adaptive kalman filter for dynamic harmonic state estimation and harmonic injection tracking," *IEEE Transactions on Power Delivery*, vol. 20, no. 2, pp. 1577–1584, 2005.

[9] C. Rakpenthai, S. Uatrongjit, N. R. Watson, and S. Premrudeepreechacharn, "On harmonic state estimation of power system with uncertain network parameters," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4829–4838, 2013.

[10] A. Medina and R. Cisneros-Magaña, "Time-domain harmonic state estimation based on the kalman filter poincaré map and extrapolation to the limit cycle," *IET Generation, Transmission & Distribution*, vol. 6, no. 12, pp. 1209–1217, 2012.

[11] I. Molina-Moreno, A. Medina, R. Cisneros-Magaña, and O. Anaya-Lara, "Time domain harmonic state estimation in unbalanced power networks based on optimal number of meters and the principle of half-wave symmetry," *IET Generation, Transmission & Distribution*, vol. 11, no. 15, pp. 3871–3880, 2017.

[12] G. D'Antona, C. Muscas, and S. Sulis, "State estimation for the localization of harmonic sources in electric distribution systems," *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 5, pp. 1462–1470, 2009.

[13] I. D. Melo, J. L. R. Pereira, A. Variz, and P. Garcia, "Harmonic state estimation for distribution networks using phasor measurement units," *Electric Power Systems Research*, vol. 147, pp. 133–144, 2017.

[14] N. Okada and K. Yukihira, "Harmonic state estimation in distribution network based on fifth harmonic current characteristic," in *16th International Conference on Harmonics and Quality of Power*, 2014, pp. 566–570.

[15] D. Bhujel, N. R. Watson, and T. S. Jalal, "Application of harmonic state estimation to a distribution system," in *PowerTech*, 2017, pp. 1–6.

[16] J. F. Breda, J. C. Vieira, and M. Oleskovicz, "Three-phase harmonic state estimation for distribution systems by using the SVD technique," in *Power and Energy Society General Meeting*, 2016, pp. 1–5.

[17] G. D'Antona, C. Muscas, P. A. Pegoraro, and S. Sulis, "Harmonic source estimation in distribution systems," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 10, pp. 3351–3359, 2011.

[18] C. F. Almeida and N. Kagan, "Harmonic state estimation through optimal monitoring systems," *IEEE Transactions on Smart Grid*, vol. 4, no. 1, pp. 467–478, 2013.

[19] K. R. Mestav, J. Luengo-Rozas, and L. Tong, "Bayesian state estimation for unobservable distribution systems via deep learning," *IEEE Transactions on Power Systems*, 2019.

[20] ——, "State estimation for unobservable distribution systems via deep neural networks," in *2018 IEEE Power & Energy Society General Meeting (PESGM)*, 2018, pp. 1–5.

[21] W. Sun, M. Zamani, M. R. Hesamzadeh, and H.-T. Zhang, "Data-driven probabilistic optimal power flow with nonparametric bayesian modeling and inference," *IEEE Transactions on Smart Grid*, 2019.

[22] W. Sun, M. Zamani, H.-T. Zhang, and Y. Li, "Probabilistic optimal power flow with correlated wind power uncertainty via markov chain quasi-monte carlo sampling," *IEEE Transactions on Industrial Informatics*, 2019.

[23] A. Bretas, N. Bretas, S. Braunstein, A. Rossoni, and R. Trevizan, "Multiple gross errors detection, identification and correction in three-phase distribution systems wls state estimation: A per-phase measurement error approach," *Electric Power Systems Research*, vol. 151, pp. 174–185, 2017.

[24] N. G. Bretas and A. S. Bretas, "The extension of the gauss approach for the solution of an overdetermined set of algebraic non linear equations," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 65, no. 9, pp. 1269–1273, 2018.

[25] A. S. Bretas, N. G. Bretas, B. Carvalho, E. Baeyens, and P. P. Khargonekar, "Smart grids cyber-physical security as a malicious data attack: An innovation approach," *Electric Power Systems Research*, vol. 149, pp. 210–219, 2017.

[26] W. Pan, Y. Yuan, J. Gonçalves, and G. B. Stan, "A sparse bayesian approach to the identification of nonlinear state-space systems," *IEEE Transactions on Automatic Control*, vol. 61, no. 1, pp. 182–187, 2016.

[27] H. Ukai, K. Nakamura, and N. Matsui, "DSP-and GPS-based synchronized measurement system of harmonics in wide-area distribution system," *IEEE Transactions on Industrial Electronics*, vol. 50, no. 6, pp. 1159–1164, 2003.

[28] A. Carta, N. Locci, and C. Muscas, "A pmu for the measurement of synchronized harmonic phasors in three-phase distribution networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 10, pp. 3723–3730, 2009.

[29] M. Chakir, I. Kamwa, and H. Le Huy, "Extended c37. 118.1 pmu algorithms for joint tracking of fundamental and harmonic phasors in stressed power systems and microgrids," *IEEE Transactions on Power Delivery*, vol. 29, no. 3, pp. 1465–1480, 2014.

[30] A. Carta, N. Locci, and C. Muscas, "GPS-based system for the measurement of synchronized harmonic phasors," *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 3, pp. 586–593, 2008.

[31] Z. Li *et al.*, "A new vibration testing platform for electronic current transformers," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 3, pp. 704–712, 2018.

[32] A. Von Meier, D. Culler, A. McEachern, and R. Arghandeh, "Micro-synchrophasors for distribution systems," in *Innovative Smart Grid Technologies Conference (ISGT), 2014 IEEE PES*, 2014, pp. 1–5.

[33] S. K. Jain, P. Jain, and S. N. Singh, "A fast harmonic phasor measurement method for smart grid applications," *IEEE Transactions on Smart Grid*, vol. 8, no. 1, pp. 493–502, 2016.

[34] I. D. de Melo, J. L. R. Pereira, A. M. Variz, and B. C. Oliveira, "Harmonic state estimation for distribution systems based on synchrophasors," in *16th International Conference on Environment and Electrical Engineering*, 2016, pp. 1–6.

[35] H. E. Mazin, E. E. Nino, W. Xu, and J. Yong, "A study on the harmonic contributions of residential loads," *IEEE Transactions on Power Delivery*, vol. 26, no. 3, pp. 1592–1599, 2011.

[36] R. Abu-Hashim *et al.*, "Test systems for harmonics modeling and simulation," *IEEE Transactions on Power Delivery*, vol. 14, no. 2, pp. 579–587, 1999.
[37] A. Bonner *et al.*, "Modeling and simulation of the propagation of harmonics in electric power networks. 1. concepts, models, and simulation techniques," *IEEE Transactions on Power Delivery*, vol. 11, no. 1, pp. 452–465, 1996.
[38] H. Hu, Q. Shi, Z. He, J. He, and S. Gao, "Potential harmonic resonance impacts of PV inverter filters on distribution systems," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 1, pp. 151–161, 2015.
[39] J. Arrillaga and N. R. Watson, *Power system harmonics*. John Wiley & Sons, 2004.
[40] R. C. Dugan, M. F. McGranaghan, and H. W. Beaty, *Electrical power systems quality*. New York, NY: McGraw-Hill, 1996.
[41] R. C. Dugan, "Reference guide: The open distribution system simulator (OpenDSS)," *Electric Power Research Institute, Inc*, vol. 7, 2012.
[42] D. Montenegro, R. Dugan, and G. Ramos, "Harmonics analysis using sequential-time simulation for addressing smart grid challenges," in *23rd Int. Conf. Electricity Distribution*, 2015, pp. 15–18.
[43] R. Dugan, R. Arrit, R. Henry, T. McDermott, and W. Sunderm, "Opendss epri distribution system simulator-harmonic load modeling documentation," https://sourceforge.net/p/electricdss/code/HEAD/tree/trunk/Doc/, 2014, [Online; accessed 10-July-2019].
[44] O. Ardakanian, *Advances in Distribution System Monitoring*. Springer International Publishing, 2018, pp. 1–3.
[45] "A Type of DPMU Produced by Power Standards Lab," https://www.powerstandards.com/product/pqube-3/highlights/, [Online; accessed 10-July-2019].
[46] W. Xu, "A multiphase harmonic load flow solution technique," Ph.D. dissertation, University of British Columbia, 1990.
[47] O. Ardakanian, Y. Yuan, V. Wong, R. Dobbe, S. Low, A. von Meier, and C. J. Tomlin, "On identification of distribution grids," *IEEE Transactions on Control of Network Systems*, 2019.
[48] K. C. Smith, K. Smith, and R. Alley, *Electrical circuits: an introduction*. Cambridge University Press, 1992.
[49] Y. Yuan, O. Ardakanian, S. Low, and C. Tomlin, "On the inverse power flow problem," *arXiv preprint arXiv:1610.06631*, 2016.
[50] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," *Neural Computation*, vol. 12, no. 10, pp. 2451–2471, 2000.
[51] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
[52] A. Gómez-Expósito, C. Gómez-Quiles, and I. Džafić, "State estimation in two time scales for smart distribution systems," *IEEE Transactions on Smart Grid*, vol. 6, no. 1, pp. 421–430, 2015.
[53] M. E. Tipping, "Sparse bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, no. Jun, pp. 211–244, 2001.
[54] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via L1 minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.
[55] B. David, E. Kuh, and R. Welsch, "Regression diagnostics: identifying influential data and sources of collinearity," 1980.
[56] W. H. Kersting, "Radial distribution test feeders," in *Power Engineering Society Winter Meeting*, vol. 2, 2001, pp. 908–912.
[57] "ADRES data set," https://www.ea.tuwien.ac.at//projects/adres_concept/EN/, [Online; accessed 10-July-2019].
[58] "IEEE PES distribution systems analysis subcommittee radial test feeders," http://sites.ieee.org/pes-testfeeders/resources/, [Online; accessed 10-July-2019].
[59] "PV-LIB toolbox," https://pvpmc.sandia.gov/applications/pv_lib-toolbox/, [Online; accessed 10-July-2019].
[60] "National renewable energy laboratory (NREL)," https://midcdmz.nrel.gov/apps/go2url.pl?site=LMU, [Online; accessed 10-July-2019].
[61] Powers Standards Lab, "Pqube events," http://map.pqube.com/ECOXplore/Events.htm, [Online; accessed 10-July-2019].
[62] IEC 61869-2, "Instrument transformers - part 2: Additional requirements for current transformers," *International Electrotechnical Commission (IEC)*, 2012.
[63] IEC 61869-3, "Instrument transformers - part 3: Additional requirements for inductive voltage transformers," *International Electrotechnical Commission (IEC)*, 2011.
[64] H. A. Pereira, F. D. Freijedo, M. Silva, V. Mendes, and R. Teodorescu, "Harmonic current prediction by impedance modeling of grid-tied inverters: A 1.4 MW PV plant case study," *International Journal of Electrical Power & Energy Systems*, vol. 93, pp. 30–38, 2017.

Wei Zhou received the B.E. degree from the Huazhong University of Science and Technology, Wuhan, China, in 2015. He is currently pursuing the Ph.D. degree with the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China. During February–August 2018, he was a Visiting Scholar in the Department of Computing Science, University of Alberta, Edmonton, Canada. His research interests include Bayesian learning and convex optimization with applications to smart grids.

Omid Ardakanian (M'15) received the B.S. degree in Computer Engineering from Sharif University of Technology, Iran, and the MMath and PhD degrees both in Computer Science from the University of Waterloo, ON, Canada. He is an Assistant Professor with the Department of Computing Science, University of Alberta, Edmonton, AB, Canada. Prior to that, he was an NSERC Postdoctoral Fellow in the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, USA, and in the Department of Electrical and Computer Engineering, University of British Columbia, Canada. His research interests include resource allocation, performance modelling and evaluation with applications to computer networks and cyber-physical systems. He served as a Guest Editor of IEEE Transactions on Smart Grid, Special Section on Theory and Application of PMUs in Power Distribution Systems. He received a number of best paper awards from IEEE PES General Meeting (2017), ACM BuildSys (2016), and ACM e-Energy (2013).

Hai-Tao Zhang (M'07–SM'13) received the B.E. and Ph.D. degrees from the University of Science and Technology of China, Hefei, China, in 2000 and 2005, respectively. During January–December 2007, he was a Postdoctoral Researcher with the University of Cambridge, Cambridge, U.K. Since 2005, he has been with Huazhong University of Science and Technology, Wuhan, China, where he was an associate professor from 2005 to 2010 and has been a full professor since 2010. His research interests include swarming intelligence, model predictive control, and multi-agent systems control. He was an Associate Editor of the *IEEE Transactions on Circuits and SystemsPart II: Express Briefs* from 2015 to 2018. He is currently an Associate Editor of the *Asian Journal of Control*.

Ye Yuan (M'13) received the B.Eng. degree (Valedictorian) from the Department of Automation, Shanghai Jiao Tong University, Shanghai, China, in September 2008, and the M.Phil. and Ph.D. degrees from the Department of Engineering, University of Cambridge, Cambridge, U.K., in October 2009 and February 2012, respectively. He has been a Full Professor at the Huazhong University of Science and Technology, Wuhan, China since 2016. Prior to that, he was a Postdoctoral Researcher at UC Berkeley, a Junior Research Fellow at Darwin College, University of Cambridge, and has been holding visiting researcher positions at California Institute of Technology, Massachusetts Institute of Technology, and Imperial College London. His research interests include system identification and control with applications to cyber-physical systems.

He is the recipient of China National Recruitment Program of 1000 Talented Young Scholars, Dorothy Hodgkin Postgraduate Award, Microsoft Research PhD Scholarship, Cambridge Overseas Student Award, Chinese Government Award for Outstanding Students Abroad, Henry Lester PhD Scholarship and a number of best paper awards in IEEE conferences.