SAND2019-13628C

Identifying Errors in Service Transformer Connections

Logan Blakely, Matthew J. Reno
Electric Power Systems
Sandia National Laboratories
Albuquerque, NM, USA
lblakel@sandia.gov

***Abstract*—Distribution system models play a critical role in the modern grid, driving distributed energy resource integration through hosting capacity analysis and providing insight into critical areas of interest such as grid resilience and stability. Thus, the ability to validate and improve existing distribution system models is also critical. This work presents a method for identifying service transformers which contain errors in specifying the customers connected to the low-voltage side of that transformer. Pairwise correlation coefficients of the smart meter voltage time series are used to detect when a customer is not physically connected to the transformer that is specified in the model. The proposed method is demonstrated both on synthetic data as well as a real utility feeder, and it successfully identifies errors in the transformer labeling in both datasets.**

Keywords*—*AMI, correlation coefficients, distribution system modeling, transformer errors

I. Introduction

Utility models of the electric grid form the basis of the simulations that inform control decisions, infrastructure investment, hosting capacity analyses, and many other grid applications. Historically, the utility models of the distribution system have a larger quantity of errors compared to the transmission system, [1], and this is especially true for the low-voltage portion of the system. [2] provides an overview, literature review, and several examples of the types of errors that are often present in utility models of the distribution system.

These errors in the model are included in grid simulations and propagate through to the results. For example, hosting capacity analyses are critical for evaluating potential distributed energy resources (DER) and reduced accuracy in those analyses is an obstacle, [3]–[5]. The connections between customers, or meters, and the low-voltage network is one area of utility models that contains errors. Due to ongoing maintenance and record keeping, the correct connection information may not be known between a particular meter and a service transformer. This meter to service transformer mapping error can affect the simulations discussed earlier, but it also has a negative impact on equipment usage. Both overloading transformers and not fully using the potential of transformers are cases that would be avoided in an optimal configuration. Thus, accurate specifications of which service transformer each meter is connected to are necessary for optimal infrastructure usage and accurate simulations.

This work leverages the recent availability of data from advanced metering infrastructure (AMI), or smart meters to identify customer to transformer mapping errors, specifically by flagging transformers which contain meters specified in their low-voltage network that are actually located under a different service transformer. The proposed algorithm produces a list of transformers which contain errors; this list can be used to efficiently direct utility resources to correct those errors. The primary contributions of this paper are as follows:

1) A method for identifying meter to transformer mapping errors that does not inject new errors into the model.

2) A straightforward, easily-interpretable, data-driven method for identifying meter to transformer mapping errors.

3) A significant reduction in required utility resources to find/correct meter to transformer mapping errors by providing a list of transformers to focus resources on.

The remainder of the paper is structured as follows. Section II provides an overview of related work in this area, Section III details the methodology used in this work, Section IV covers results on a synthetic dataset; examples from a utility dataset; and a brief comparison with similar algorithms, Section V details future work items, and Section VI provides a conclusion.

II. Related Work

This section provides an overview of related research in this area. The use of correlation coefficients for this type of model correction is well-documented in literature.

The work in [1] and [6] provides some of the foundational work and inspiration for the proposed algorithm. The authors used the voltage time series collected from AMI meters, calculated a point-of-coupling (POC) voltage using the line impedance and current, and calculated pairwise correlation coefficients from the POC time series. The correlation coefficients are then used both to identify customer to transformer errors and identify the correct placement of those customers, and both aspects must be successful for the method to work. This work was field validated on a 700-customer feeder in Vancouver, Canada.

[7] uses correlation coefficients combined with a two step clustering process to solve the meter to transformer pairing problem. First, customers are clustered spatially, either using DBSCAN or using pre-existing knowledge about the feeder laterals, and then customers are clustered using K-means with

This material is based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under Solar Energy Technologies Office (SETO) Agreement Number 34226. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

correlation as a distance metric. The authors report 80%-90% accuracy in their results. Given the importance of the simulations using these models, better accuracy is desirable.

[8] proposes a method based on linear regression for the meter to transformer mapping task. The method uses the POC approach, grouping pairs customers with the highest R^2 fit value from the linear regression in a hierarchical fashion, combining paired customers into a POC until a complete tree is built. This work was validated on a small dataset of 36 transformers.

[9] uses an approach based on calculating the POC voltage for each customer labeled on a transformer and comparing the resulting profiles for irregularities. The results in this work are proof-of-concept examples. The authors note that this approach could potentially be automated and be done hierarchically, resulting in a tree structure, but that is left to future work.

In [10], the authors calculate a pairwise 'concentration matrix', which is a type of correlation, and use that to build a minimum spanning tree that represents the radial structure of the distribution system. This work does not explicitly discuss meter to transformer mapping, however the tree structure represents similar knowledge. [11] uses phasor measurement unit (PMU) data with the Chow-Liu algorithm for the topology estimation problem. Utilization of this method requires access to PMU data.

In [12] the authors use a two stage method for pairing distribution transformers with the correct feeder label; they separate the detection of an error from the correction step, as we also propose to do in our method. The first stage flags suspicious transformers based an R^2 fit value from a linear regression using voltage time series, and the second stage corrects the pairing label.

One key consideration in all of these methods is the question of whether the method potentially injects new error into the model. The body of research discussed in this section does not touch on this topic, and this may be a hinderance in applying these methods in the field. A major advantage of the method proposed in this paper is that it is incapable of adding additional error to the utility model and only requires AMI voltage time series data. For further experimental results from this work comparing it with the proposed method to methods from [1], [8] please see Section IV C.

III. Methodology

The proposed method leverages the concept that customers connected to the same service transformer will have voltage time series that are more correlated than two customers that are connected to different transformers. This fact is well demonstrated in the literature, both for customer to transformer pairing research, as well as for customer phase identification [7], [8], [13]. Pearson correlation coefficients were calculated between the voltage time series of each pair of customers to produce a pairwise correlation coefficient matrix. Although AMI meters may record other information, only the voltage measurements are used in this method. POC voltage could be explored as future work, but the results shown here do not implement the POC technique described above. This work focuses on the problem of flagging service transformers that contain an incorrectly specified customer (i.e. a customer that is actually connected to a different service transformer). Figure 1 shows a conceptual illustration of the proposed method. There are four customers that are specified as being connected in the low voltage network of a single-phase transformer, and the table shows the pairwise correlation coefficients of the voltage time series for this set of customers. Customers 1-3 are highly correlated with each other, while Customer 4 is not well correlated with any of the other customers, suggesting that Customer 4 is in a different low voltage network, connected to a different service transformer. In this case, the transformer would be flagged for further analysis by utility personnel. This method is currently designed for single-phase customers, and the full set of customers is down selected to include only single-phase customers as a pre-processing step. The same algorithm could work for identifying 3-phase customers on the same transformers if voltage measurements from all three phases (or average phase voltage) is provided. Future work will investigate identifying combinations of 3-phase and single-phase customers on the same transformers.

	Cust #1	Cust #2	Cust #3	Cust #4
Cust #1	1	0.9	0.9	0.2
Cust #2	0.9	1	0.9	0.2
Cust #3	0.9	0.9	1	0.2
Cust #4	0.2	0.2	0.2	1

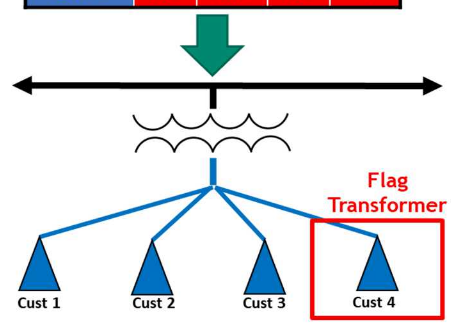

Figure 1 - Conceptual example of the proposed method

Two other pre-processing steps are implemented prior to beginning the transformer error flagging process. First, the voltage time series are normalized around a mean of 1. Second, the voltage time series are converted into a 'voltage difference' representation. The difference is taken between adjacent, time-consecutive measurements to produce the transformed time series. The resulting time series are reduced in length by one measurement and can now be interpreted as the (normalized) voltage change between time steps. The efficacy of these steps has been demonstrated in [7], [13].

Figure 2 shows a flowchart of the proposed methodology in more detail. In Step 1, the pairwise correlation coefficient matrix is calculated. This methodology uses a 'window' methodology to calculate the correlation coefficients, [7], [13]. A 'window' of available data, 4-days in this case, is selected, any customers with missing data during this window are removed, and the pairwise correlation coefficients are then calculated for the remaining customers. This process is repeated for subsequent windows until all available data has been utilized. This approach has several advantages. First, it allows a way to deal with datasets containing missing data; second, it enables the algorithm to be more scalable in the case of large datasets; and finally, it permits flexibility in the calculation of the final pairwise correlation coefficient. In this algorithm, the median of all pairwise values across the available windows is used as the final correlation coefficient, but the window approach allows for the choice to use the mean value, do outlier detection before using the values, etc.

In Step 2, a group of customers specified as being on the same transformer are selected, and the correlation coefficients are analyzed. If any of the pairwise correlation coefficients are below a previously determined threshold, then the transformer is flagged for further analysis. Further discussion of the choice of threshold can be found in the Results section. Steps 2 and 3 are repeated until all transformers on the feeder have been analyzed. Note that using this methodology, transformers with a single customer are omitted as this type of analysis cannot be used on those transformers.

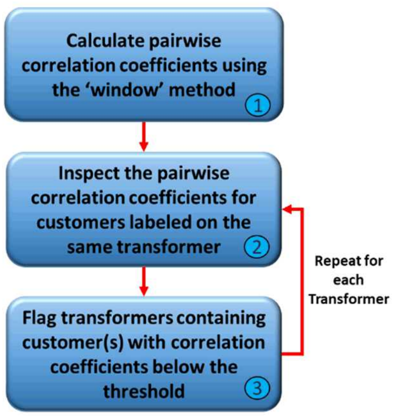

Figure 2 - Flowchart of the proposed methodology

IV. Results

The proposed algorithm was tested on both a synthetic dataset and a utility dataset. The bulk of the testing was conducted on the synthetic dataset where the ground truth of customers on a transformer is known, and realistic data concerns can be controlled and understood. The algorithm was also demonstrated in a proof-of-concept test on a real utility feeder.

A. Synthetic Data Results

1) Dataset

The synthetic dataset consists of one year of 1-minute measurement interval AMI measurements for 1379 residential customers and 581 service transformers. The average real power was extracted from Pecan Street [14] to create load profiles for the customers. OpenDSS [15] was then used with EPRI's Test circuit 5 [16] to calculate voltage time series. A uniformly distributed range of power factors was used (0.79-0.99), varied every 30 minutes. The data was then averaged to 15-minute intervals for use in this work. This dataset has also been used in [13], [17], [18], and more details on the data generation can be found in those references.

2) Experimental Results

A series of experiments were conducted to test the robustness of the proposed algorithm under different data conditions. The work in [17] details a selection of data considerations of interest.

This work focuses on the number of customers which have an incorrect transformer label and the amount of measurement noise within the voltage measurements. The incorrect transformer labels are injected by percentage of the customers, thus 1% of customers mislabeled means that 13 of the 1369 customers were given incorrect transformer labels. In practice, this is roughly equivalent to the number of transformers that contain an error, until the percentage of customers mislabeled becomes large. The voltage measurement noise was injected into all customers uniformly at random up to a specified maximum percentage. Thus, if the maximum percentage of noise is 0.2% then for each measurement in the time series, a value is selected uniformly at random from the range [-0.48, +0.48] where 0.48 is 0.2% of 240V, the mean voltage.

There are two primary metrics of interest in these experiments. The first is the number of transformers that should have been flagged (meaning that they are specified as having a customer in their low voltage network that is not connected to that transformer) but were not flagged; those are referred to as the 'false negative' transformers. Second, the number of transformers which were incorrectly flagged (meaning that they have the correct group of customers assigned to them); that group of transformers is referred to as the 'false positive' transformers. Ideally the set of false negative transformers and the set of false positive transformers will both be empty sets.

The threshold for flagging a transformer based on the correlation coefficients is the primary parameter that requires selection in advance using this methodology. The following figures demonstrate the sensitivity of that parameter on the false negative and false positive results. Figure 3 shows the results for 1% of customers mislabeled, without injecting any noise into the dataset. The x-axis shows the value of the voltage correlation coefficient threshold; for example, given a threshold of 0.6, if there are pairwise correlation coefficients less than 0.6 in a group of customers labeled on a particular transformer, then that transformer would be flagged. On the blue (left) y-axis is the number of false positive transformers, and on the red (right) y-axis is the number of false negative transformers. We can see the tradeoff inherent in the choice of the threshold value: too small of a value and the false negative rate increases, and too large of a value and the false positive rate increases. There is a range of acceptable values from ~0.5 to ~0.78 where both the false negatives and false positives are 0 and the algorithm achieves 100% accuracy in flagging transformers with incorrect customers and does not flag any correct transformers.

The next figures demonstrate the sensitivity of the algorithm to the quantity of customers that have incorrect transformer labels and varying levels of measurement noise injected into the voltage measurements. Figure 4 shows the results when a maximum of 0.15% noise is injected into the dataset and the quantity of customers which are given incorrect transformer labels is varied. We can see that the range of acceptable threshold values varies only slightly among the three simulations. In fact, if the y-axis were given in percent instead of number of transformers, the lines would be plotted nearly on top of one another.

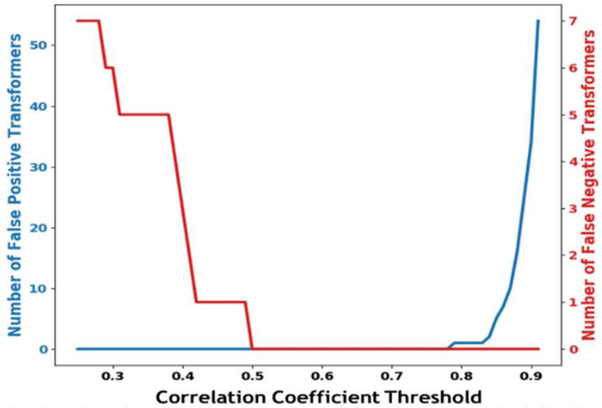

Figure 3 - Results using 1% of customers mislabeled without noise injection

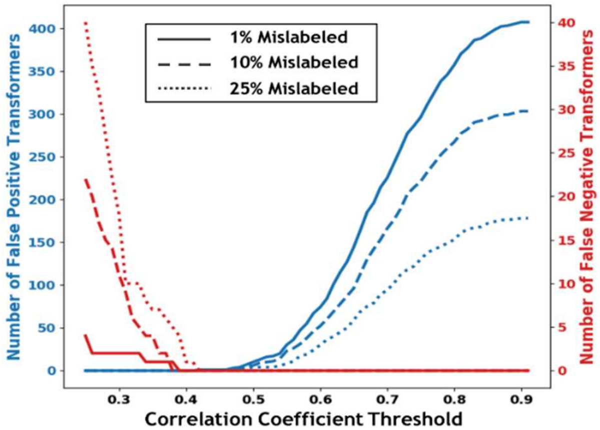

Figure 4 - Results using 0.15% maximum noise with varying quantities of mislabeled customers

Figure 5 shows the results using 10% of customers with incorrect transformer labels with varying levels of injected noise. Increasing the level of measurement noise shifts the plots to the left, although note that in each case, there is still separation where there are acceptable values for the threshold. This shift is intuitive because the addition of measurement noise forces all the correlation coefficient to be less correlated, but the customers on different transformers remain less correlated than customers on the same transformer.

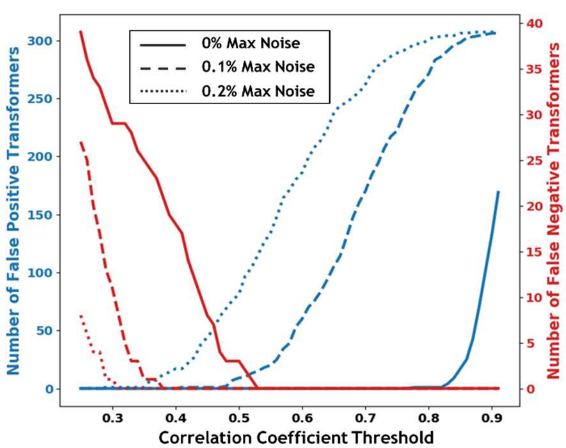

Figure 5 - Results using 10% of customers mislabeled with varying levels of injected measurement noise

B. *Utility Feeder Results*

1) Dataset

The utility dataset used in this work is approximately 15 months of data, measured at 15-minute intervals, using the averaging method, from the northeastern United States. This dataset is also used in the [2], [13], and [2] gives examples of the types of errors commonly found in distribution system models. There are not ground truth labels for this dataset, thus the following example is shown as a proof-of-concept that the proposed method works given real data. In the absence of ground truth labels, publicly available Google Street View images can be used to validate certain algorithm predictions. Further examples of this from the same dataset can be seen in [2].

Figure 6 shows satellite imagery of two transformers and four customers, and the original model shows that all four customers are connected to the southern (bottom) transformer. However, this transformer was flagged by the proposed algorithm, and inspection of Google Street View imagery confirms the configuration in Figure 7. Two customers are connected to the south transformer and two are connected to the north transformer. Table 1 shows the pairwise correlation coefficients for this set of four customers, and the two groupings of two can be clearly seen. Note that the correlation between Customer 1 and Customer 2 is only 0.77, demonstrating that real data can often contain factors that lower the correlation coefficients even between customers on the same transformer. The algorithm also correctly identified several other known transformer labeling issues on this feeder that had been previously identified in other work.

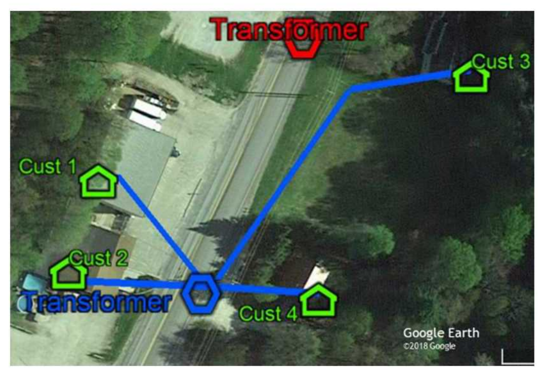

Figure 6 - Original utility labeling

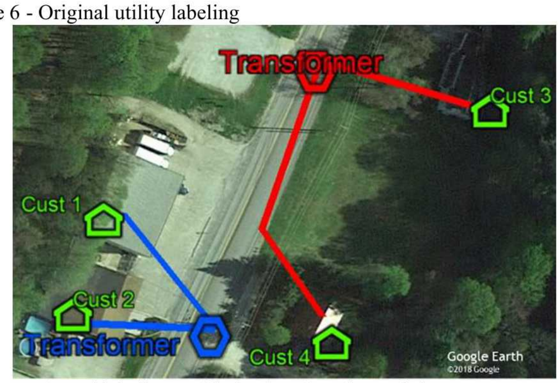

Figure 7 - Actual labeling verified using Google Earth imagery

TABLE 1 – VOLTAGE CORRELATION COEFFICIENT MATRIX FOR THE CUSTOMERS IN FIGURE 6

	Cust #1	Cust #2	Cust #3	Cust #4
Cust #1	1	0.777	0.434	0.575
Cust #2	0.777	1	0.344	0.446
Cust #3	0.434	0.344	1	0.958
Cust #4	0.575	0.446	0.957	1

C. Comparison with Similar Algorithms

The proposed algorithm takes a similar approach to other works in literature, particularly [1]. During initial work, variations of the methods proposed in [1] (correlation coefficients) and [8] (linear regression) were tested on our synthetic dataset. We were able to show that the linear regression methodology was not robust to the injected noise perturbation and the correlation coefficient methodology was not robust to increasing levels of mislabeled customers. Note that some of the errors produced by these methods were injecting *new* errors into the utility model. For the correlation coefficient methodology, most errors were occurring in the second stage of the process, assigning a customer known to have a transformer labeling error to its correct transformer. This fact inspired the direction taken in this work. Although this work focused on flagging the error, this remains of great use to utilities because the number of transformers to be inspected is greatly constrained.

V. FUTURE WORK

There are several aspects of future work suggested by the proposed method. Firstly, there are some elements that require further investigation due to their random affects. The configuration of which customers happen to be mislabeled is likely to have a role in the efficacy of the algorithm; secondly, the assignment of noise to the voltage time series is done in a random fashion, thus each simulation would perform slightly differently. Further testing is also required on other utility feeders to determine how the correlation coefficients change under differing conditions. Finally, although this work presents a novel method for identifying customers labeled on incorrect transformers, in some sense it solves an 'abridged' version of the complete customer to transformer pairing problem. This work focuses on identifying where the errors occur in the utility model, and work is ongoing in correcting those errors, which is a much more challenging problem.

VI. CONCLUSION

This work presents a methodology to identify service transformers in distribution system models that have customers which are not physically connected to that transformer, leveraging the information provided by the correlation coefficients between customers' AMI voltage time series. The proposed algorithm achieved 100% accuracy in flagging on the synthetic dataset of 581 transformers, with varying quantities of injected measurement noise and varying percentages of mislabeled customers. It is possible to correctly flag all incorrect transformers and avoid flagging any transformers that have the correct grouping of customers. The method was also tested as a proof of concept on a real utility feeder and successfully flagged several of the known transformer labeling errors within that feeder. This method shows excellent promise in enabling utilities to intelligently direct their personnel and resources towards transformers that need further analysis.

REFERENCES

[1] W. Luan, J. Peng, M. Maras, B. Harapnuk, and J. Lo, "Smart Meter Data Analytics for Distribution Network Connectivity Verification," *IEEE Transactions on Smart Grid*, vol. 6, p. 1, Jul. 2015.

[2] Blakely, M. J. Reno, and J. Peppanen, "Identifying Common Errors in Distribution System Models," *Photovoltaic Specialists Conference (PVSC)*, Jun. 2019.

[3] B. Palmintier *et al.*, "On the Path to SunShot: Emerging Issues and Challenges in Integrating Solar with the Distribution System," *National Renewable Energy Laboratory*, vol. NREL/TP-5D00-65331, 2016.

[4] A. Nguyen *et al.*, "High PV Penetration Impacts on Five Local Distribution Networks Using High Resolution Solar Resource Assessment with Sky Imager and Quasi-steady State Distribution System Simulations," *Solar Energy*, vol. 132, pp. 221–235, Jan. 2016.

[5] M. Ebad and W. M. Grady, "An Approach for Assessing High-Penetration PV Impact on Distribution Feeders," *Electric Power systems Research*, vol. 133, pp. 347–354, Apr. 2016.

[6] W. Luan, J. Peng, M. Maras, and J. Lo, "Distribution network topology error correction using smart meter data analytics," in *2013 IEEE Power Energy Society General Meeting*, 2013, pp. 1–5.

[7] R. Mitra *et al.*, "Voltage Correlations in Smart Meter Data," *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1999–2008, 2015.

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

[9] A. J. Berrisford, "A Tale of Two Transformers: An Algorithm for Estimating Distribution Secondary Electric Parameters Using Smart Meter Data," *26th IEEE Canadian Conference on Elctrical and Computer Engineering (CCECE)*, May 2013.

[10] S. Bolognani, N. Bof, D. Michelotti, R. Muraro, and L. Schenato, "Identification of Power Distribution Network Topology Via Voltage Correlation Analysis," *52nd IEEE Conference on Decision and Control*, Dec. 2013.

[11] Y. Liao, Y. Weng, G. Liu, and R. Rajagopal, "Urban MV and LV Distribution Grid Topology Estimation via Group Lasso," *IEEE Transactions on Power Systems*, vol. 34, no. 1, Jan. 2019.

[12] Y. Chen, J. Chen, H. Jiao, Y. Guo, W. Jiang, and H. Tang, "Two-Stage Topology Identification Method for Distribution Network Via Clustering Correction," *IEEE PES Innovative Smart Grid Technologies (ISGT) Asia*, 2019.

[13] L. Blakely, M. J. Reno, and W. Feng, "Spectral Clustering for Customer Phase Identification Using AMI Voltage Timeseries," *Power and Energy Conference at Illinois (PECI)*, Feb. 2019.

[14] "Pecan Street Database," *Pecan Street*. [Online]. Available: http://www.pecanstreet.org.

[15] D. Montenegro, R. C. Dugan, and M. J. Reno, "Open Source Tools for High Performance Quasi-Static-Time-Series Simulation Using Parallel Processing," *IEEE Photovoltaic Specialists Conference*, 2017.

[16] J. Fuller, W. Kersting, R. Dugan, and S. C. Jr., "Distribution Test Feeders," *IEEE PES AMPS DSAS Test Feeder Working Group*, 2013. [Online]. Available: http://sites.ieee.org/pes-testfeeders/.

[17] L. Blakely, M. J. Reno, and K. Ashok, "AMI Data Quality And Collection Method Consideration for Improving the Accuracy of Distribution System Models," *Photovoltaic Specialists Conference (PVSC)*, 2019.

[18] K. Ashok, M. J. Reno, D. Divan, and L. Blakely, "Systematic Study of Data Requirements and AMI Capabilities for Smart Meter Connectivity Analytics," *Smart Energy Grid Engineering (SEGE)*, 2019.