Where to Install Distribution Phasor Measurement Units to Obtain Accurate State Estimation Results?

Wenyu Wang, Nanpeng Yu
Department of Electrical and Computer Engineering
University of California, Riverside
wwang032@ucr.edu, nyu@ece.ucr.edu

Farnoosh Rahmatian
NuGrid Power Corp
Vancouver, Canada
frahmatian@nugridpower.com

Shikhar Pandey
Commonwealth Edison
Chicago, USA
shikhar.pandey@comed.com

***Abstract*—With the rapid expansion of distributed energy resources in the distribution network, system operators need accurate and real-time estimations of system states in order to actively manage the distribution system. However, existing metering infrastructure does not provide adequate support for real-time monitoring in the distribution system. In this paper, we propose a real-time, data-driven approach to estimate the nodal voltage magnitudes in the distribution system. In this approach, we first train a neural network that maps historical measurements from distribution phasor measurement units (DPMUs) to smart meter data. The trained neural network is then used to estimate real-time nodal voltage magnitudes based on streaming DPMU data. In addition, we design an approach to identify the optimal locations to install DPMUs based on the facility location function selection method. Numerical study results on an IEEE test feeder show that our proposed method produces accurate nodal voltage magnitude estimates. The facility location function selection method also recommends DPMU candidate locations that yield lower voltage magnitude estimation error.**

***Index Terms*—Distribution system, state estimation, distribution phasor measurement units, facility location.**

I. Introduction

The rapid expansion of distributed energy resources (DERs) is changing the operational regime of the electric distribution system. DERs introduce bidirectional power flows and the higher likelihood of voltage violations. To actively control the power distribution network, system operators need advanced tools to monitor the distribution network states in real-time.

The existing metering infrastructure of the distribution system does not provide adequate support for real-time monitoring. Although the increasing penetration of smart meters improves situation awareness, their measurements are usually not transmitted back to the control center in real-time. The distribution phasor measurement units (DPMUs) provide real-time time synchronized measurements at a higher sampling frequency. However, the penetration rate of DPMUs is much lower than that of the smart meters due to the high costs. The goal of this paper is to develop a data-driven, real-time distribution system state estimation (DSSE) algorithm that uses a large number of smart meters and a small number of DPMUs. In addition, we propose an algorithm to select installation locations for DPMUs that improves state estimation accuracy.

The topic of state estimation has been extensively studied for power transmission systems. In the conventional approach, the state estimation task is often formulated as a weighted least square (WLS) optimization problem, and various enhancements to this approach have been proposed to improve the state estimation performance and robustness [1]. However, these methods perform poorly when directly applied to the power distribution system. This is because distribution systems do not have the high level of metering data redundancy in transmission systems and are thus unobservable [2].

One way to deal with the unobservability is to use "pseudo-measurement" data, which are generated from historical load data of the power distribution system [3]. To model the distribution system loads and their uncertainty, probabilistic and statistical approaches such as expectation maximization (EM) [4], correlation analysis [5], and spatial-temporal modeling [6] have been proposed. To generate pseudo load data, machine learning methods such as artificial neural networks (ANNs) [7] and clustering algorithms [8] are proposed. If the pseudo-measurements appropriately capture the statistical distributions of the actual field data, the state estimation results with augmented data could be satisfactory. The disadvantage of data augmented state estimation approach is that it relies heavily on accurate modeling of the distribution network. However, it is difficult for electric utilities to maintain a precise distribution network model [9], especially for secondary feeders [10].

Several approaches based on machine learning and data mining are proposed to monitor the unobservable distribution system. Matrix completion [11] and tensor completion [12] are proposed to recover unobserved states from available sensor data. However, accurate feeder models and parameters are needed to form physical constraints such as power flow equations. A Bayesian state estimation approach using deep learning is proposed in [13], which requires accurate feeder models to prepare training data. Neural networks are trained to estimate nodal voltages of a single-phase [14] and three-phase distribution systems [15] using measurements from the supervisory control and data acquisition (SCADA) system. However, the topic of optimal sensor location selection for improving DSSE accuracy has not been thoroughly studied.

Several algorithms are proposed to find the optimal locations of time synchronized measurements for the purpose of state estimation. Integer programming, exhaustive search, and their variations have been adopted to find optimal locations of phasor measurement units (PMUs) in transmission systems [16] and distribution systems [17], [18]. However, these approaches are designed for observable systems with redundant measurements from PMUs and DPMUs. Thus, they do not work in distribution systems with limited sensor coverage. Techniques such as depth of unobservability and observability propagation are proposed for incomplete observability [19], but they are not designed for extremely low PMU coverage.

In this paper, we propose a real-time, data-driven approach to estimate nodal voltages in the distribution system. In this approach, we first train a feedforward neural network (FNN) based on historical data of DPMU and smart meters. The trained FNN is then used to estimate real-time nodal voltages with DPMU data. In addition, we design an approach to find the optimal locations of DPMU installation, based on facility location analysis. Compared with the existing methods, our proposed method is directly applicable to practical unobservable distribution systems with limited real-time measurements.

The rest of the paper is organized as follows. Section II describes the problem setup and assumptions. Section III presents the machine learning methodology for state estimation and the optimal location selection for DPMU. Section IV evaluates the performance of the proposed state estimation method and DPMU location selection with a numerical study. Section V states the conclusion.

II. Problem Setup

A. Objective of Selecting Installation Location of DPMU for Data-Driven State Estimation

It is assumed that the distribution feeders already have high penetration of smart meters. To further improve real-time situational awareness, the system operator tries to install a small number of DPMUs. In real-time distribution system state estimation, we try to infer the voltage magnitudes recorded by smart meters based on DPMU measurements. The objective of selecting the installation location of DPMUs is to minimize the estimation error of the data-driven state estimation algorithm.

B. Assumptions

1) Data and Model Availability: First, for a single-phase load on phase i, the smart meter measures the voltage magnitude of phase i. Second, for a two-phase delta-connected load between phase i and j, the smart meter measures the voltage magnitude across phase i and j. Third, for a three-phase load on phase i, the smart meter measures the voltage magnitude of one of three phases. Fourth, a DPMU is installed at one of the nodes in the distribution network. The DPMU measures the voltage of that node and the current of a primary line connecting that node. The voltage measurement includes three phases' voltage magnitude and phase angle; the current measurement includes three phases' current magnitude and phase angle. Thus, each DPMU has 12 input features. Since the DPMU has a much higher sampling frequency than the smart meter, the DPMU data is represented by the average reading of the DPMU during the same measurement period of the smart meter, so that the DPMU data and smart meter match in terms of time instances. Fifth, we know the timestamp of the smart meter and DPMU, including hour, month, weekday/weekend, holiday, and month. Sixth, the connectivity model and the parameters of the primary feeder are known. This assumption is used to identify the optimal installation location of DPMU. All the measurement data of smart meters and DPMUs are converted to per unit.

2) Input and Output of the State Estimation: As mentioned earlier in this Section, the output of the state estimation is the smart meter voltage magnitude. The input includes two types of data. The first type is the DPMU measurement data. The second type is the timestamp of the state estimation, which includes hour, weekday/weekend, holiday, and month.

III. The Data-Driven State Estimation and Optimal Location Selection for DPMU

A. Machine Learning-based DSSE Method

We adopt a supervised machine learning model, feedforward neural network (FNN), to learn the nonlinear mapping between DPMU, timestamp information and the smart meter voltage magnitudes. The development and validation of the machine learning-based DSSE algorithm consists of three steps. First, we preprocess the dataset and split it into training, validation, and testing datasets. The training dataset and validation dataset represent the historical dataset recorded by the data management system of the distribution network. The testing dataset is used to perform and evaluate the real-time state estimation. Second, we train the FNN based on the training dataset and determine when to stop the training by using the validation dataset. Third, we evaluate the performance of the trained model using the testing dataset. Technical details of the machine learning-based DSSE method are provided below.

1) Data Preprocessing and Split: We apply z-score normalization on the DPMU data, i.e., the data is centered and normalized by their standard deviation. This improves the convergence in training. The weekday/weekend information is represented by a binary variable. The binary variable is 1 if it is a weekend, and 0 otherwise. Similarly, the holiday information is represented by a binary variable, which is 1 if it is a holiday and 0 otherwise. The cyclical encoding is used to encode hour and month information. The k-th hour ($k = 1, 2, ..., 24$) is encoded by $[cos\frac{2\pi k}{24}, sin\frac{2\pi k}{24}]$. Similarly, the k-th month ($k = 1, 2, ..., 12$) is encoded by $[cos\frac{2\pi k}{12}, sin\frac{2\pi k}{12}]$. The dataset is split randomly into three parts. 64% of the samples are used as the training dataset. 16% of the samples are used as the validation dataset for early stopping. The remaining 20% of the samples are used as the testing dataset to evaluate the FNN's voltage estimation performance.

2) The Architecture and Training of FNN: The FNN consists of three components: an input layer, three hidden layers of 200 neurons, and an output layer. The number of neurons in the input layer is equal to the number of input features, and the number of neurons in the output layer is equal to the number of smart meters. Each neuron has directed connections to the neurons of the subsequent layer and each connection has a corresponding weight. In the input layer, each neuron corresponds to an input variable. In the hidden layer, each neuron takes in the weighted sum of neurons from the previous layer (plus a bias term) and produces an output value by the ReLU activation function. The output layer is a linear function of the neurons in the last hidden layer. The FNN learns the mapping from the input (DPMU data ad timestamp data) to the output (smart meter voltage magnitudes) by minimizing the mean squared error (MSE) between the estimated value and the true value of smart meter voltage magnitudes.

In the training process, to avoid being trapped in a local minimum, the training dataset is often randomly grouped into small mini-batches. The FNN parameters (weights and biases) are then updated based on samples of one mini-batch at a time. When the model goes through all the mini-batches for one time, it is called an "epoch".

We use "early stopping" to determine when to stop training the FNN model. When training FNN, the model may be overfitting after too many epochs. In the early stopping scheme, we use the validation dataset to evaluate the accuracy of the model on unseen data. If the performance on the validation dataset is not improved after a number of epochs (called "patience"), then the training is stopped. The model with the best performance on the validation dataset is then chosen as the trained model.

3) Evaluating the State Estimation Performance: After training the machine learning model, its performance is evaluated based on the testing dataset. The trained model takes DPMU and timestamp data as inputs to produce real-time voltage magnitude estimates. Mean absolute percentage error (MAPE) is used to measure the estimation accuracy.

B. Selection of Optimal DPMU Locations

Due to the high cost of DPMU, only a limited number of DPMUs can be installed in a distribution system. We propose a method to find the optimal locations of DPMUs that minimizes voltage magnitude estimation errors of a data-driven DSSE.

Our proposed approach is based on the facility location selection method [20]. The working principle is as follows. Suppose we select a subset of locations $\mathcal{D}_{\text{select}} \subseteq \mathcal{D}$, where $\mathcal{D}$ is the set of all potential sensor locations. $\mathcal{S}$ is the set of smart meters, and a_{ij} is the similarity or score between sensor location d_i and smart meter s_j. When the size of $\mathcal{D}_{\text{select}}$ is limited, the optimal selection of $\mathcal{D}_{\text{select}}$ is derived by maximizing the facility location selection function (1).

$$r(\mathcal{D}_{\text{select}}) = \sum_{s_j \in \mathcal{S}} \max_{d_i \in \mathcal{D}_{\text{select}}} a_{ij} \tag{1}$$

We need to solve two problems when selecting the optimal DPMU locations based on the facility location selection method. The first problem is how to combine the selection of voltage measurement and current measurement locations. For each DPMU, the node whose voltage is measured must be connected to the branch whose current is measured. The second problem is that the optimizing the facility location function is in general NP-hard.

We design an algorithm to find the optimal DPMU locations. In this algorithm, to combine the selection of voltage and current measurement locations, we first rank the voltage measurement nodes and current measurement branches separately and then select the node-branch combination that has the lowest sum of rankings. To make the optimization easier to solve, we use a greedy algorithm, i.e., the DPMU locations are selected one by one, and each time the best available DPMU location is selected. Such a greedy optimization method has been shown to provide a good approximation to the optimal solution [21]. The details of the DPMU selection method are shown in Algorithm 1. In the algorithm, step 1 initializes the set of selected DPMU locations to empty sets. Step 2 calculates the similarity between voltage measurement nodes/current measurement branches and the smart meters. Step 3–8 represent the greedy algorithm and it stops when N DPMU locations are selected. Step 9 returns the selected DPMU locations in terms of the selected nodes and branches. The calculation of ϕ_{ij} and ψ_{ij} in step 2 is explained below.

Algorithm 1 DPMU Location Selection

Input: Set of potential DPMU voltage measurement nodes $\mathcal{V}$ and current measurement branches $\mathcal{C}$, set of smart meters $\mathcal{S}$, number of DPMU installations N.

1: Set $n = 0$, $\mathfrak{V} = \emptyset$, $\mathfrak{C} = \emptyset$.
2: Calculate the similarity ϕ_{ij} between each node $v_i \in \mathcal{V}$ and each smart meter $s_j \in \mathcal{S}$. Calculate the similarity ψ_{ij} between each branch $c_i \in \mathcal{C}$ and each smart meter $s_j \in \mathcal{S}$.
3: **repeat**
4: Calculate $r_v(i) = \sum_{s_j \in \mathcal{S}} \max_{v_k \in (\mathfrak{V} \cup \{v_i\})} \phi_{kj}$, $\forall v_i \in \mathcal{V}$ and $r_c(i) = \sum_{s_j \in \mathcal{S}} \max_{c_k \in (\mathfrak{C} \cup \{c_i\})} \psi_{kj}$, $\forall c_i \in \mathcal{C}$.
5: Rank all $v_i \in \mathcal{V}$ in descending order of $r_v(i)$, and rank all $c_i \in \mathcal{C}$ in descending order of $r_c(i)$.
6: For each pair of $v_i \in \mathcal{V}$ and $c_j \in \mathcal{C}$, such that the v_i is connected to c_j in the distribution network, calculate the the sum of the ranking of $r_v(i)$ and the ranking of $r_c(j)$. The pair with the lowest sum of ranking is the chosen voltage and current measurement location.
7: Suppose the chosen pair is v_k and c_l. Then $\mathfrak{V} = \mathfrak{V} \cup \{v_k\}$, $\mathfrak{C} = \mathfrak{C} \cup \{c_l\}$, $\mathcal{V} = \mathcal{V} \setminus \{v_k\}$, $\mathcal{C} = \mathcal{C} \setminus \{c_l\}$, $n = n+1$.
8: **until** $n = N$.
9: **return** The selected DPMU location $\mathfrak{V}$ and $\mathfrak{C}$.

1) Similarity Score ϕ_{ij} Between Node v_i and Smart Meter s_j: Meter s_j measures the voltage magnitude of 1 of the 6 phase connections: A, B, C, AB, BC, or CA. First, using power flow, we can derive historical voltage magnitudes of these 6 phase connections at node v_i based on the connectivity model, the parameters of the primary feeder model, and the historical smart meter data. Second, we calculate s_j's voltage's absolute correlation coefficient with each of the 6 phases' voltage. These correlation coefficients are different due to the unbalanceness in the feeder. The highest absolute correlation coefficient is then defined as the similarity score ϕ_{ij}.

2) Similarity Score ψ_{ij} Between Branch c_i and Smart Meter s_j: Calculating ϕ_{ij} is similar to ψ_{ij} in Section III-B1. First, using power flow, we can derive historical current magnitudes in phase A, B, and C of branch c_i based on the primary feeder model and historical smart meter data. Second, we calculate meter s_j's voltage's absolute correlation coefficient with each of the 3 phases' current. The highest absolute correlation coefficient is defined as the similarity score ψ_{ij}.

IV. Numerical Study

A. Setup for Numerical Tests

We evaluate the performance of our proposed real-time, data-driven state estimation method and DPMU location selection method with a modified IEEE 37-bus test feeder, which is shown in Fig. 1. We modify the standard 37-bus test feeder by introducing loads with all 7 types of phase connections, AN, BN. CN, AB, BC, CA, and ABC. The test circuits' primary feeder contains 21 line segments and 22 nodes, which serve 25 loads. The smart meter of the ABC-phase load is assumed to measure the voltage of phase AN. We assume only one DPMU is installed, which measures the voltage of one node and the current of one primary line connecting the node. Thus, there are 42 possible different DPMU locations.

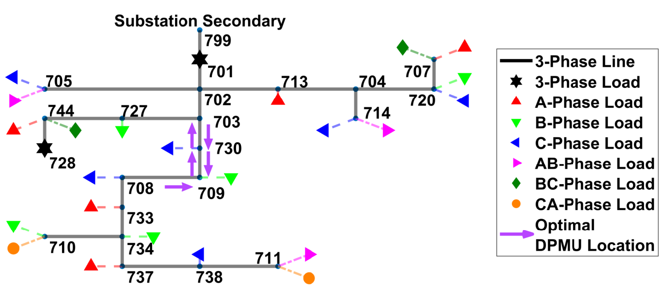

Fig. 1. Schematic of a modified IEEE 37-bus test feeder.

We aggregate the half-hourly electricity load data (kWh) from the London households [22] to simulate the instantaneous real power (kW) of each load. The length of the real power consumption time series is 17520, which represents 365 days of measurements. The reactive power time series are calculated by assuming a lagging power factor, which follows a uniform distribution $\mathcal{U}(0.9, 1)$. The peak load of the 37-bus test feeders is 2.4MW respectively. After running the power flow of the test feeder using OpenDSS, we obtain the measurement of smart meter and DPMU. To simulate the smart meter measurement noise, we use zero-mean normal distributions with three standard deviation matching 0.1% of the nominal values. The 0.1% accuracy class smart meters established in ANSI C12.20-2015 represents the typical noise level in real-world advanced metering infrastructure. To simulate the DPMU measurement noise, we add simulated noise phasor to the measurement phasor. The phasor's angle noise follows a uniform distribution $\mathcal{U}(0, 2\pi)$, and its magnitude follows a half-normal distribution whose original normal distribution has a standard deviation matching 1% of the nominal values. This 1% total vector error (TVE) established in IEEE standard C37.118 represents the typical noise level in real-world DPMUs. Note that the simulated DPMU data is the average of measurements from the device with a 60 Hz sampling frequency.

To evaluate the DSSE performance with different DPMU locations, we build an FNN model for each of the 42 possible DPMU locations. The input of FNN includes 12 DPMU features (voltage and current magnitude and angle in three phases), 1 binary variable representing weekday/weekend, 1 binary variable for holiday, 2 features for hour, and 2 features of month. Thus, the input layer of FNN has 18 nodes. The FNN's output layer has 25 nodes corresponding to 25 smart meters. The 17520 samples are randomly split into training (64%), validation (16%), and testing (20%) datasets. The size of mini-batches is 10 samples, the early stopping patience is 100 epochs. The FNN is trained using the Adam algorithm.

B. Performance of the State Estimation Method

We use MAPE to quantify the error of the proposed data-driven DSSE algorithm. The average MAPE of the 42 tests using different DPMU locations is 0.056%. The distribution of the MAPE of each of the 42 tests is shown in Fig. 2. From Fig. 2, we can see that the overall MAPE of these tests are very low and the proposed data-driven DSSE is very accurate. Fig. 3 shows two sample smart meters' voltage estimations versus the true values in 50 hours. Fig. 3 (a) is one of the least accurate scenarios with MAPE=0.141%; Fig. 3 (b) has the average performance with MAPE=0.055%. From Fig. 3 we can see that on average the voltage estimation is highly accurate and even the worst case has decent estimation result.

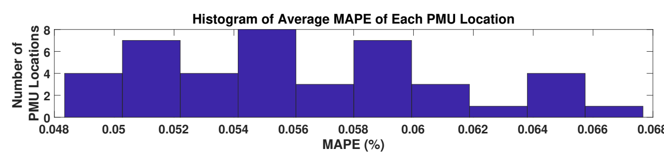

Fig. 2. Histogram of the MAPE of 42 tests using different DPMU locations.

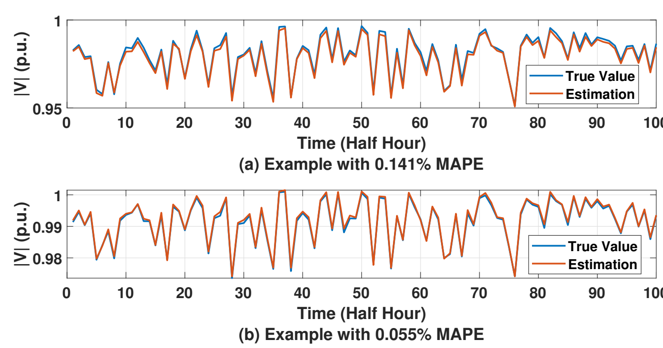

Fig. 3. Examples of voltage estimation.

C. Performance of the Optimal DPMU Location Selection

We rank the DPMU locations using our proposed selection method in Section III-B. The selection method is evaluated by examining the DSSE MAPE of using the top and bottom ranked DPMU locations. Fig. 4 shows the DSSE MAPE of using the top 5 and bottom 5 ranked DPMU locations. The box plot represents the distribution of the MAPE of all the 42 tests with different DPMU locations. The green diamond and the red bar represent the mean and median MAPE of all tests. From Fig. 4, we can see that the proposed facility location selection method can find DPMU locations that have higher DSSE accuracy. Although the selected top 5 locations do not match exactly to the lowest 5 MAPE results of the tests, the selection method still produces significantly lower MAPE than random selection. The lowest MAPE of all the 42 tests is 0.048%, the average MAPE of using the top 5 ranked DPMUs is 0.052%, the average MAPE of using the bottom 5 ranking DPMUs is 0.057%, and the highest MAPE of all tests is 0.068%. We highlight the top 5 DPMU locations in Fig. 1 using arrows. The arrow body aligns with the current-measurement line segment and the arrow head points to the voltage-measurement node.

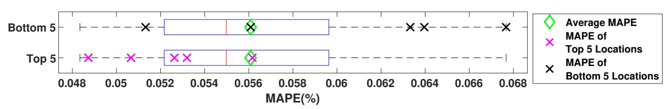

Fig. 4. MAPE of top and bottom ranked DPMU locations.

D. Distribution of Nodal Voltage Magnitude Estimation Error

System operators are also interested in the voltage magnitude estimation error of specific smart meters. To illustrate this, we choose one of the top 5 DPMU locations identified in Section IV-C and then choose a smart meter that has low state estimation accuracy. Fig. 5 shows the distribution of the absolute percentage error (APE) of this smart meter's voltage magnitude estimates. In Fig. 5, the estimation error of 3504 time samples in the testing dataset is organized by hour. We can see that except for a small number of outliers (marked by red plus signs), the absolute percentage error of the voltage magnitude estimates are below 0.2%. This result shows that the voltage magnitude estimates are extremely accurate.

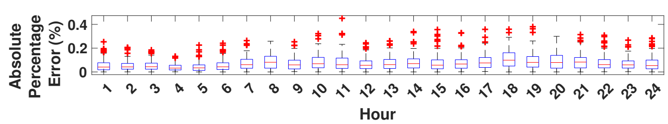

Fig. 5. Distribution of APE of a smart meter's voltage magnitude estimates.

V. Conclusion

In this paper, we develop a real-time, data-driven state estimation method for the distribution system. Our proposed method is broadly applicable as it uses the readily available smart meter data and limited DPMU data. The machine learning model is trained to estimate nodal voltage magnitudes using historical data of DPMU and smart meters. We also design an approach to find the optimal locations of the DPMUs based on the facility location selection method. Numerical study results on an IEEE test feeder show that our proposed method estimates the voltage states accurately, and the location selection method can find the DPMU locations that have higher estimation accuracy.

References

[1] K. Dehghanpour, Z. Wang, J. Wang, Y. Yuan, and F. Bu, "A survey on state estimation techniques and challenges in smart distribution systems," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 2312–2322, 2018.

[2] Y. Gao and N. Yu, "State estimation for unbalanced electric power distribution systems using AMI data," in *2017 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. IEEE, 2017, pp. 1–5.

[3] M. Baran and T. McDermott, "Distribution system state estimation using AMI data," in *2009 IEEE/PES Power Systems Conference and Exposition*. IEEE, 2009, pp. 1–3.

[4] R. Singh, B. C. Pal, and R. A. Jabr, "Statistical representation of distribution system loads using Gaussian mixture model," *IEEE Trans. Power Syst.*, vol. 25, no. 1, pp. 29–37, 2009.

[5] D. T. Nguyen, "Modeling load uncertainty in distribution network monitoring," *IEEE Trans. Power Syst.*, vol. 30, no. 5, pp. 2321–2328, 2014.

[6] J. Shi, Y. Liu, and N. Yu, "Spatio-temporal modeling of electric loads," in *2017 North American Power Symposium (NAPS)*. IEEE, 2017, pp. 1–6.

[7] E. Manitsas, R. Singh, B. C. 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, pp. 1888–1896, 2012.

[8] Y. R. Gahrooei, A. Khodabakhshian, and R.-A. Hooshmand, "A new pseudo load profile determination approach in low voltage distribution networks," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 463–472, 2017.

[9] B. Foggo and N. Yu, "Improving supervised phase identification through the theory of information losses," *IEEE Trans. Smart Grid*, vol. 11, no. 3, pp. 2337–2346, 2019.

[10] W. Wang, N. Yu, B. Foggo, J. Davis, and J. Li, "Phase identification in electric power distribution systems by clustering of smart meter data," in *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2016, pp. 259–265.

[11] P. L. Donti, Y. Liu, A. J. Schmitt, A. Bernstein, R. Yang, and Y. Zhang, "Matrix completion for low-observability voltage estimation," *IEEE Trans. Smart Grid*, vol. 11, no. 3, pp. 2520–2530, 2019.

[12] R. Madbhavi, H. S. Karimi, B. Natarajan, and B. Srinivasan, "Tensor completion based state estimation in distribution systems," in *2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. IEEE, 2020, pp. 1–5.

[13] K. R. Mestav, J. Luengo-Rozas, and L. Tong, "Bayesian state estimation for unobservable distribution systems via deep learning," *IEEE Trans. Power Syst.*, vol. 34, no. 6, pp. 4910–4920, 2019.

[14] M. Ferdowsi, A. Löwen, P. McKeever, A. Monti, F. Ponci, and A. Benigni, "New monitoring approach for distribution systems," in *2014 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*. IEEE, 2014, pp. 1506–1511.

[15] M. Pertl, K. Heussen, O. Gehrke, and M. Rezkalla, "Voltage estimation in active distribution grids using neural networks," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*. IEEE, 2016, pp. 1–5.

[16] N. M. Manousakis, G. N. Korres, and P. S. Georgilakis, "Taxonomy of PMU placement methodologies," *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 1070–1077, 2012.

[17] Z. Wu, X. Du, W. Gu, Y. Liu, P. Ling, J. Liu, and C. Fang, "Optimal PMU placement considering load loss and relaying in distribution networks," *IEEE Access*, vol. 6, pp. 33 645–33 653, 2018.

[18] R. S. Biswas, B. Azimian, and A. Pal, "A micro-PMU placement scheme for distribution systems considering practical constraints," in *2020 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE, 2020, pp. 1–5.

[19] X.-C. Guo, C.-S. Liao, and C.-C. Chu, "Enhanced optimal PMU placements with limited observability propagations," *IEEE Access*, vol. 8, pp. 22 515–22 524, 2020.

[20] B. Foggo and N. Yu, "Analyzing data selection techniques with tools from the theory of information losses," *arXiv preprint arXiv:1902.09602*, 2019.

[21] A. Krause and D. Golovin, "Submodular function maximization." *Tractability*, vol. 3, pp. 71–104, 2014.

[22] UK Power Networks, "Smartmeter energy consumption data in London households," Feb 2014. [Online]. Available: https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households