Final Technical Report: Improving Solar and Solar+Storage Screening Techniques to Reduce Utility Interconnection Time and Costs

Jeremy Keen,[1] Wenbo Wang,[1] Julieta Giraldez,[2]
Dan Morris,[2] Robert Broadwater,[3] Jason Bank[4]

1. *National Renewable Energy Laboratory*
2. *Kevala*
3. *National Information Solutions Cooperative*
4. *Electrical Distribution Design*

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-6A40-83942
September 2022

Final Technical Report: Improving Solar and Solar+Storage Screening Techniques to Reduce Utility Interconnection Time and Costs

Jeremy Keen,[1] Wenbo Wang,[1] Julieta Giraldez,[2] Dan Morris,[2] Robert Broadwater,[3] Jason Bank[4]

1. *National Renewable Energy Laboratory*
2. *Kevala*
3. *National Information Solutions Cooperative*
4. *Electrical Distribution Design*

Suggested Citation
Keen, Jeremy, Wenbo Wang, Julieta Giraldez, Dan Morris, Robert Broadwater and Jason Bank. 2022. *Final Technical Report: Improving Solar and Solar+Storage Screening Techniques to Reduce Utility Interconnection Time and Costs*. Golden, CO: National Renewable Energy Laboratory. NREL/TP-6A40-83942. https://www.nrel.gov/docs/fy22osti/83942.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-6A40-83942
September 2022

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 Solar Energy Technologies Office. The views expressed herein do not necessarily represent the views of the DOE or the U.S. Government.

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

Agency/Office/Program	Solar Energy Technology Office
Award Number	36470
Project Title	Improving Solar and Solar+Storage Screening Techniques to Reduce Utility Interconnection Time and Costs
Principal Investigator	Jeremy Keen, Researcher III, Jeremy.Keen@nrel.gov, 717-580-5767
Business Contact	Jeremy Keen, Researcher III, Jeremy.Keen@nrel.gov, 717-580-5767
Submission Date	9/19/2022
DUNS Number	80-594-8051
Recipient Organization	National Renewable Energy Laboratory
Project Period	**Start:** 2/3/2020 / **End:** 6/30/2022
Project Budget	Total $1,030,000 (DOE: $1,030,000; C/S: $350,000)
Submitting Official Signature	Signature

1. **Acknowledgement**: This material is based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Solar Energy Technologies Office Fiscal Year 2019 Funding Program, Award Number DE-FOA-0002064.

2. **Disclaimer**: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. The authors would like to thank the project utility partners Pepco (Exelon) and Xcel for providing distribution system information and technical advice on the research conducted.

3. **List of Acronyms**

AMI	advanced metering infrastructure
DERs	distributed energy resources
DERMS	distributed energy resource management system
DEW	Distributed Engineering Workstation
DOE	U.S. Department of Energy

DUNS	Data Universal Numbering System
EDD	Electrical Distribution Design
EPRI	Electric Power Research Institute
GEMINI	Grid-Enhanced, Mobility-Integrated Network Infrastructures
GIS	geographic information system
MAE	mean absolute error
MBE	mean bias error
PEPCO	Potomac Electric Power Company
PV	photovoltaic(s)
RWO	real world object
SAVMVI	System Average Voltage Magnitude Violation Index

4. Executive Summary

Residential solar photovoltaic (PV) installations have increased rapidly over the last decade, and the increased application volume has caused permitting delays and lower overall adoption rates (Fekete, et al. 2022). One cause of delays and low adoption rates are interconnection modeling shortcuts that lead to a failed screening study. For example, in the absence of high-quality distribution network models, utilities may fast-track applications only if aggregate PV capacity is less than 15% of feeder peak load. Alternatively, they might use models that assume customers are directly connected to service transformers. In this project, we evaluated whether data-driven secondary modeling and screening techniques could help utilities assess customer rooftop PV interconnection applications more accurately than traditional screening shortcuts. Throughout the project, feedback was solicited from seven utilities, national laboratories, solar developers, the U.S. Department of Energy (DOE), grid technology vendors, regulators, colleagues, and trade organizations. Data-driven screening techniques were described by one utility as a “right-sized” approach for residential customers given the low-risk of small errors and the high-cost of accurate modeling.

Data-driven solar screening required two new capabilities. First, utilities rarely have secondaries in their distribution models. These are the “grid edge” wires connecting customers to the primary distribution network. Data-driven “supervised learning” methods were developed to create the synthetic secondary network model. Second, a data-driven “random forest” method was developed to predict the likelihood of a customer passing a screening study based on characteristics of their interconnection request (e.g., solar size) and the distribution network (e.g., penetration of solar). These capabilities were developed by the National Renewable Energy Laboratory (NREL), Kevala, and Electrical Distribution Design (EDD). Data and feedback were provided by Xcel Energy, Potomac Electric Power Company (Pepco), and a utility technical advisory committee.

To model secondary networks, methods were developed to predict secondary topologies and secondary conductors. Secondary topologies are predicted using decision trees and commonly available information, such as service transformer, customer, building parcel data and street locations. Decision trees are a transparent, rule-based method for predicting common residential topologies, and for this project, they were validated using Kevala’s “Trace Assist” tool. Conductors were predicted using

a logistic regression method based on real world object (RWO) types, service transformer ratings, conductor length, and distance to transformer.

The project used voltage comparisons derived from geographic information system (GIS)-derived secondary models, advanced metering infrastructure (AMI), and hosting capacity to study the accuracy of the predicted secondary networks. Pepco synthetic secondary network voltages, relative to circuit models with GIS-derived secondary models, had a mean absolute error (MAE) of 0.35 volts. Pepco synthetic secondary network voltages, relative to AMI voltages, was 1.65 volts MAE, with the additional error likely being caused by error in the primary distribution network model. In the hosting capacity studies with Pepco data, we found utility screening shortcuts can lead to 30-50 additional percentage points of screening failures and predicted secondaries led to 20 percentage points of additional failures (see Figure 14). To further reduce the error associated with the predicted secondary networks, we also developed a calibration tool with Pepco data that modified conductor line lengths to minimize error. The average error improvement was 0.25 volts. Larger improvements are expected on distribution networks with fewer data quality issues, and the project team is exploring several such opportunities to continue work on developing synthetic secondary networks.

After developing the combined primary and secondary distribution network model, Pepco hosting capacity results were used to train a random forest model to predict the pass/fail likelihood of a customer application. The number of PV customers, primary solar penetration, distribution voltage class, and several other features were found to be good predictors of customer application success (Figure 16, page 23). Power-flow based models using synthetic secondaries and data-driven methods both increased the screening success rate, relative to common utility heuristics, by as much as 55 percentage points (Figure 18). Further validation work is needed determine whether the screening tool results have sufficient tolerance to account for unforeseen network operational changes.

5. Table of Contents

1. Acknowledgement 1
2. Disclaimer 1
3. List of Acronyms 1
4. Executive Summary 2
5. Table of Contents 4
6. Background 5
7. Project Objectives 6
8. Project Results and Discussion 8
9. Significant Accomplishments and Conclusions 26
10. References 28

6. Background

Distribution network analysis is often limited by data quality and availability. Manually collecting and cleaning data from a geographic information system (GIS) or assets in the field is expensive. The large number of assets and the dynamic nature of distribution networks also make it challenging to maintain quality models. These problems are exacerbated when adding secondary networks[1] to distribution network models, which have been shown to be important for understanding impacts of DERs (distributed energy resources). Developing secondary network models would be even more resource-intensive than primary distribution network models because GIS data are often missing and data collection would require utilities to manually inspect sites for thousands of customers. The alternative has been to use heuristic (i.e., shortcut) models that lack validation.

This project has two unique contributions. The first contribution is the development of data-driven credible synthetic secondary networks connecting service transformers to residential loads. This contribution includes methods for creating and validating the synthetic secondary networks. The second contribution leverages the synthetic secondary networks to develop a data-driven residential solar application screening tool. To our knowledge, the screening tool is the first attempt to develop statistical rules characterizing the relative importance of different secondary network and solar characteristics, and it is the first to apply these rules to predict customer screening success.

Power systems analysis has a long development history of representative and synthetic network models (Billinton Jonnavithula 1996) (IEEE 2022). NREL developed "Synthetic Models for Advanced Realistic Testing: Distribution Systems"[2] (Smart-DS) use geospatially accurate parcel and road data to model distribution secondary networks (NREL 2022). In the absence of utility network models, Smart-DS is sometimes used for location-specific analysis. For example, the GEMINI (Grid-Enhanced, Mobility-Integrated Network Infrastructures) project co-simulates Smart-DS feeders power-flow and electric vehicle charging demand in the San Francisco Bay Area (Muratori, et al 2020). However, despite their use in regional analysis, representative or synthetic networks have never been designed for specific customers or to inform specific utility investments.

Other research has tried to use voltage measurements to predict or improve existing distribution network models. Several attempts have been made to estimate power distribution network topologies, including secondary networks, using voltage correlations in AMI data (Lave, Reno Peppanen 2019). These methods were explored by the project team but were unsuccessful because they are not robust to poor AMI coverage or poor customer-to-transformer mappings, which are pervasive in utility GIS. Methods that use AMI measurements to clean network models, either through phase

[1] In this report, the term "secondary networks" refers to the radial low-voltage connections between service transformers and customers. The project scope did not include meshed networks or downtown networks with closed loops.

[2] https://www.nrel.gov/grid/smart-ds.html

identification or customer-to-transformer mappings are more successful (Short 2013) (Blakely Reno 2020) (Padullaparti, 2022). However, in our experience, data-driven techniques to clean customer-to-transformer mappings cannot achieve 100% accuracy and perform poorly when GIS lack customer geospatial coordinates. In the absence of high-quality models, researchers and utilities have frequently used prototype secondaries or heuristics to model secondaries. Prototype secondaries are based on common topologies and conductors. This approach is used often by researchers without validation or an impact assessment. Based on the project's advisory board's feedback, utilities most commonly use heuristics with minimal secondary network representation. For example, utilities have commonly used a solar PV screen that fast-tracks applications if the aggregate PV capacity (including behind-the-meter secondary-connected solar and front-of-the-meter solar) is less than 15% of feeder peak load. Reno and Broderick show that this simplified assessment has become overly conservative as customer solar adoption increases (2015). Other common modeling approaches include direct customer-to-transformer connections and direct customer-primary-connections with assumed secondary network voltage drops.

The development and maintenance of distribution network models is resource-intensive. In recent years, research into data-driven hosting capacity has explored the possibility of circumventing network models using machine learning (Zhang 2017) (Taheri, 2021). To our knowledge, data-driven approaches have not been applied directly to interconnection screening processes.

7. Project Objectives

Residential PV installations have increased rapidly over the last decade, and the increased application volume has caused permitting delays and lower overall adoption rates (Fekete, et al. 2022). These delays increase soft costs of solar installation and jeopardize national goals to reduce the levelized cost of solar electricity (EERE 2011). Data-driven secondary modeling and screening techniques can help utilities assess customer applications more quickly and with greater certainty. Additionally, screening guidelines based on common distribution network attributes can be used to identify inequities in the grid that prevent customers from benefiting from solar (Brockway, Conde and Callaway 2021).

The goal of the project is to show the value of improving the currently low-fidelity representation of the grid-edge to 1) improve interconnection evaluation of solar PV and solar plus storage (S+S) and propose new and improved technical screens, and 2) enable innovative mitigation strategies (e.g., advanced inverters, residential battery dispatch) with high-spatial-resolution load and DER forecasting to be considered against traditional wires alternatives when alleviation of solar PV and S+S impacts are required.

Table 1 summarizes the tasks completed by the project to achieve these goals. Table 2 summarizes the project milestones and go/no-go decision points.

Table 1. Summary of Tasks

Task	Task Name	Description
1	*Stakeholder engagement*	*Create project advisory groups throughout the project to receive and integrate the feedback of the users, technical peers, and broader stakeholders throughout the project.*
2	*Development of secondary modeling method*	*Develop a methodology to approximate secondary circuits in at least two utility territories, incorporating the feedback received from the utility user and technical peer groups*
3	*Value of improved secondary modeling*	*Compare improvements in the evaluation of solar and S+S with detailed secondary topology. Evaluate and quantify metrics identified with the utility user and technical peer review groups, such as voltage and thermal overload that drive the technical evaluation of solar and S+S applications in the utility interconnection process.*
4	*Secondary scenario modeling and refined technical screens*	*Use scenario modeling techniques to create a data-driven approach to identify lessons learned from the secondary topologies that can inform utilities in the evaluation of the impact of solar and S+S customer applications. Use the data-driven approach to derive new and more sophisticated technical screens for the evaluation of solar and S+S interconnections. Disseminate lessons learned from the secondary scenario modeling hosting capacity results.*
5	*Screening guidelines with innovative mitigation strategies*	*Evaluate innovative mitigation strategies on improving the interconnection screening study success. Mitigation strategies will include advanced inverter grid support functions and storage. Load and PV conditions will use high spatial resolution forecasting techniques.*

Table 2. Summary of Project Milestones and Midpoint Go/no Go Decision

Year #	Task #	Milestone	Milestone Name/Description	Planned End Date
1	1	1	Create expert groups	03/31/2020
1	1	2	Demonstrate preliminary methods to approximate secondary circuits to the utility user and technical peer groups and establish metrics to measure the midpoint go/no-go goal	6/30/2020
1	2	1	Validate the final algorithms with real secondary circuit data from at a least two utilities	09/30/2020
1	3	1	Compare improvements in evaluating impacts from solar and S+S in Synergi and Distribution Engineering Workstation (DEW)	12/31/2020
1	3	2	*Midpoint Go/No-Go: Inability to show the improvement in the metrics established with the technical peer review group would result in a project stop.*	12/31/2020
2	1	1	Conduct workshop with project advisory group to incorporate feedback from scenario modeling preliminary results and on priorities for the cost-benefit analysis of mitigation strategies	6/30/2021
2	3	3	Compare improvements in evaluating impacts from solar and S+S in OpenDSS	6/30/2021
2	4	1	Compile lessons learned from secondary scenario modeling and improved technical screens, and dissemination of results	6/30/2021
2	5	1	Develop high-spatial-resolution load and DER forecast for the cost-benefit analysis and integrate in OpenDSS	6/30/2021
2	5	2	Calibrate secondary networks with AMI	3/31/2022
2	5	3	Validate Secondary networks at scale	12/31/2022
2	5	4	Develop advanced screening guidelines with innovative mitigation strategies	3/31/2022

8. Project Results and Discussion

Task 1: Stakeholder Engagement

Task 1 covered Milestones 1.1.1, 1.1.2, 2.1.1, and 2.4.1. Feedback was solicited from stakeholders throughout the project. Advisory board meetings were held in June 2020 and November 2021, a workshop was held April 2021, and technical peer feedback was received in June 2020. The advisory board meetings solicited feedback from seven utilities. Peer review was received from four colleagues from the National Renewable Energy Laboratory, Sandia National Laboratories, and EPRI (the Electric Power Research Institute). The workshop had approximately 30 attendees from utilities, national laboratories, solar developers, DOE, grid technology vendors, regulators, and other trade organizations.

As part of Milestone 1.1.2 and the project's midpoint go/no-go decision, the secondary methods and metrics were shared with utility and peer group advisors. Metrics shared with advisors included voltage error metrics, line length metrics, customer-to-transformer mapping, and customer degree. Voltage error metrics were then calculated using GIS models and AMI data.

Key findings from utility advisory board members included the following:

- The project should focus on older overhead secondaries. Older secondaries lack design standards, GIS data are less useable, and cables tend to be undersized with higher resistivity and more shared customers.
- Most utilities do not model secondary networks in their interconnection process and instead check only the primary network for thermal and voltage constraints.
- Most utilities do consider secondary networks when PV clustering is present, but heuristics are used to understand PV impacts. Heuristics include modeling only the secondary transformer or assuming a fixed voltage drop.
- Data-driven screening techniques are a "right-sized" approach for residential customers given the low-risk of small errors and the high-cost of accurate modeling.

These findings validated the projects approach. One ongoing challenge is choice of secondary types to model. Though older overhead secondaries are a good starting point, in future work, the project team hopes to provide a solution with more coverage of secondary topology types.

Key findings from peer reviewers included the following:

- The SAVMVI (System Average Voltage Magnitude Violation Index)[3] metric is not recommended because it can be hard to interpret and with it large temporary violations are difficult to see.
- Inaccuracies in primary models will drive secondary errors.
- Inaccuracies in secondary customer-to-transformer mappings and service transformer parameters will drive calculated secondary errors.
- It will be important to better understand the size of the ground-truth data set needed for validating the project's data-driven approaches.
- The project should consider using a metric for the number of customer connections to a transformer.
- In the future, a sequence of methods should be applied to validate primary and secondary portions of the distribution network model.

These peer review findings predicted many project challenges. Unfortunately, data quality in available primary models and secondary models prevented the project team from following all recommendations. The suggestion to apply a sequence of methods at all distribution network levels was out of scope for this project, but it is being proposed for future work with interested utilities. Though the SAVMVI metric was not applied, the

[3] For definition and related metrics see: https://www.nrel.gov/docs/fy17osti/67296.pdf

project team still relied on average voltage error scores that would not reveal large temporary deviations.

Key findings from the workshop included the following:

- Autonomous inverter settings and traditional upgrades (e.g., reconductoring) are the most commonly used DER grid integration strategies used by utilities (see Figure 1).
- In the next 10 years, DERMS (distributed energy resource management systems), net-load controls, and PV curtailment are also expected to be more widely used (see Figure 2).

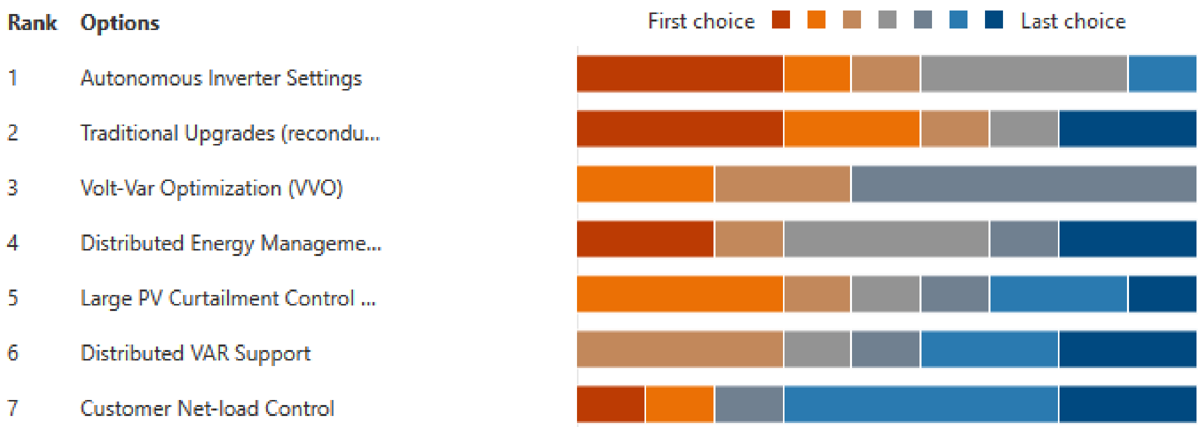

Figure 1. Workshop question: Please order the following DER grid integration strategies by current order of importance to your company. Based on nine responses from workshop participants.

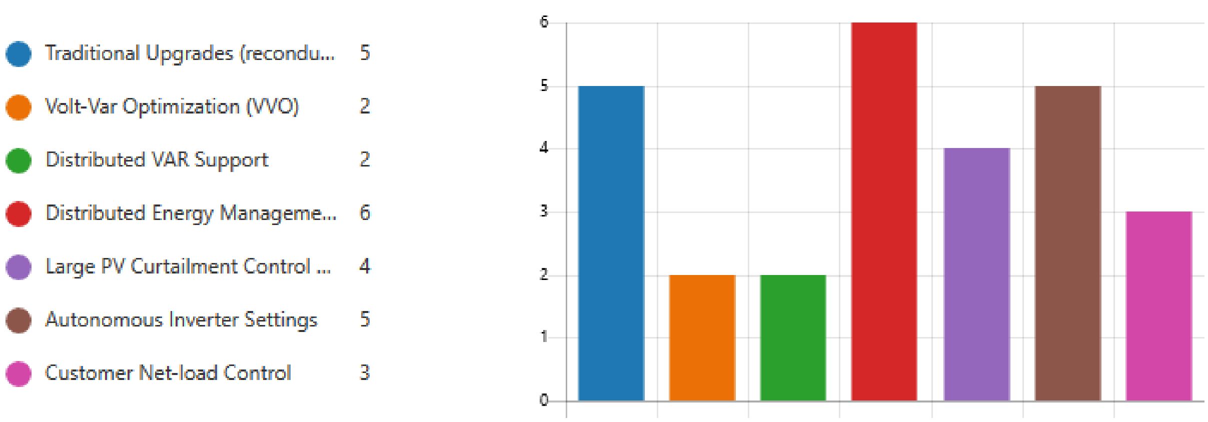

Figure 2. Please select the three grid integration strategies/technologies that will be the most important in the next 10 years for integration of DERs in the distribution system. Based on nine responses from workshop participants.

During closeout (Milestone 2.4.1), the project team is continuing to engage utility project partners and other stakeholders interested in synthetic secondary modeling. On June 27, 2022, final project results were shared with Pepco. The NREL, Kevala, and Electrical Distribution Design (EDD) project team members were requested to write a proposal for Pepco, which is in progress. The project team is also sharing results with New York utilities. The objective is to find utility partners with high-quality primary models and develop secondary models. This strategy will allow improved error characterization and secondary calibration (discussed under Task 2).

Task 2: Development of Secondary Modeling Method

Task 2 covered Milestones 1.1.2, 1.2.1, 2.5.2, and 2.5.3. The data-driven synthetic secondary models require methods to predict secondary topology and conductors. Details of the method were published by the project team (Wang, et al. 2022). Figure 3 summarizes the flow of data and methods to develop the secondary models. Open-source map data, utility customer data, and GIS data were used to train a decision tree to predict secondary topologies. Topology and GIS information was then used to predict conductor types on the secondary models. The secondary power-flow models could then be integrated with the feeder primary models.

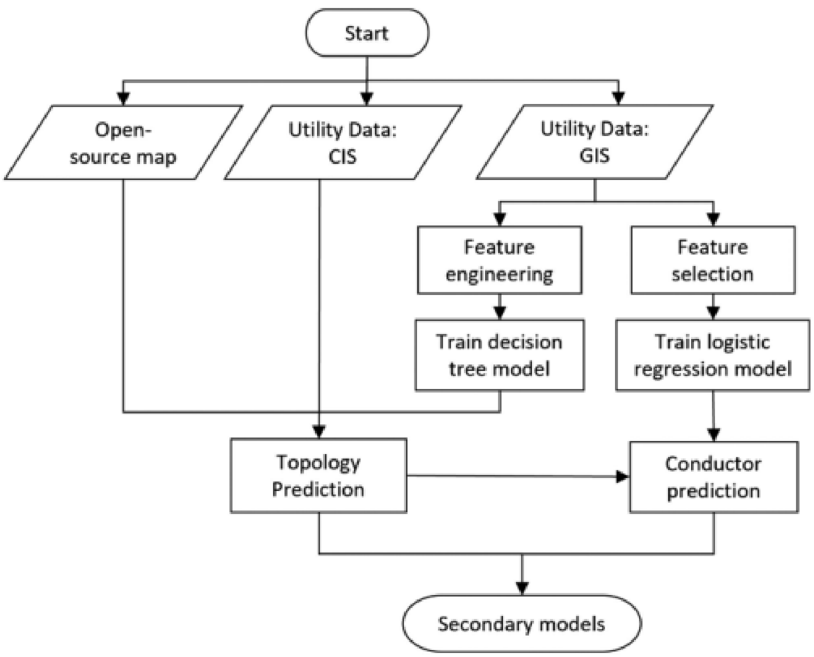

Figure 3. Secondary topology and conductor prediction workflow

Secondary Topology Prediction

Secondary topologies were predicted using decision trees with Pepco and Xcel data. The project team first analyzed utility secondary networks in GIS to determine common types of secondary networks. The objective was to develop a decision tree to predict common residential topologies. Figure 4 summarizes the decision tree logic for selecting secondary network topologies. Figure 5 shows the three types of topologies that the decision tree can predict: 1) direct connections linking customers to a service transformer with a secondary conductor and service drop, 2) a branch connection with secondary conductors that follow poles and connects customers to nearby poles with a service drop, and 3) intermediate branch connection that follow the logic of Topology 2 but connect customer with a “Y” branch if the customers are close.

The topology branching rules used were as follow:

- If a service point is within X meters of a service transformer, the conductor takes the shortest path from the service point to the service transformer.
- If a service point is within Y meters of a branching node, the conductor takes the shortest path to this node before following the main branch to the transformer
- If two service points are within Z meters of one another, they are joined by a small intermediate branch before connecting to a main branch node.

The X, Y, and Z parameters were trained on utility secondary data and selected to minimize path length error with actual topologies. Figure 6 and Figure 7 show actual and decision-tree predicted secondary topologies.

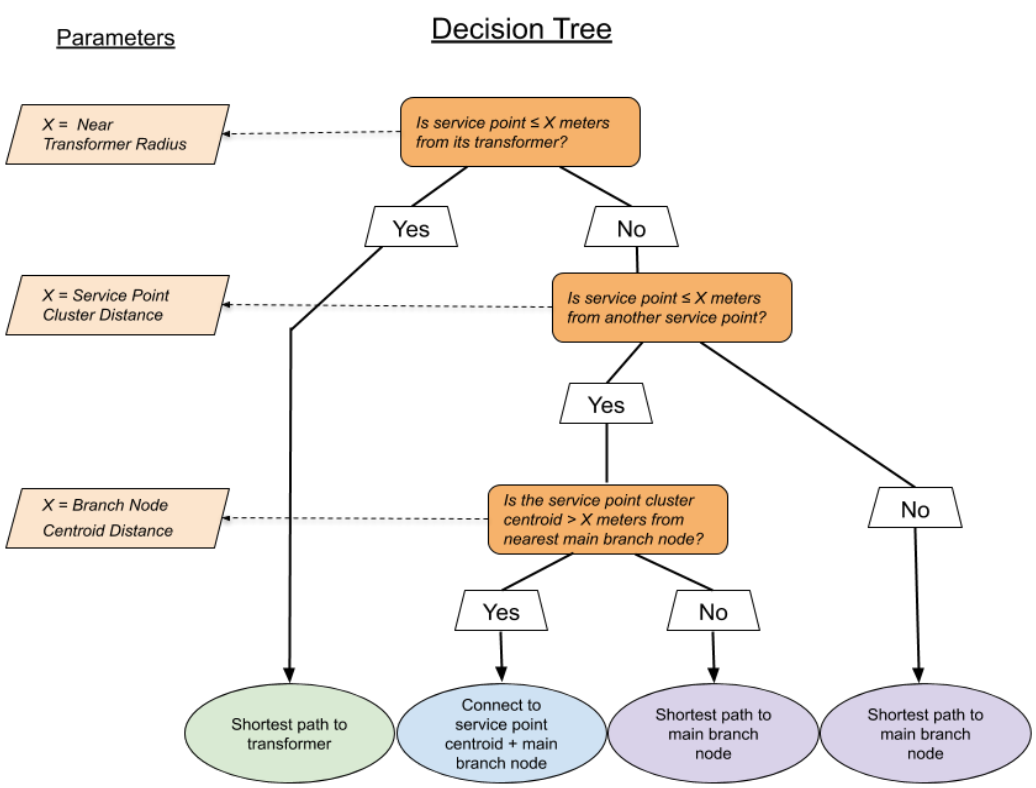

Figure 4. Secondary network decision tree

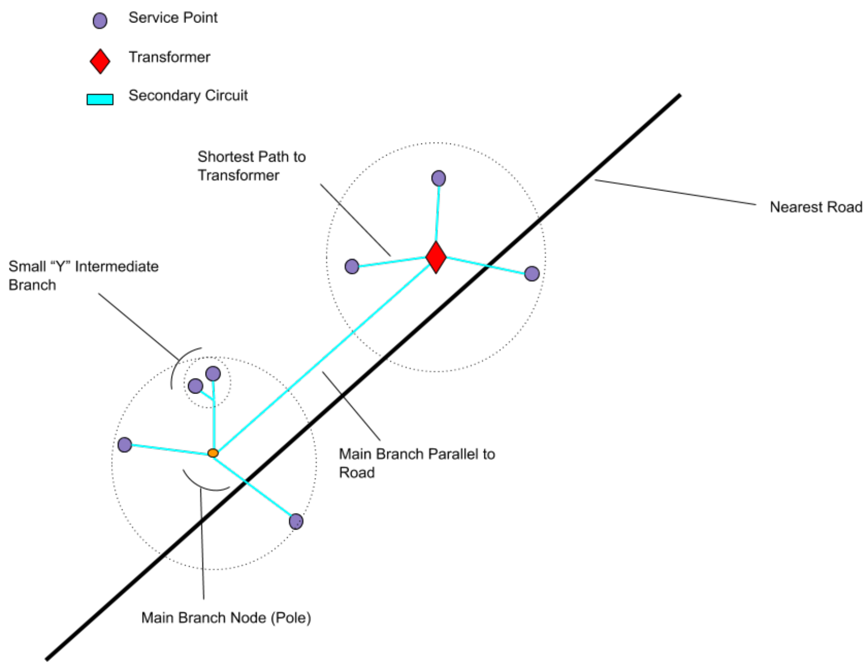

Figure 5. Secondary network topology options

Figure 6. Predicted and true secondary topologies for a direct customer connection

Figure 7. Predicted and true secondary topologies for a topology with several branches

More than 500 secondary networks on seven feeders were created and validated for Pepco and Xcel. The secondary network topologies were validated using Kevala's Trace Assist tool. It analyzes Google Street View imagery to determine secondary topologies. It could also be used to validate existing utility models.

Two topology error metrics were calculated. The path length error metric is defined as the error in distance between the customer service point and service transformer. The path length error ranged from 6.9 to 20.5 meters on feeders that were examined. Upstream load count metric looks at error in the number of secondary customers upstream from a service point. The upstream load count error ranged from 0.1 to 1.3 customers. These metrics were shared with stakeholders for the June 30, 2020, midpoint go/no-go decision.

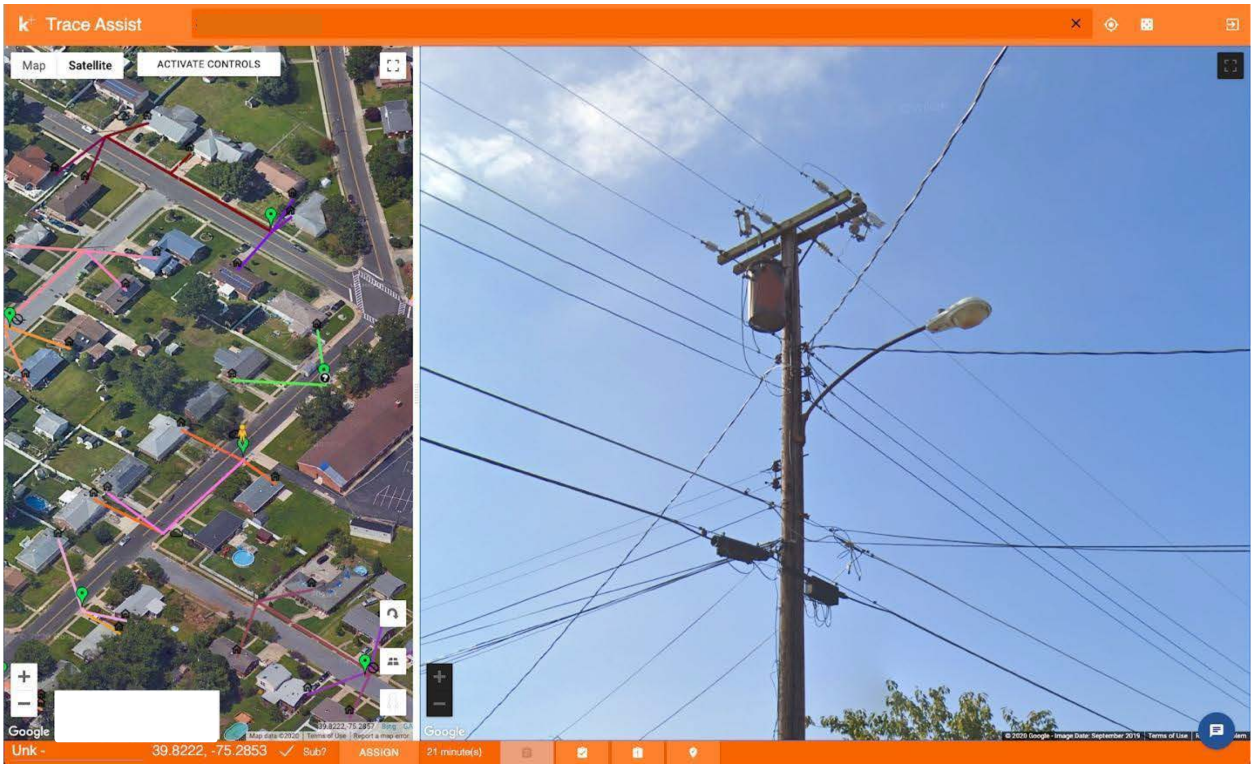

Figure 8. Screenshot of Kevala Trace Assist Tool

Secondary Conductor Prediction

After predicting secondary topologies, a logistic regression model was used to predict conductor types on the secondary network. Several features were used to predict conductor type, including real world object (RWO) types, service transformer ratings, conductor length, and distance to transformer. Of all the attributes, the RWO type is the best predictor of conductor type. This is consistent with the engineering understanding that secondary main lines have higher capacity than service drop lines. The logistic regression prediction achieves approximately 50% accuracy as shown in the confusion matrix in Figure 9. In future work, the project team hopes to include construction year in the logistic regression model. Construction year is known to affect conductor type and will likely increase the accuracy of the logistic regression model.

Ultimately, a primary objective of synthetic secondary network models is to accurately predict voltage. An incorrect conductor prediction may not cause significant error in predicted voltages if the conductor length is small or if the impedance of the predicted conductor is close to the impedance of the actual conductor. Thus, metrics that summarize voltage error are a better descriptor of the efficacy of the synthetic secondaries.

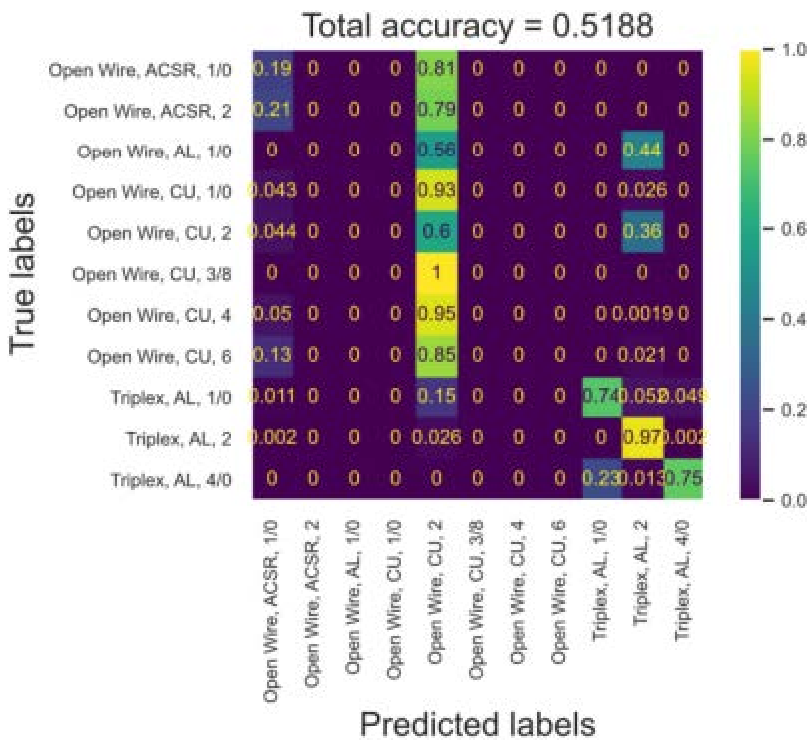

Figure 9. Confusion matrix for conductor prediction using secondary logistic regression with Pepco data

Secondary Network Validation

Two validation tests were performed to estimate the voltage accuracy of the synthetic secondary networks. In the first test, service point voltages from synthetic secondary networks were compared to voltages from GIS secondary networks on eight feeders. The results are summarized Table 3. The average MAE and the mean bias error (MBE) are 0.35 volts and -0.14 volts respectively. These results suggest the synthetic secondary add approximately 0.35 volts of error to customers' modeled voltages, and they are fairly unbiased (i.e., the average bias is less than 0.15 volts).

Table 3. Synthetic Pepco Secondary Voltage Error, Synthetic Compared to GIS Secondary Networks

Feeder	MAE	MBE
1	0.33	0.05
2	0.41	-0.12
3	0.25	-0.03
4	0.18	-0.02
5	0.57	-0.37
6	0.49	-0.42
7	0.57	-0.54
8	0.4	-0.21
All	0.35	-0.14

In the second validation test, AMI voltages were compared with modeled feeders on 35 Pepco utility feeders. MBE and MAE are shown in Table 4. Figure 10 shows a box-and-whiskers plot describing the error found at all service points. It is important to note that one reason the modeled AMI voltage errors are higher than GIS errors is that the AMI errors include error introduced at the primary model level. It was not in the project scope to validate utility primary models and minimize this this error. Nonetheless, AMI comparisons provide insight into potential improvements.

Table 4. Modeled Primary and Synthetic Secondary versus AMI, Voltage Error on 34 Pepco Feeders

MBE (V)			MAE (V)		
Phase A	Phase B	Phase C	Phase A	Phase B	Phase C
0.67	0.77	0.62	1.65	1.74	1.68

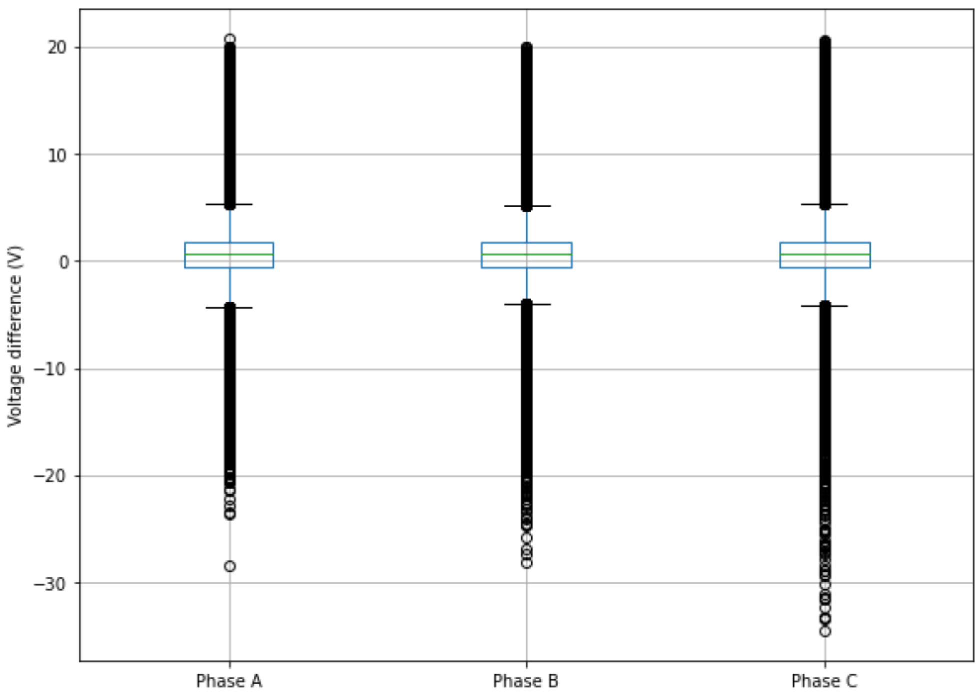

Figure 10. Modeled Pepco primary and synthetic secondary versus AMI, voltage error box-and-whiskers plot

Though the mean errors were within the project's 2-volt error target, a significant number of service points had larger errors. An assessment was done to characterize the source of error. Table 5 summarizes the error characterization. There were several key findings. First, large error variation among feeders suggests large variability in the error introduced by the primary feeder model. Second, data quality on the secondary networks also introduced error. A large number of service points did not have known geospatial coordinates and had to be imputed. Location imputation added on average 0.5 volts of MBE. Third, topologies directly connecting customers to service transformers added on average 0.5 volts of MBE. Larger numbers of customers and larger distances between customer and transformer showed a clear trend in increased

MBE. Neither the presence of solar nor the type of line (overhead or underground) impacted error.

Table 5. Modeled Pepco Primary and Synthetic Secondary versus AMI, Voltage Error Source Characterization

Source	Error Characterization
Feeder	MBE, MAE ranges from -0.4 to 5.0 volts
Original location known	Average 0.5-volt impact on MBE
Branching rule	Significant impact on MBE, especially class I
Distance to transformer	MBE ranges from 1.0 to -4.0 volts with distance
Customer count	MBE ranges from 1.0 to -4.0 volts with customer count
Overhead or underground	No Significant impact
PV on parcel	No Significant impact

Secondary Network Calibration

For the final milestone (2.5.2) associated with secondary network development, techniques were deployed to calibrate secondary line lengths to reduce error between modeled Pepco voltages and AMI. The project team targeted demonstration on three or more secondaries and a MAE of 0.5 volts. The project exceeded the target number of secondaries to test the method on. Four-hundred secondaries were tested, exceeding the target, but the MAE target was not achieved on all of the secondaries. Though the calibration methods were successful on many secondaries, on average, the methods were unable to bring secondary voltage errors within 0.5 volts. This underperformance was likely caused by model errors upstream of the service transformer (which secondary network calibration cannot address). Figure 11 shows an example where secondary calibration was able to reduce voltage error significantly. On average, for all the secondaries, the reduction in voltage error was approximately 0.25 volts.

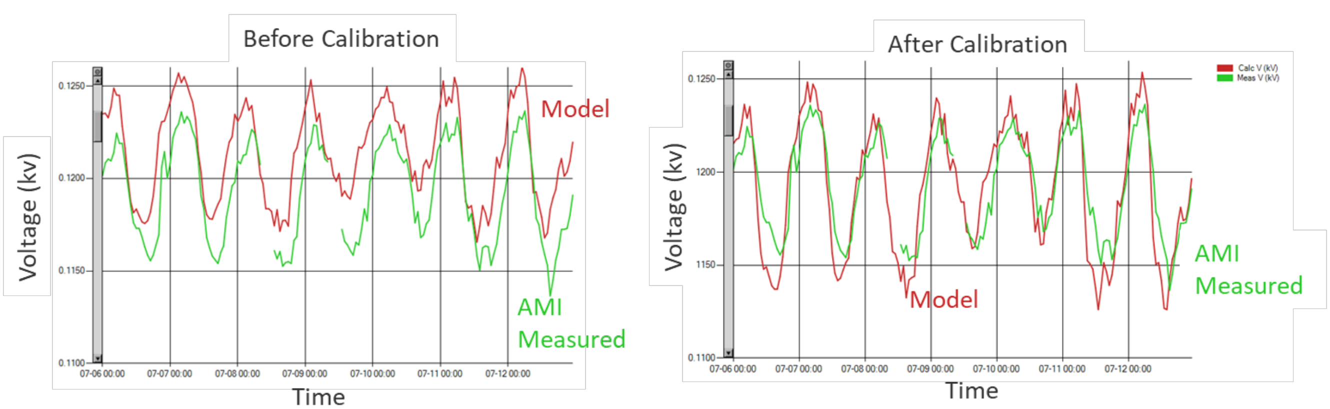

Figure 11. Customer service voltage error before and after calibration. Pepco data were used for calibration.

A simple example of the secondary calibration process is shown in Figure 12 for a secondary topology with direct connection between a service transformer and customer load. Mathematically, secondary calibration translates to finding the value of α that minimize errors between power-flow-calculated voltage Vm1’ and AMI voltage measurement, E_{m1}^{t}. First, the voltage drop between node n_o and n_1 was calculated using power-flow. Then, the tuning parameter α was identified to minimize the voltage error at n_1. In the example in Figure 12, with a direct customer connection, identifying alpha was trivial.

Several factors complicate α identification on secondaries with more nodes. First, the tuning parameter, α, can be interpreted as an adjustment to line length and constraints are placed on the maximum and minimum values of α that correspond to acceptable changes in secondary topology. Modeling errors upstream of n_o, for example, in the secondary transformer impedance, could result in unrealistic corrections to the secondary lines and should be avoided.

Second, realistic secondaries often have branches and nodes without AMI measurements. In Figure 13, “type I tuning” was used to minimize voltage errors at n_2 and n_4. A value of α was identified to reduce error on both nodes. The calibration procedure was as follows. The nodes were tuned in tiers, describing the number of branches between the node and service transformer. The closest tiers were turned first. First, power-flow was run. Then, α on Tier 1 (branch n_o to n_1) was tuned, followed by Tier 2 (branch n_o to n_1 and n_o to n_2). After all tiers were tuned, power-flow was run again. If nodal voltage errors exceeded defined limits, the tiers were retuned.

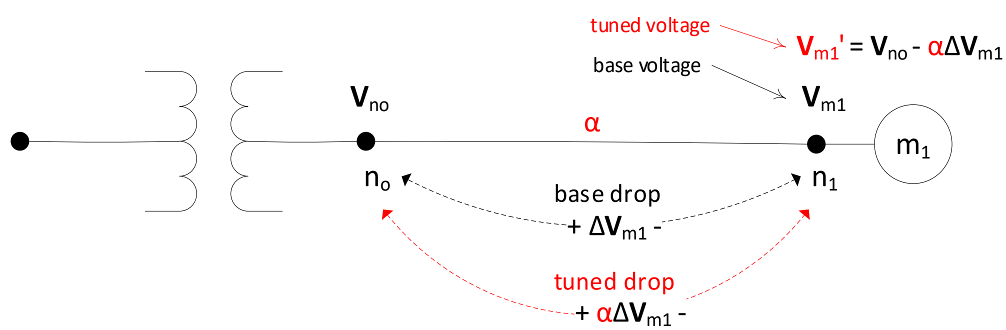

Figure 12. AMI secondary calibration with a simple direct connection secondary topology, where V_{no}: Base power flow voltage at parent node n_0. ΔV_{m1}: Base voltage drop between parent node n_0 and measurement location m_1

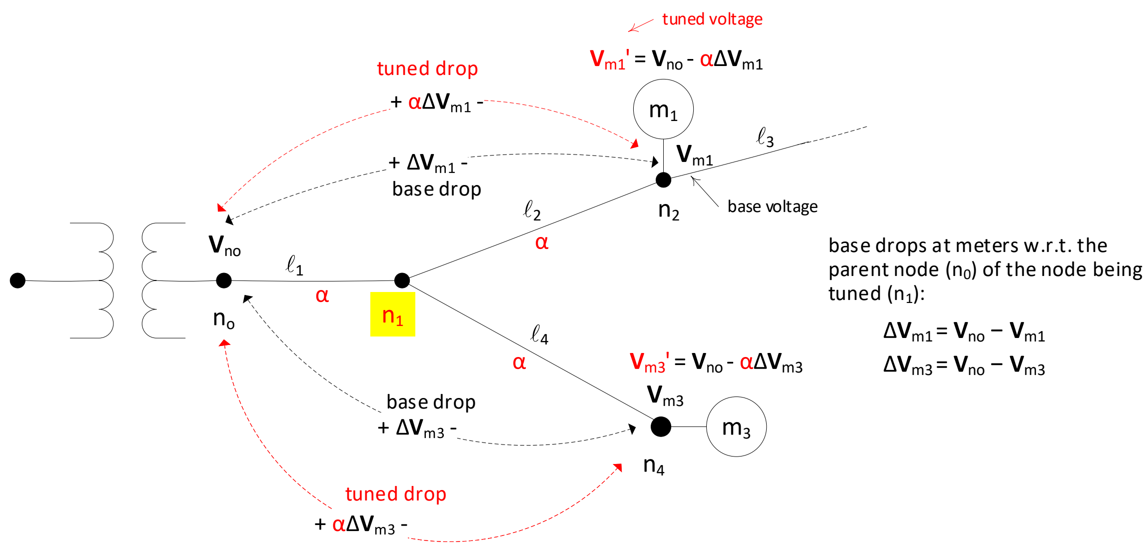

Figure 13. AMI secondary calibration with a realistic multibranch secondary topology

Task 3: Value of Improved Secondary Modeling

Task 3 covered Milestones 1.3.1, 1.3.2, and 2.3.3. The project demonstrated the value of improved secondary models relative to common utility heuristics. The objective was to show that improved secondary models can lead to greater confidence in the customer application process. Thus, utilities should be able to accept more applications without increased risk of distribution network voltage violations. Hosting capacity analyses were used on feeders from Pepco and Xcel to estimate the efficacy of the synthetic secondary models. Note, this approach differs from that of the original project scope, where the value of synthetic secondaries was to be determined by the ability to evaluate the improvements of at least 0.01 volts per unit in customer voltage and 20% in thermal loading accuracy. These earlier metrics were not used because they could not be used to evaluate the impact on synthetic secondary models on the customer application process.

Hosting capacity analyses with synthetic secondaries are shown for Pepco in Figure 14 and Xcel for Figure 15. In Figure 14, the number of voltage violations is shown relative to newly added solar PV. The project's synthetic secondary is shown in pink. Results using GIS models are shown in brown (S3b) and purple (S3). Scenario S3b is more conservative than S3 because it assumes primary voltages of 1.0375. Scenarios S1 (blue), S1b (orange), S2 (green) and S2b (red) are the most conservative. These scenarios assume direct connections to the primary and secondary transformer with assumed voltage drops. According to hosting capacity analysis on these eight feeders, the synthetic secondary scenario, KB (pink), increases the number of customer applications that will be accepted, relative to utility heuristics. In the most conservative scenario (orange), customers are directly attached to the service transformer with an assumed primary voltage of 1.0375. In this scenario, no customer applications would be accepted.

Similar results were found for the second utility. In Figure 15, the hosting capacity was tailored to mimic the customer application process. Unlike hosting capacity analyses, if a customer failed an application, the customer was not included in later power-flow simulations. In scenarios with direct customer connections to the primary network (orange) and service transformer (red) with assumed voltage drops, more customers were rejected. The number of passing applications for the synthetic applications (brown) causes slightly more failed applications than the GIS model (pink). Secondary networks with assumed star topologies (i.e., all customers are directly attached to the service transformer) accepts more customers than the GIS network and thus, would likely results in secondary network voltage violations. Similar trends were observed for nine feeders.

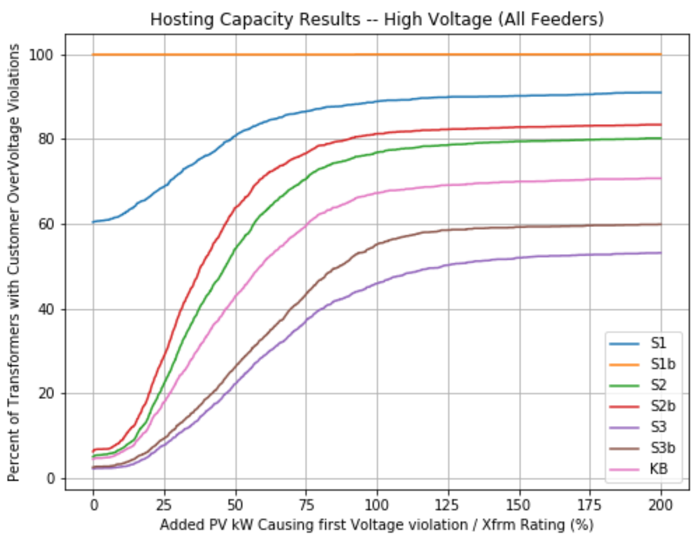

Figure 14. Hosting capacity results for Pepco with synthetic secondary networks

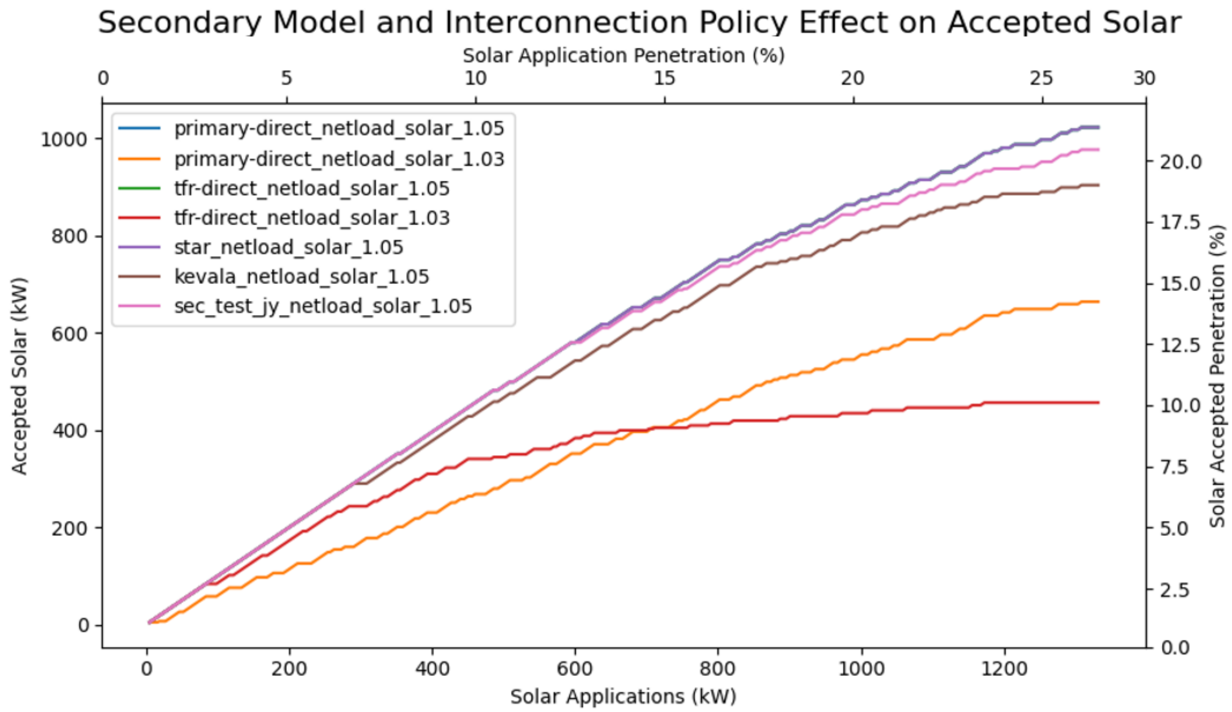

Figure 15. Hosting capacity results for Xcel with synthetic secondary networks

Task 4: Secondary Scenario Modeling and Refined Technical Screens, *and*

Task 5: Screening Guidelines with innovative mitigation strategies

Tasks 4 and 5 cover Milestones 2.5.1, and 2.5.4 respectively.

Data-Driven Screening

In Tasks 4 and 5, a data-driven method was developed using Pepco data to predict customer interconnection success. The objective of this task was to provide utilities with a transparent data-driven method for predicting customer interconnection success. A random forest (RF) model was developed to screen customer applications based on commonly available PV application information and network data as inputs, such as application size and solar penetration to provide screening decisions.

Figure 16 shows features that are used in the screening tool. Some of the features (e.g., number of regulators and capacitors) are grouped because they were highly correlated in the available data. Utilities may choose one feature per group based on data availability. Mitigation strategies can also be used as features in the tool. Figure 17 shows the impact of several mitigation strategies on customer voltage and serves as intuition for expected results from the random forest model. All training and testing data were collected hosting capacity analysis on 34 Pepco feeders.

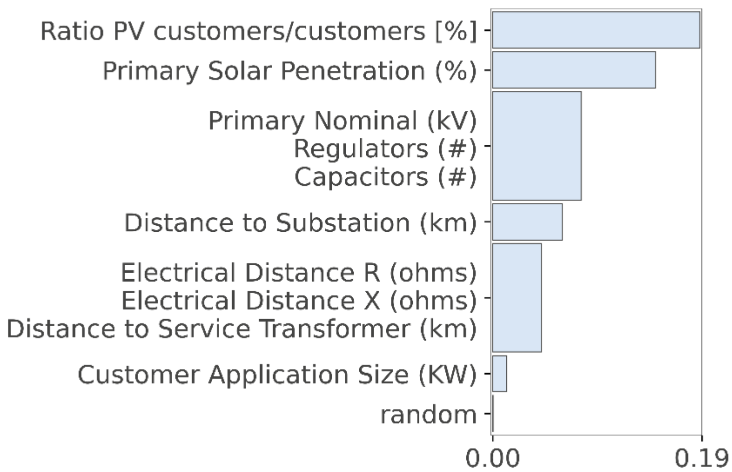

Figure 16. Feature importance and correlation using Pepco feeder models

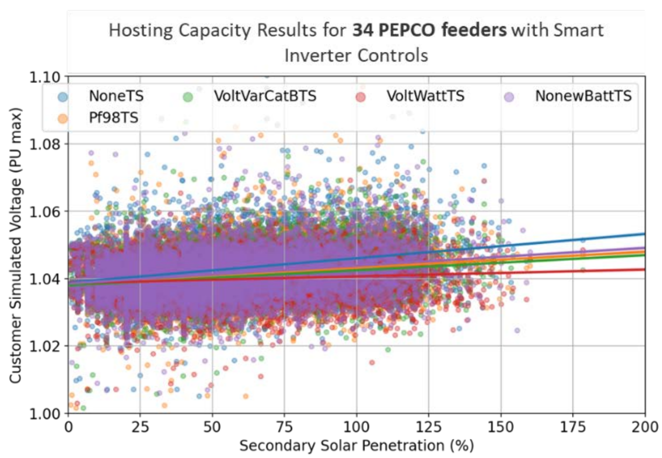

Figure 17. Impact of mitigation strategies on customer voltage using Pepco feeder models

The data-driven screening tool utility implementation is used with tabular inputs (Figure 17). Utilities enter information about the primary and secondary network and the type of mitigation strategies being used on customers. The screening tool outputs the probability that the customer will pass.[4] Representative results are shown with two customers. Though mitigation strategies were able to consistently improve the likelihood of passing, mitigation strategies did not typically increase the passing likelihood. The tool may need to be trained on more distribution circuits and higher-quality models to be able to capture the smaller voltage improvements from the mitigation strategies (see Figure 17) translate to improvements in customer pass rates.

The data-driven screening tool was more successful than traditional screening methods. Figure 18 shows pass rates for customers using a full power-flow model, the screening tool, and traditional heuristic screening methods. The traditional methods are very conservative and result in many customers failing the screening. In contrast, the random forest screening tool pass rate is comparable to the full power-flow model. For

[4] For details on how this probably is calculated, see Olson and Wyner (2018)

customers where primary and secondary models do not exist, the random forest screening tool offers a low-cost option for screening customers.

Table 6. Utility Data-Driven Screening Tool, based on Pepco data

Customer ID	Cust 1	Cust 2
Application Size (kw)	7	5
Secondary Solar Penetration (%)	102	57
Primary Solar Penetration (%)	113	70
Primary Nominal (kV)	13.2	13.2
Capacitors (#)	0	0
Regulator (#)	1	1
Distance to Substation (km)	5.04	5.35
Distance to Service Transformer (km)	0.04	0.05
Probability of Passing without Control	0.11	0.69
Probability of Passing with Battery	0.16	0.72
Probability of Passing with Volt-Var(CatB)	0.66	0.79

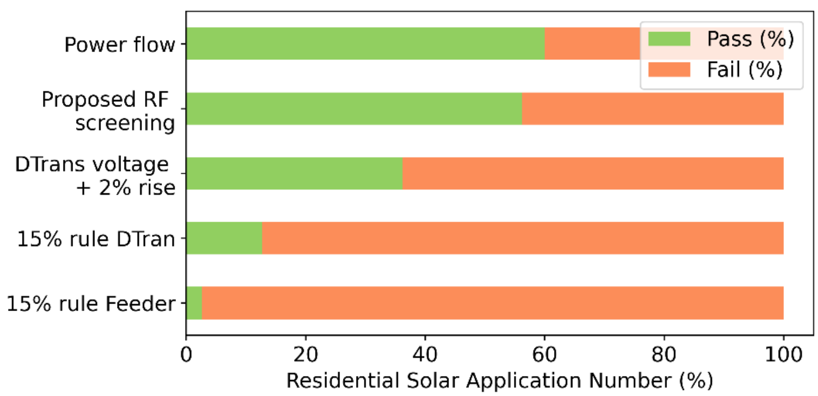

Figure 18. Customer pass rate with power-flow model, proposed screening tool, and traditional screens without secondary representation

The data-driven screening tool has several advantages over conventional fast-track screens. It was found to be more accurate than the conventional fast-track screens. It was also found to be faster than detailed power flow studies and nearly as accurate. Today, many utilities lack detailed power flow models that include secondaries, so the option to use the proposed method could bring high accuracy results at low cost.

The data-driven screening tool can also complement the conventional fast-tract screens and detailed power flow studies. Utilities can triangulate results among the methods. They can also use the data-driven screening tool to do preliminary or what-if analysis (e.g., changing the voltage class of a feeder) that can inform detailed power flow studies.

Implementation of a data-driven screening tool requires comprehensive and systematic curation of training data. Ideal training data would come from utility feeder models that include secondary networks and are validated with AMI. Additionally, the feeder models would have a broader range of key features, including but not limited to the features used in this paper.

Here, we found that some design parameters (e.g., the number of regulators) did not improve the interconnection pass rates as expected, and we theorize that this is due to confounding variables. For example, the presence of capacitors and regulators on a feeder may significantly help the interconnection pass rate, but if they are found on older feeders with other design constraints, data-driven techniques may inadvertently associate them with poor interconnection pass rates.

Solar and Battery Forecasting

As part of Milestone 2.5.1, the project team developed a high-spatial-resolution load and DER forecast. The purpose of this milestone was to inform the cost-benefit analysis of the data-driven secondary and screening tools. However, in September 2021, the project team modified the technical working plan, shifting focus from the cost-benefit analysis to Milestones 2.5.2, 2.5.3, and 2.5.4. Nonetheless, results for the high-spatial-resolution load forecast are shown.

Kevala data scientists developed a demand and PV generation simulation methodology to create a time series of electrical loading for each service point associated with the Xcel feeders provided by the project team. The core of the PV and storage adoption calculation is a stochastic simulation, which assumes parcels will adopt behind-the-meter resources with a likelihood that is inversely proportional to the time that it takes to pay back the cost of the system (payback period). The first step is to estimate a reasonable PV system capacity and battery size for all parcels where PV does not already exist. The methodology estimates that each adopted PV system will be sized to generate approximately the same amount of energy as is consumed by the parcel over the course of a year.

Different approaches are applied for estimating storage system sizes for commercial and residential parcels. For commercial systems, the storage system is sized to manage peak demand. The model selects the smallest storage system that would provide enough power to shave the monthly peak demand, with the greatest difference between the first and second greatest demand peaks. For residential systems, batteries are sized for optimal resiliency (e.g., during outages). Selected storage system sizes cover 50% of the energy usage on the largest energy usage day at the residence.

Once the time series of PV and storage system generation/load for each parcel is estimated, the hypothetical bill savings that could be achieved using PV and storage can be calculated. The methodology uses the rate schedule associated with the individual customer and the customer's predicted demand curve to calculate the annual bill. With annual bill savings estimated for all parcels, adoption can be simulated. The simulation proceeds by randomly selecting which parcels will adopt each year. The likelihood of adoption for each parcel is determined by the adoption probability; parcels with a higher adoption probability are more likely to install behind-the-meter resources. Figure 19 shows the aggregate load time series for the Pepco service territory. Storage does not have a large impact on the net load because storage growth assumptions are modest in the service territory.

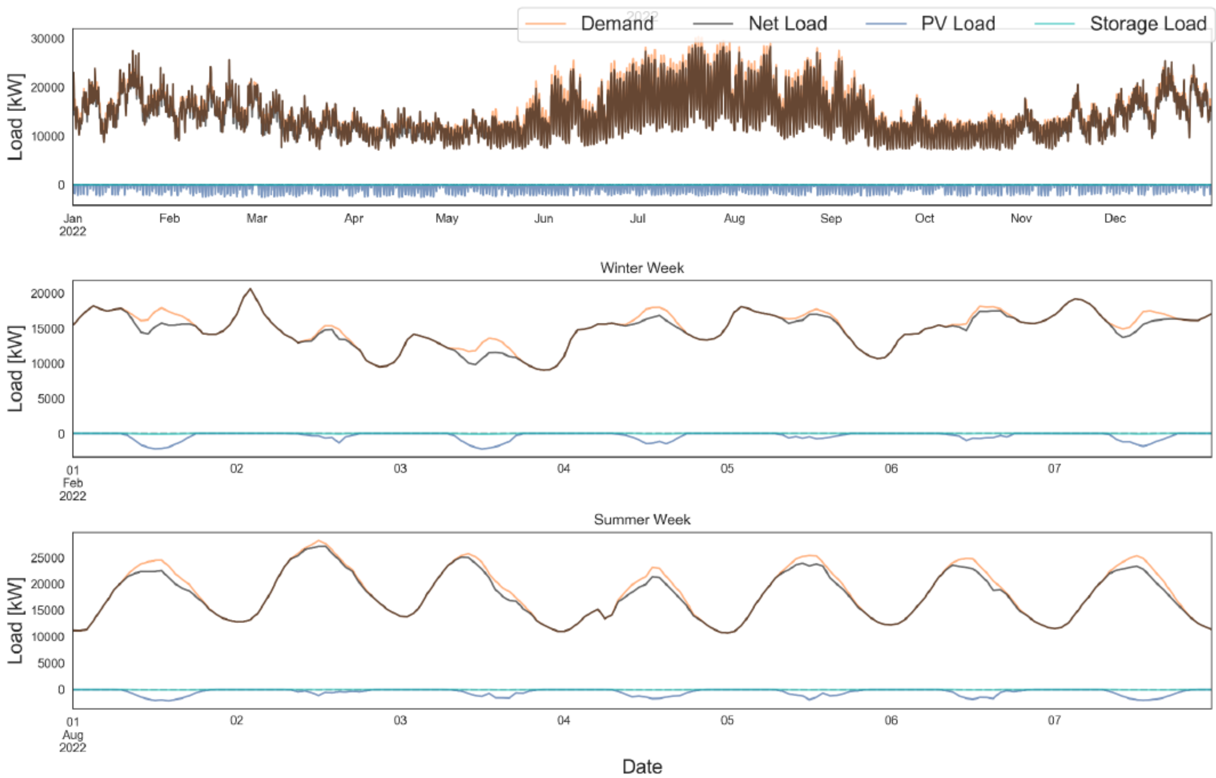

Figure 19. Aggregate load time series for Xcel customers. Note: Demand data are estimated and not from AMI.

9. Significant Accomplishments and Conclusions

The project produced significant accomplishments and lessons for both the data-driven secondary networks and the data-driven screening method. Of the two methods, the data-driven secondary network method has a higher technology readiness level, and the project team is discussing a pilot projects proposal with Pepco. Part of the proposal is intended to showcase advantages and account for disadvantages of the data-driven secondary methods.

One advantage of the proposed decision tree method is its robustness with bad data. Engineering judgement is used to develop decision tree rules for the most common secondary topologies, thus excluding topologies that are caused by bad GIS data. This rule-based approach helps make predictions intuitive, which is a weakness of some machine learning methods. In addition, some GIS data can be verified by tracing overhead lines on satellite images.

A disadvantage is that the proposed method does depend on some accurate inputs. Secondary transformer location, customer location, and customer-to-transformer data are critical. During data cleaning, AMI can be used to improve these customer-to-transformer mappings. The proposed method can be efficiently scaled to predict a large number of feeder secondaries, but it is less extensible to many types of feeder secondary topologies. Secondary networks today are based on decades of heterogeneous design practices. In the authors' experience and the experience of the utility technical advisory committee, design manuals are not available for older secondary networks.

A utility considering the proposed method should explore two options for adding new secondary types to the decision tree. First, if the topologies are generally representative of their territory, the decision tree would only need to be reparametrized. Second, new topologies can be added to the decision tree. The second option is more time-consuming. An analysis should weigh the accuracy of the improved interconnection study and the network prevalence. Doing so can help utilities identify a "right-sized" approach that helps them manage the risks of increased deployment of electric vehicles and DERs

Increasing the data-driven screening tool's technology readiness level would require more-comprehensive and systematically curated training data than used in this project. Ideal training data would come from feeder models that include secondary networks and are validated with AMI. Additionally, the feeder models would have a broader range of key features including voltage level and regulator settings. To avoid confounding variables, comprehensive variations on design principals on each feeder model (i.e., getting training data from the same feeder but varying voltage level, capacitor number, other parameters) should also be included. In this research, we found some design parameters (e.g., the number of regulators) did not improve interconnection pass rates as expected, and we theorize this is due to confounding variables. For example, the presence of capacitors and regulators on a feeder may significantly help interconnection pass rate, but if they are found on older feeders with other design constraints, data-driven techniques may inadvertently associate them with poor interconnection pass rates.

Traditional scenario-based hosting capacity analysis also has limitations. A customer interconnection will depend on not only on the inverter settings of that individual customer but also on the inverter settings of customers sharing the feeder. In this report, we assumed all customers shared common inverter settings. In reality, customers may have a combination of legacy inverters without any controls and smart inverter controls with various setpoints. It is not feasible for hosting capacity analyses to test all permutations of inverter settings, so new hosting capacity training methods may be needed

10. References

Billinton, Roy, and Satish Jonnavithula. 1996. "A Test System for Teaching Overall Power System Reliability Assessment." *IEEE Transactions on Power Systems* 1670-1676. doi:10.1109/59.544626.

Blakely, Logan, and Matthew J. Reno. 2020. "Identifying Errors in Service Transformer Connections." *IEEE Power & Energy Society General Meeting (PESGM).* doi:10.1109/PESGM41954.2020.9282035.

Bolognani, Saverio, Nicoletta Bog, Davide Michelotti, Riccardo Muraro, and Luca Schenato. 2013. "Identification of Power Distribution Network Topology via Voltage Correlation Analysis." 52nd IEEE Conference on Decision and Control. doi:10.1109/CDC.2013.6760120.

Brockway, Anna, Jennifer Conde, and Duncan Callaway. 2021. "Inequitable Access to Distributed Energy Resources Due to Grid Infrastructure Limits in California." *Nature Energy* 892-903. doi:10.1038/s41560-021-00887-6.

EERE. 2011. *The SunShot Initiative.* https://www.energy.gov/eere/solar/sunshot-initiative.

Fekete, Emily S., Jesse R. Cruce, Shiyuan Dong, Eric O'Shaughnessy, and Jeffrey J. Cook. 2022. *A Retrospective Analysis of Distributed Solar Interconnection Timelines and Related State Mandates.* Golden: National Renewable Energy Laboratory. doi:10.2172/1841350.

Gotseff, Peter, Nick Wunder, Andy Hoke, Earle Ifuku, and Reid Ueda. 2018. "Residential Advanced Photovoltaic Inverter Pilot Study Results for Select Distribution Secondaries in Hawai'i." *IEEE 7th World Conference on Photovoltaic Energy Conversion.* Waikoloa. doi:10.1109/PVSC.2018.8548182.

IEEE. 2022. *IEEE PES Test Feeder.* https://cmte.ieee.org/pes-testfeeders/resources/.

Lave, Matthew, Matthew J. Reno, and Jouni Peppanen. 2019. "Distribution System Parameter and Topology Estimation Applied to Resolve Low-Voltage Circuits on Three Real Distribution Feeders." *IEEE Transactions on Sustainable Energy.* doi:10.1109/TSTE.2019.2917679.

Muratori, Matteo, Colin Sheppard, Tim Lipman, Bryan Palmintier, N. Panossian, P. Jadun, H. Laarabi, R. Waraich, R. Desai, and A. Von Meier. 2020. "Grid-Enhanced, Mobility-Integrated Network Infrastructures for Extreme Fast Charging (GEMINI-XFC)." June. https://www.nrel.gov/docs/fy20osti/76718.pdf.

NREL. 2022. *SMART-DS: Synthetic Models for Advanced, Realistic Testing: Distribution Systems and Scenarios.* https://www.nrel.gov/grid/smart-ds.html.

Olson, Matthew, and Abraham Wyner. 2018. "Making Sense of Random Forest Probabilities: a Kernel Perspective." https://arxiv.org/pdf/1812.05792.pdf.

Padullaparti, Harsha, Santosh Veda, Jiyu Wang, Martha Symko-Davies, and Tom Bialek. 2022. *Phase Identification in Real Distribution Networks with High PV Penetration Using Advanced Metering Infrastructure Data.* Golden: National Renewable Energy Laboratory. https://www.nrel.gov/docs/fy22osti/81336.pdf.

Reno, Matthew, and Robert J. Broderick. 2015. "Technical Evaluation of the 15% of Peak Load PV Interconnection Screen." *IEEE 42nd Photovoltaic Specialist Conference (PVSC).* New Orleans. doi:10.1109/PVSC.2015.7356168.

Schneider, K P, Y Chen, D P Chassin, R G Pratt, and D W Engel. 2008. *Modern Grid Initiative Distribution Taxonomy Final Report.* PNNL. doi:10.2172/1040684.
Short, Tom A. . 2013. "Advanced Metering for Phase Identification, Transformer Identification, and Secondary Modeling." *IEEE Transacations on Smart Grid.* doi:10.1109/TSG.2012.2219081.
Taheri, Sina, Mana Jalali, Vassilis Kekatos, and Lang Tong. 2021. "Fast Probabilistic Hosting Capacity Analysis for Active Distribution Systems." *IEEE Transactions on Smart Grid* 2000-201. doi:10.1109/TSG.2020.3038651.
Wang, Wenbo, Jeremy Keen, Julieta Giraldez, Kyle Baranko, Brent Lunghino, Dan Morris, Frances Bell, et al. 2022. "Supervised Learning for Distribution Secondary Systems Modeling: Improving Solar Interconnection Processes." *IEEE Transactions on Sustainable Energy* 13: 948-956. doi:10.1109/tste.2022.3140650.
Zhang, Xiaochen. 2017. "Machine Learning and Big Data Analytics for the Smart Grid." Georgia Institute of Technology. https://smartech.gatech.edu/bitstream/handle/1853/58729/ZHANG-DISSERTATION-2017.pdf.