A Machine Learning-Based Method to Estimate Transformer Primary-Side Voltages with Limited Customer-Side AMI Measurements

Preprint

Jiyu Wang,[1] Harsha Padullaparti,[1] Santosh Veda,[1] Murali Baggu,[1] Martha Symko-Davies,[1] Amin Salmani,[2] and Tom Bialek[2]

1 National Renewable Energy Laboratory
2 San Diego Gas & Electric Company

Presented at the 2021 IEEE Power and Energy Society General Meeting
July 26–29, 2021

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-5D00-78299
August 2021

A Machine Learning-Based Method to Estimate Transformer Primary-Side Voltages with Limited Customer-Side AMI Measurements

Preprint

Jiyu Wang,[1] Harsha Padullaparti,[1] Santosh Veda,[1] Murali Baggu,[1] Martha Symko-Davies,[1] Amin Salmani,[2] and Tom Bialek[2]

1 National Renewable Energy Laboratory
2 San Diego Gas & Electric Company

Suggested Citation
Wang, Jiyu, Harsha Padullaparti, Santosh Veda, Murali Baggu, Martha Symko-Davies, Amin Salmani, and Tom Bialek. 2021. *A Machine Learning-Based Method to Estimate Transformer Primary-Side Voltages with Limited Customer-Side AMI Measurements: Preprint*. Golden, CO: National Renewable Energy Laboratory. NREL/CP-5D00-78299. https://www.nrel.gov/docs/fy21osti/78299.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-5D00-78299
August 2021

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

NOTICE

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 and by Electric Program Investment Charge (EPIC) Program in San Diego Gas & Electric Company under a cooperative research and development agreement (CRADA) # CRD-17-712. The views expressed herein do not necessarily represent the views of the DOE or the U.S. Government.

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

U.S. Department of Energy (DOE) reports produced after 1991 and a growing number of pre-1991 documents are available free via www.OSTI.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.

A Machine Learning-based Method to Estimate Transformer Primary-Side Voltages with Limited Customer-Side AMI Measurements

Jiyu Wang[1], Harsha Padullaparti[1], Santosh Veda[1], Murali Baggu[1], Martha Symko-Davies[1], Amin Salmani[2], Tom Bialek[2]
[1]National Renewable Energy Laboratory (NREL), Golden, Colorado.
[2]San Diego Gas & Electric Company (SDG&E), San Diego, California.
Santosh.Veda@nrel.gov

Abstract— **Distribution control applications such as volt/var optimization, network reconfiguration, and distribution automation require accurate knowledge of the distribution system state. The lack of sufficient sensors on the primary side of distribution networks often limits the accuracy of the control decisions by these applications. The deployment of advanced metering infrastructure (AMI) provides utilities an opportunity to translate the AMI data on the secondary onto the primary so that it can be used as pseudo-measurements to augment the limited existing measurements on the primary. This paper develops a machine learning based approach for estimating service transformer primary-side voltages by using limited secondary-side AMI measurement. The machine learning model is developed by using random forest algorithm. The estimated primary-side voltages can be used by utilities as pseudo-measurements for distribution control applications. The detailed secondary model topology, which is an essential input data for many existing algorithms, is not required for the proposed method. The performance of the proposed method is validated by using AMI measurements from the field and an actual distribution feeder model of San Diego Gas & Electric Company.**

Index Terms— **Advanced metering infrastructure (AMI), distribution system, service transformer, smart grid, voltage estimation.**

I. INTRODUCTION

Obtaining accurate information of the system state is critical for many distribution control applications including volt/var optimization, network reconfiguration, and distribution automation in today's modernized grid [1]. These control applications can help improve grid operations by solving problems caused by increasing integration of distributed energy resources (DERs) [2]–[4]. Nevertheless, the lack of enough sensors on the feeder primary side might limit the accuracy of these control decisions. To avoid these issues during operation, electric utilities require accurate feeder models for planning and real-time data measurement for control implementations [5]. Therefore, utilities in the United States have deployed advanced metering infrastructure (AMI) to further modernize the grid [6].

The deployment of AMI on the customer side can enable the collection of smart meter measurements from the grid edge to improve the efficiency of utility system operation and reduce the cost for metering and billing. The AMI measurement can also help with feeder analysis, such as phase identification [7], transformer identification, identifying energy thefts [8], and load forecasting [9]. In addition, because the AMI measurements are usually from the customer side, the secondary model and service transformer primary-side voltage can be estimated by the collected data as well. The estimated voltage at the service transformer primary side is important to electric utilities for network model validation, state estimation, and control implementation.

The methods using AMI measurements to estimate secondary model topology and primary-side voltage have been studied by researchers. Bottoms-up regression method was used in [8] to group the AMI data from two transformers that are electrically close and estimate the primary voltage; however, the estimated primary voltage is for the constructed mock circuit based on AMI data instead of the actual distribution feeder. The works reported in [10] and [11] developed a linear regression method to estimate the distribution system secondary-side parameters and topology; however, this method requires AMI measurements from all customers on the feeder. In reality, utilities might not be able to obtain these measurements because of customer privacy concerns. An algorithm to improve the network connectivity topology was proposed in [12]. One assumption of this algorithm is that the voltage magnitude decreases downstream along the feeder, which means that it will not work when DERs are integrated. In [13], the authors presented a simplified DER-integrated distribution system secondary circuit model estimation method when limited AMI measurements were available. However, it assumes that each secondary circuit had one or more PV systems, which is not realistic in all distribution systems.

In this context, the contributions of this paper are:

(1) A machine learning-based method is developed to estimate the service transformer primary-side voltages with limited AMI measurements from the secondary side. The estimated voltages on the primary can be used as pseudo-measurements for distribution control applications.

(2) The proposed approach does not require detailed topological information of the secondary networks which is often complex and not available with the utilities.

The rest of this paper is organized as follows. Section II describes the models and data used in this study. Section III introduces the proposed machine learning-based methodology to estimate the transformer primary-side voltage. The simulation and method validation results are shown and analyzed in Section IV. Section V concludes this paper and presents potential future work.

II. Feeder Model and Field Data

This section details the distribution feeder modeling and the data sets used in this study.

A. *Feeder Model*

An actual distribution feeder model developed in the Synergi software platform is received from San Diego Gas & Electric Company (SDG&E) for this study. This is a 12-kV feeder with a peak load of 10.3 MW. The topology of the feeder plotted using the GridPV toolbox [14] is shown in Fig. 1. The substation transformer is equipped with a load tap changer. Three capacitor banks are available on the feeder for reactive power support. The feeder serves more than 5,000 customers using 341 service transformers. Solar generation of approximately 70% relative to the peak load is present in this feeder at the locations highlighted in Fig. 1.

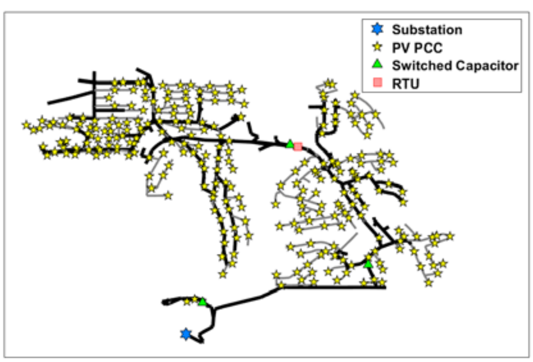

Fig. 1. Topology of the distribution feeder

The original Synergi feeder model is converted to OpenDSS using the Distribution Transformation Tool (DiTTo) model conversion tool to perform quasi-static time-series (QSTS) simulation [15], [16]. The conversion is validated by comparing the voltage mismatches between the Synergi and OpenDSS power flow results, which is shown in Fig. 2. It can be observed that the mismatches are very low, which confirm the accuracy of the model conversion process.

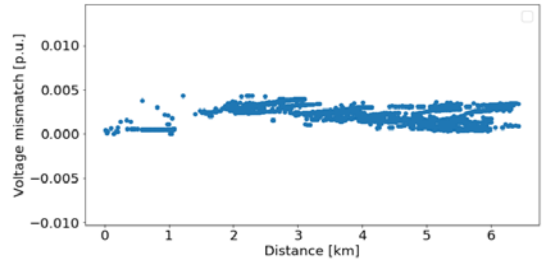

Fig. 2. Bus voltage mismatches between Synergi and OpenDSS power flow results.

B. *AMI Data*

AMI data are recorded at the secondary side of each service transformer in the SDG&E feeder. The AMI data set includes the voltage and real power measurements from two customers and the total real power consumption at the secondary side of each service transformer. The AMI data recorded for a period of 107 days from October 2018 to January 2019 are used. The data resolution is 1-hour for the real power measurements and 5-minute for the voltage measurements at the load locations. The voltage distribution for all AMI measurements on a selected day is shown in Fig. 3.

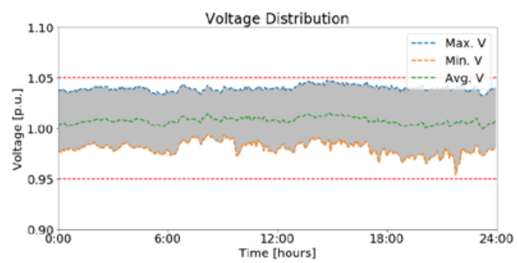

Fig. 3. Voltage distribution for all AMI measurements for a selected day.

C. *SCADA Data*

The historical SCADA measurements at the feeder head include line-to-line voltages, line currents, and three-phase real and reactive power. Additionally, primary-side voltage measurements from a remote terminal unit (RTU) installed on the feeder are available in the SCADA system. The RTU location is highlighted in Fig. 1.

III. Methodology

A machine learning-based approach is proposed in this work to estimate the service transformer primary-side voltage by using the AMI measurements on the secondary side. The details of this approach are presented in this section.

A. *Synthetic Primary-Side Voltage Generation*

Machine learning approaches typically require a training data set that contains the features to be estimated. In this application, the inputs include the AMI measured power and voltages at two customers under each service transformer and the total power consumption of all customers under the service transformer. The output is the transformer primary-side voltage. Therefore, the transformer primary voltage data must be included in the training data set in addition to the other specified feature data; however, because no primary-side measurements are available in this feeder except the RTU voltage, as described in Section II, time-series voltage data recorded from the simulations in OpenDSS are used to form the required training data set. The feeder is simulated in OpenDSS with QSTS mode for the whole 107 days to obtain the primary-side voltages. In the QSTS simulation, the time resolution is hourly to follow the AMI load time resolution. The load profile of each secondary-side measured load is set to be the AMI measured total power under that transformer. The simulated primary-side synthetic voltage from the QSTS simulation and the actual measured secondary-side voltages at the two AMI measured loads recorded are used to train the machine learning model.

B. *Machine Learning Model*

Multiple machine learning algorithms—random forest, adaptive boosting, and gradient boosting—are tested to find the relationship between the primary-side voltages and the AMI measurements under each service transformer [17]-[19]. The data from each service transformer (341 in total) will be trained separately to account for their unique characteristics, i.e., separate models are constructed for each service transformer. The input of each model is the hourly load measurement from two AMI meters under that service transformer, the average hourly AMI voltage measurement, and the total load of that service transformer. The output of the model is the voltage on the primary side of the service transformer.

The data from first month are selected to compare the estimation accuracy of different algorithms. K-fold cross validation is used to validate the machine learning models, and the validation is repeated 30 times. In each test, 80% of the monthly data are randomly drawn from the data set to train the model, and the remaining 20% are used for testing. The mean absolute percentage error (MAPE) and maximum absolute percentage error between the synthetic primary voltage and estimated primary voltage is used to evaluate the performance of each machine learning method. The performance comparison is shown in TABLE I.

TABLE I. PERFORMANCE OF DIFFERENT METHODS

	Machine Learning Method		
	Random Forest	**AdaBoost**	**Gradient Boost**
MAPE	0.12%	0.75%	0.48%
Maximum	0.46%	1.08%	0.95%

The results summarized in TABLE I show that random forest model performs better than the other two models in the selected performance criteria; therefore, it is selected to estimate the primary-side voltages in this study. Another advantage of using random forest algorithm is as random forest is an ensemble learning method that integrates multiple decision trees, it will combine these decision trees and use average or voting schemes to calculate the results. Therefore, the outliers in the AMI measurements can be well handled with this algorithm. Further, an exhaustive search is conducted to determine the model parameters (number of decision trees and maximum depth). These two parameters are varied from 1 to 500 and 1 to 30, respectively, to test the estimation performance. Considering both estimation accuracy and training time, the number of decision trees are selected to be 80 and the maximum depth to be 10. The time to build the machine learning model for each service transformer is around 5 seconds, and the total time for building the models for all service transformers is 30 minutes. As the process of training model is usually developed for the distribution system planning studies, it meets the run-time requirement.

IV. CASE STUDIES

This section presents the case studies for validating the proposed approach by using both simulated and actual voltage data. The training and test dataset for each case study is summarized in TABLE II.

TABLE II. SUMMARY OF DATA USED FOR EACH CASE STUDY

		Case 1	Case 2	
			2a	**2b**
Train	**Primary voltage**	Simulated	Actual	Simulated
	Secondary voltage	Simulated	Actual	Actual
	Load	Actual	Actual	Actual
Test	**Primary voltage**	Simulated	Actual	Actual
	Secondary voltage	Simulated	Actual	Actual
	Load	Actual	Actual	Actual

A. *Case 1: Validation with Simulated Data*

The proposed machine learning model is first validated by the synthetic primary-side voltage generated from the QSTS simulation. A secondary model is built for each service transformer in OpenDSS. Each secondary model includes the two loads with voltage measurements and a load without voltage measurement. The load profiles of each secondary measured load are set to be one AMI measured power under that transformer. The load profile of the unmeasured load is set to be the AMI measured total power at that service transformer minus the two measured loads. The primary-side and secondary-side voltages at the two AMI measured loads recorded from the QSTS simulation are used to train the machine learning model. The data from the first 1,000 hours are used as training data set to train the model for each service transformer, and the next 1,568-hour data are used to test the performance of each machine learning model.

The MAPEs for the estimation of all service transformer primary-side voltages are shown in Fig. 4. All of them are less than 0.07%. Although the largest estimation error is around 0.65%, the number of such occurrences is very small. For most estimations, the error is less than 0.02%. Overall, the MAPE for all predictions in the feeder is 0.012%, and the MAPE for the service transformer with maximum error is

0.056%. The comparison between estimated and actual voltages (synthetic voltage, in this case) for one example service transformer is shown in the two subplots of Fig. 5. The first subplot shows the voltage comparison, and the second subplot shows the estimation absolute percentage error at each time step. Generally, the shape of the estimated voltages follows the actual voltages. The mismatch between the estimated and actual voltages is within 0.2%, which is very small. The model is also tested when using the first 2000-hour data as training dataset and test with the rest 568-hour data. The performance is similar to the previous case, which means over-fitting problem does not exists for the model.

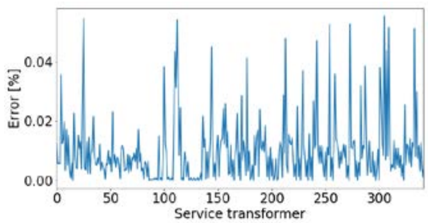

Fig. 4. MAPE for the voltage estimation of each service transformer.

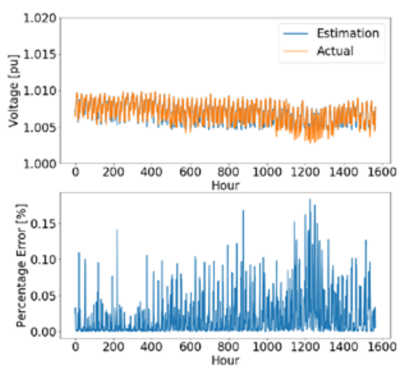

Fig. 5. Comparison between estimated and synthetic voltages for one example service transformer.

B. *Case 2: Validation with Actual Data*

In Section IV-A, the trained machine learning model is validated with synthetic primary-side voltage data. However, the synthetic voltages are from the simulation and because there are differences between the OpenDSS feeder model and the operation in reality, the machine learning method still needs to be validated with actual data. The SCADA voltage measurement from the RTU location will be used as the primary-side voltage measurement. A service transformer that has the shortest distance to the RTU location is selected for the validation.

1) *Case 2a: Model trained by Actual Data*

In this validation, we use the first 1,000-hour SCADA measured primary-side voltages to train the model, and we validate it with the rest of the 1,568-hour data. The validation results are shown in Fig. 6 and Fig. 7. The plots show that all estimation errors are within 1% range, and most errors are smaller than 0.5%. Because the operation condition in reality is much more complex than the simulation, these errors are larger than the result in the previous subsection. It means that if the electric utilities can record the primary-side voltages for a period of time for each service transformer by using a movable meter, however, the trained models could estimate the voltages with small errors.

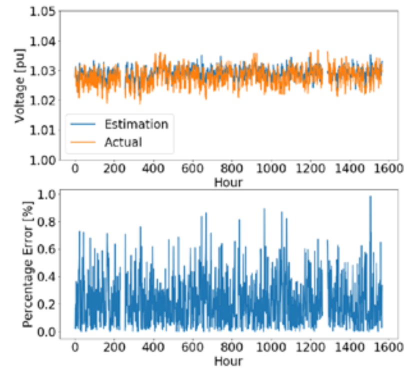

Fig. 6. Comparison between actual voltages and estimated voltages from actual data trained model

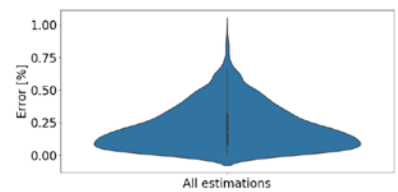

Fig. 7. Violin plot of all estimation errors with actual data trained model

2) *Case 2b: Model trained by Simulated Data*

In this validation, we use the SCADA measured primary-side voltages to validate the model trained by the synthetic data. Sometimes the utilities will not collect any primary voltage measurements. Under this situation, we can use the model trained by synthetic primary-side voltages and actual AMI measurements to estimate the actual primary-side voltages. We input the 2,568-hour measurements from the two AMI meters and the total load consumption under that service transformer into the model trained by the synthetic data, and the estimated voltages are compared with the actual SCADA data. The comparison results are shown in the two subplots in Fig. 8. The first subplot shows the voltage comparison. The shape of the estimated voltages follows the actual voltages, but the deviations are much larger. The percentage errors are shown in the second subplot, and there are some estimation errors larger than 2%. Overall, the average percentage error is 0.57%, and the largest error is 2.20%.

The error distribution is shown in Fig. 9. Although the overall estimation errors are larger compared with the validation in the previous two cases, most errors are still less than 1%. For more than 85% of the time, the estimation error is within 1%, and there is only 5% of time the error is larger than 1.5%. The results show that when field data measurements are not available, the machine learning models can be trained by synthetic data to estimate actual primary voltages with reasonable accuracy. If more sensors are available on the primary to measure actual primary voltages could be measured for a period of time, the estimation accuracy can be improved further. The addition of available AMI measurements under each service transformer would also help improve the results.

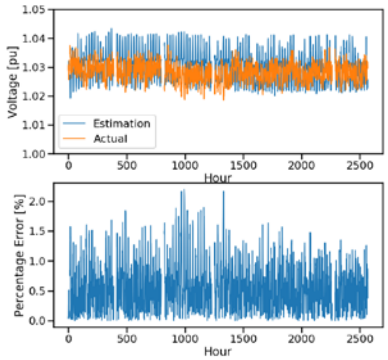

Fig. 8. Comparison between actual voltages and estimated voltages from synthetic data trained model

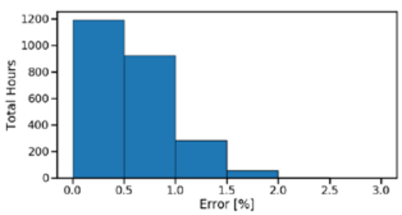

Fig. 9. Estimation error distribution with synthetic data trained model

V. CONCLUSION

This paper presented a machine learning-based method to estimate the service transformer primary-side voltage by using limited AMI measurements from the secondary side. A realistic feeder model and AMI measurements were used in this study. The performances for different machine learning methods were compared, and the random forest model was used in the estimation. The parameters in the random forest model were selected based on an exhaustive search. Three case studies were conducted, and the proposed method was validated with the model trained by synthetic and actual primary voltage data. The validation results demonstrated the advantage of the proposed method. As part of the future work, we will compare the proposed method with the other state-of-the-art methods to show the effectiveness.

ACKNOWLEDGMENT

This work is authorized 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 and by Electric Program Investment Charge (EPIC) Program in San Diego Gas & Electric Company under a cooperative research and development agreement (CRADA) # CRD-17-712. 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.

REFERENCES

[1] J. R. Aguero, E. Takayesu, D. Novosel and R. Masiello. "Modernizing the grid: challenges and opportunities for a sustainable future." *IEEE Power and Energy Magazine*, vol. 15, no. 3, pp. 74-83, 2017.

[2] X. Zhu, J. Wang, N. Lu, N. Samaan, R. Huang and X. Ke, "A Hierarchical VLSM-Based Demand Response Strategy for Coordinative Voltage Control Between Transmission and Distribution Systems," in *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 4838-4847, Sept. 2019.

[3] F. Ding, H. Padullaparti, M. Baggu, S. Veda, and S. Meor Danial. *"Data Enhanced Hierarchical Control to Improve Distribution Voltage with Extremely High PV Penetration."* No. NREL/CP-5D00-72755. National Renewable Energy Lab.(NREL), Golden, CO (United States), 2019.

[4] M. Karimi, H. Mokhlis, K. Naidu, S. Uddin, and A. H. A. Bakar. "Photovoltaic penetration issues and impacts in distribution network–A review." *Renewable and Sustainable Energy Reviews,* vol. 53, pp.594-605, 2016.

[5] J. Wang, X. Zhu, M. Liang, Y. Meng, A. Kling, D. Lubkeman and N. Lu, "A Data-Driven Pivot-Point-Based Time-Series Feeder Load Disaggregation Method," in *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 5396-5406, Nov. 2020, doi: 10.1109/TSG.2020.3008603.

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

[7] H. Padullaparti, S. Veda, S. Dhulipala, M. Baggu, T. Bialik and M. Symko-Davies. "Considerations for AMI-Based Operations for Distribution Feeders". National Renewable Energy Lab.(NREL), Golden, CO (United States), 2019.

[8] T. A. Short, "Advanced Metering for Phase Identification, Transformer Identification, and Secondary Modeling," in *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 651-658, June 2013.

[9] T. Hong, J. Wilson and J. Xie, "Long Term Probabilistic Load Forecasting and Normalization With Hourly Information," in *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 456-462, Jan. 2014.

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

[11] M. Lave, M. J. Reno and J. Peppanen, "Distribution System Parameter and Topology Estimation Applied to Resolve Low-Voltage Circuits on Three Real Distribution Feeders," in *IEEE Transactions on Sustainable Energy*, vol. 10, no. 3, pp. 1585-1592, July 2019.

[12] W. Luan, J. Peng, M. Maras, J. Lo and B. Harapnuk, "Smart Meter Data Analytics for Distribution Network Connectivity Verification," in *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 1964-1971, July 2015.

[13] J. Peppanen, M. J. Reno, R. J. Broderick and S. Grijalva, "Distribution System Model Calibration With Big Data From AMI and PV Inverters," in *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2497-2506, Sept. 2016.

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

[15] T. Elgindy, N. Gensollen, D. Krishnamurthy, M. Rossol, E. Hale & B. Palmintier. *DiTTO (Distribution Transformation Tool)* (No. DiTTO). National Renewable Energy Lab.(NREL), Golden, CO, 2018

[16] IEEE P1547.7 D5.1 Draft Guide to Conducting Distribution Impact Studies for Distributed Resource Interconnection, IEEE Standard, Aug. 2011.

[17] A. Liaw, and M. Wiener. "Classification and regression by randomForest." *R news* 2.3 (2002): 18-22.

[18] Y. Freund and RE. Schapire. "Experiments with a new boosting algorithm." *icml.* Vol. 96. 1996.

[19] L. Breiman. "Arcing the edge.", *Technical Report 486*, Statistics Department, University of California at Berkeley, 1997.