Considerations for AMI-Based Operations for Distribution Feeders

Preprint

Harsha Padullaparti,[1] Santosh Veda,[1] Surya Dhulipala,[2] Murali Baggu,[1] Tom Bialek[3] and Martha Symko-Davies[1]

[1] *National Renewable Energy Laboratory*
[2] *University of Florida*
[3] *San Diego Gas & Electric Company*

Presented at the 2019 IEEE Power & Energy Society General Meeting (IEEE PES GM)
Atlanta, Georgia
August 4–8, 2019

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contract No. DE-AC36-08GO28308

Conference Paper
NREL/CP-5D00-72773
September 2019

Considerations for AMI-Based Operations for Distribution Feeders

Preprint

Harsha Padullaparti,[1] Santosh Veda,[1] Surya Dhulipala,[2] Murali Baggu,[1] Tom Bialek[3] and Martha Symko-Davies[1]

[1] National Renewable Energy Laboratory
[2] University of Florida
[3] San Diego Gas & Electric Company

Suggested Citation
Padullaparti, Harsha, Santosh Veda, Surya Dhulipala, Murali Baggu, Tom Bialek, and Martha Symko-Davies. 2019. *Considerations for AMI-Based Operations for Distribution Feeders: Preprint*. Golden, CO: National Renewable Energy Laboratory. NREL/CP-5D00-72773. https://www.nrel.gov/docs/fy19osti/72773.pdf.

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contract No. DE-AC36-08GO28308

Conference Paper
NREL/CP-5D00-72773
September 2019

National Renewable Energy Laboratory
15013 Denver West Parkway
Golden, CO 80401
303-275-3000 • www.nrel.gov

Cover Photos by Dennis Schroeder: (clockwise, left to right) NREL 51934, NREL 45897, NREL 42160, NREL 45891, NREL 48097, NREL 46526.

NREL prints on paper that contains recycled content.

Considerations for AMI-Based Operations for Distribution Feeders

Harsha Padullaparti[1], Santosh Veda[1], Surya Dhulipala[3], Murali Baggu[1], Tom Bialek[2], Martha Symko-Davies[1]
[1]National Renewable Energy Laboratory (NREL), Golden, Colorado.
[2]San Diego Gas & Electric (SDG&E), San Diego, California.
[3]University of Florida, Gainesville, Florida.
Santosh.Veda@nrel.gov

Abstract— **More than $5 billion in investments in advanced metering infrastructure (AMI) technologies, AMI deployments, as pervasive secondary network voltage monitoring systems, provide opportunities for utility operations and controls. This paper focuses on the considerations for AMI-based tools and techniques as the industry moves toward operationalizing such large data sets. Phase identification is a first such tool. Numerous distribution network analysis, monitoring, and control applications—including volt/volt-ampere reactive control, state estimation, and distribution automation—require accurate phase connectivity information in the system models. The phase connectivity database maintained by utilities is inaccurate because of a significant amount of missing data, restoration activities, and network reconfiguration. Existing phase identification techniques that estimate phase connectivity work well in distribution feeders that have low or no photovoltaic (PV) generation; however, they fail to identify the phases accurately when considerable PV generation is present. This work addresses the phase identification problem in the presence of high PV generation using statistical analysis methods. Further, insights into the AMI data requirements for this application in terms of data window length and resolution are provided using sensitivity analysis performed on an actual distribution feeder model of San Diego Gas & Electric Company. The results of this study show that the phase connectivity, even in the presence of high PV generation, can be accurately identified using statistical analysis of AMI data of 1 day.**

Index Terms—**advanced metering infrastructure, correlation coefficient, phase identification, power distribution lines, regression analysis.**

I. Introduction

The utility industry has invested more than $5 billion in the deployment of advanced metering infrastructure (AMI) and customer solution technologies in the United States [1]. This investment has brought about important benefits to the utilities, including reduced costs for metering and billing, improved operational efficiencies leading to enhanced revenue collection, and improved outage management and restoration. Automated customer metering and billing is the prevalent use case for the AMI deployments. Remote connect/disconnect and outage management are also being widely used. Utilities in California use AMI data to estimate if the voltages are within the American National Standards Institute voltage limits for normal operations, especially in the presence of smart inverters [2].

AMI deployments can provide measurements from the grid edge that can be indicative of system health and hence provide important insights into system operations. AMI-based analytics and controls present interesting challenges that need to be addressed for integration with traditional utility operations.

Although utility operations (especially distribution automation controls) typically use supervisory control and data acquisition measurements from utility-owned meters, AMI measurements represent a very different class of meters. The size of AMI deployments (several hundreds of thousands of meters) and the long interval between meter reporting (5-minutes for bellwether meters, even longer for non-bellwether meters) require its own class tools and techniques for utility operations and controls. Increasing levels of photovoltaic (PV) penetration add more challenges.

This paper presents the challenges and considerations for AMI-based operational tools by studying the impact of factors such as length of data window, measurement resolution, and PV penetration. The paper focuses on phase identification algorithms and how their accuracy is affected by the factors identified. The study required developing frameworks for generating AMI-like data and for evaluating the effectiveness of different phase identification algorithms (identified through the literature survey in Section II). Section III presents the feeder model under study and the framework for generating synthetic AMI data. Section IV presents in detail the multiple linear regression technique for phase identification that has been prevalently used. Section V presents the sensitivity study, and Section VI concludes the paper with a discussion on future work.

II. Literature Survey

AMI is a major milestone toward the vision of achieving a modernized electric grid. Substantial investments are being made by utilities to improve metering and the communications infrastructure. Outage management, power quality issues, integration of demand response, and distributed energy resources (DER) [3] are some of the benefits of AMI. At the consumer level, smart meters communicate consumption data to both users and the service provider. Time-based pricing, consumption data, net metering, loss of power notification, remote turn ON/OFF, demand response, power quality monitoring, and communication with other intelligent devices are some of the functions of smart meters [4], [5]. Certain advanced functions on smart meters can be programmed to maintain a schedule for the operation of home appliances.

An important application for AMI deployment is enabling the utility to maintain good models of the network. Traditional utility operations use network models for several operational and planning processes. Model accuracy is especially critical for utilities that use advanced distribution automation systems, such as an advanced distribution management system and an outage management system. Although model quality consists of several aspects, customer phase connectivity is of importance in numerous distribution network analysis, monitoring and control applications such as volt-var control, state estimation, and distribution automation.

The operation of distribution systems with increased DER penetration requires accurate feeder models down to the point of interconnection. There is an increasing need to model the secondary low-voltage distribution circuits because the DERs are located at this level. An approach based on linear regression and basic voltage drop equations for phase identifications using smart meter data was proposed in [6], [7]. References [8], [9] proposed k-means clustering algorithm and hybrid clustering for phase identification using AMI data, respectively. Clustering based on the Pearson correlation coefficient and geographical location for phase identification was used in [10]. Principal component analysis and its graph-theoretic interpretation was proposed by [11]. The phase identification problem formulated as an integer programming problem was proposed in [12]. Harmonic voltage correlation for phase identification was proposed by [13]. The majority of these works considered passive distribution networks to determine the phase connectivity. In this work, we demonstrate that the errors are introduced in the phase identification when PV generation is present in the distribution feeder. Further, we show that the statistical analysis can be used to determine the meter phase connectivity accurately even when highly intermittent PV generation is present with the AMI data of 1 day.

III. Distribution Feeder Model and Synthetic AMI Data Generation Process

The study of phase identification accuracy using different statistical methods requires the availability of AMI data at desired locations, capturing the feeder response under different operating conditions, and reporting required feature data. It is important to identify the capabilities of each method of analysis because one method giving accurate results might not work reasonably well in a different system or under other operating conditions of the same system. However, since it is not practical in a distribution feeder to have AMI meters at all the locations providing required measurements at desired resolutions, the project team developed a framework to generate synthetic AMI data using a detailed distribution feeder model from a real feeder from San Diego Gas & Electric Company (SDG&E). The synthetic AMI data set is then used as input to different statistical methods to assess their phase identification performance. This approach is helpful in identifying the AMI data requirements and effective methods that work well under ideal conditions; therefore, the identified methods have the potential to use for this application. This section presents the details of the feeder model used and the procedure followed to generate the synthetic AMI data.

It is assumed that all the smart meters are bellwether meters that can be polled every 5 minutes to obtain i) 5-minute average active power (kW), ii) 5-minute average reactive power (kvar), iii) 5-minute average per-phase voltage. They can be polled every 4 hours to obtain energy register data, energy interval data, and event data. Data collected by smart meters are a combination of parameters such as a unique meter identifier, time stamp of data, and the data mentioned previously. Non-bellwether meters typically report every few hours with measurements of voltage and real and reactive power averaged over 15-minute time intervals.

A. Distribution Feeder Details

The topology of the distribution feeder model used in this work, plotted using the GridPV tool [14], is shown in Fig. 1. This is a 12-kV feeder serving 548 customers with a peak load demand of 10.3 MW. Distributed PV generation of 30% relative to the peak load is present in this feeder. One load tap changer (LTC) at the substation and three fixed capacitor banks are available for voltage regulation. The locations of the fixed capacitor banks and the PV systems are highlighted in Fig. 1.

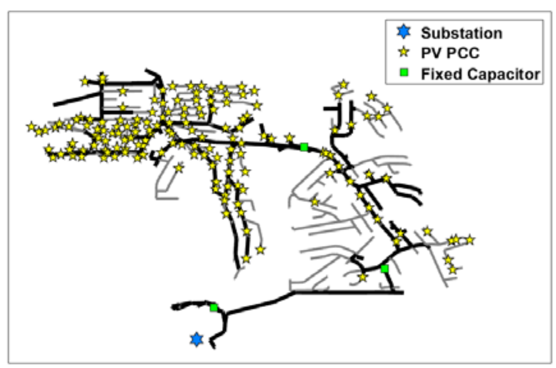

Fig. 1. Topology of the distribution feeder.

Each phase at a given bus in this feeder is referred to as a node in this paper. Thus, a given bus on the feeder might have one, two, or all three phases/nodes. This feeder includes 4,213 nodes total, excluding the nodes at the substation. The percentage distribution of nodes among the phases is 31%, 34.5%, and 34.5% for phases A, B, and C, respectively. Identification of all the 4,213 nodes correctly represents 100% phase identification accuracy.

B. Synthetic AMI Data Generation and Statistical Analysis

The synthetic AMI data generation process is depicted in Fig. 2. The process begins with the simulation of the feeder model in OpenDSS in quasi-static time-series (QSTS) simulation mode for a required time period (1 week/1 month) at a 1-minute time step resolution. The node voltage magnitudes and node real and reactive power demands including those at the feeder head (substation) are saved from the QSTS simulation as the synthetic AMI data. Thus, the original data set contains the AMI data at a 1-minute resolution. These data are then averaged over the required time intervals to generate the AMI data at different time resolutions as needed in the different scenarios discussed in Section V. For example, for data analysis with AMI data of the 5-minute resolution, the 1-minute AMI data are averaged over every five time steps.

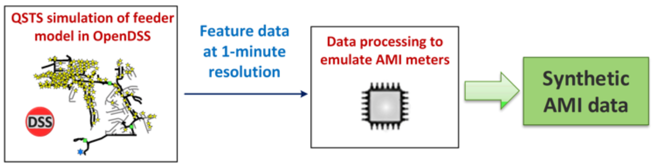

Fig. 2. Synthetic AMI data generation process.

In QSTS mode, all the loads and PV systems are assumed to vary according to the load and PV profiles shown in Fig. 3. The load and PV profile measurement data are provided by SDG&E. Because the synthetic AMI data generated for a period of 1 week are used for most of the analysis presented in this paper, the corresponding profiles are plotted in Fig. 3.

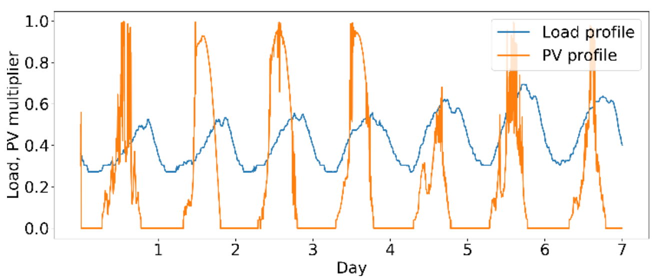

Fig. 3. Load and PV profiles of the 1-week period.

Several scenarios are created by varying PV penetration levels and AMI data resolutions, and the corresponding data are stored in different data sets. The AMI data set of a given scenario is then passed to the statistical analysis module as input to apply multiple statistical methods using different feature sets for phase identification, as illustrated in Fig. 4. At all the nodes in the feeder, the identified phases are compared with the actual phasing information from the simulation model stored separately to check the accuracy of the results.

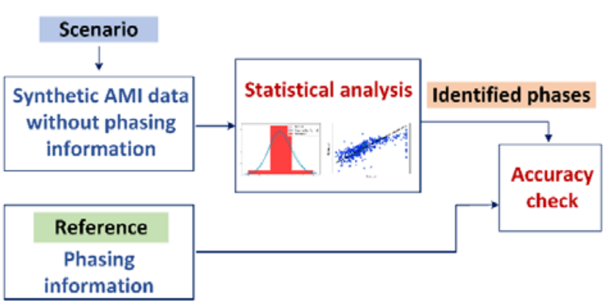

Fig. 4. Phase identification through statistical analysis of AMI data.

IV. Phase Identification Using Multiple Linear Regression

Initially, the multiple linear regression technique with the features reported in [6] is applied to the 1-month data set for the phase identification. In this, the 5-minute averaged measurement data of the node voltage, active power demand at the node, and active power demand at the feeder head are used as the independent variables to predict the voltage at the feeder head (dependent variable), as in (1):

$$V_{sub} = k_0 + k_1 V_n + k_2 P_n + k_3 P_{sub} \qquad (1)$$

where V_{sub} and P_{sub} are the average voltage and average active power demand at a node at the substation; V_n and P_n are the average voltage and average active power demand at a node on the feeder (recorded by a fictitious AMI meter connected to that node); and k_0, k_1, and k_2 are the regression coefficients.

At each feeder node, three linear regression models are formed using the V_{sub} and P_{sub} data of the three nodes (phases A, B, and C) at the substation in (1). The regression models are then used to predict the corresponding substation node voltages, and the coefficient of determination (R2) for each fit is computed. The substation phase having the highest R2 is identified as the phase of that feeder node. This process is repeated for all 4,200+ nodes in the feeder for phase identification.

The scatter plot of measured (data recorded from the QSTS results) substation node voltages and measured voltages at a selected feeder node are shown in Fig. 5 as blue dots. The substation voltages predicted by the regression model are marked by the red dots. Because there are more than two independent features in (1), only one feature is used for this plot for simplicity. Because both the feeder and the substation nodes pertaining to this plot are the same (Phase B), the red dots overlapping the blue dots represent an accurate fit. The R2 value of this fit is 0.99. The distributed PV systems are not included in generating the AMI data used in this scenario.

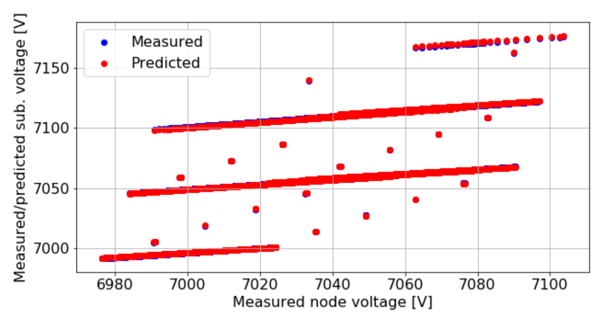

Fig. 5. Predicted substation voltages using multiple linear regression accurately matching with the measured values.

Next, the AMI data generated by including the 30% existing PV generation are used for the regression analysis. The measured and predicted voltages of the same substation node are plotted against the measured feeder node voltages in Fig. 6. Compared to Fig. 5, we observed more errors in the predicted substation voltages in the presence of 30% PV generation. The prediction errors are primarily due to not capturing the PV generation in terms of measurable features. Thus, the regression models formed using (1) become inaccurate. As a result, the R2 value is reduced to 0.74 causing phase identification mismatch at this feeder node. Due to similar prediction errors at the other nodes, the phase identification mismatch increased from 0% in the no PV scenario to 15% in the 30% PV scenario when the multiple linear regression is used.

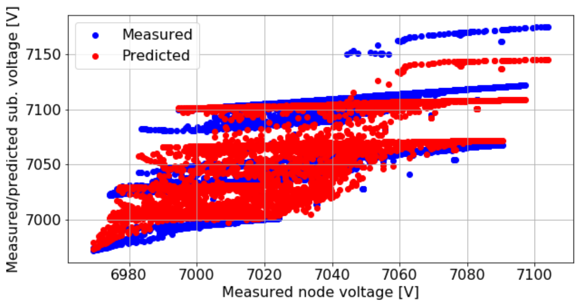

Fig. 6. Errors in the predicted substation voltages using multiple linear regression in the presence of 30% PV generation.

V. Sensitivity of Phase Identification Accuracy to Various Parameters

The AMI data analysis for accurate phase identification involves the employment of an appropriate analysis method, the selection of suitable features, and the use of a sufficient amount of data with good resolution. The impact of each parameter on the phase identification accuracy is studied to develop insights into the AMI data requirements for this application. In addition to the regression analysis discussed in Section IV, four statistical methods (M1 through M4) are tested with appropriate feature data for this study. The results are discussed here.

A. How Much Data is Needed?

The phase identification accuracy of a given method depends on the length of the data window used for the analysis. To study the amount of data required for each statistical method, the data window length used for the analysis is varied in 1-hour increments, and the phase identification accuracy of different methods for each variation is plotted in Fig. 7. It is observed that all the methods except Method 4 could identify all the phases correctly with only a data window length of only 3 hours. Table I. shows the results when the window lengths of 1 week and 1 month are used, and it confirms the robustness of these methods to the data set containing normal system events such as load variations and LTC tap changes. Further, it is evident that 1-day AMI data at a 5-minute resolution can be sufficient under ideal conditions for accurate phase identification when PV is not present.

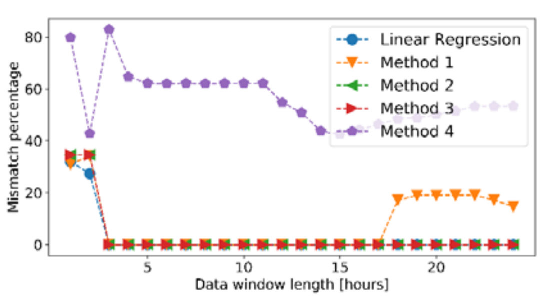

Fig. 7. Sensitivity of phase identification accuracy to data window length.

TABLE I. Phase Identification Mismatches for Different Data Window Lengths

Window Length	LR	M1	M2	M3	M4
1 week	0%	15%	0%	0%	53%
1 month	0%	15%	0%	0%	53%

Next, the existing PV generation in the feeder, which is 30% relative to the peak load, is added to the system to study the impact on the performance of the selected methods. The results, shown in Fig. 8, indicate that the phase identification accuracy is significantly reduced when the feeder has considerable levels of PV generation. Specifically, the linear regression with the feature set reported in [6] no longer provides accurate results with the 1-day data set; however, Method 3 works well with the same data set even in the presence of PV generation. To study if the supply of more data results in the improvement of phase identification accuracy, the data window length is increased to 1 week and 1 month. The results, shown in Table II. reveal that the increase in data window length does not necessarily improve the accuracy of these methods as the mismatches did not reduce when 1-month data window length is used compared to when 1-week data set is used. Method 3 consistently provided accurate results with the longer data windows.

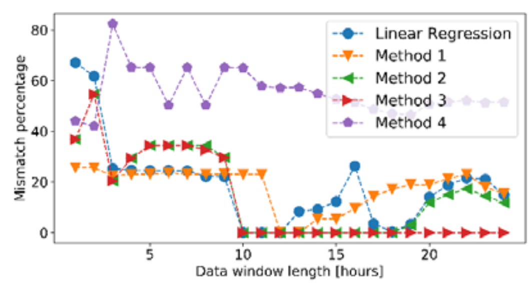

Fig. 8. Impact of the existing PV generation in the feeder on phase identification accuracy.

TABLE II. Phase Identification Mismatches with 30% Existing PV Generation in the Feeder

Window Length	LR	M1	M2	M3	M4
1 week	15%	16%	12%	0%	52%
1 month	15%	16%	12%	0%	52%

B. How Do PV Generation Levels Impact Accuracy?

Sections IV and V-A showed that PV generation can cause phase identification errors. This section studies the performance of the different methods as the PV penetration levels are varied.

The feeder model without any PV systems is considered as the baseline scenario. Then PV systems are added to the customers selected randomly from the set of all customers. The size of the PV systems at each location is selected randomly between 0 kW and 50 kW. For the 10% PV penetration level, the PV systems are added in this way until the total active power output of the PV systems is less than or equal to 10% of the peak load demand of the feeder. This process is repeated for the other PV penetration levels.

The AMI data are generated for each PV penetration level through the QSTS simulation of a 1-week period. Statistical analysis is performed on the 5-minute averaged data of this period for phase identification. The results are shown in Fig. 9. It is observed that the linear regression provided accurate results up to 20% PV penetration level beyond which errors are introduced in the phase identification. Method 3 gave 100% accurate phase identification results for all the PV penetration levels. Method 2, with small percentage of errors at some PV penetration levels, also found to be promising with low phase identification mismatches.

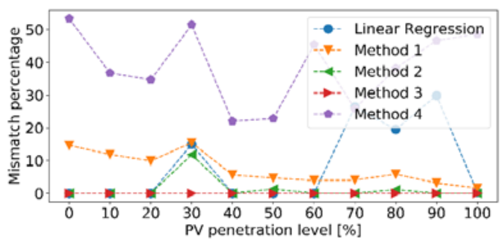

Fig. 9. Sensitivity of phase identification accuracy to PV penetration levels.

C. *How Granular Should the AMI Data Be?*

To study the requirements of AMI data granularity, the impact of different AMI data resolutions on phase identification accuracy is studied. To generate synthetic AMI data at different resolutions, data averaging is performed at different time intervals on the AMI data set generated without the PV systems in the feeder. The time intervals considered are 1 minute, 5 minute, 10 minute, 15 minute, and 30 minute. Because the AMI data of the 1-week window length was found to be sufficient for obtaining accurate results, as shown in Section V-A, 1-week data are used for this analysis. The phase identification mismatches for data with different resolutions are plotted in Fig. 10. The results show that all the methods are mostly robust to the changes in the AMI data resolution. The phase identification mismatches remained constant as the data resolution reduced from 1 minute to 30 minutes, except for Method 2 which showed slight increase in the errors.

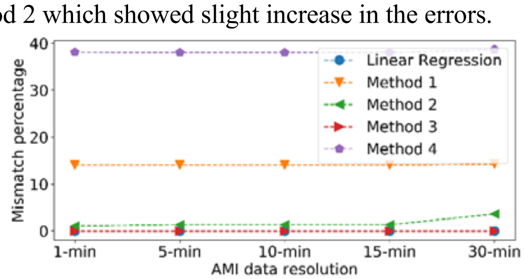

Fig. 10. Sensitivity of phase identification accuracy to AMI data resolution.

VI. Conclusions and Future Work

Existing phase identification techniques work well in the distribution feeders having low or no PV generation. They fail to identify the phases accurately when considerable PV generation is present as the features used in these methods ignore the impact of PV generation in forming the voltage prediction models. In this work, the phase identification problem in the presence of high PV generation is addressed. The feasibility of accurate phase identification in the presence of high PV generation using statistical analysis is demonstrated. Further, insights into the AMI data requirements for phase identification in terms of data window length and resolution are provided. Future work involves the application of the studied methods on the AMI data recorded in the field for accurate phase identification.

VII. Acknowledgment

This work was authored in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by U.S. Department of Energy Office of Energy Efficiency and Renewable Energy. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

VIII. References

[1] "Advanced Metering Infrastructure and Customer Systems," US Department of Energy, Technical Report, Sep. 2016.

[2] California Public Utilities Commission (CPUC), "Resolution E-4898. Approval, with Modifications, of Request for Modifications to Electric Rule 21 Tariff to Incorporate Smart Inverter Phase 3 Advanced Functions in Compliance with Decision 16-06-052.," Apr. 2018.

[3] B. Seal, "Advanced metering infrastructure (AMI) for distributed solar (PV) integration : Utility survey results," Electric Power Research Institute, 2009.

[4] S. S. S. R. Depuru, L. Wang, V. Devabhaktuni, and N. Gudi, "Smart meters for power grid — Challenges, issues, advantages and status," in *IEEE/PES Power Systems Conference and Exposition*, 2011, pp. 1-7.

[5] R. R. Mohassel, A. Fung, F. Mohammadi, and K. Raahemifar, "A survey on Advanced Metering Infrastructure," *International Journal of Electrical Power & Energy Systems,* vol. 63, pp. 473 - 484, 2014.

[6] T. A. Short, "Advanced metering for phase identification, transformer identification, and secondary modeling," *IEEE Transactions on Smart Grid,* vol. 4, pp. 651-658, 2013.

[7] J. Peppanen, S. Grijalva, M. J. Reno, and R. J. Broderick, "Distribution system low-voltage circuit topology estimation using smart metering data," in *2016 IEEE/PES Transmission and Distribution Conference and Exposition (T&D)*, 2016, pp. 1-5.

[8] 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 *Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference on*, 2016, pp. 259-265: IEEE.

[9] W. Wang and N. Yu, "Advanced Metering Infrastructure Data Driven Phase Identification in Smart Grid," 2017.

[10] V. Arya and R. Mitra, "Voltage-based clustering to identify connectivity relationships in distribution networks," in *2013 IEEE International Conference on Smart Grid Communications*, 2013, pp. 7-12.

[11] J. P. Satya, N. P. Bhatt, R. Pasumarthy, and A. Rajeswaran, "Identifying Topology of Power Distribution Networks Based on Smart Meter Data," *CoRR,* vol. abs/1609.02678, 2016.

[12] V. Arya *et al.*, "Phase identification in smart grids," in *2011 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2011, pp. 25-30.

[13] J. D. Watson, J. Welch, and N. R. Watson, "Use of smart-meter data to determine distribution system topology," *The Journal of Engineering,* vol. 2016, no. 5, pp. 94-101, 2016.

[14] M. J. Reno and K. Coogan, "Grid integrated distributed PV (GridPV) version 2," *Sandia National Labs SAND2014-20141,* 2014.