Application of Advanced Metering Infrastructure Data to Advanced Utility System Operations

Cooperative Research and Development Final Report

CRADA Number: CRD-17-00712

NREL Technical Contact: Murali Baggu

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

Technical Report
NREL/TP-5C00-83877
Revised November 2022

Application of Advanced Metering Infrastructure Data to Advanced Utility System Operations

Cooperative Research and Development Final Report

CRADA Number: CRD-17-00712

NREL Technical Contact: Murali Baggu

Suggested Citation
Baggu, Murali. 2022. Application of Advanced Metering Infrastructure Data to Advanced Utility System Operations: *Cooperative Research and Development Final Report, CRADA Number CRD-17-00712.* Golden, CO: National Renewable Energy Laboratory. NREL/TP-5C00-83877. https://www.nrel.gov/docs/fy22osti/83877.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

Technical Report
NREL/TP-5C00-83877
Revised November 2022

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

NOTICE

This work was authored 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 the U.S. Department of Energy Office of Electricity Delivery and Energy Reliability and the U.S. Department of Energy Grid Modernization Laboratory Consortium. 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.

Errata

This report, originally published in August 2022, was revised in November 2022 to include acknowledgments.

Cooperative Research and Development Final Report

Report Date: August 9, 2022

In accordance with requirements set forth in the terms of the CRADA agreement, this document is the CRADA final report, including a list of subject inventions, to be forwarded to the DOE Office of Scientific and Technical Information as part of the commitment to the public to demonstrate results of federally funded research.

Parties to the Agreement: San Diego Gas & Electric Company (SDG&E)

CRADA Number: CRD-17-00712

CRADA Title: Application of Advanced Metering Infrastructure Data to Advanced Utility System Operations

Responsible Technical Contact at Alliance/National Renewable Energy Laboratory (NREL):

Murali Baggu | Murali.Baggu@nrel.gov (for Santosh Veda)

Name and Email Address of POC at Company:

Frank Goodman | FGoodman@sdge.com (for Tom Bialek)

Sponsoring DOE Program Offices:

Office of Electricity Delivery and Energy Reliability; Grid Modernization Laboratory Consortium

Joint Work Statement Funding Table showing DOE commitment:

Estimated Costs	NREL Shared Resources a/k/a Government In-Kind
Year 1 Year 2, Year 3, Modification #1	$250,000.00 $0.00 $150,000.00
TOTALS	$400,000.00

Acknowledgments:

This material is based on work supported by the U.S. Department of Energy Office of Electricity and by California's Electric Program Investment Charge (EPIC) program in San Diego Gas & Electric Company under a cooperative research-and-development agreement (CRADA) # CRD-17-712.

NREL
Murali Baggu, Harsha Padullaparti, Santosh Veda, Martha Symko-Davies, Jiyu Wang, Marcos Netto, Jun Hao, and Valerie Rose

San Diego Gas & Electric Company
Frank Goodman, Tom Bialek, Jay Bick, Chippy Impreso, Julian Jones, Kyle Kewley, Gina Lindsay, William O'Brien, Amin Salmani, Subburaman Sankaran, Matt Smith, Tyson Swetek, Catarino Vargas, Stacy Williams, and William Wood

Executive Summary of CRADA Work:

The primary objective of the proposed project is to evaluate the capabilities of Advanced Metering Infrastructure (AMI) based controls for grid operations. This includes monitoring and control at the secondary transformer level using existing AMI infrastructure. In addition, some of the challenges of using a model-less control will also be identified. The lessons learned will be disseminated to other utilities around the country through peer-reviewed publications with the consent of SDG&E. The scope of the work includes using data from SDG&E's existing AMI infrastructure and working with third party vendors selected by SDG&E. The grid edge monitoring systems and controls provided by the selected vendors will perform grid operations using AMI data. The National Renewable Energy Laboratory (NREL) will perform a software and controller-based evaluation of the grid edge controller, intelligent devices and sensors, at the Energy System Integration Facility (ESIF) laboratory.

CRADA Benefit to DOE, Participant, and US Taxpayer: Assists laboratory in achieving programmatic scope

Summary of Research Results:

Task 1: Develop feeder models:

This study used two SDG&E's feeders, namely Feeder A and Feeder B. Two AMI datasets are used for Feeder A. The first dataset has the AMI data of the three-month period between October 1, 2018 to December 31, 2018 (2018 dataset) and the second dataset has the AMI data for the entire 2019-year period. For Feeder B, the AMI data for the entire year of 2019 was used. The phase identification results of the two feeders are documented in this section.

Feeder A is a 12-kV feeder with a peak load of 10.3 MW. The topology of the feeder is shown in Figure 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.

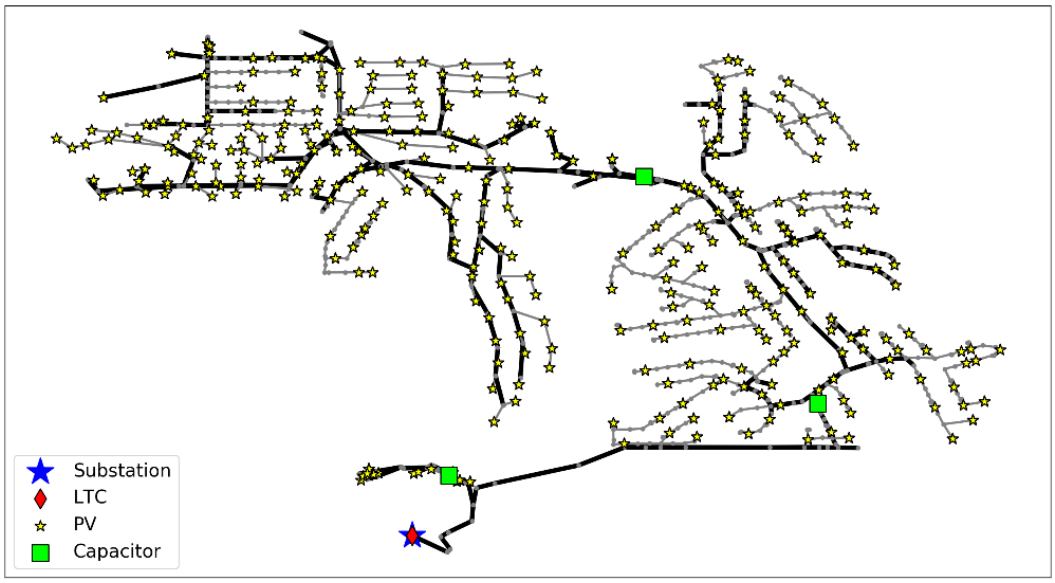

Figure 1. Topology of Feeder A.

The second feeder used for this study is Feeder B. This is also a 12-kV feeder with a peak load of 13.29 MW. The topology of the feeder is shown in Figure 2. The substation transformer is equipped with a load tap changer. Two capacitor banks are available on the feeder for reactive power support and there are no line voltage regulators. This feeder has 657 service transformers. Solar generation of 3.14 MW is present in this feeder which is approximately 24% relative to the peak load.

Figure 2. Topology of Feeder B.

Both feeders were provided by SDG&E with Synergi format. The feeder models are converted into OpenDSS format and the OpenDSS feeders are used in this study.

Task 2: Develop use case scenarios (Task 2 of original CRADA and Tasks 1, 2, and 4 of modified CRADA):

The documentation in this section relates to the following tasks:

- PV Smart Inverter Study: Task 2 of original CRADA, Task 1 of modified CRADA (Analyze the impact of different levels of PV penetration on feeder voltage profile using AMI data)
- Utility Planning Network Model Anomaly Detection Tool: Task 2 of original CRADA
- Phase Identification Tool: Task 2 of original CRADA, Task 2 of modified CRADA (Validate and Implement Algorithms for Phase Identification)
- Meter-to-Transformer Mapping: Task 2 of original CRADA
- Modified Task 4 (Validate using field verification)

Four use cases are developed in this study and the detail of each use case is introduced as follows:

PV Smart Inverter Study

The first use case developed is studying the effectiveness of PV smart inverter. High PV penetration levels in distribution feeders can cause operational challenges including voltage issues, reverse power flow, and protection issues. Standards recommend using PV smart inverters to support the distribution grid services, specifically voltage regulation. However, there are many smart inverter settings recommended by the standards and their performance on the SDG&E feeders has not been reported in the literature. In this study, the impact of various smart inverter settings including California Rule 21 (CA 21), Hawaii Rule 14, IEEE 1547, CA 21 with no deadband, hockey stick, and volt-var-watt control on the selected SDG&E feeders was examined.

Methods:

The distribution feeder model used in this study is Feeder A. The AMI data of this feeder was provided for the period between October 1, 2018, to January 15, 2019 (107 days). AMI load measurements from SDG&E included the net load consumption of each customer, therefore, a disaggregation was required to extract the PV profile and load profile for each load location. From the load definition in the feeder model, and peak power generation of each PV system, the determination was made that the PV penetration for this feeder is around 70% relative to peak load. The irradiance profile of the feeder area during the selected period of 107 days was downloaded from the National Solar Radiation Database (NSRDB) [1]. By using the ratings of each PV system, the irradiance profile, and the net load profile of each load node, the PV profile and load profile at each load location were disaggregated. After the disaggregation, the scenario of 100% PV penetration was modeled with the load and PV profiles. This was the case used in the following simulations.

In OpenDSS, the volt-var-watt smart inverter control function is not yet fully operational, therefore a Python function to implement the volt-var-watt control was developed. The inputs of the function include inverter rated kVA, solar irradiance at current time step, and measured per-unit (p.u.) voltage at the previous time step. First, the volt-var and volt-watt curves were predefined. Then based on the voltage and volt-watt curve, the function determined the required real power output and the maximum available reactive power. After that, the maximum available reactive power, and volt-var curve, and the reactive power output was calculated based on the measured voltage. The outputs of this function were the real and reactive power outputs of the PV system. These outputs were used to update the PV system output in OpenDSS. The smart inverter curves for all cases are summarized in this section. Several smart inverter curves, both from the standards and the custom curves of interest were studied in this work. These curves are depicted in Figure 3.

California Rule 21 (CA 21): The maximum and minimum percentage of available reactive power is +/-30%. This percentage is zero when the voltage is within 0.967-1.033 p.u., and reaches maximum/minimum when the voltage is below/over 0.92/1.07 p.u.

Hawaii Rule 14 (HI 14): The maximum and minimum percentage of available reactive power is +/-44%. This percentage is zero when the voltage is within 0.97-1.03 p.u., and reaches maximum/minimum when the voltage is below/over 0.94/1.06 p.u.

IEEE 1547: The maximum and minimum percentage of available reactive power is +/-44%. This percentage is zero when the voltage is within 0.98-1.02 p.u., and reaches maximum/minimum when the voltage is below/over 0.92/1.08 p.u.

California Rule 21 without deadband: The maximum and minimum percentage of available reactive power is +/-30%. This percentage reaches maximum/minimum when the voltage is below/over 0.92/1.07 p.u.

Hockey stick curve without compensation in low voltage region: The minimum percentage of available reactive power is -30%. This percentage is zero when the voltage is below 1.033 p.u., and reaches maximum when the voltage is above 1.07 p.u.

Hockey stick curve with deeper Q absorption: The minimum percentage of available reactive power is -75%. This percentage is zero when the voltage is below 1.033 p.u., and reaches maximum when the voltage is above 1.07 p.u.

Volt-VAR-Watt: The volt-var curve is the same as California Rule 21. For its volt-watt curve, the maximum available real power starts to decrease from 100% when the voltage is above 1.06 p.u. and reaches zero when the voltage is above 1.1 p.u.

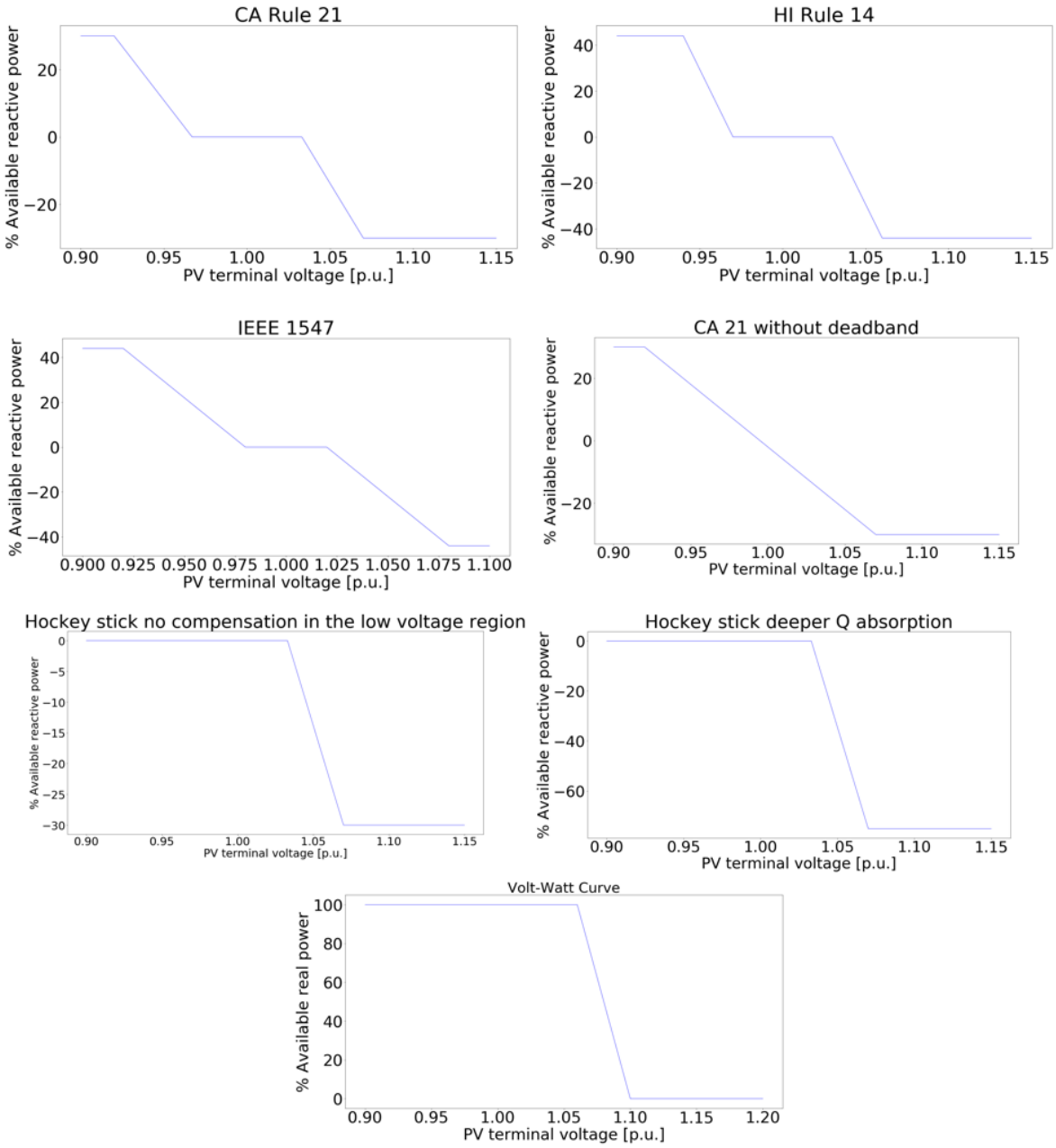

Figure 3. Volt-var and volt-watt curves for PV smart inverter.

The performance of each smart inverter function was evaluated by using multiple metrics: number of capacitor changes, number of LTC operations, average voltage, voltage fluctuation index, voltage unbalance index, and number of voltage exceedances nodes. Voltage fluctuations is described by as repetitive or random variations of the voltage envelope due to sudden changes in the real and reactive power drawn by the load.

Let the T stand for the total time steps in the simulation and N stand for the total number of nodes in the feeder, the average voltage is calculated by:

$$V^{mean} = \frac{1}{N} \times (\frac{1}{T} \sum_{i=1}^{N} \sum_{t=1}^{T} V^i(t)) \quad (1)$$

The voltage fluctuation index (VFI) measures how the nodal voltage is changing between time steps, i.e., voltage fluctuations all across the circuit. It is calculated by:

$$VFI = \frac{1}{N} \times (\frac{1}{T} \sum_{i=1}^{N} \sum_{t=1}^{T} \left| V^i(t+1) - V^i(t) \right|) \quad (2)$$

The voltage unbalance index (VUI) measures the unbalance level of nodal phase voltages all across the circuit. It is calculated by:

$$VUI = \frac{1}{N} \times (\frac{1}{T} \sum_{i=1}^{N} \sum_{t=1}^{T} V_{imb}^i(t)) \quad (3)$$

where $V_{imb}^i(t)$ is calculated by using the maximum deviation from average voltage over the average voltage.

The voltage exceedance is defined as voltage out of range 0.94-1.06. The exceedance node is defined as a node with more than 12 hours exceedance in the three-month period.

Utility Planning Network Model Anomaly Detection Tool

In the second use case, an automated tool was demonstrated that uses the AMI measurement data to identify the inaccuracies in the network model used for distribution planning. Numerous distribution network analysis, monitoring, and control applications; including volt/var optimization, state estimation, and distribution automation, require accurate distribution network models. The GIS maintained by utilities can be inaccurate because of a significant amount of missing data, restoration activities, and network reconfiguration which can lead to network model inaccuracies. The utility planning network model anomaly detection tool used the AMI data to identify network model issues. It accomplished this by building the approximated secondary network models from the AMI data and using them to estimate the primary voltages. The estimated primary voltages were then compared with the primary voltages obtained from the simulations of the utility planning network model to identify model anomalies.

Methods and Results:

The software tool uses the combination of a physics-based method and a machine learning method to estimate the primary bus voltages from the AMI voltage measurements on the distribution secondary network.

Physics-based Method

In the physics-based primary voltage estimation method, the voltage magnitudes on the primary side of the distribution service transformers were estimated using only two smart meters per secondary network. The smart meters were strategically placed on the closest and farthest load from the transformer, in the electrical sense. This method relies exclusively on smart meter data, and therefore it is fully data driven.

The physics-based primary voltage estimation method has two stages:

First Stage – Linear Regression: The first stage performed a linear regression on the latest data window available at the control center. A data window of 288 points was used, which is equivalent to a day for five-minute sampling resolution. The first stage was executed only once.

The equivalent circuit shown in Figure 4 was used for each service transformer. In this circuit, r_p and r_s denote the losses in the primary and the secondary winding of the service transformer, respectively. The variables, v_p' and v_s' denote the voltage magnitudes at the primary and the secondary of an ideal transformer, respectively and n_t is the transformer turns ratio. The variables, v_1 and p_1 denote the voltage magnitude and the active power measured at the closest load from the service transformer while v_2 and p_2 denote the voltage magnitude and the active power measured at the farthest load from the transformer. The variables r_1 and r_2 account for cable impedance; and v_u and p_u are unknown. A constrained linear least-squares minimization problem was solved to estimate the resistance r_2 and the equivalent resistance between the first meter and the primary bus.

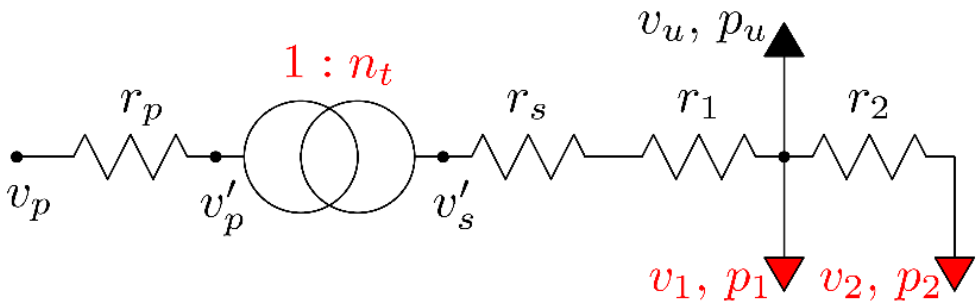

Figure 4. Equivalent circuit used for each secondary in physics-based voltage estimation method.

Second Stage – Kalman Filtering: The second stage used a Kalman filter to update the primary voltage magnitude estimates continuously, based on new data points. The processing steps in both stages are discussed in [2].

Machine Learning Method

Machine learning approaches typically require a training data set that contains the features to be estimated. In this application, the inputs included AMI measured power and voltages of 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 were available for this feeder, except the RTU voltage, time-series voltage data recorded from the simulations in OpenDSS were used to form the required training data set in the algorithm development stage. The quasi-static time-series (QSTS) simulation of Feeder A was performed in OpenDSS for the period between October 1, 2018 to January 15, 2019 (107 days) to obtain primary-side voltages. In the QSTS simulation, time resolution was set to hourly to follow AMI load time resolution. The load profile of each secondary-side measured load was set to be the AMI measured total power under that transformer. The simulated primary-side synthetic voltages from the QSTS simulation and the actual measured secondary-side voltages at the two AMI measured loads were used to train the machine learning model. The machine learning method for this application is discussed in [3].

After validating the performance of the machine learning-based algorithm using simulation data, it was applied to actual AMI measurement data recorded in the field. In this stage, the machine learning models were trained using the primary voltages estimated by the physics-based method instead of the simulation data.

Combined Method

Both the physics-based and the machine learning-based methods have limitations in estimating the primary-side voltages when applied individually. The physics-based method can use the available AMI data to conduct the estimation, but the accuracy is lower than desired. The machine learning-based method can have a higher estimation accuracy, but requires measurements of service transformer primary voltages, which was not included in the dataset provided. A combined method was developed to leverage the advantages of both these methods. In the combined method, primary voltages were estimated by the physics-based method for a given time duration, a primary voltage correction was applied, and corrected primary voltages were used to train a machine learning model. The trained machine learning models can be used to estimate primary voltages for any time duration.

Voltage drop across a service transformer typically varies from two to 13 V. The average voltage drop between the two AMI meters in all the secondaries of SDG&E Feeder A is 0.58 V. This implies that the voltage drop on a service transformer would be four to 20 times of the voltage drop between the AMI meters in the corresponding secondary. Accordingly, a correction factor was applied to the primary-side voltages estimated by the physics-based method before using them as training data for the machine learning models.

Phase Identification Tool

The third use case in this study is developing a phase identification tool. The phase identification tool performs automated phase mapping of the AMI meters based on AMI data. The GIS database maintained by the utility is known to have phase connectivity errors due to restoration activities, network reconfiguration, human error, and missing data. Traditionally, the phase connectivity database is periodically updated by field verification which is expensive and time-consuming. With the availability of AMI data, the phase connectivity can be identified through data analytics. The existing phase identification techniques work well in distribution feeders that have low or no PV generation; however, they fail to identify the phases accurately when considerable PV generation is present. The phase identification tool demonstrated in this project uses supervised learning to determine the phase connectivity accurately even when significant PV generation is present.

Methods:

The key assumption in this study is that voltage profiles from AMI meters pertaining to each phase connectivity are highly correlated with each other. Thus, the voltage magnitude time series of the AMI meters that are on the same phase tend to exhibit similar variations in the voltage measurements which are different from the meters on the other phases.

Phase Identification using Supervised Learning

Phase identification was performed using the random forest classifier [4], [5]. The random forest classifier is a supervised machine learning model. In the phase identification process, first the voltage magnitude time series from each meter in the AMI dataset was obtained for a selected duration of time. Next, a preset percentage of meters were selected for each phase connectivity as a training dataset for the supervised machine learning algorithm. The phase connectivity of these meters must be accurate since this is part of the training process for the machine learning algorithm. Then a random forest classifier was constructed, which is a function that predicts the phase connectivity of each meter in the training dataset based on the voltage magnitude time series data. Finally, the trained random forest classifier was used to identify the phase connectivity of the rest of the meters in the AMI dataset based on their voltage magnitude time series.

The phase identification algorithm steps are given below:

1. Data preprocessing: Load the AMI dataset with the voltage magnitude time series data and perform data standardization. A small number of meters with consistently reported bad data or empty data were removed from the AMI dataset in this step.
2. Training the random forest classifier: Select 30% of the AMI meters for each phase connectivity for training the random forest classifier. The phase connectivity of these AMI meters, obtained through field validation or some other means, was supplied to the random forest classifier in this step.
3. Phase identification: Input the voltage magnitude time series data of the rest of the meters to the random forest classifier model trained in Step 2 to identify the phase connectivity of the rest of the meters in the AMI dataset.

Data Requirements

The phase identification algorithm in this demonstration used the voltage magnitude time series data of the AMI meters and the validated phase connectivity information for 30% of the meters in the AMI dataset for training. The AMI dataset had average, maximum, and minimum voltages for each meter at five-minute intervals. The five-minute average voltage magnitude data was used for phase identification. The inputs and output of this algorithm are summarized below:

Inputs:

1. Average voltage magnitude time series data (preferably three months or more) for each AMI meter at five-minute resolution
2. Accurate phase connectivity information for 30% of the AMI meters for each type of phase connectivity

Output:

1. A table with AMI meter ID and associated phase connectivity for all the AMI meters

Meter-to-Transformer Mapping

The fourth use case in this study is a meter-to-transformer mapping algorithm. SDG&E has a meter-to-transformer connectivity mapping database. However, the records in this database do not always reflect the latest field conditions due to routine meter field change and human data entry errors. Accurate meter-to-transformer mapping information is needed for load balancing, service order work, and transformer load management. A solution that can check and correct the service transformer and meter mapping records is required to address this need. The goal of this use case was to demonstrate a proof-of-concept AMI meter-to-service transformer mapping solution that identified incorrect records based on AMI measurement data.

Methods:

The methodology for meter-to-transformer mapping solely uses the voltage data recorded by the AMI. There is no requirement for the length of the data and the methodology is also robust to missing data which can be observed frequently in the voltage dataset.

The main idea is the voltage measurements from the AMIs connected to the same service transformer secondary should be highly correlated and have a high correlation coefficient. Therefore, the key is to find a threshold to identify the potential incorrect records. If the correlation coefficient is lower than the threshold, that means records are incorrect. The AMI meter to service transformer mapping procedure consists of the following steps:

1. Calculate correlation coefficient between meters connected to the same service transformer for the records in the existing database
2. Rank the calculated correlation coefficients and pick a threshold
3. Loop through all the correlation coefficients and select the AMI records whose correlation coefficients are lower than the threshold

4. Calculate the correlation coefficients with the rest of the dataset and choose the service transformer with the highest score
5. Perform Step 4 for each meter in the selected set of records in Step 3. If the score is higher than the threshold, then correct the record to the new service transformer and meter pair, otherwise keep the record the same

For instance, consider that the correlation coefficient of AMI1 and AMI2 are lower than threshold τ in Figure 5, and they are connected to the service Transformer 1 in the original record. Then the algorithm will check the rest of the dataset to find a transformer with the highest correlation coefficient for both AMI1 and AMI2. If the score is higher than τ , then the algorithm will update the records. In this case, the mapping of AMI1 is changed to associate with Transformer 2 after running the algorithm.

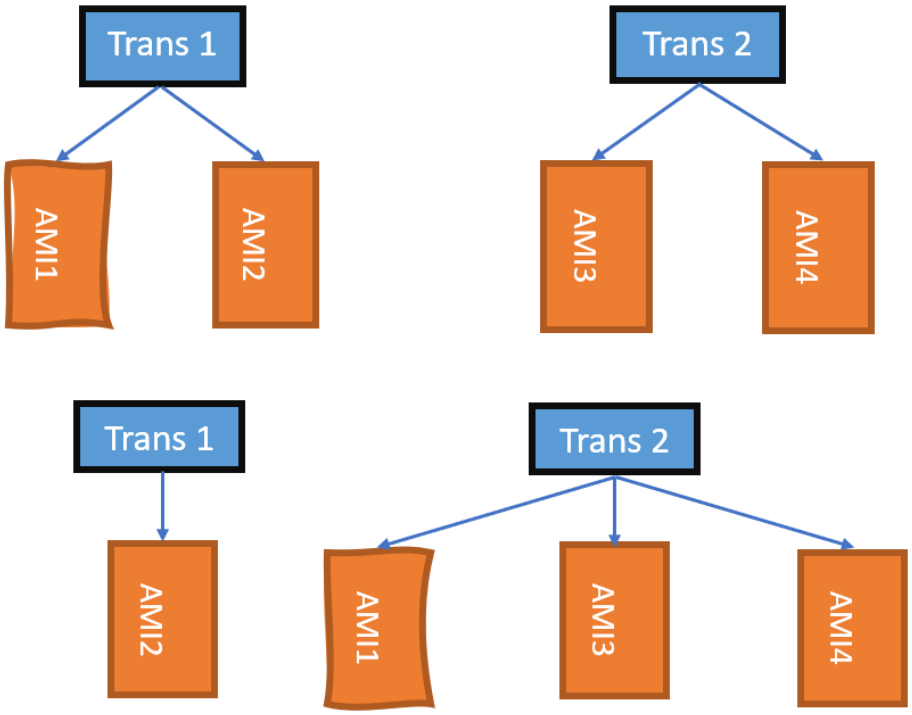

Figure 5. Illustration of the AMI meter to service transformer mapping algorithm.

Recall and precision are used to evaluate the performance of the proposed methodology. Recall evaluates the overall accuracy of the algorithm, and precision evaluates the accuracy of the correction. The equations below show the definition of recall and precision in mathematical formulas. In our case, the true positive is the right correction, the false positive is the wrong correction, and the false negative means the records remain the same as before.

$$R(recall) = \frac{TP(True\ Positive)}{TP + FN(False\ Negative)} \tag{4}$$

$$(precision) = \frac{TP(True\ Positive)}{TP + FP(False\ Positive)} \tag{5}$$

Three correlation coefficients namely Pearson [6], Kendall's rank [7], and Spearman's rank [8] are used in this work. These are defined as follows:

The Pearson correlation coefficient is a measure of linear correlation between two sets of data. It is the ratio between the covariance of two variables and the product of their standard deviations. The Kendall rank correlation coefficient is a statistic used to measure the ordinal association between measured quantities. It is a measure of rank correlation and the similarity of the orderings of the data when ranked by each of the quantities. Spearman's rank correlation coefficient is a nonparametric measure of rank correlation. It assesses how well the relationship between two variables can be described using a monotonic function.

Task 3: Develop system control evaluations plan and prepare the testbed (Task 3 of original CRADA and Tasks 3 of modified CRADA "Implement SDG&E's feeders on ADMS test bed to evaluate AMI-based algorithms"):

The documentation in this section relates to Task 3 of original CRADA and Task 3 of modified CRADA.

Data-centric Grid Operations

The integration of ADMS and AMI measurements offers a unique opportunity to further modernize grid control. In this task, an AMI-based, data-driven, volt/var control algorithm, and its synergies with ADMS for distribution grid operations, were evaluated using SDG&E feeder and AMI data. The inputs of this algorithm were AMI power and voltage measurements. The algorithm controls the substation transformer load tap changer (LTC) tap position, capacitor bank switch positions, and PV inverter setpoints to ensure voltage regulation. This new paradigm for grid operations was demonstrated using NREL's ADMS Test Bed capability wherein the feeders and the controls were implemented and evaluated in a realistic utility environment.

Methods and Results:

In the voltage prediction task that attempts to predict the voltage issues in the distribution network, an AMI voltage forecasting model was prototyped for the distribution feeder. The secondaries were modeled as having three AMI data points, with the closest and furthest to the transformer as individual data points and a third one that is an aggregate of all other power consumed on the secondary. Synthetic average hourly voltage data was simulated for three and a half months. Two machine learning algorithms were used in modeling voltage time series data which can be used for forecasting s . The models are learned globally and simultaneously process all AMI time series data. Simulations with various scenarios of available historical data (60, 30, and 15 days) were performed which were explicitly incorporated into the model and evaluated for performance. The performance of a model hyper parameter set when forecasted 24-hours ahead is shown in Figure 6. The plot represents an averaging over five-folds in the validation set. The model is evaluated globally on all simulated meters on the distribution feeder.

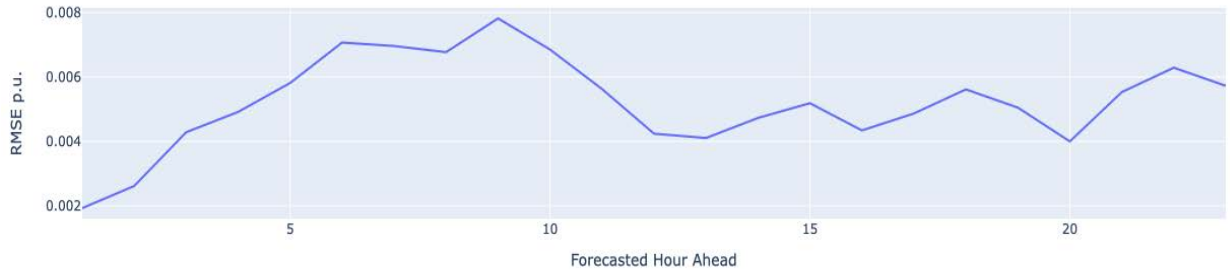

Figure 6. Twenty-four hour forecast performance of a hyperparameter run averaged over five folds.

The integration of ADMS and AMI measurements offers a unique opportunity to further modernize distribution system control. In this task, an AMI-based data-driven volt/var control algorithm and its synergies with ADMS for distribution grid operations were evaluated using the SDG&E feeder and the AMI data. The inputs of this algorithm were AMI power and voltage measurements. The algorithm controls the LTC tap position, capacitor banks switches, and PV inverter setpoints to ensure voltage regulation.

Figure 7 and Figure 8 show the results from the evaluation of this algorithm. In the base case, the LTC and capacitor banks follow their local controllers, and the PV smart inverters inject power at unity power factor. In the unity power factor operation, the PV smart inverters inject active power only and no reactive power is injected or absorbed. As observed in Figure 7, many customer voltages on the secondary are experiencing high voltage exceedances in the base case. In the next scenario in which the data-driven control algorithm is enabled, the voltage exceedances are significantly reduced, and the average voltages are closer to 1.0 p.u. Once the voltage deviates from the preset voltage regulation set point (selected 1.0 p.u. in this case), the algorithm primarily raises or lowers the LTC tap position to regulate the voltages as observed in Figure 8.

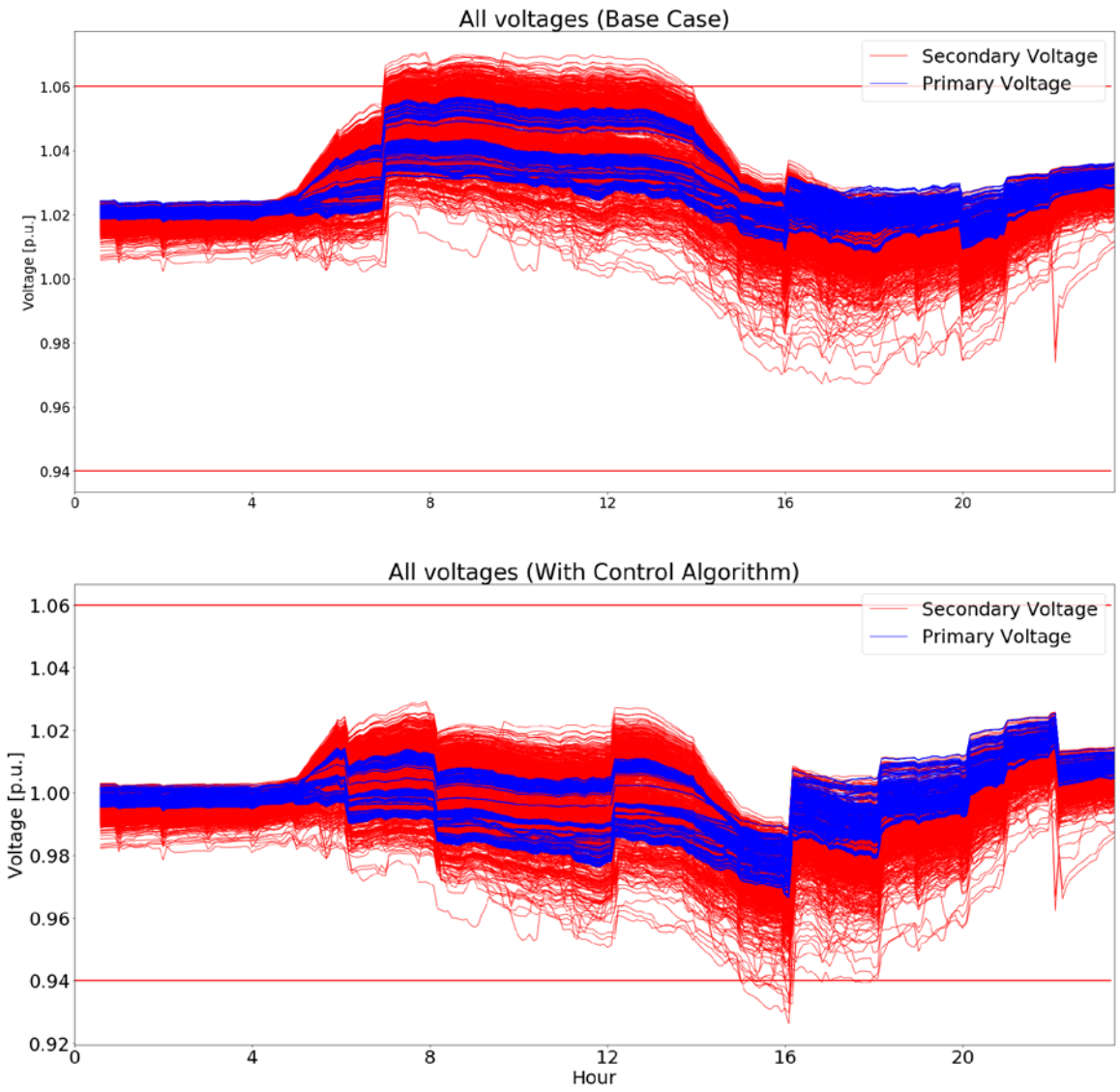

Figure 7. Comparison of bus voltages.

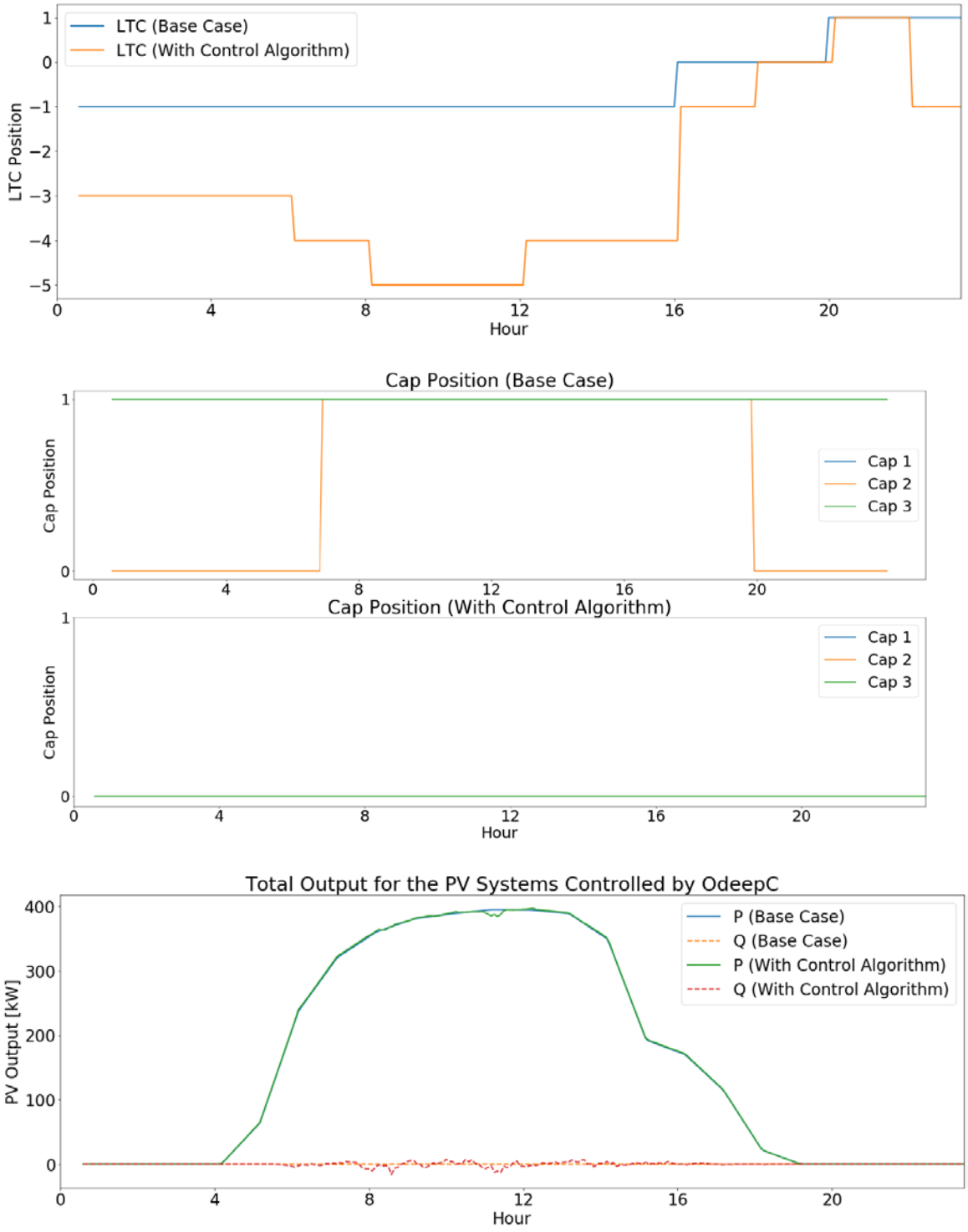

Figure 8. Comparison of LTC, capacitor bank statuses, and total PV generation.

Task 4: Execute use case scenarios and analyze results (Task 4 of original CRADA, Task 4 of modified CRADA "Validate using field verification"):

The documentation in this section relates to the following tasks:

- PV Smart Inverter Study Results: Task 4 of original CRADA
- Utility Planning Network Model Anomaly Detection Results: Task 4 of original CRADA
- Phase Identification Results: Task 4 of original CRADA, Task 4 of modified CRADA
- Meter-to-Transformer Mapping Results: Task 4 of original CRADA

PV Smart Inverter Study Results:

The PV and load profiles were interpolated to five-minute resolution and the simulation was run using the data from October 2018 to January 2019 (107 days, 30816 data points in total). There were 1560 PV systems in the model in total, all with ratings between 5-10 kVA. For the baseline, the power outputs of PV systems were determined by the irradiance and inverter rating. For the case with smart inverter function enabled, the power outputs of these PV systems follow the corresponding curves. The voltage plots in this section are presented for one selected day i.e., October 1, 2018. The metrics were computed for the three-month period and summarized in Table 1 and 2.

Table 1. Summary of Feeder Operations

	Capacitor bank status changes	LTC tap changes	Average Voltage (V)	V fluctuation index score	V unbalance index score
Baseline	562	1291	249.93	9.67	9.90
CA 21	646	1335	249.20	9.68	9.86
HI 14	538	1471	248.60	9.67	9.82
IEEE 1547	580	1478	248.66	9.66	9.81
No Deadband	606	1394	248.41	9.67	9.77
HS-no compensation	646	1335	249.20	9.68	9.86
HS-deeper Q	624	1506	248.60	9.62	9.83
Volt-Var-Watt	158	784	250.27	9.61	9.81

Table 2. Summary of Voltage Exceedances

	Secondary		Primary	
	Voltage exceedances hours per node	Number of voltage exceedances nodes	Voltage exceedance-hours per node	Number of voltage exceedance nodes
Baseline	23.52	752	42.83	481
CA 21	0.55	16	0.61	0
HI 14	0.21	9	0.76	12
IEEE 1547	0.47	28	0.96	14
No Deadband	1.05	37	2.84	42
HS-no compensation	0.55	16	0.61	0
HS-deeper Q	0.09	3	0.91	12
Volt-Var-Watt	4.45	110	2.95	53

The results show the implementation of smart inverter settings improves the feeder voltage profile by reducing the voltage exceedances. Based on the results from the study, the Rule 21 curve showed superior results in terms of the number of voltage regulation device actions and eliminating the primary voltage exceedances. The Rule 14 curve showed superior results in terms of eliminating the secondary voltage exceedances. The voltage exceedances for the volt-var-watt function are higher than the others, but it has the lowest number of voltage regulation device actions. The numbers of voltage regulation device changes are similar for all other smart inverter functions, and the average voltages are all near 249 V. Based on different purposes of controlling the feeder, the corresponding smart inverter functions can be selected by using the results from this study. For example, if the utility wants to minimize the action times of the voltage regulation devices, Rule 21 can be set for the smart inverters on this feeder.

Utility Planning Network Model Anomaly Detection Tool:

Physics-based Method Results

Five service transformer locations as shown in Figure 9 were selected to perform the validation of the physics-based method. The corresponding secondaries were modeled in detail with the realistic topology and load data. Each secondary model comprises a few loads including the two loads for which the AMI load consumption data was available. The load profiles of the two loads were set to be the same as AMI load consumption data for those two loads. The aggregated power consumption data at the service transformer level (minus the sum of the two loads) were distributed evenly across the rest of the loads of that secondary. The primary voltages of the selected secondaries were estimated using the physics-based method. The estimation mismatch results are summarized in Table 3. It can be observed that the estimation mismatches are all around 4%, which is larger than expected. The physics-based method usually will overestimate the primary voltages. Therefore, a machine-learning based method and a combined method are developed to improve the estimation accuracy.

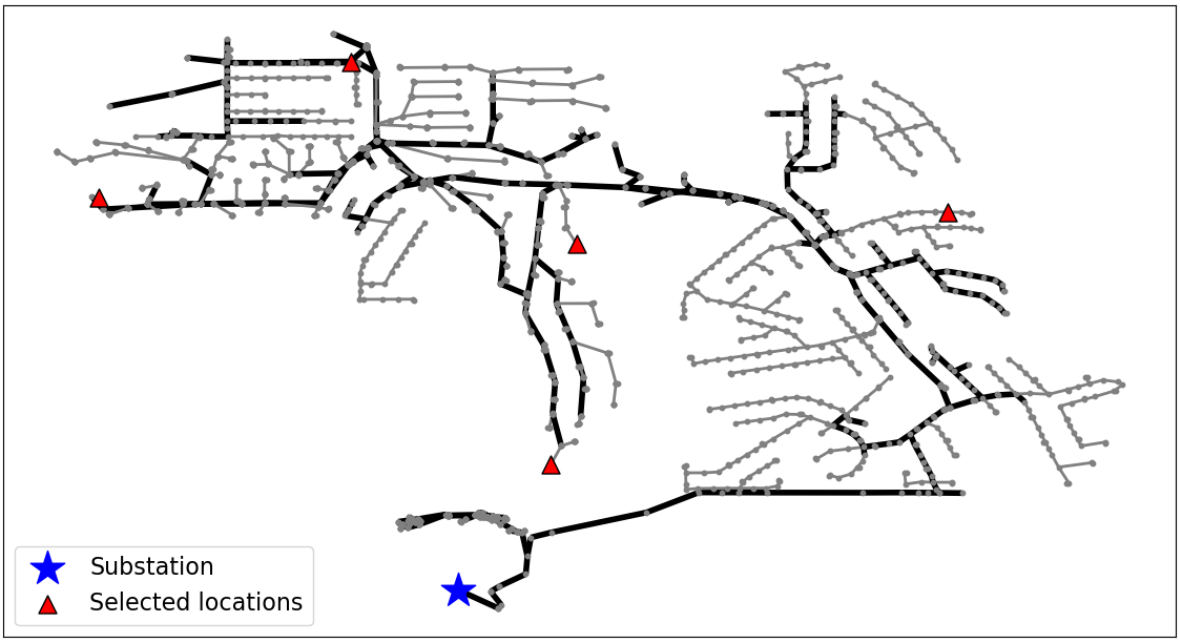

Figure 9. Five selected locations for the validation of physics-based method.

Table 3. Physics-based method validation results

	Estimation Mismatch
Secondary 1	3.74%
Secondary 2	4.06%
Secondary 3	3.53%
Secondary 4	4.28%
Secondary 5	3.84%

Machine Learning Method Results

Multiple machine learning algorithms: random forest, adaptive boosting, and gradient boosting [9]-[11] were tested to find the relationship between the primary-side voltages and the AMI measurements under each service transformer. The data from each service transformer (341 in total) was trained separately to account for their unique characteristics, i.e., separate models were constructed for each service transformer. The input of each model was 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 was the voltage on the primary side of the service transformer.

The data from the first month were selected to compare the estimation accuracy of different algorithms. K-fold cross validation was used to validate the machine learning models, and the validation was repeated 30 times. In each test, 80% of the monthly data was randomly drawn from the data set to train the model, and the remaining 20% was used for testing. The mean absolute percentage error (MAPE) and maximum absolute percentage error between the synthetic primary voltage and estimated primary voltage was used to evaluate the performance of each machine learning method. The performance comparison is shown in Table 4.

Table 4. 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 4 show that the random forest model performs better than the other two models in the selected performance criteria; therefore, it was selected to estimate the primary-side voltages in this study. Another advantage of using the random forest algorithm (as random forest is an ensemble learning method that integrates multiple decision trees) is that it will combine these decision trees and use average, or voting, schemes to calculate the results. Therefore, any outliers in the AMI measurements can be handled with this algorithm. Further, an exhaustive search was conducted to determine the model parameters (number of decision trees and maximum depth). These two parameters are varied from one to 500 and one to 30, respectively, to test the estimation performance. Considering both estimation accuracy and training time, the number of decision trees were selected to be 80 and the maximum depth to be 10. The time to build the machine learning model for each service transformer was around five seconds, and the total time for building the models for all service transformers was 30 minutes. As the process of training the model is usually developed for the distribution system planning studies, it meets the run-time requirement.

The performance of the machine learning-based approach was validated by the synthetic primary-side voltages generated from the QSTS simulation of the feeder model in OpenDSS. A secondary model was built for each service transformer in OpenDSS. Each secondary model included the two loads with voltage measurements and a load without voltage measurement. The load profiles of each secondary measured load were set to be one AMI measured power for one meter under that transformer. The load profile of the unmeasured load was 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 were used to train the machine learning model. The data from the first 1,000 hours were used as a training data set to train the model for each service transformer, and the next 1,568-hour data were 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 Figure 10. 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%. Comparing with the 4% error from the physics-based method, the estimation accuracy has been improved a lot, however, this method requires some primary voltage data to train the machine learning model for each node. The comparison between estimated and actual voltages (synthetic voltage, in this case) for one example service transformer is shown in the two subplots of Figure 11. 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 was also tested when using the first 2,000-hour data as the training dataset and tested with the remaining 568-hour data. The performance is similar to the previous case, which means the over-fitting problem does not exist for the model.

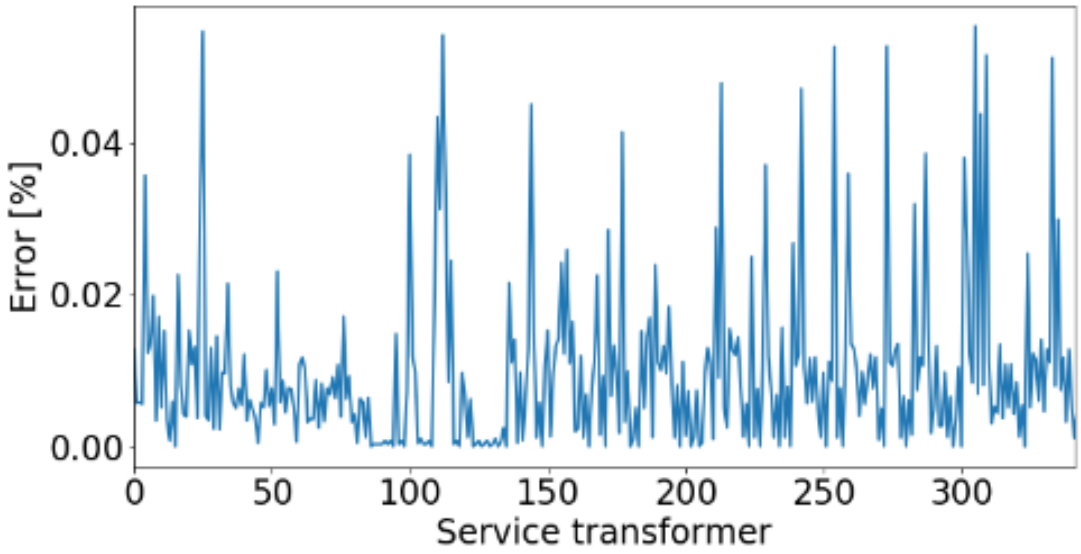

Figure 10. MAPE for the voltage estimation of each service transformer

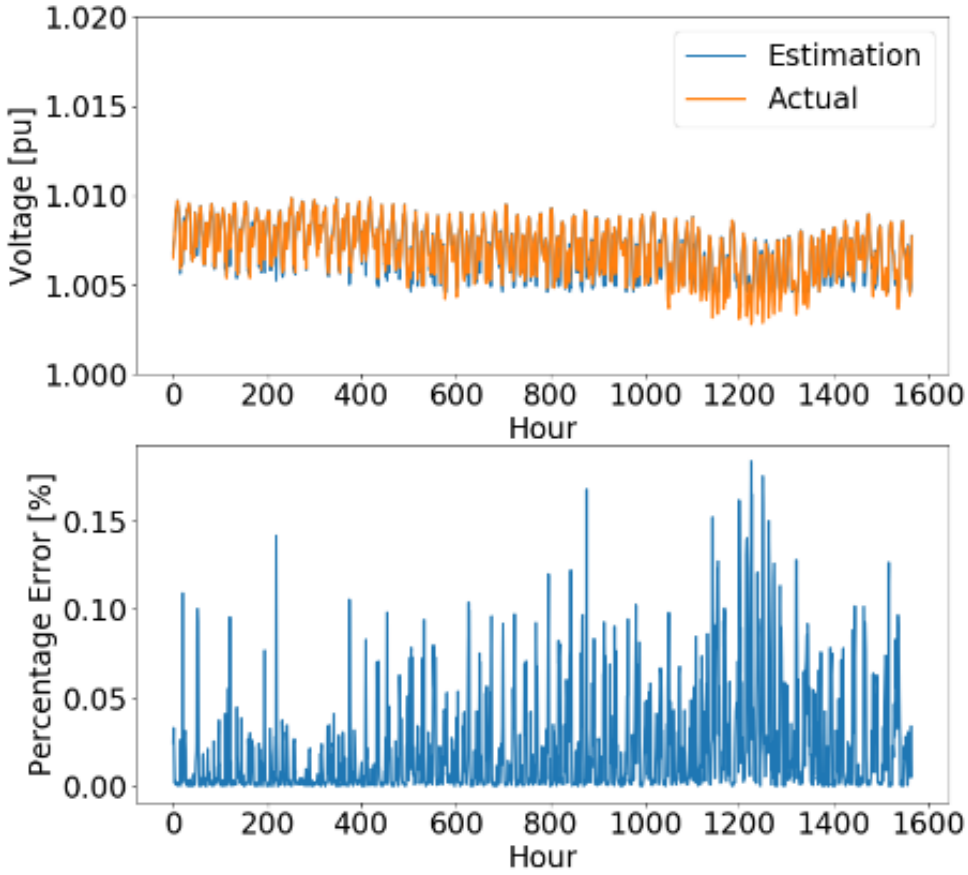

Figure 11. Comparison between estimated and synthetic voltages for one example service transformer.

Combined Method Results

The performance of the combined method was validated by the synthetic primary-side voltages generated from the QSTS simulation of the OpenDSS model, since the primary-side voltages were not recorded in the field. The QSTS simulation was performed and the primary-side and secondary-side voltages at the two AMI measured loads for the five selected secondaries were recorded. These measurements from the secondary were used as AMI data to validate the combined methods. The estimation results are summarized in Table 5. From the results it can be observed that the estimation error decreased from 4% to 1% in the 2000-hour testing period. For most of the time the errors are within 1%. If some other information is available, for example, some secondary topologies or the reactive power measurement, we can integrate them in our existing method to improve the estimation accuracy. This combined method has been developed to a tool to estimate the primary voltages by using secondary AIM measurements.

Table 5. Combined method validation results

	Estimation Mismatch	
	Physics-based	**Combined**
Secondary 1	3.74%	0.90%
Secondary 2	4.06%	0.59%
Secondary 3	3.53%	0.43%
Secondary 4	4.28%	0.79%
Secondary 5	3.84%	1.40%

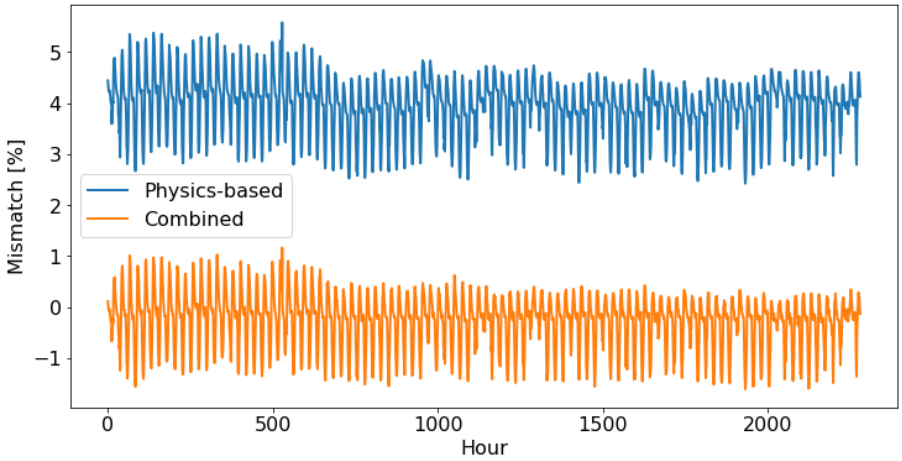

Figure 12. Estimation mismatch from one example secondary model.

Identifying Planning Model Anomalies

The combined method was used to identify the anomalies in the distribution network i planning model of Feeder A first. For this, the primary voltages estimated by the combined method for a selected duration were compared with those obtained from the time-series simulation of the distribution network planning model for the same duration. The peak load and minimum load days in December 2018 and January 2019 (four days) were selected for this process. The average estimation mismatches for all primary buses are shown in Figure 13 and the histogram of all estimation mismatches is shown in Figure 14. The geographic plot is shown in Figure 15.

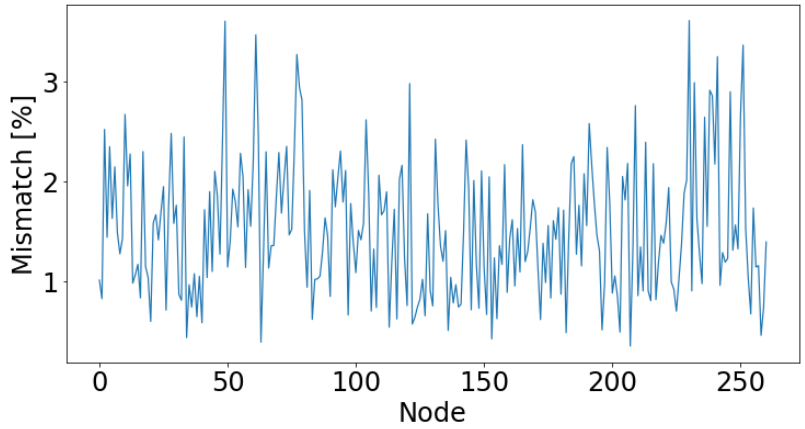

Figure 13. Estimation mismatch for all primary buses on Feeder A.

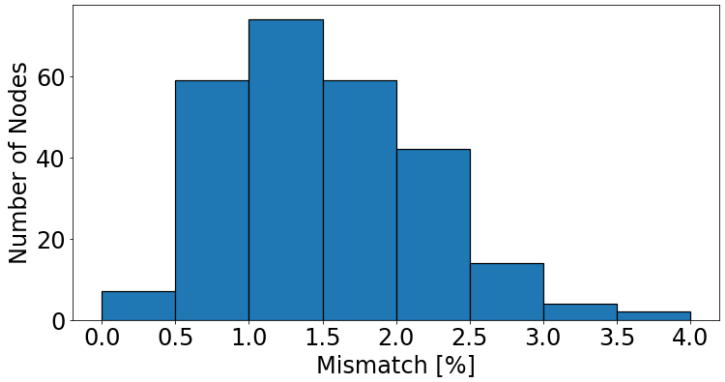

Figure 14. Histogram of all estimation mismatches in Feeder A.

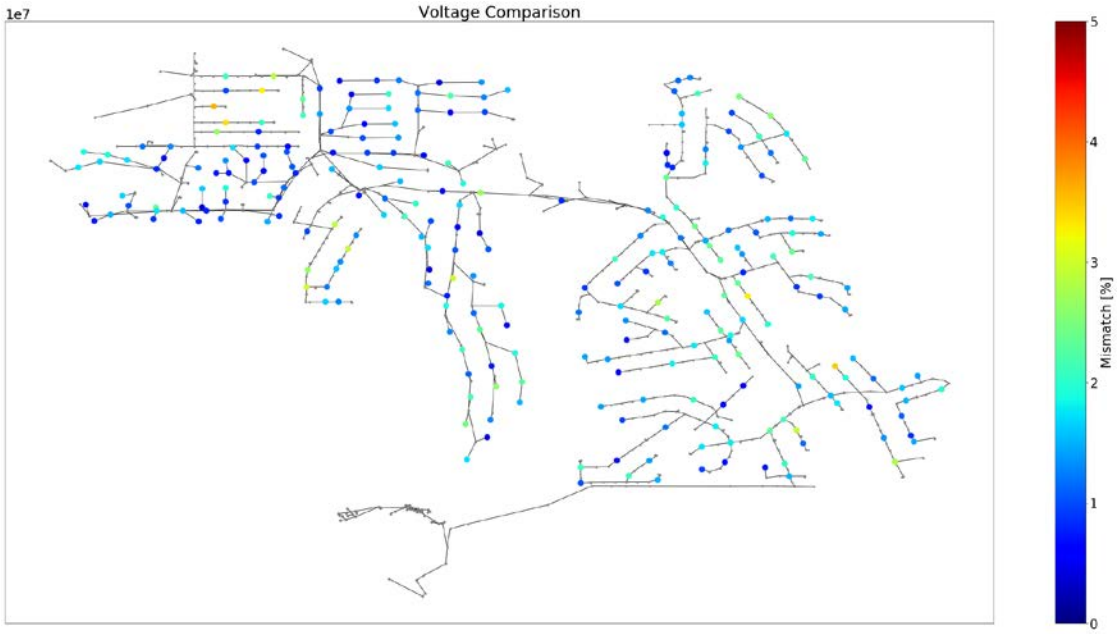

Figure 15. Geographic plot with the mismatch distribution on Feeder A.

The combined method was also used to identify the anomalies in the distribution network planning model of Feeder B first. For this, the primary voltages estimated by the combined method for a selected duration were compared with those obtained from the time-series simulation of the distribution network planning model for the same duration. The peak load and minimum load days in August and September 2019 (four days) were selected for this process. The average estimation mismatches for all primary buses are shown in Figure 16 and the histogram of all estimation mismatches is shown in Figure 17. The geographic plot is shown in Figure 18.

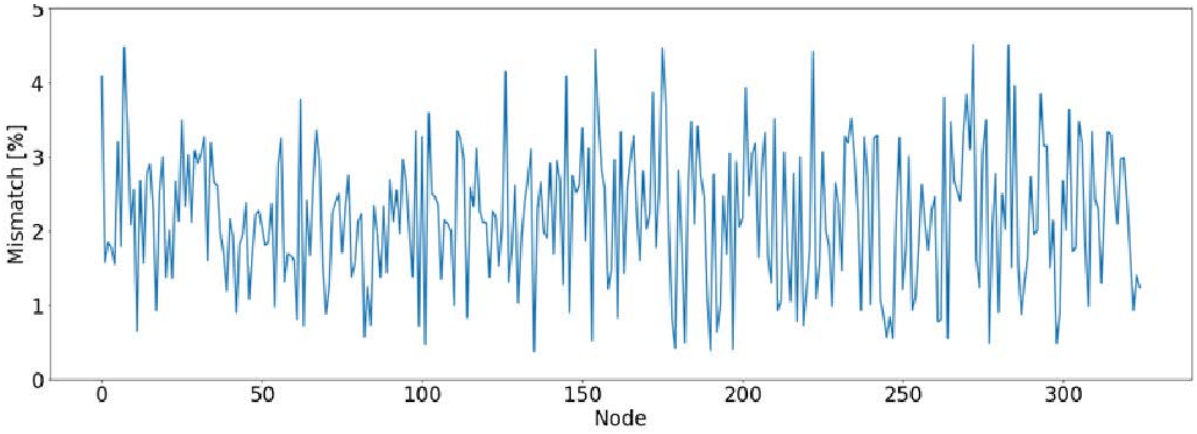

Figure 16. Estimation mismatch for all nodes on Feeder B.

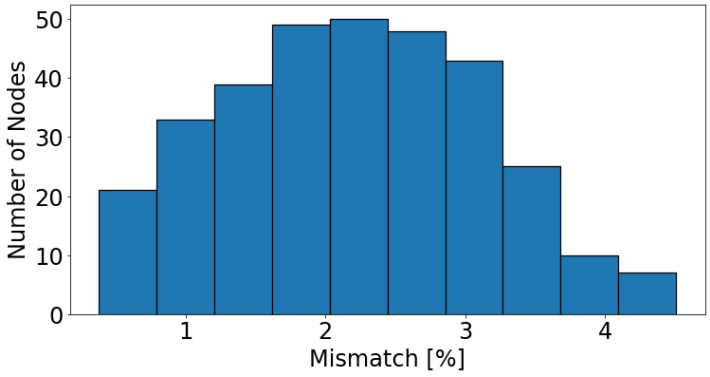

Figure 17. Histogram for all estimation mismatches for Feeder B.

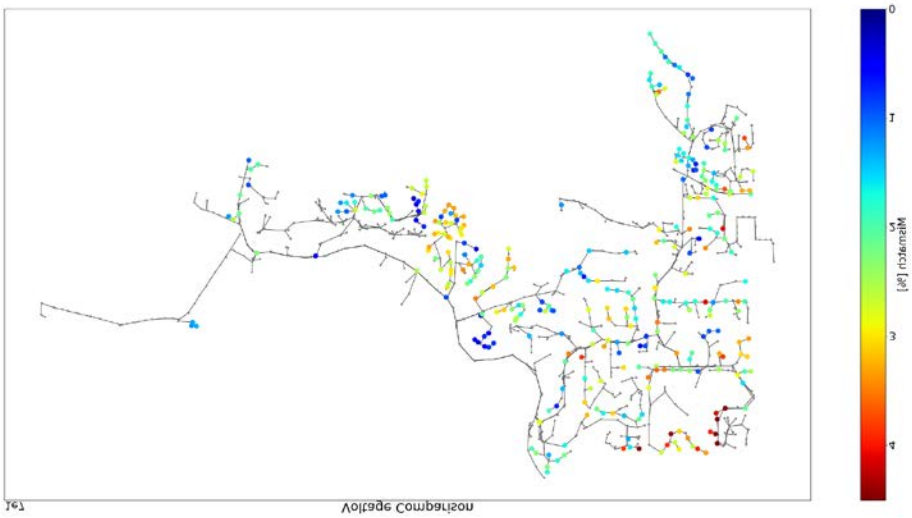

Figure 18. Geographic plot with the mismatch distribution on Feeder B.

Phase Identification Tool:

Phase Identification in Feeder A using 2018 AMI Dataset

Phase identification was performed on Feeder A first using the 2018 AMI dataset. This dataset has the average voltage magnitude time series data for 561 AMI meters for the three-month period between October 1, 2018, to December 31, 2018. The AMI data for two meters per service transformer were available in this dataset. Additionally, the field validated phasing information was also available for all the meters. Based on this information, the distribution of the phasing for the AMI meters is shown in Figure 19.

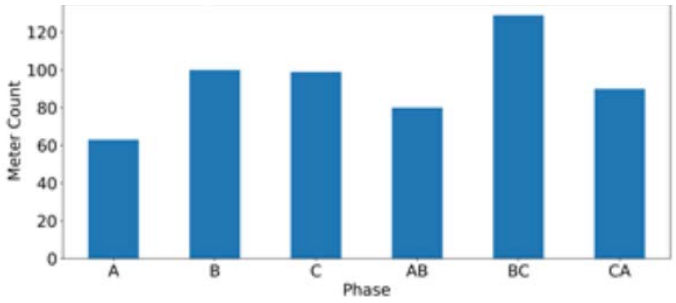

Figure 19. AMI meter phasing distribution in 2018 AMI dataset of Feeder A.

The phase identification results are shown in Figure 20. The field validated phasing information is considered as the ground truth. For each type of phase connectivity, the number of meters the algorithm identified as pertaining to that connectivity is shown against the ground truth. The results show the phase identification algorithm can identify all the types of phase connectivity accurately.

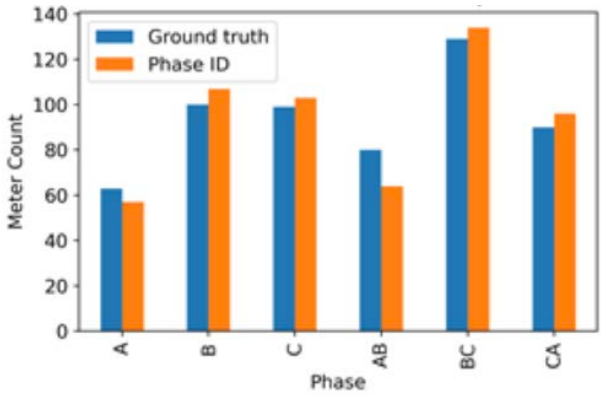

Figure 20. Phase identification results of Feeder A using 2018 AMI dataset.

The detailed breakdown of the AMI meter counts in each of the training, testing, and full datasets of phase identification are shown in Table 6. For each type of phase connectivity, 30% of the meters are selected randomly along with their ground truth phase connectivity for the training dataset. The full dataset includes both the training and testing datasets together. With the phase connectivity identified accurately for 335 out of 391 meters in the testing set alone, the phase identification accuracy is 85.7% on the testing set. The phase identification accuracy on the training and full datasets are 100% and 90%, respectively.

Table 6. Summary of phase identification results of Feeder A using 2018 AMI dataset

Dataset		Phase Connectivity						Total	Accuracy
		A	B	C	AB	BC	CA		
Full	Ground truth	63	100	99	80	129	90	561	90%
	Phase identification	53	96	96	60	122	78	505	
Testing	Ground truth	43	67	70	57	92	62	391	85.7%
	Phase identification	33	63	67	37	85	50	335	
Training	Ground truth	20	33	29	23	37	28	170	100%
	Phase identification	20	33	29	23	37	28	170	

The geographic distribution of the AMI meters for which the phase connectivity identified by the algorithm matched the ground truth is shown in Figure 21. The meters are distributed all over the feeder; thus, the algorithm can detect the correct phase connectivity in all the feeder neighborhoods.

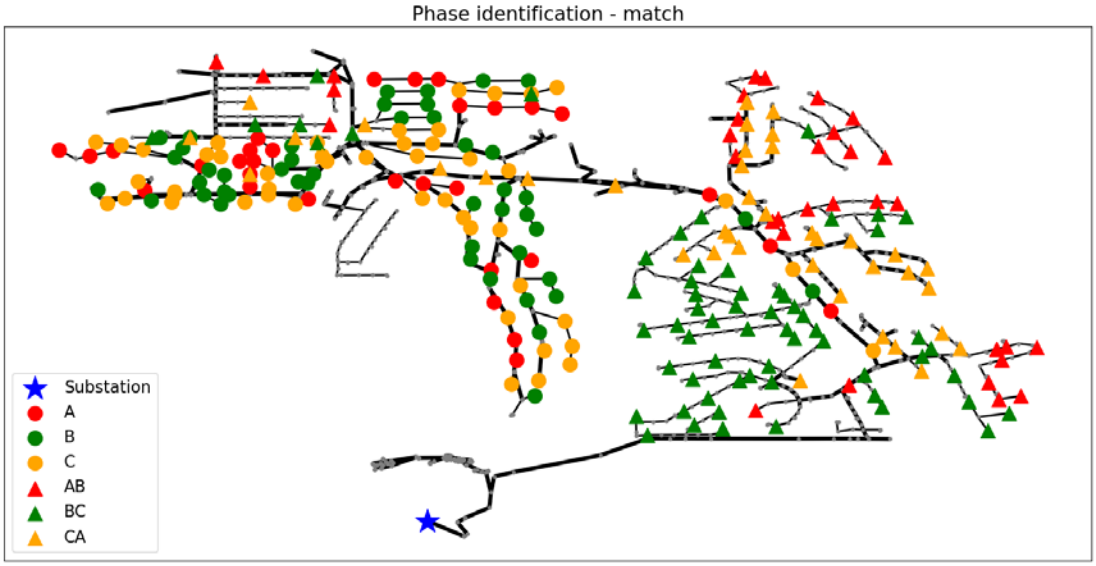

Figure 21. Locations of AMI meters for which the phase connectivity is identified correctly.

The locations of the AMI meters where the identified phase connectivity does not match the ground truth are shown in Figure 22. The correct phase connectivity according to the ground truth is shown in this figure at these locations. The mismatches are generally not clustered or constrained to any specific feeder neighborhoods.

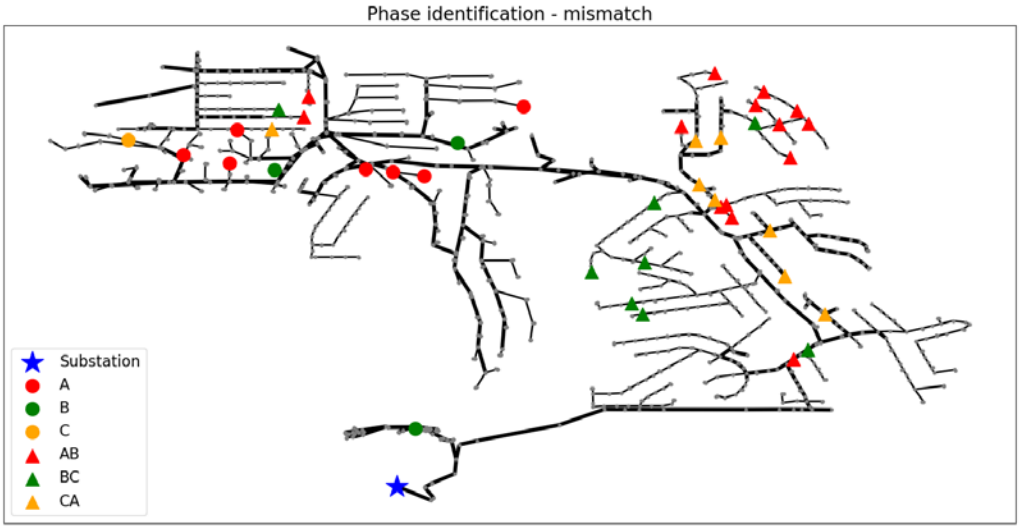

Figure 22. Locations of AMI meters for which the phase connectivity is identified incorrectly.

Phase Identification in Feeder A using 2019 AMI Dataset

The phase identification was also performed on Feeder A using the 2019 AMI dataset. This dataset has the average voltage magnitude time series data for 568 AMI meters for the full 2019-year period. The AMI data for two meters per service transformer were available in this dataset in addition to the field validated phasing information. The phase identification results are shown in Figure 23. The results are similar to those obtained using the 2018 AMI dataset and show that the phase identification algorithm can identify all the types of phase connectivity accurately.

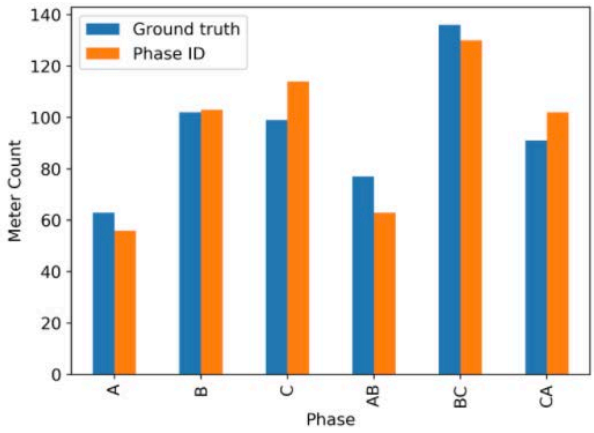

Figure 23. Phase identification results of Feeder A using 2019 AMI dataset.

The detailed breakdown of the AMI meter counts in each of the training, testing, and full datasets of phase identification are shown in Table 7. For each type of phase connectivity, 30% of the meters were selected randomly along with their ground truth phase connectivity for the training dataset. The phase identification accuracies on the testing, training, and full datasets are 86.5%, 100%, and 90.5%, respectively.

Table 7. Summary of phase identification results of Feeder A using 2019 AMI dataset.

Dataset		Phase Connectivity						Total	Accuracy
		A	B	C	AB	BC	CA		
Full	Ground truth	63	102	99	77	136	91	568	90.5%
	Phase identification	55	98	98	56	126	81	514	
Testing	Ground truth	45	72	70	54	96	64	401	86.5%
	Phase identification	37	68	69	33	86	54	347	
Training	Ground truth	18	30	29	23	40	27	167	100%
	Phase identification	18	30	29	23	40	27	167	

The geographic distribution of the AMI meters for which the phase connectivity identified by the algorithm match the ground truth is shown in Figure 24. The meters are distributed all over the feeder; thus, the algorithm can detect the correct phase connectivity in all the feeder neighborhoods.

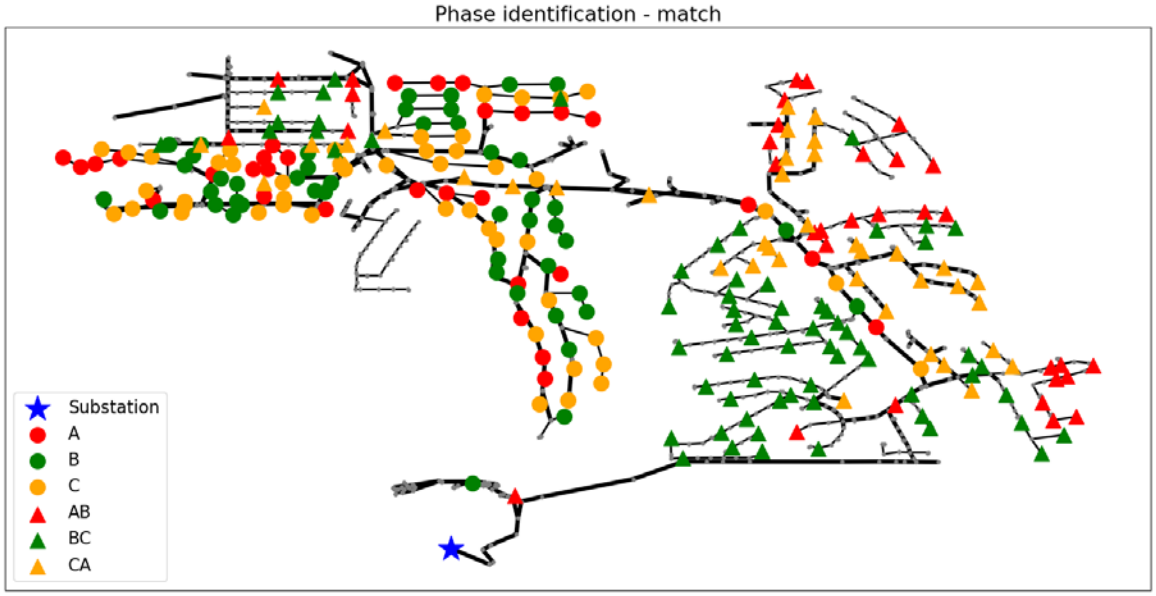

Figure 24. Locations of AMI meters for which the phase connectivity is identified correctly.

The locations of the AMI meters where the identified phase connectivity does not match the ground truth are shown in Figure 25. The correct phase connectivity according to the ground truth is shown in this figure at these locations.

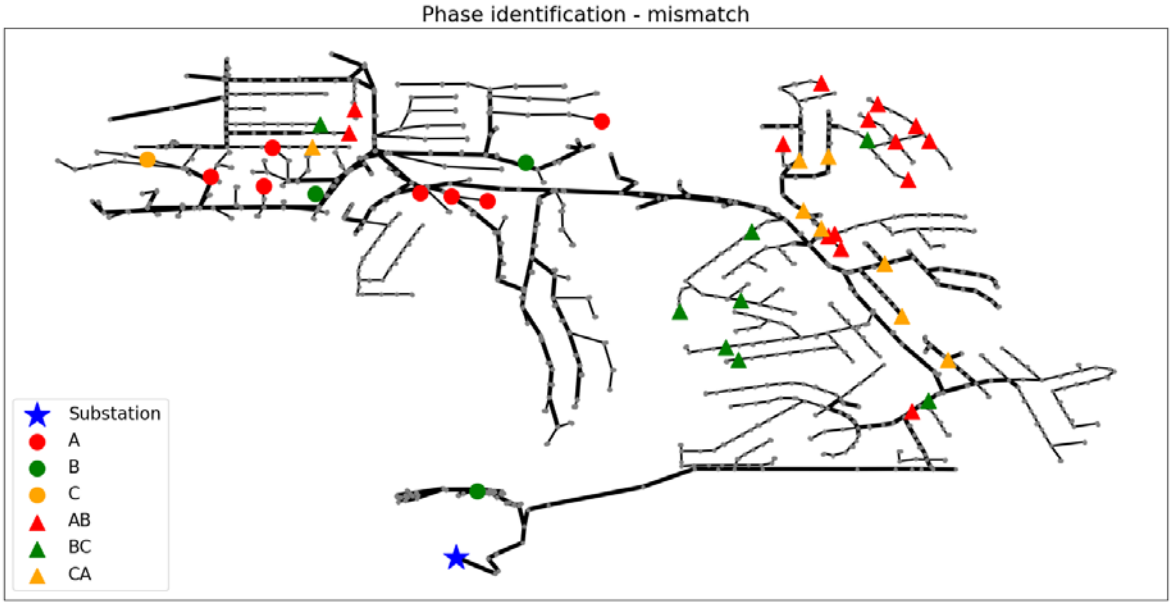

Figure 25. Locations of AMI meters for which the phase connectivity is identified incorrectly.

Phase Identification in Feeder B using 2019 AMI Dataset

The phase identification algorithm was applied to the 2019 AMI dataset of Feeder B. This dataset has the average voltage magnitude time series data for 857 AMI meters for the full 2019-year period. The AMI data for two meters per service transformer were available in this dataset in addition to the field validated phasing information for these meters. The field validated phasing information was considered the ground truth. The phase identification results are shown in Figure 26. The ground truth phasing distribution in this figure indicates this feeder primarily has phase-to-neutral AMI phase connectivity. A small number of meters are connected to phase-to-phase.

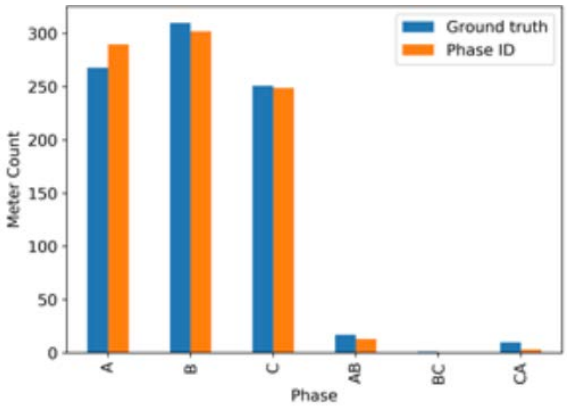

Figure 26. Phase identification results of Feeder B using 2019 AMI dataset.

The detailed breakdown of the AMI meter counts in each of the training, testing, and full datasets of phase identification are shown in Table 8. For each type of phase connectivity, 30% of the meters are selected randomly along with their ground truth phase connectivity for the training dataset. With the phase connectivity identified accurately for 809 out of 857 meters in the testing set alone, the phase identification accuracy is 94.4% on the testing set. The phase identification accuracy on the training and full datasets are 100% and 92%, respectively.

Table 8. Summary of phase identification results of Feeder B using 2019 AMI dataset

Dataset		Phase Connectivity						Total	Accuracy
		A	B	C	AB	BC	CA		
Full	Ground truth	268	310	251	17	1	10	857	94.4%
	Phase identification	260	293	241	12	0	3	809	
Testing	Ground truth	188	217	176	12	1	7	601	92%
	Phase identification	180	200	166	7	0	0	553	
Training	Ground truth	80	93	75	5	0	3	256	100%
	Phase identification	80	93	75	5	0	3	256	

The geographic distribution of the AMI meters for which the predicted phase connectivity matches the ground truth is shown in Figure 27. The meters whose phase connectivity is identified correctly are distributed all over the feeder. This indicates the algorithm can detect the correct phase connectivity in all the feeder neighborhoods.

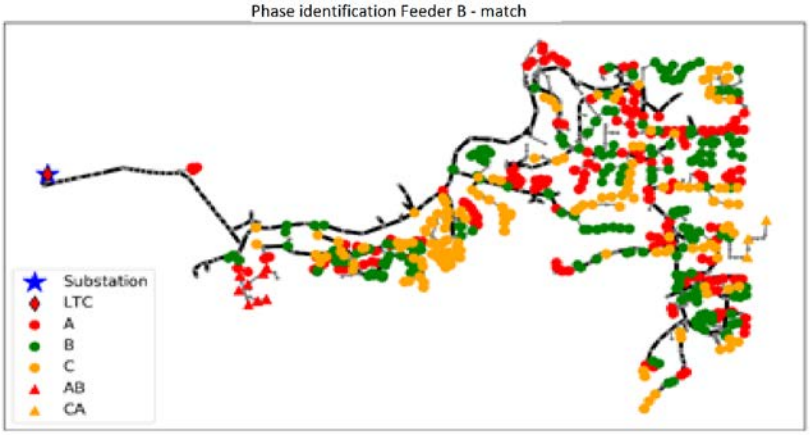

Figure 27. Locations of AMI meters for which the phase connectivity is identified correctly.

The locations of the AMI meters where the identified phase connectivity does not match the ground truth are shown in Figure 28. The correct phase connectivity according to the ground truth is shown in this figure at these locations. The mismatches are generally not clustered or constrained to any specific feeder neighborhoods.

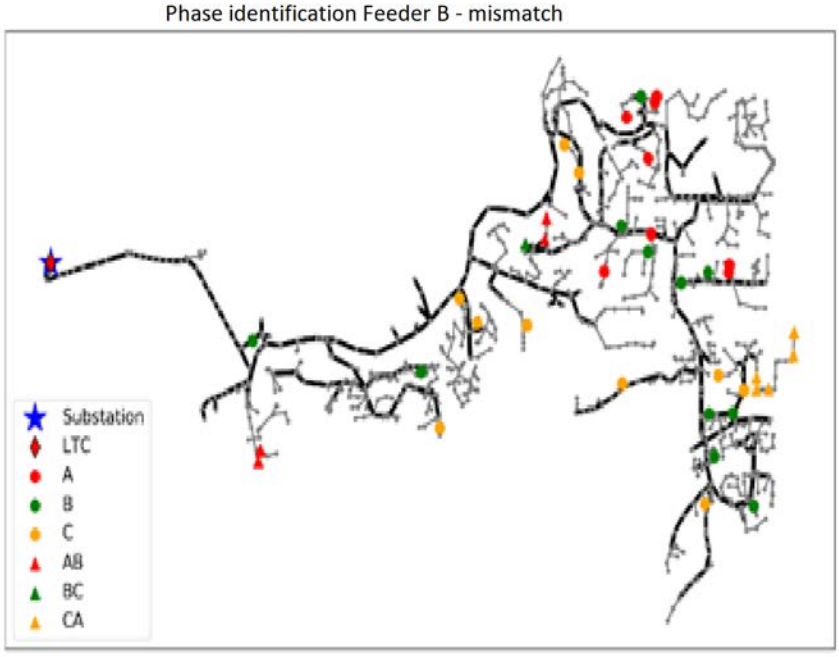

Figure 28. Locations of AMI meters for which the phase connectivity is identified incorrectly.

Meter-to-Transformer Mapping Results:

Tables 9, 10, and 11 below, show the results of the methodology with three different correlation coefficient calculation methods. Table 12 summarizes the overall comparison results. The category, "# of swap" means the total number of incorrect records which are randomly created to test the methodology.

Table 9. Pearson correlation coefficient

Test	# of Swap	Detection	Correction	Right Correction
1	10	10	7	5
2	15	15	8	8
3	20	20	11	10
4	25	24	15	12
5	30	28	22	16
Summary	100	97%	63%	51%

Table 10. Spearman correlation coefficient

Test	# of Swap	Detection	Correction	Right Correction
1	10	10	7	5
2	15	15	10	9
3	20	20	12	10
4	25	24	15	13
5	30	28	23	17
Summary	100	97%	67%	54%

Table 11. Kendall correlation coefficient

Test	# of Swap	Detection	Correction	Right Correction
1	10	10	7	5
2	15	15	11	10
3	20	20	12	10
4	25	24	16	13
5	30	28	23	18
Summary	100	97%	69%	56%

Table 12 indicates that the Spearman correlation coefficient has the highest score in both recall and precision, and therefore fits the methodology best. All methods detect 97% of the incorrect records.

Table 12. Method comparison

Method	Recall	Precision
Pearson	63%	80.95%
Kendall	67%	80.60%
Spearman	69%	81.16%

Task 5: Project Management:

Final Task: CRADA Final Report: Preparation and submission in accordance with Article X.

The EPIC report is prepared to summarize all studies in this project [12].

Subject Inventions Listing:

None

ROI #:

None

References:

[1]Y. Xie, M. Sengupta, Y. Liu, H. Long and A. Habte, "Progress on the National Solar Radiation Data Base (NSRDB): A new DNI computation," *2020 47th IEEE Photovoltaic Specialists Conference (PVSC)*, pp. 0330-0332, 2020.

[2]M. Netto, J. Hao, H. Padullaparti, and V. Krishnan, "On the Use of Smart Meter Data to Estimate the Voltage Magnitude on the Primary Side of Distribution Service Transformers," presented at the IEEE Power and Energy Society General Meeting, 2021

[3]J. Wang et al., "A Machine Learning-based Method to Estimate Transformer Primary-Side Voltages with Limited Customer-Side AMI Measurements," presented at the IEEE Power and Energy Society General Meeting, 2021

[4]G. G. Moisen, "Classification and regression trees," Encyclopedia Ecology, vol. 1, pp. 582–588, 2008.

[5]Y. Ren, L. Zhang, and P. N. Suganthan, "Ensemble classification and regression-recent developments, applications and future directions [review article]," IEEE Comput. Intell. Mag., vol. 11, no. 1, pp. 41–53, Feb. 2016.

[6]Benesty, Jacob, et al. "Pearson correlation coefficient." Noise reduction in speech processing. Springer, Berlin, Heidelberg, 2009. 1-4.

[7]McLeod, A. Ian. "Kendall rank correlation and Mann-Kendall trend test." R Package Kendall (2005).

[8]Mukaka, Mavuto M. "A guide to appropriate use of correlation coefficient in medical research." Malawi medical journal 24.3 (2012): 69-71.

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

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

[11]L. Breiman. “Arcing the edge.”, Technical Report 486, Statistics Department, University of California at Berkeley, 1997.

[12] https://www.sdge.com/sites/default/files/EPIC-3%20Project%203-AMI-Module%201%20Final%20Project%20Report.pdf