SAND2020-10631PE

# Data-Driven Calibration of Electric Power Distribution System Models

Matthew J. Reno, and Logan Blakely

IEEE PES Subcommittee on Big Data & Analytics for Power Systems
IEEE Big Data & Analytics Tutorial Series
September 30, 2020

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

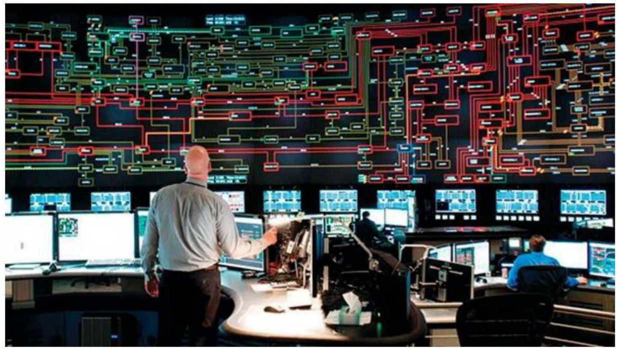
# Outline

- Introduction
- Overview of Data-Driven Calibration of Distribution System Models
- Use Cases
  - Parameter and Topology Estimation of Low-Voltage Secondary Systems
  - Phase Identification
  - Identification of Service Transformer Connections
- Future Directions of Data-Driven Modeling
- Conclusions

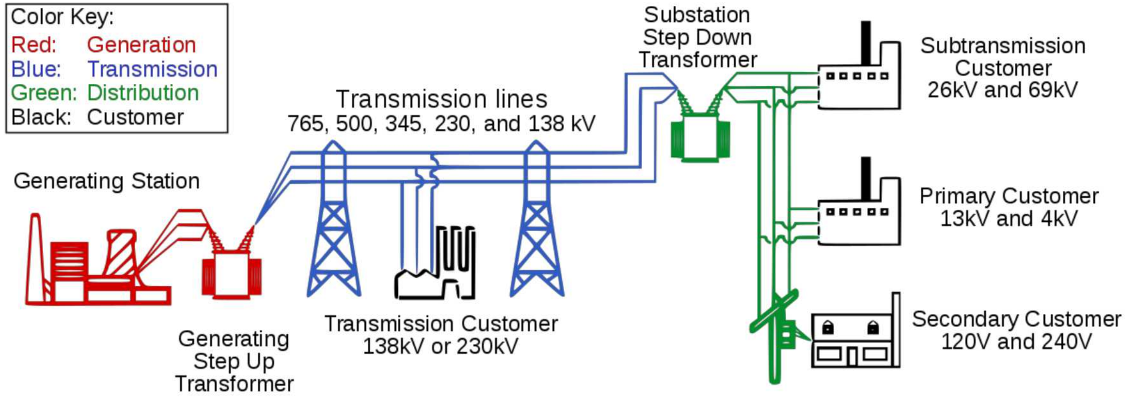
# Introduction

The electric power grid provides roughly 4000 TWh of power per year, delivering power to critical aspects of America's economy, transportation, water, emergency services, telecommunications, manufacturing, defense facilities, and residences.

Transmission System (up to 500kV) transfers power throughout the U.S.
- Real-time measurements (1-3 second SCADA, PMU, etc.)
- State Estimation, Optimization, and Control



Distribution System (4kV – 35kV) connects to the customers
- Much less monitoring or control due to the size
  - ~300,000 miles of transmission lines vs. ~6,000,000 miles of distribution lines
  - ~20,000 substation transformers vs. ~200,000,000 service transformers
- Visibility into distribution system operations is limited, and models are prone to errors




http://www.ferc.gov/industries/electric/indus-act/reliability/blackout/ch1-3.pdf

# Electric Power Distribution Systems

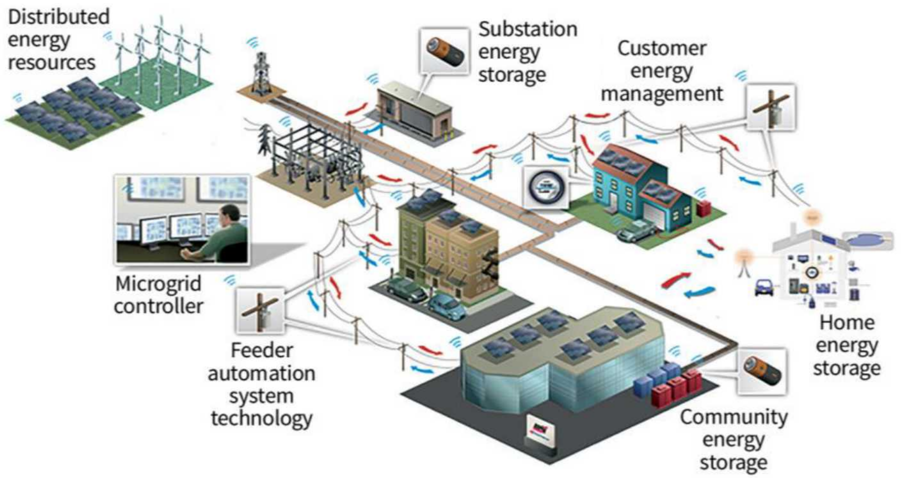
Historically, distribution system model accuracy was of little concern, rarely validated, and had limited measurements.

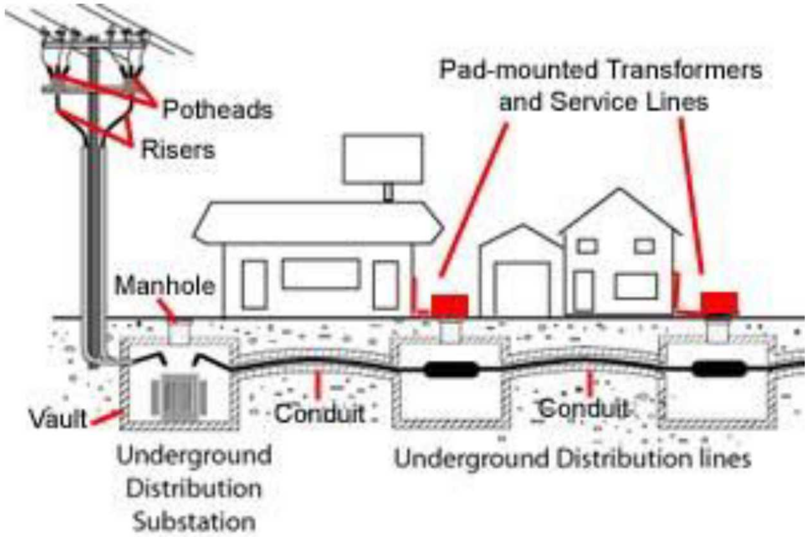
Many of the recent advances in smart grid technologies, proliferation of distributed energy resources (DER), and new control strategies are on the distribution system – electric vehicles, rooftop PV, energy storage, microgrids, etc.

With new smart grid technologies, accurate models are critical

- Accurate PV interconnection analysis and screening
- Optimal operations and control
- Investment planning and decisions
- Improved reliability and resilience (fault location, isolation, and service restoration)

**Modern distribution system algorithms and tools are continually improving, but their functionality is only as good as the utility's model of their grid**.



# Distribution System Models

The distribution system has been built over many decades, historically recorded with paper schematics for installations, upgrades, and maintenance.

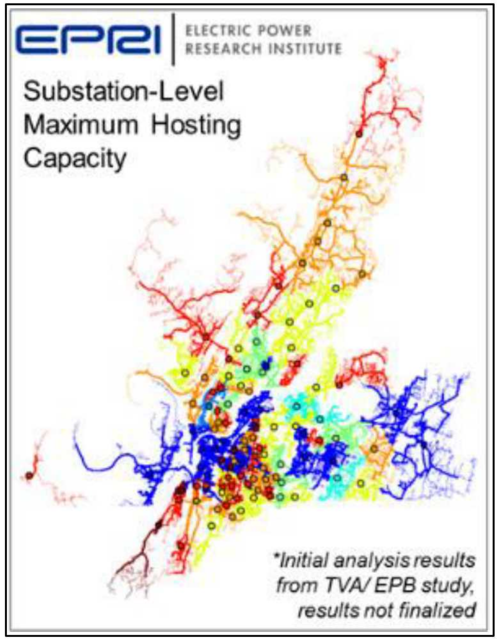
## Distribution System Models

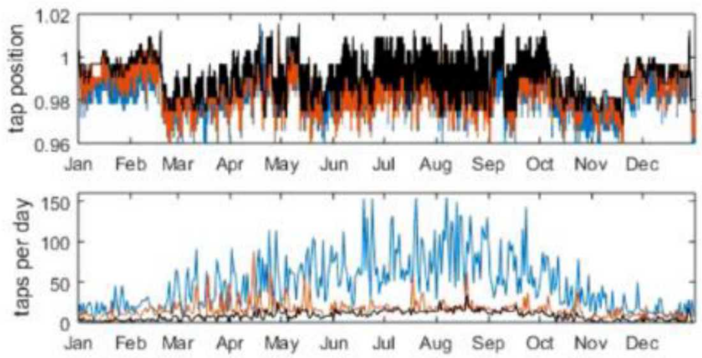
- Are based on manual data entry that is prone to error and often out of date
- Contain additional complexity because they are multi-phase unbalanced with single-phase customers. Cannot use symmetrical component single-line models from transmission system modeling

## Sources of Error

- Unlogged or erroneous maintenance reports
- Information not initially recorded in the model

Recent additions of Advanced Metering Infrastructure (AMI), or smart meters, provide measurements of each customer's power consumption, and possibly other quantities, such as voltage, that provide new insights and levels of accuracy in distribution system modeling



# Common Distribution System Model Errors

- **System State and Setting Errors**
  - **State of switches**
  - Capacitor States
  - Voltage Regulator Settings
  - Switching Capacitor Settings
- **Phase Label Errors**
  - **Individual transformer phase label error**
  - **Lateral Phase Label Error**
  - **Three-phase Customer Labeled as Single-Phase**
  - **Single-phase Customer Labeled as Three-phase**
- **Data Missing From the Model**
  - **Missing or Incorrect GIS coordinates**
  - **Unmarked Transformers**
  - **Unmarked PV Installations**
  - Unmetered Load
- **PV Installations**
  - **PV kW Rating**
  - **Tilt**
  - **Azimuth**
  - **Volt/VAr Settings**
  - Connection (LN or LL)
  - Connect or Disconnect Dates
- **Meter Configuration**
  - PT or CT Ratios
  - Units (kW vs MW)
  - Time Zone
  - Measurement Location
  - Collection Type (time-avg vs instantaneous)
- **Model Parameters**
  - **Wire types and Lengths**
  - **Overhead Line Configuration**
  - Underground Cable Insulation
  - Transformer Rating, connection, or Turns Ratio
  - Substation Short Circuit Impedance
  - **Meter to Transformer Connection Errors**
- **Equipment Addition or Replacement**
  - New Home Construction
  - Reconductoring of Lines
  - New Voltage Regulation Equipment
  - Service Transformer Replacement

L. Blakely, M. J. Reno, and J. Peppanen, "Identifying Common Errors in Distribution System Models," IEEE Photovoltaic Specialists Conference (PVSC). 2019.

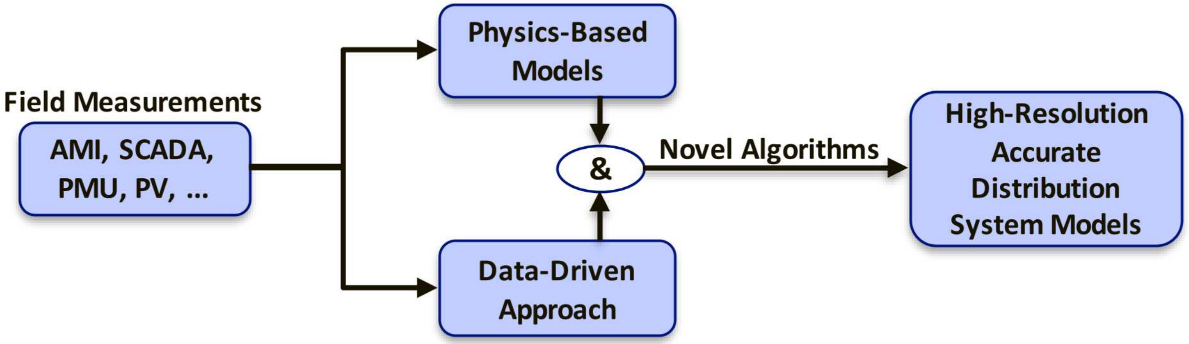
# Project Overview

DOE EERE Solar Energy Technologies Office (SETO) funded project "Physics-Based Data-Driven Grid Modelling to Accelerate Accurate PV Integration"

- Physics-based – using known electrical equations and models that work with today's power systems simulation software (not black box)

FY19 - FY21 project to efficiently process grid measurements and Big Data to provide a more granular understanding of the distribution system and to substantially increase the precision and accuracy of distribution system models – creating a fundamental change from models based on manual entry to data-driven modeling.

Project led by Sandia National Labs, partnering with Lawrence Livermore National Laboratory, Electric Power Research Institute, Georgia Tech, and CYME/Eaton




U.S. DEPARTMENT OF ENERGY | Energy Efficiency & Renewable Energy

SOLAR ENERGY TECHNOLOGIES OFFICE U.S. Department Of Energy

Sandia National Laboratories | Lawrence Livermore National Laboratory | EPRI ELECTRIC POWER RESEARCH INSTITUTE | Georgia Tech | CYME INTERNATIONAL T&D Eaton's CYME Power Engineering Software

# Comparison of Distribution Modeling Methods

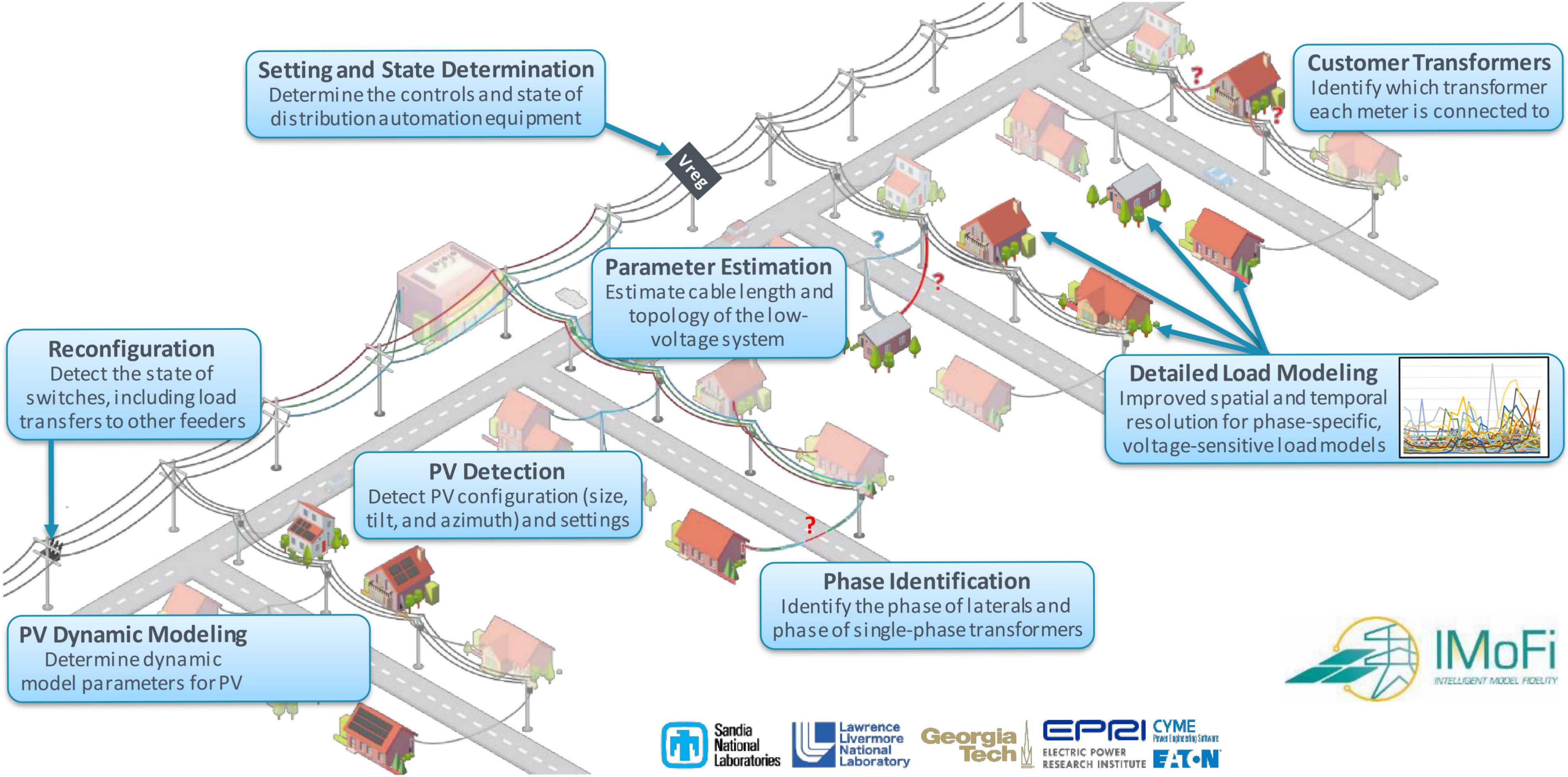
## Conventional Methods

- Manual data entry – compiling records of installations, upgrades, and maintenance over decades
- Prone to errors – unlogged or erroneous maintenance reports or entry into the model
- Little validation with measurements
- Often out of date with a list of changes to add to the model

## Physics-Based Data-Driven Modeling

- Leveraging AMI data and other grid edge sensing to derive and validate system models
- High accuracy and fidelity – a reproduction faithful to the original
- Granular and high resolution, multi-phase model down to the low-voltage system
- Model dynamically adapts and automatically updates based on system conditions

# Data-Driven Distribution Model Calibration



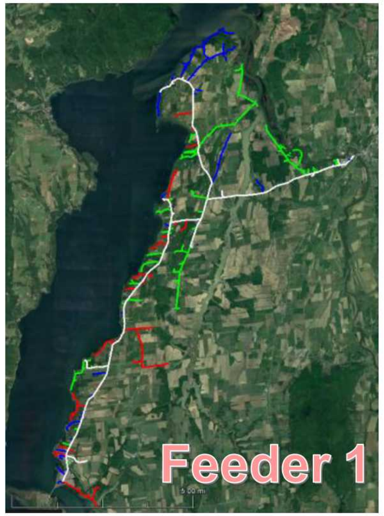

IMoFi
INTELLIGENT MODEL FIDELITY

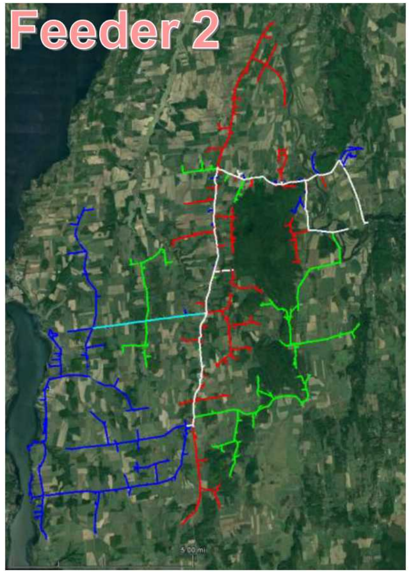
Sandia National Laboratories | Lawrence Livermore National Laboratory | Georgia Tech | EPRI Electric Power Research Institute | CYME Power Engineering Software | EATON

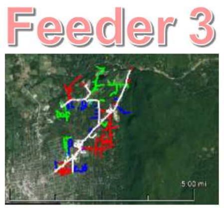
# Advantages of Different Data Sources

## Demonstration with Utility Partners

- AMI (Real Power, Reactive Power, and Voltage) measurements from all customers (~3500) on feeders
- 15-minute measurement interval for a year or more
- Physical validation of the model is difficult and expensive. Results can be confirmed using Google StreetView, satellite images, and network topology.

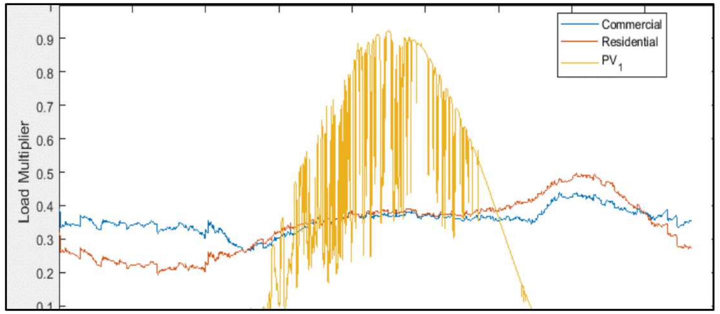


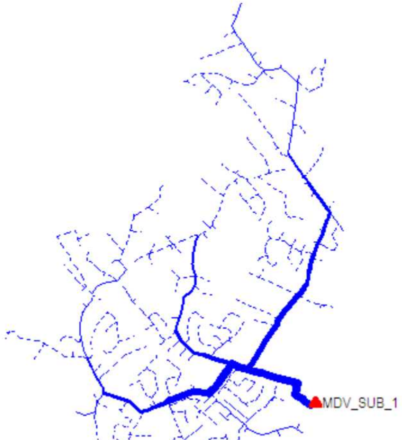




## Testing with Synthetic Data

- Actual measurements from customers of real and reactive power are inserted into a distribution system model to simulate the voltages.
- Measurement noise is added to the simulation results of power and voltage
- Known errors are injected into the distribution system model to measure the accuracy of the algorithms to detect the model errors.



# Practical Considerations for Real Data

Real data has issues such as bad data or events that should be filtered

- Data cleaning, filtering, and denoising due to failing meters or anomalous readings around outages
- Changepoint detection to filter certain events such as load transfers from other feeders, CVR events, etc..
- Handling missing data – imputation

It is important to replicate these issues in synthetic data for algorithm development

- Analysis of utility data provides general range of the amount of frequency of missing AMI data
- Experimental testing of meters in Sandia's AMI Lab provides expected accuracy and measurement noise

Data/metering considerations that are evaluated in synthetic data:

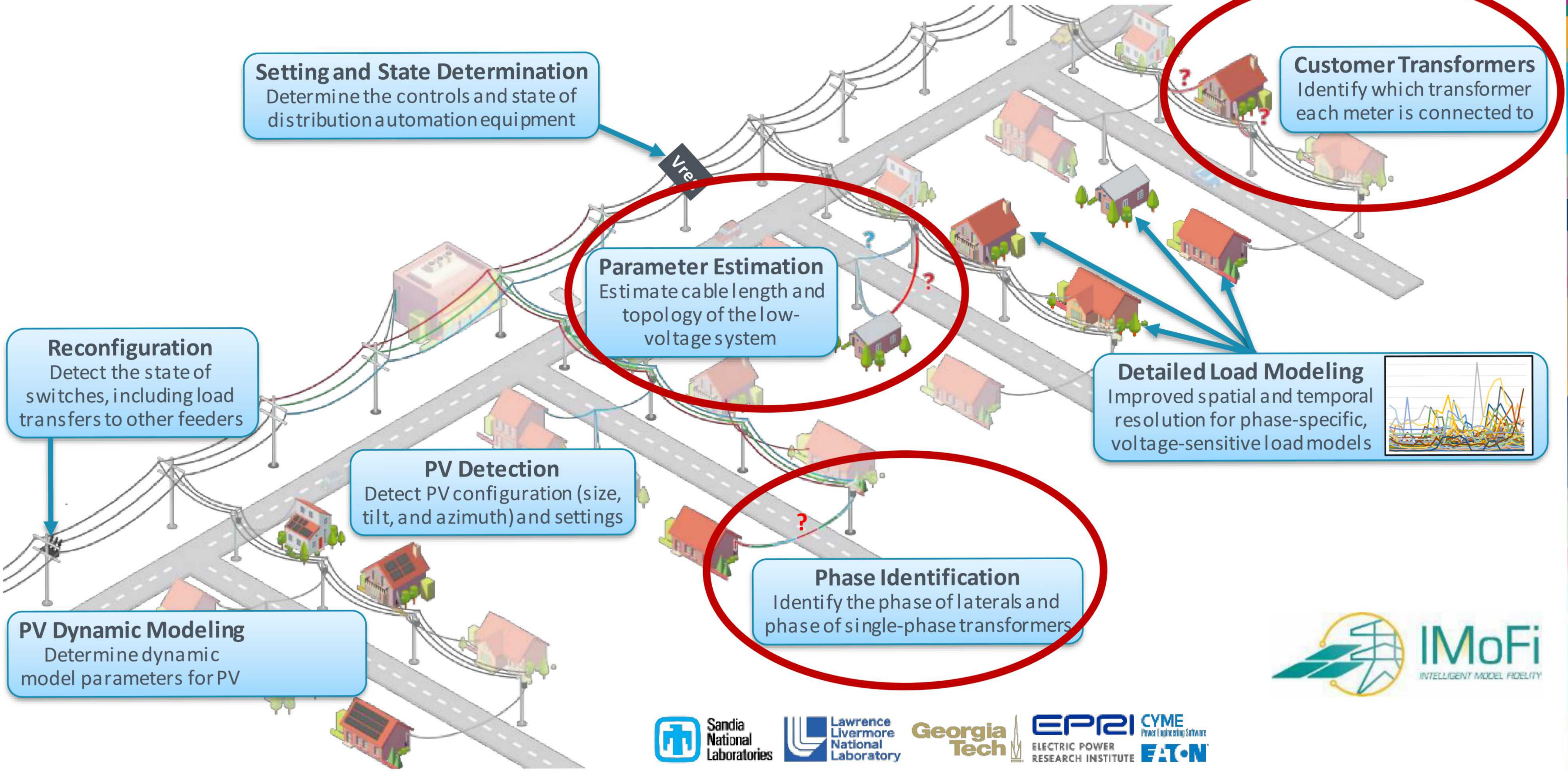
- Measurement Interval
- Data Resolution
- Meter Precision
- Meter Bias
- Time Synchronization
- Missing data
- Data Availability

**AMI Data Quality and Collection Recommendations.**
**Based on analysis of synthetic data with varying amount of error**

| | AMI Requirement |
|---|---|
| Measurement Interval | 15 - 30 minute intervals |
| Data Resolution | At least 1 decimal on voltage and power measurements (0.1V, 0.1kW) |
| Meter Precision | < 0.25% maximum noise |
| Data Availability | > 4 months of AMI data |

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

# Data-Driven Distribution Model Calibration



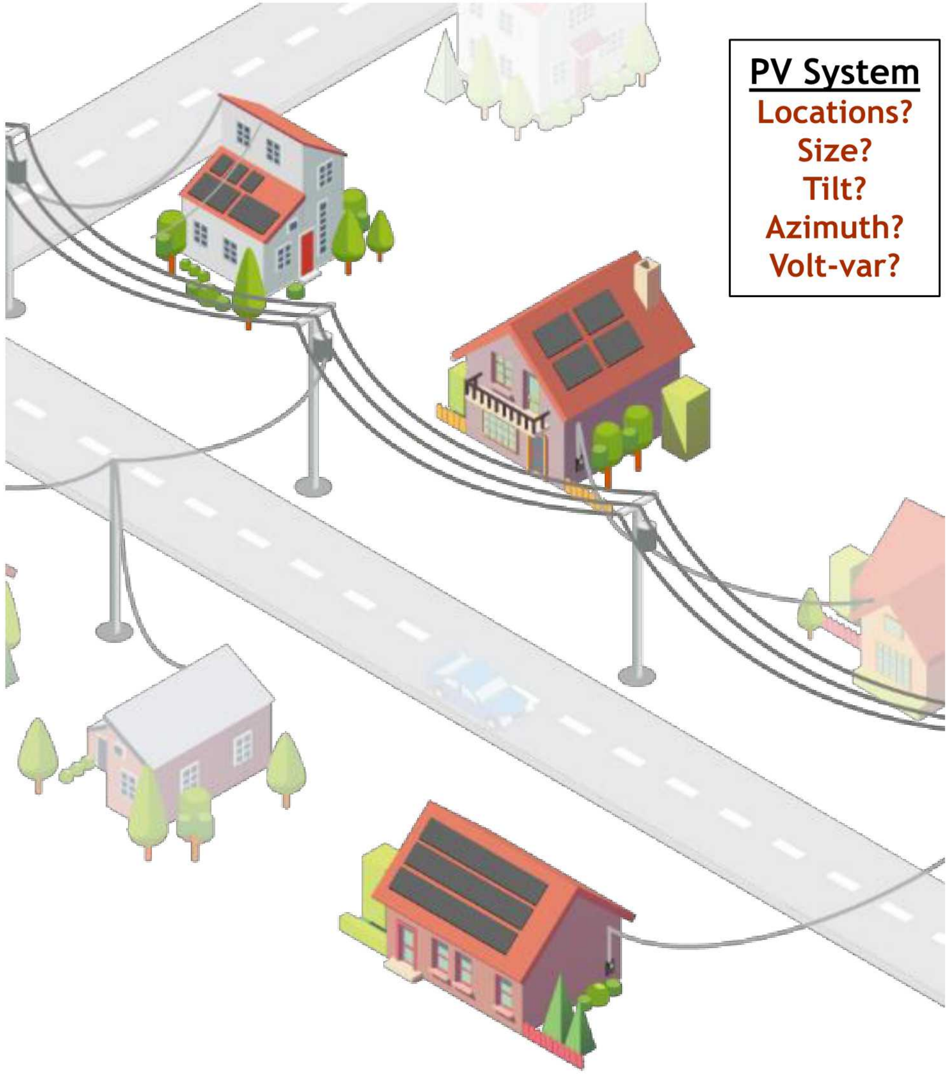
# PV System Identification

**Background**: PV systems may vary from the interconnection plan - not interconnected, project delayed, changed size, shading issues, gradual soiling, or module/string failures

**Problem**: Keeping PV interconnection databases updated is a major challenge. Many utilities do not record parameters for distributed PV such as their DC power rating, tilt, or azimuth. Residential solar PV systems are generally behind-the-meter (BTM), lacking direct measurements or observability.

For BTM PV, solar disaggregation methods can separate the PV from the load measurements. Deep Neural Network used to learn the signature of BTM PV to detect if there is PV, along with its size, tilt, and azimuth.

Parameter estimation of behind-the-meter parameters that do not have direct measurement. Partial observability requires leveraging multiple data streams – nearby voltage measurements, weather data, irradiance measurements, etc.



# Medium-Voltage Model Calibration

Data-driven calibration of the medium-voltage primary system distribution system:

**Voltage Regulator**

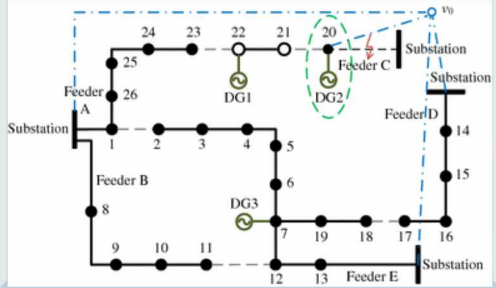
Estimate the regulator control settings and tap position

**Switching Capacitors**

Estimate the capacitor control settings and position

**Reconfiguration Detection**

Detect the state of switches and when they changed, including load transfers to other feeders




**Meter to Transformer Pairing**

Identify which transformer each meter is connected to

# Switching Capacitor Controls and State

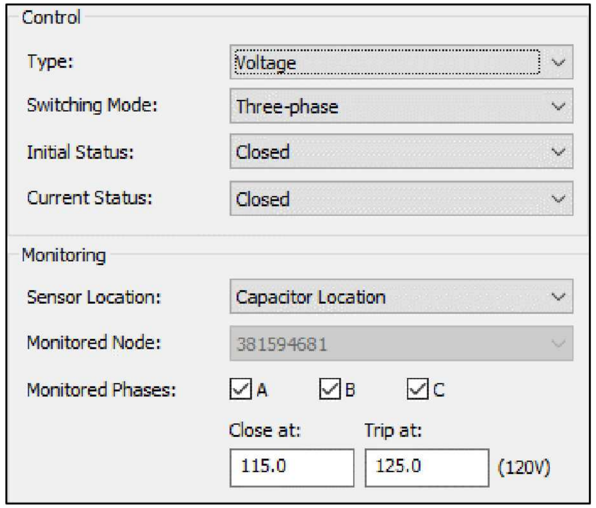
**Background**: On the distribution system, voltage regulators and switching capacitor banks control the voltage by switching taps or switching on/off depending on their controls

**Problem**:

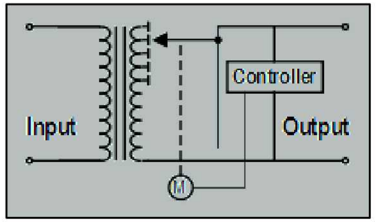
- Most distribution voltage regulation equipment (VRE) do not have remote login capabilities, so verifying their settings in planning models requires sending a crew to the device.
- For state estimation or power flow results, knowing the state of VRE is required, but this information is often not available in historical data

By combining distribution system state estimation (DSSE), machine learning, and Big Data from grid edge measurements, we can identify the 1) VRE settings, and 2) the state of the VRE at a given time

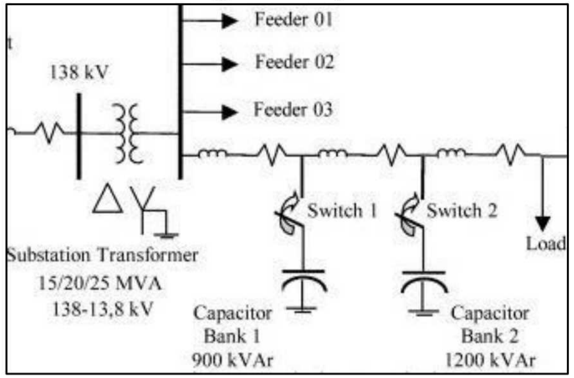
**Determining the Settings of the Switching Capacitor**




**Determining the Tap Position**




**Determining the Capacitor State**



# Improved Spatial and Temporal Load Modeling

**Background**: Distribution planning and operations uses a combination of load measurements and load allocation methods, which have significant influence on the system performance

**Problem**: 1) It is challenging to fully leverage the increased visibility to loads provided by AMI and other emerging data streams. 2) Even with "100% penetration" of AMI, there are many unmetered load in the system. 3) Reactive power measurements of loads are rare and typically very inaccurate.

**Objective**: Develop improved spatial and temporal load modeling methods leveraging AMI and other emerging data streams that are robust to incomplete data sets.

**Improved Spatial & Temporal Resolution for Phase-Specific Voltage-Sensitive Load Modeling**

# Outline

- Introduction
- Overview of Data-Driven Calibration of Distribution System Models
- Use Cases
  - Parameter and Topology Estimation of Low-Voltage Secondary Systems
  - Phase Identification
  - Identification of Service Transformer Connections
- Future Directions of Data-Driven Modeling
- Conclusions

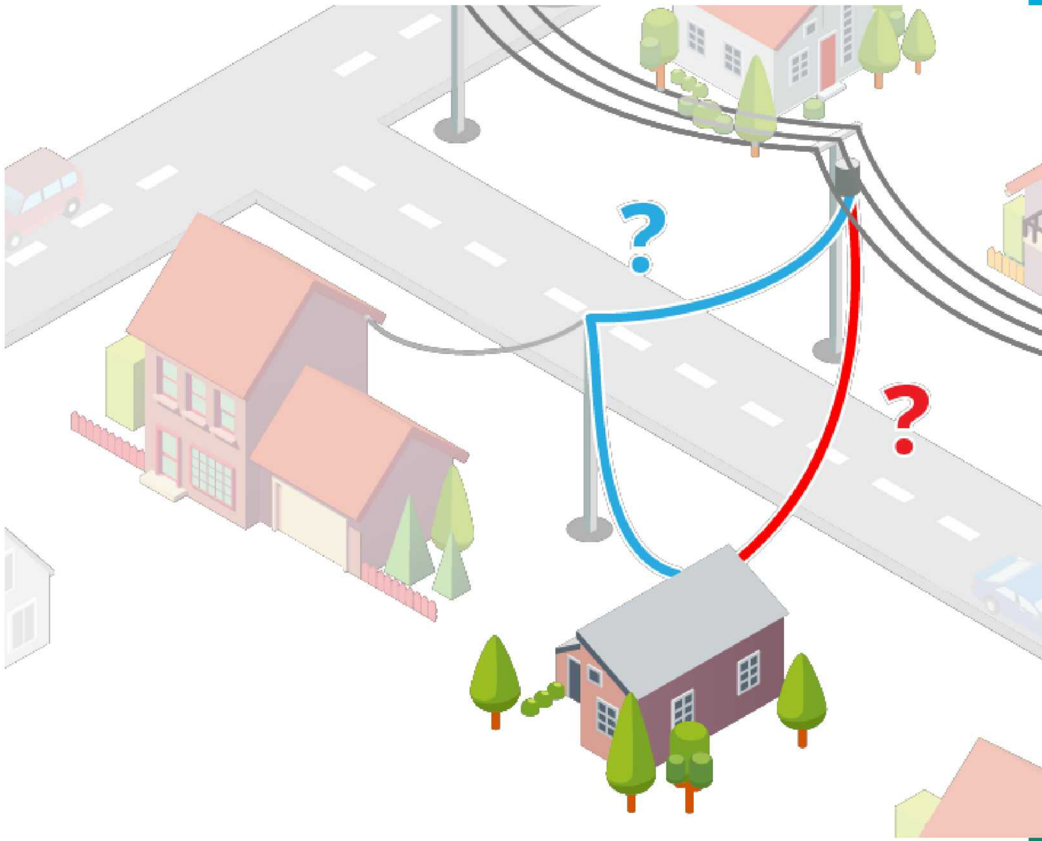
# Secondary System Parameter and Topology Estimation

**Background**: Multiple customers are generally connected to the distribution feeder through a service transformer and a low-voltage secondary system

**Motivation**: A large portion of the per-unit voltage drop/raise occurs over the secondaries. A large number of DERs and sensors are connected to the secondary circuits

**Problem**: Secondary circuits are typically not modeled or modeled with limited detail. Manual inspections require considerable man hours and extra resources => not cost effective, and may be hard to perform in urban areas with wiring underground and in buildings

**Objective**: Use customer meter (AMI) voltage and power measurement to resolve secondary system parameters and topology



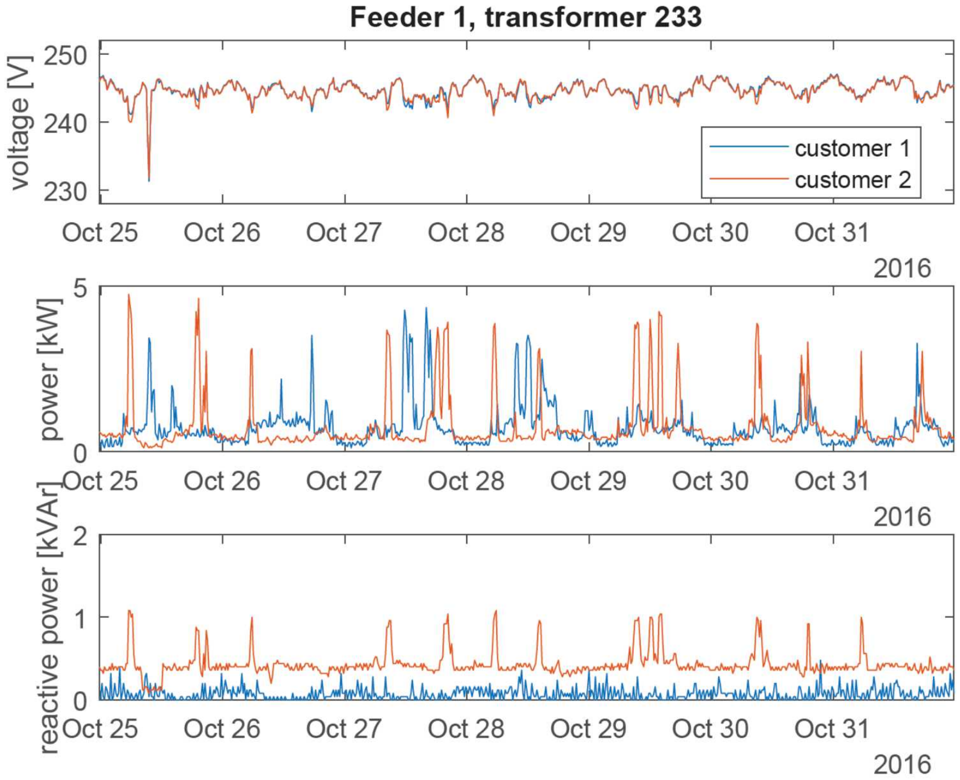
# Secondary System Parameter and Topology Estimation

**Inputs**: AMI data (voltage, real and reactive power) at 15-minute intervals for 6-months to 1-year, transformer connection

**<u>Procedure:</u>**

1. Resolve the parameters and topology for all transformers with 2+ customers.
2. Resolve the parameters for transformers with only a single customer by pairing them with other single-customer transformers.
3. Pair transformers resolved in step 1 with one another to resolve any additional parameters between the virtual nodes where the customers meet and the transformers.

**Output**: Models for each low-voltage secondary system. Compare to utility's unverified, manually-entered secondary model



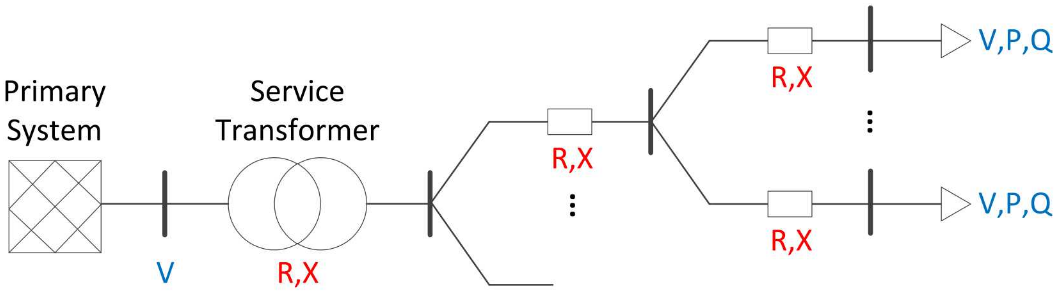
# Step 1

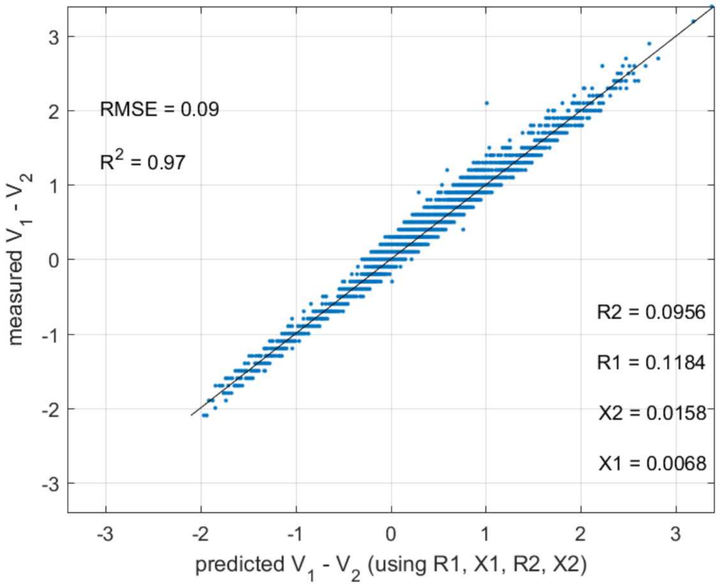
For all customers on a transformer, find R1, R2, X1, X2

$$\boldsymbol{V}_1 - \boldsymbol{V}_2 = \boldsymbol{I}_{R1}R_1 + \boldsymbol{I}_{X1}X_1 - \boldsymbol{I}_{R2}R_2 - \boldsymbol{I}_{X2}X_2 + \boldsymbol{\epsilon}$$

Known Unknown

- Basic concept
  - Fit R1, R2, X1, X2 values which best fit the V1-V2 fluctuations

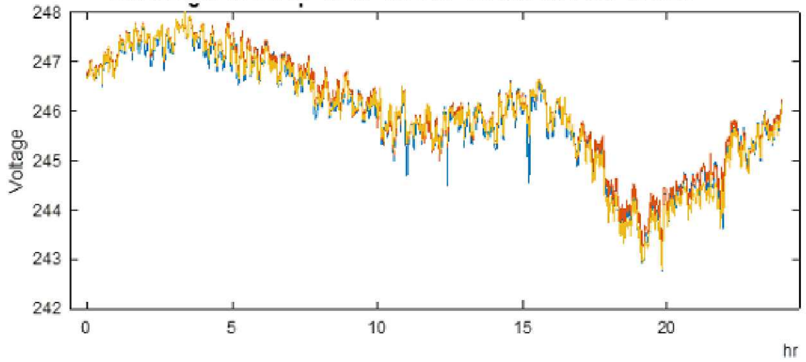


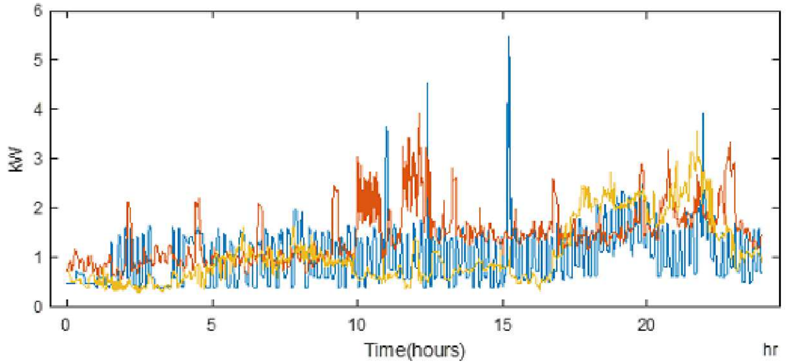



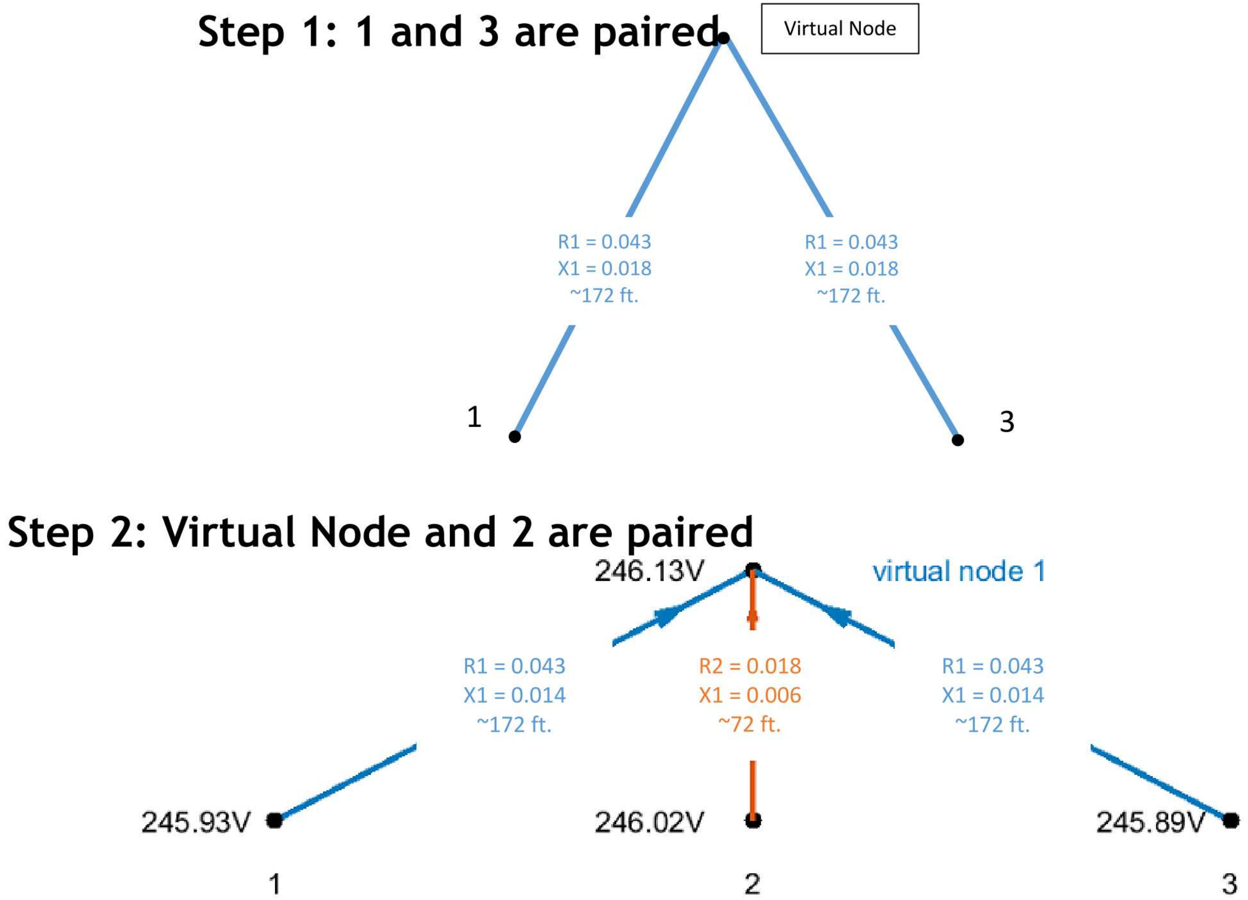

- For comparison to utility model
  - R values were used to compute a distance in feet of triplex cable for various types of cable (#2, 2/0, 4/0)

# Secondary System Example

**Example voltage and load for 24-hours for the 3 customers on the transformer**



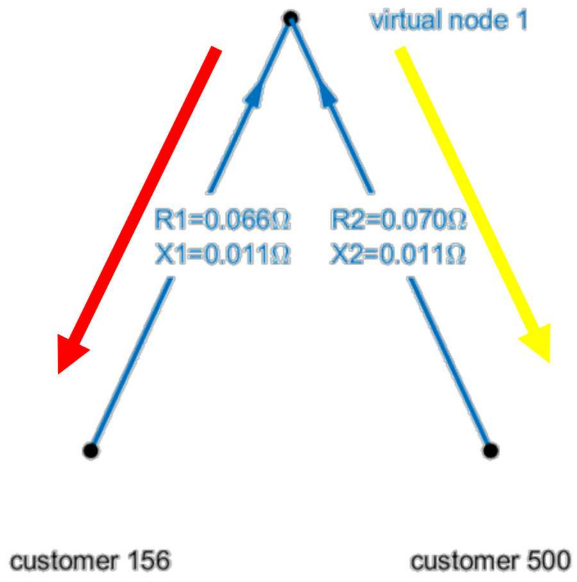




**Parameter Estimation Results**

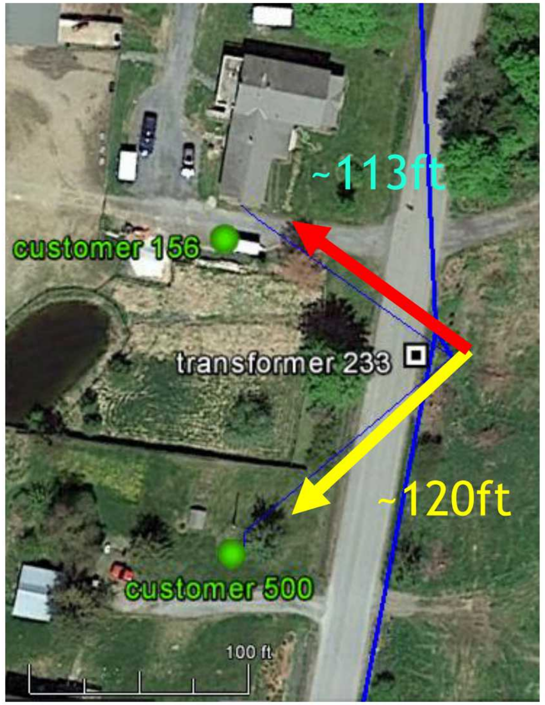
| Meter # | Estimated | | | Actual | | | % Error | |
|---|---|---|---|---|---|---|---|---|
| | R | X | Length (ft) | R | X | Length (ft) | R | Length |
| 1 | 0.043 | 0.014 | 172 | 0.0425 | 0.0129 | 170 | 1.18% | 1.18% |
| 2 | 0.018 | 0.006 | 72 | 0.0175 | 0.0053 | 70 | 2.7% | 2.7% |
| 3 | 0.043 | 0.014 | 172 | 0.0425 | 0.0129 | 170 | 1.18% | 1.18% |

# Transformer 233 on Feeder 1

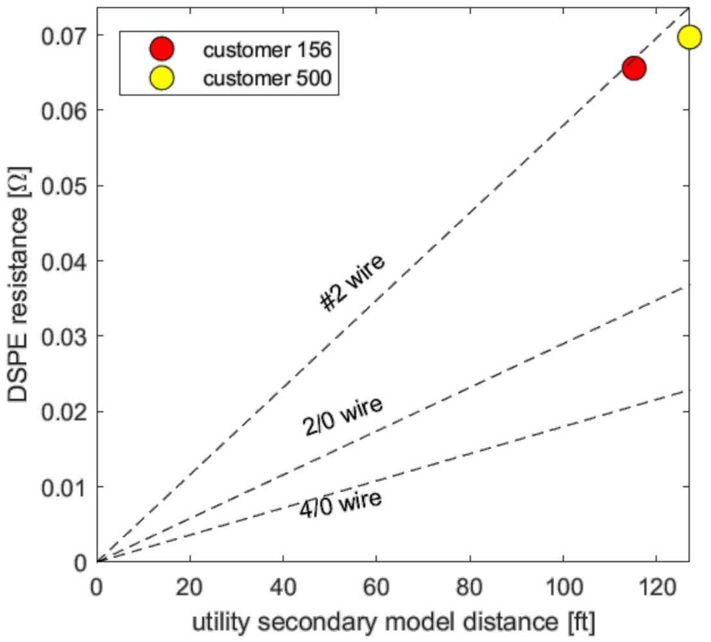
Distribution System Parameter Estimation




Imagery/Model



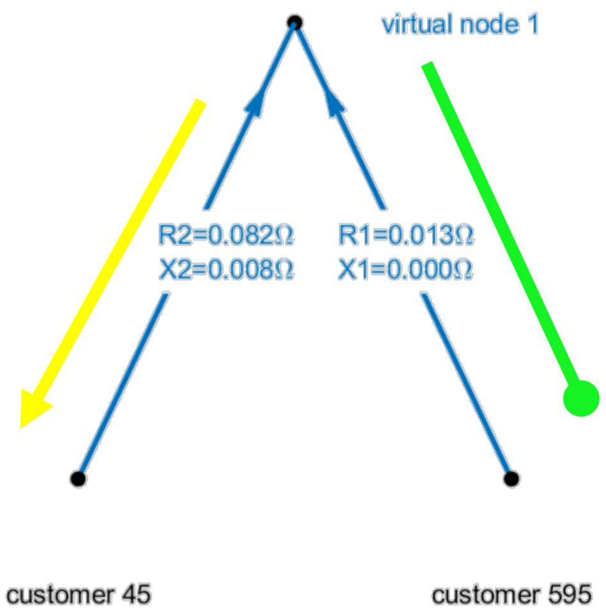

DSPE vs. Model



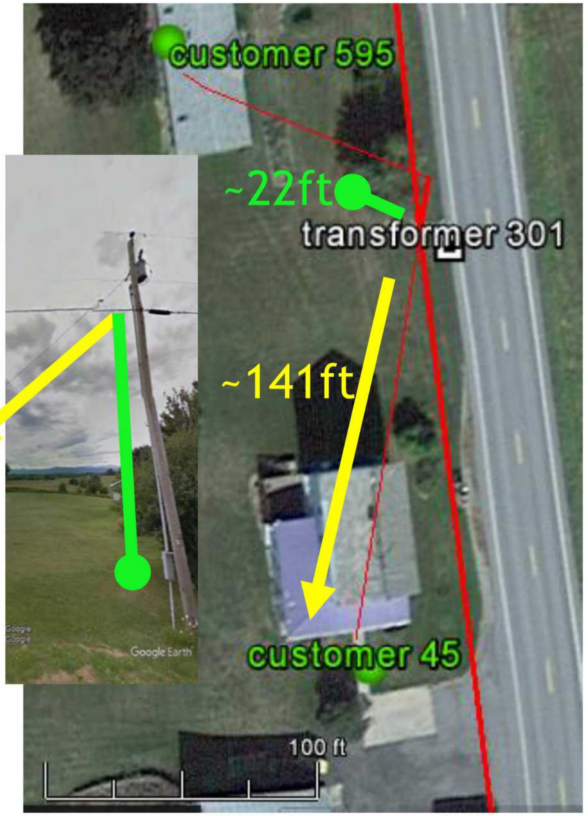

DSPE results match utility model well, consistent with #2 wire.

# Transformer 301 on Feeder 2

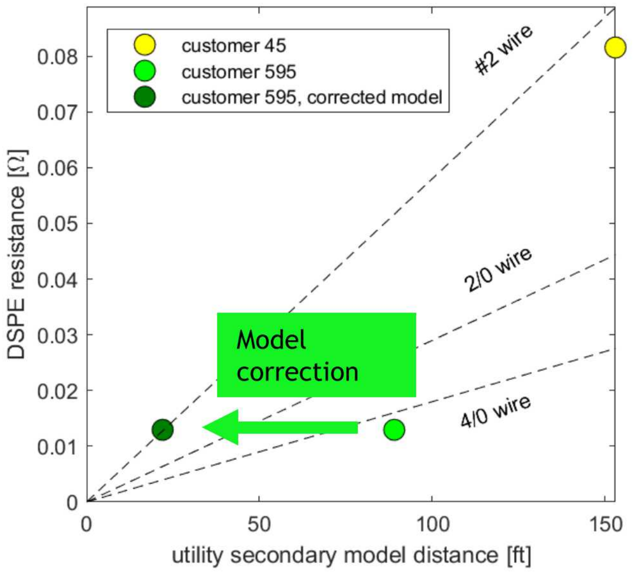
Distribution System Parameter Estimation




Imagery/Model

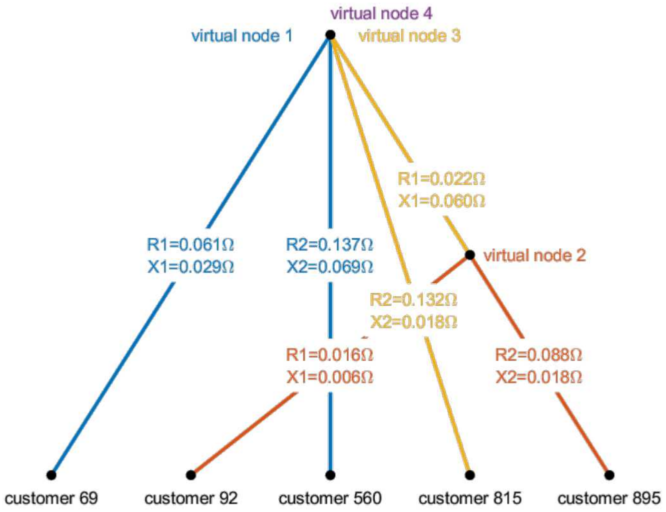


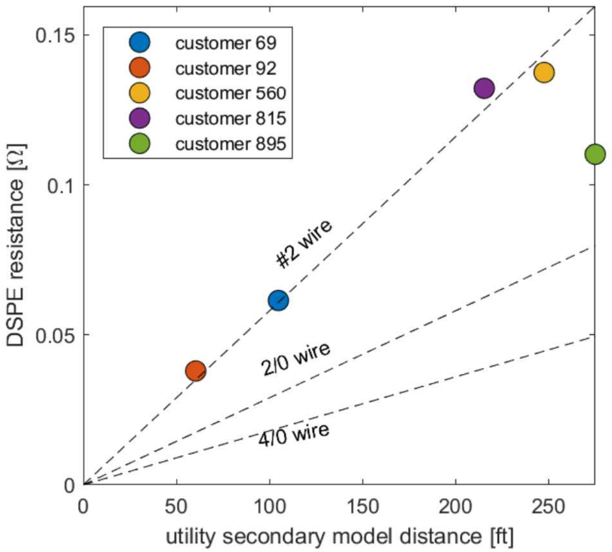

DSPE vs. Model

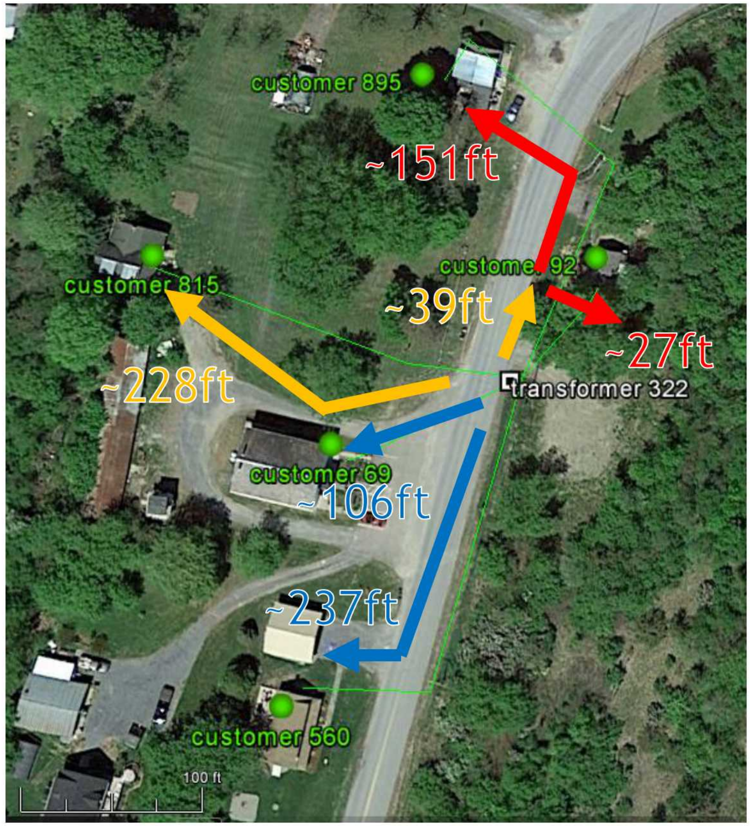



DSPE results indicate error in utility model: customer 595's meter is actually at the bottom of the utility pole, not at the house.

# Transformer 322 on Feeder 1



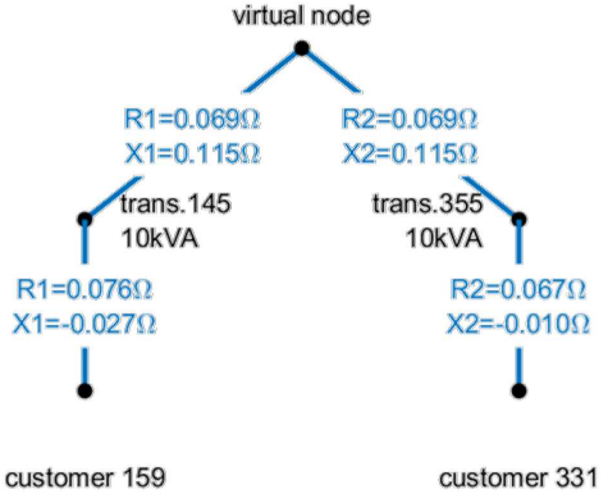




DSPE results consistent with utility model for several customers with complicated topology.

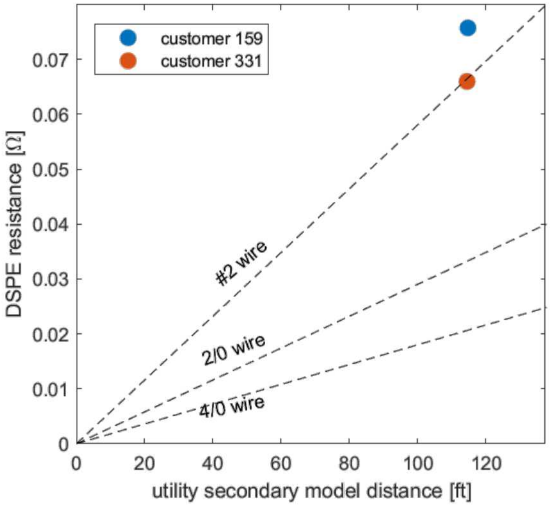
# Step 2

Pair customers on transformers with only one customer with other solo customers

- Topology is always parallel – step 3 virtual node is on primary
- Should always be additional resistance beyond the transformer due to the customer being located away from the transformer







# Step 3

Pair transformers with one another, run parameter estimation on virtual nodes created in step 1

- Topology is always parallel – step 2 virtual node is on primary
- Most likely scenario is that virtual node from step 1 is at transformer low side and any found impedance will be due to transformer impedance
- In some cases, step 1 virtual node will be away from transformer
  - Serial connection between customers
  - Parallel connection that meets before the transformer
- It is important to derive the additional impedance to fully resolve the secondary circuit

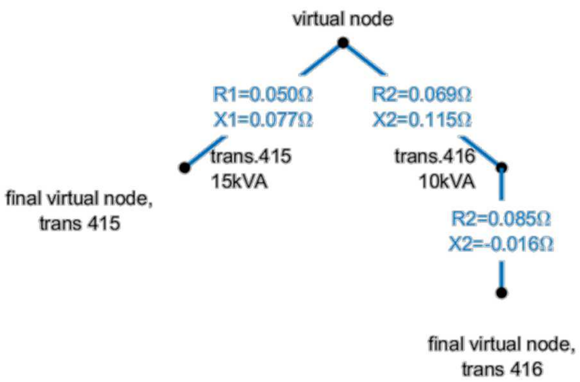
| Transformer size (kVA) | 3 | 5 | 10 | 15 | 25 | 37.5 | 50 | 75 |
|---|---|---|---|---|---|---|---|---|
| Assumed resistance | 1.5% | 1.5% | 1.2% | 1.3% | 1.16% | 0.96% | 1% | 0.87% |

# Transformers 415 and 416

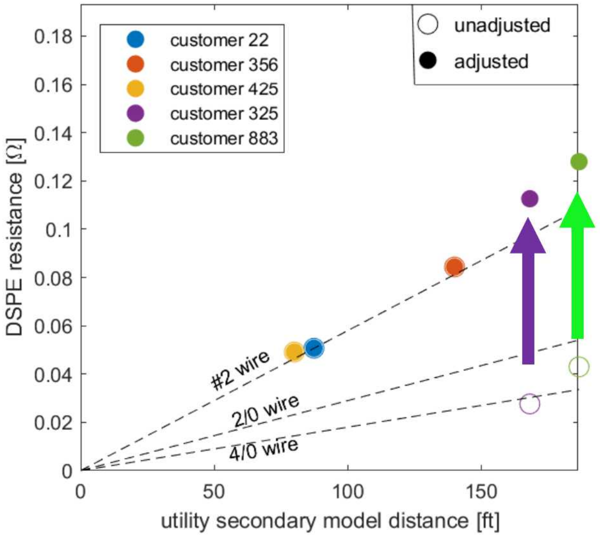
Distribution System Parameter Estimation

Imagery/Model

DSPE vs. Model

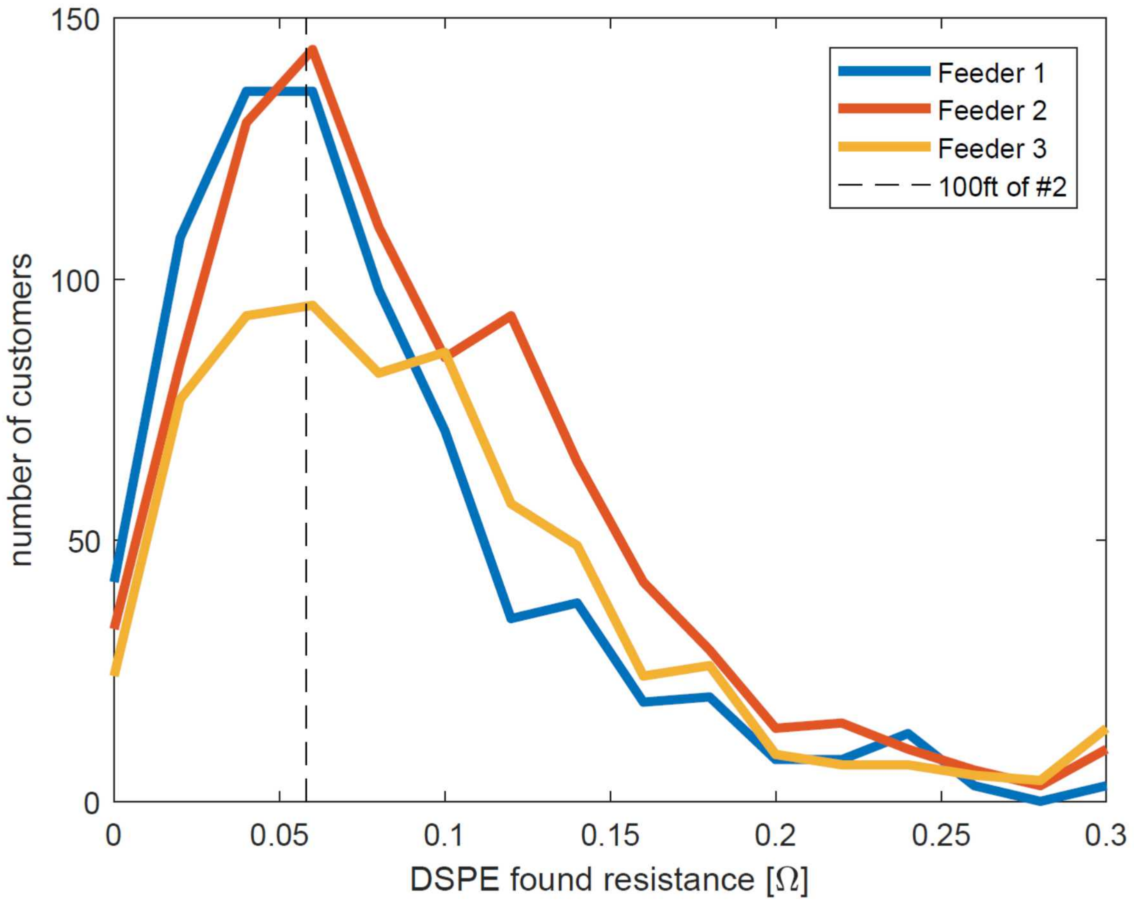







Customers 325 and 883 (on transformer 416) had a virtual node away from the transformer, which is accounted for by pairing transformer 415 with 416.

# Results for Entire Feeders

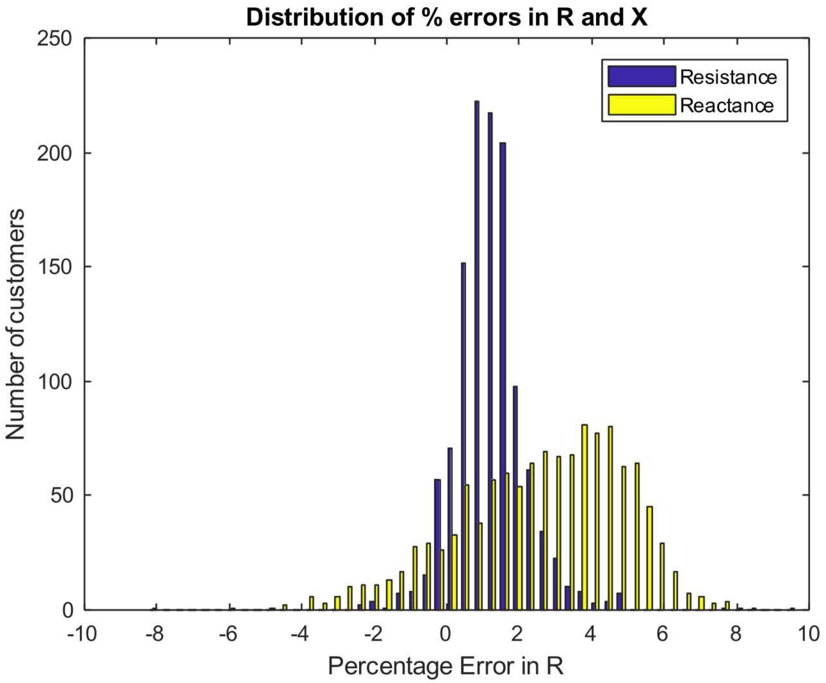


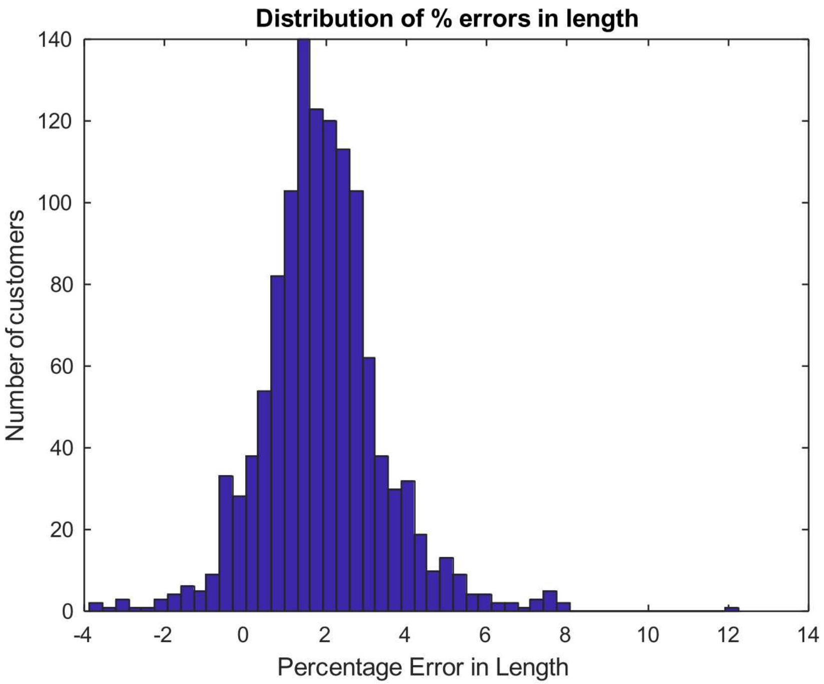

Customers often vary significantly from a simple 100ft of #2 assumption: up to three times this value was common.

# Parameter Estimation Results

The results for the test circuit (1209 residential customers):

- Error is defined as *Estimated_Value – Actual_Value*

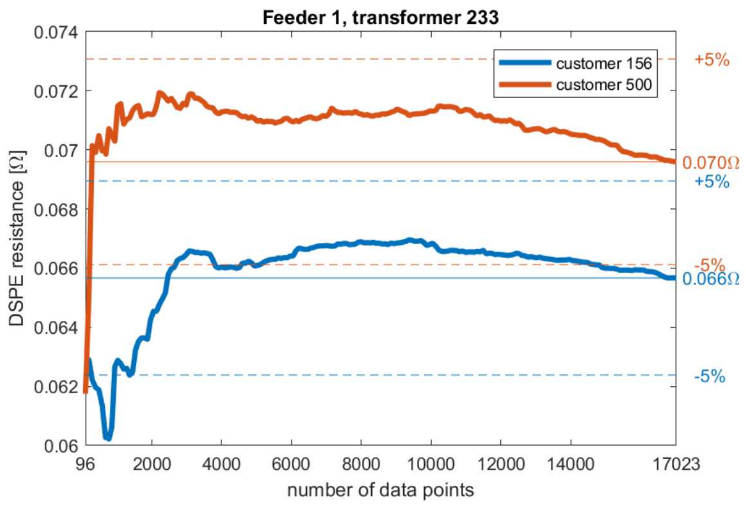


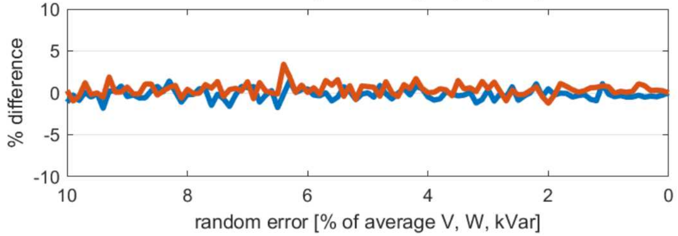

| | R % Error | X % Error | Length (ft) |
|---|---|---|---|
| Mean Error | 1.2064 % | 2.6187% | 1.26 |
| Mean Absolute Error | 1.5400 % | 3.284% | 1.4 |
| Root Mean Square Error | 1.653 % | 3.5010% | 1.5 |

# Sensitivity Analysis

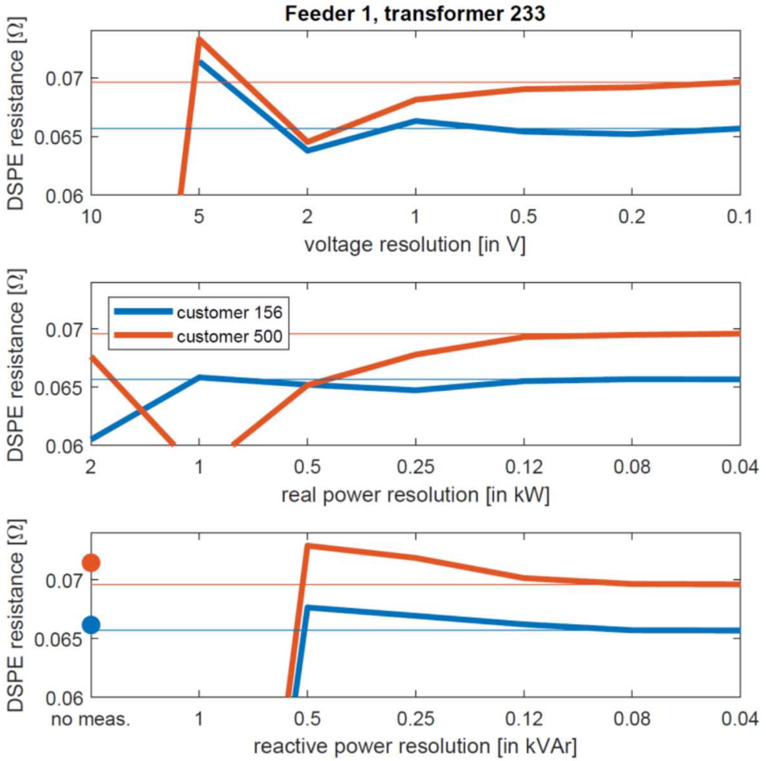
## Amount of Data




## Random Errors




## Data Resolution




- Over all customers, found that ~8,000 data points (<3 months of 15-min data) was sufficient to accurately derive parameters and topology.
- Need about 2V and 0.25kW or better resolution; low kVAr sensitivity
- Random errors in measurements => random errors in DSPE

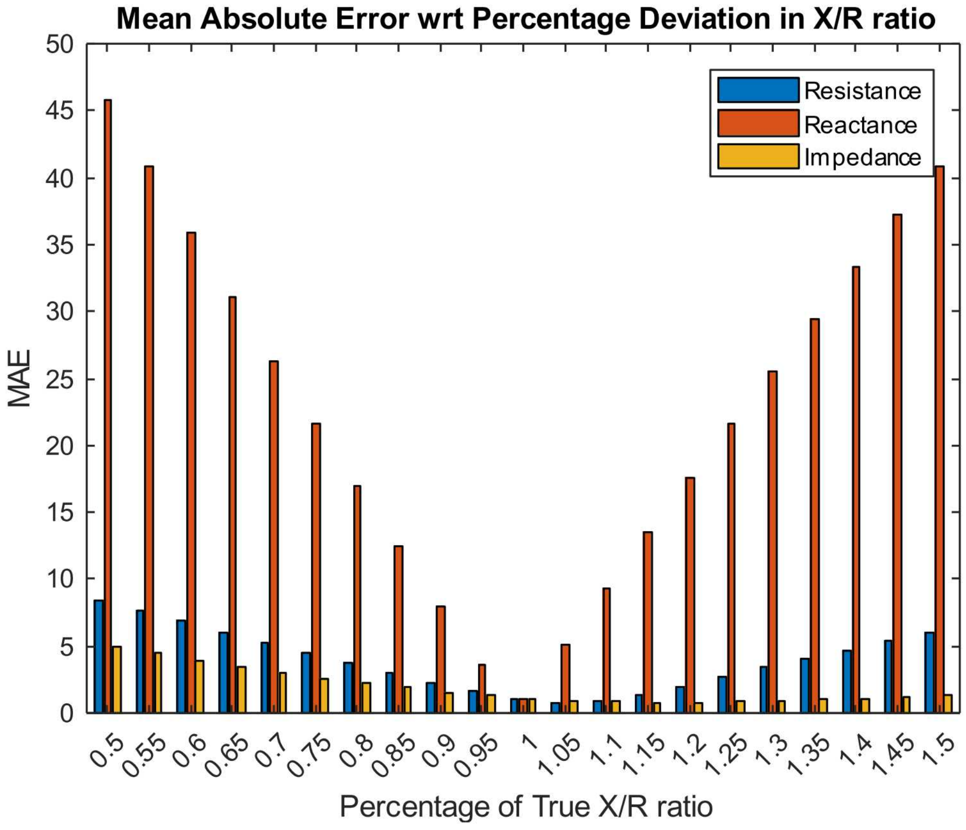
# Unknown AMI Power Factor

Utilities may not record reactive power measurements from AMI

Parameter estimation can still be performed without reactive power, if the X/R ratio of the conductors is known (based on cable type)

- As a side note, using the correct X/R of the cable improves the parameter estimation results – 1.0% MAE vs. 1.5% MAE
- Accuracy is very sensitive to having the correct X/R ratio

It also only works if the X/R is low and power factor is high. Errors are large for parameter estimation across transformers if we do not have reactive power measurements.



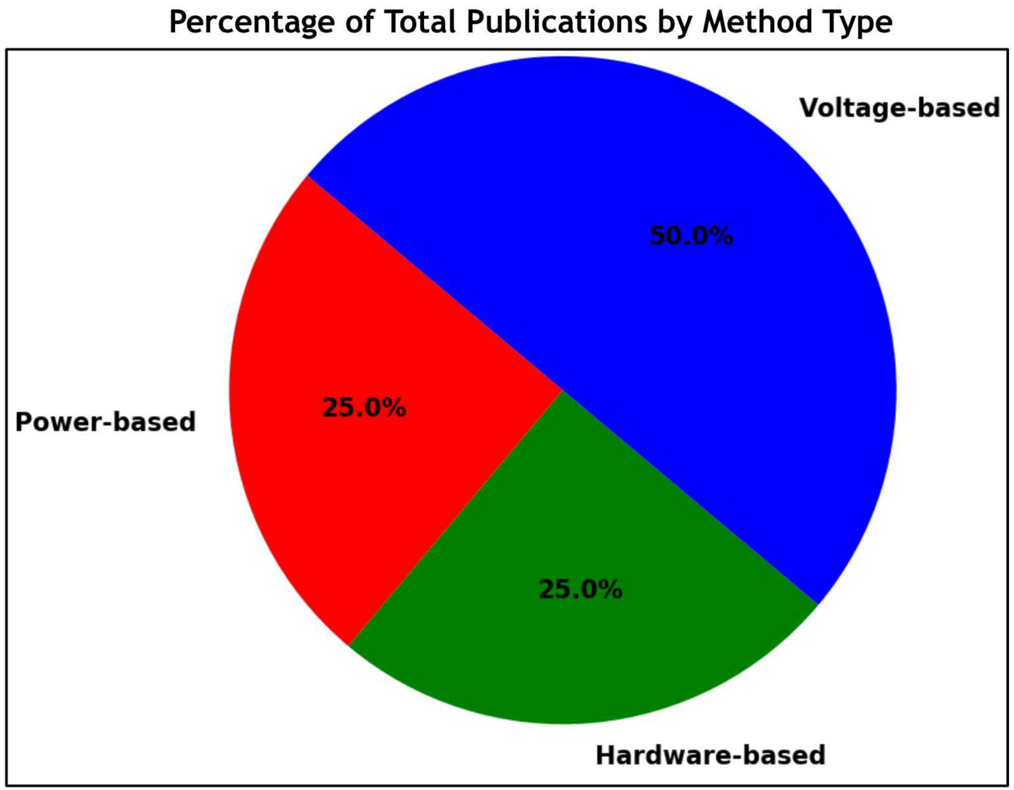
# Outline

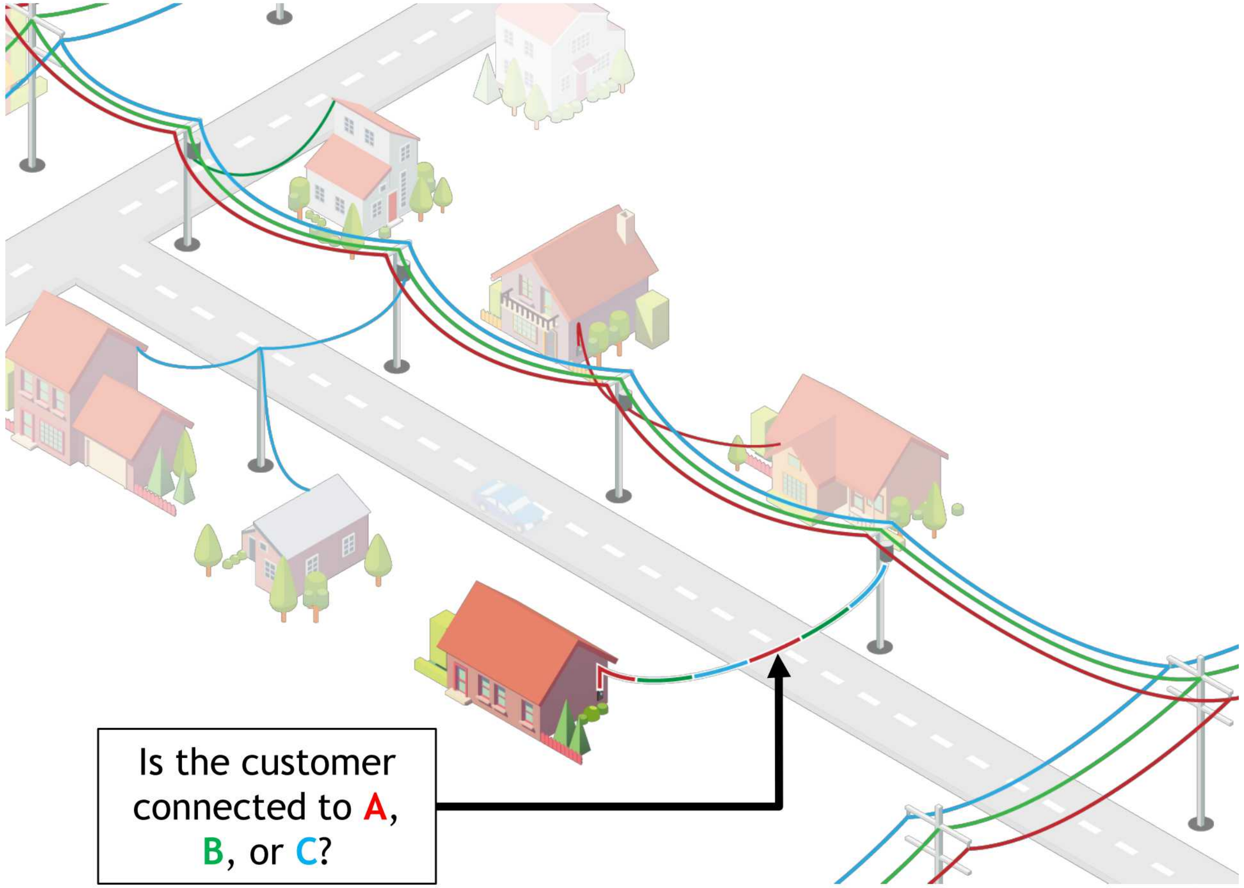
- Introduction
- Overview of Data-Driven Calibration of Distribution System Models
- Use Cases
  - Parameter and Topology Estimation of Low-Voltage Secondary Systems
  - Phase Identification
  - Identification of Service Transformer Connections
- Future Directions of Data-Driven Modeling
- Conclusions

# Phase Identification - Introduction

Problem Statement: Given a set of customers, identify the correct phasing for each customer

- Phase Identification methods generally fall into three categories, hardware-based, power-based, or voltage-based methodologies
- ~40 publications on phase identification methods

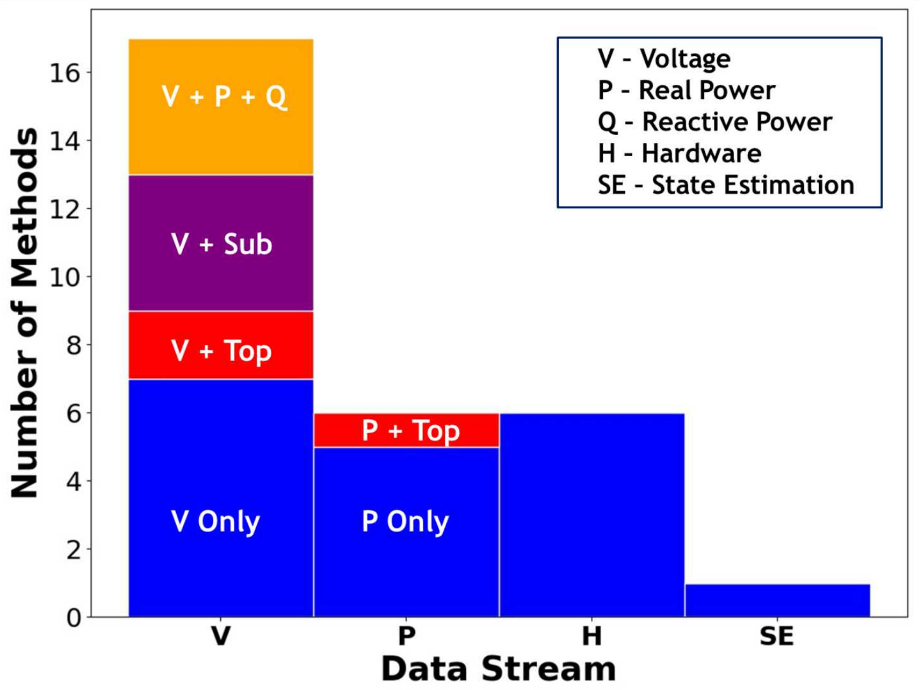


# Phase Identification - Introduction

**Hardware Methods:**
**($\mu PMU$ devices, Signal injection)**
- Advantages:
  - Highly accurate and well-established
- Disadvantages:
  - Expensive equipment
  - Large numbers of man-hours are required

Literature Review Methods and Required Data

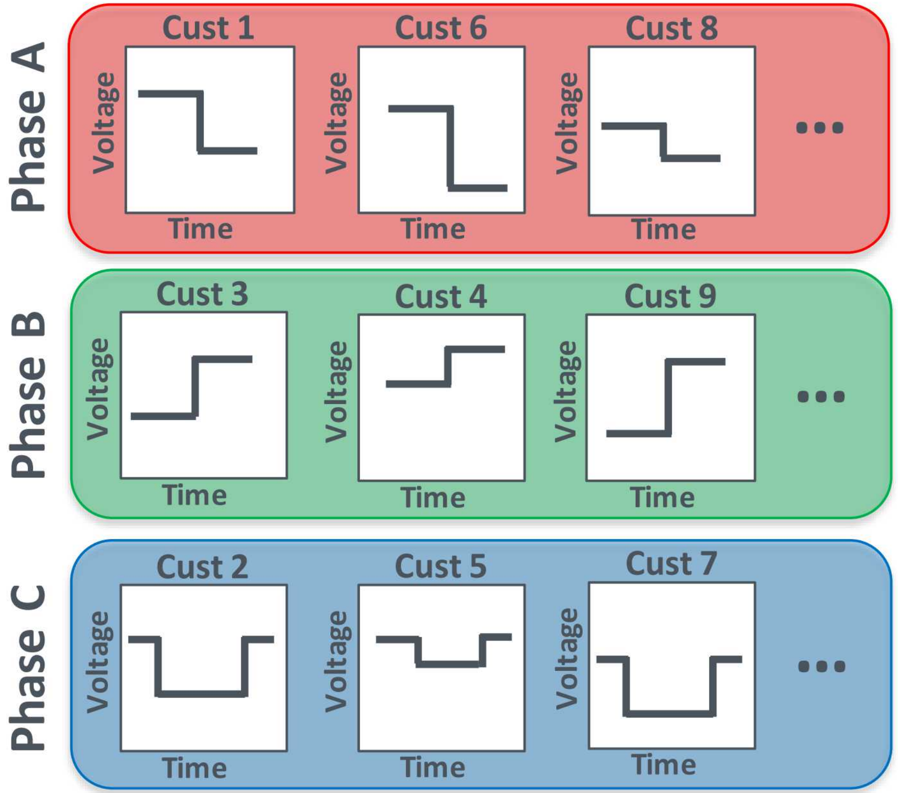


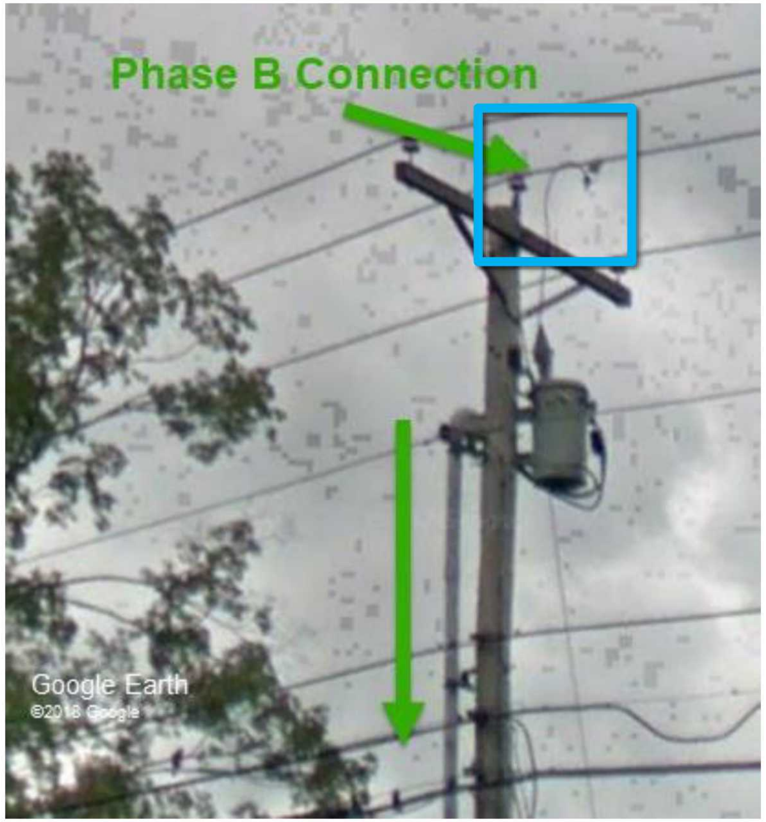

**Power-Based Methods:**
**(Load summing methods, Salient features)**
- Advantages:
  - Most utilities record this data
  - Do not require extensive man-hours
- Disadvantages:
  - Often sensitive to less than 100% AMI coverage
  - Recent work shows these methods to be less accurate in general than hardware-based or voltage-based methods

**Voltage-Based Methods:**
**(Correlations between customers)**
- Advantages:
  - Robust to less than 100% AMI coverage
  - Shown to be more accurate in general than power-based methods
  - Do not require extensive man-hours
- Disadvantages:
  - Fewer utilities are currently recording this data

# Phase Identification - Methodology

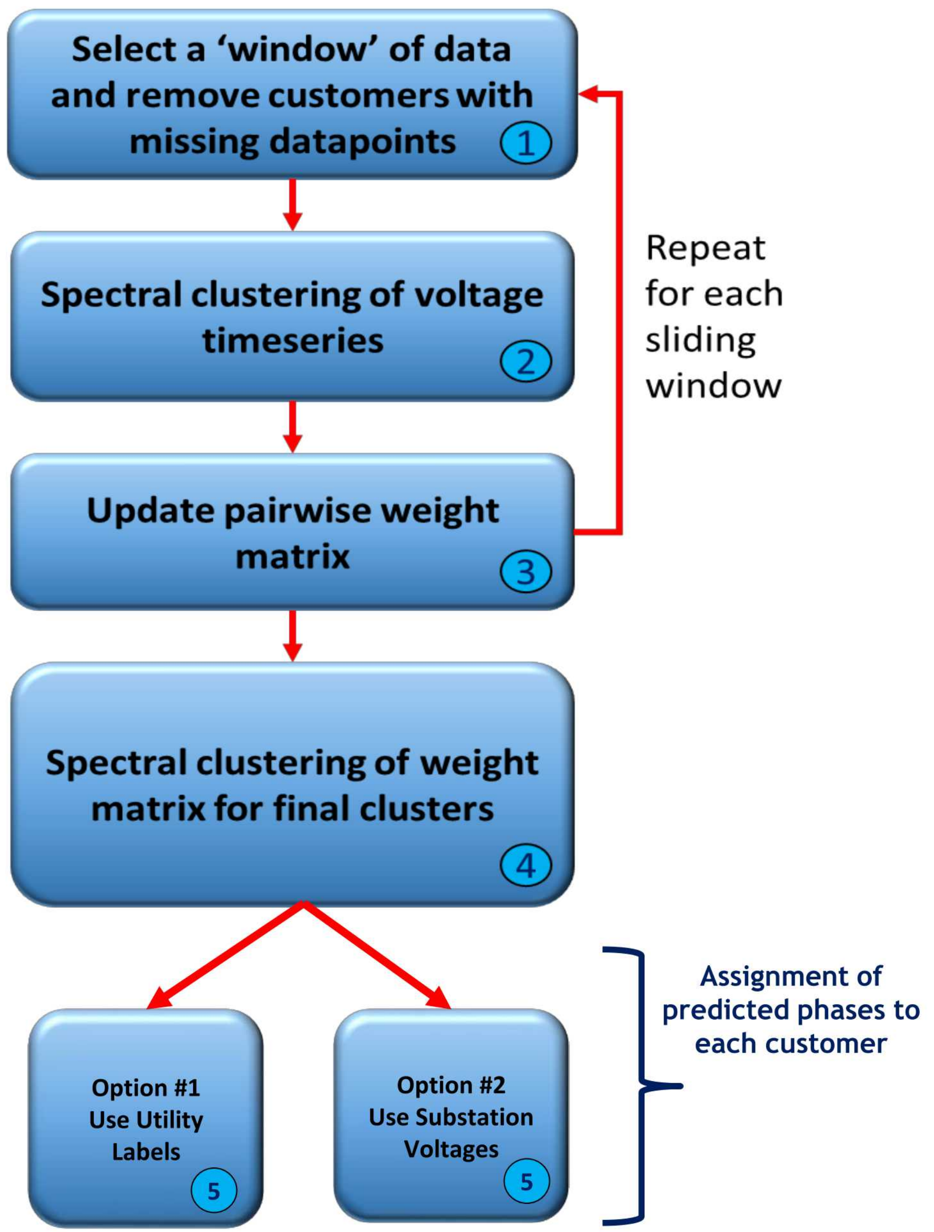
**Conceptually we understand from experience and the physical design of the system, that customers connected to the same wire (Phase A, Phase B, or Phase C) probably vary together.**



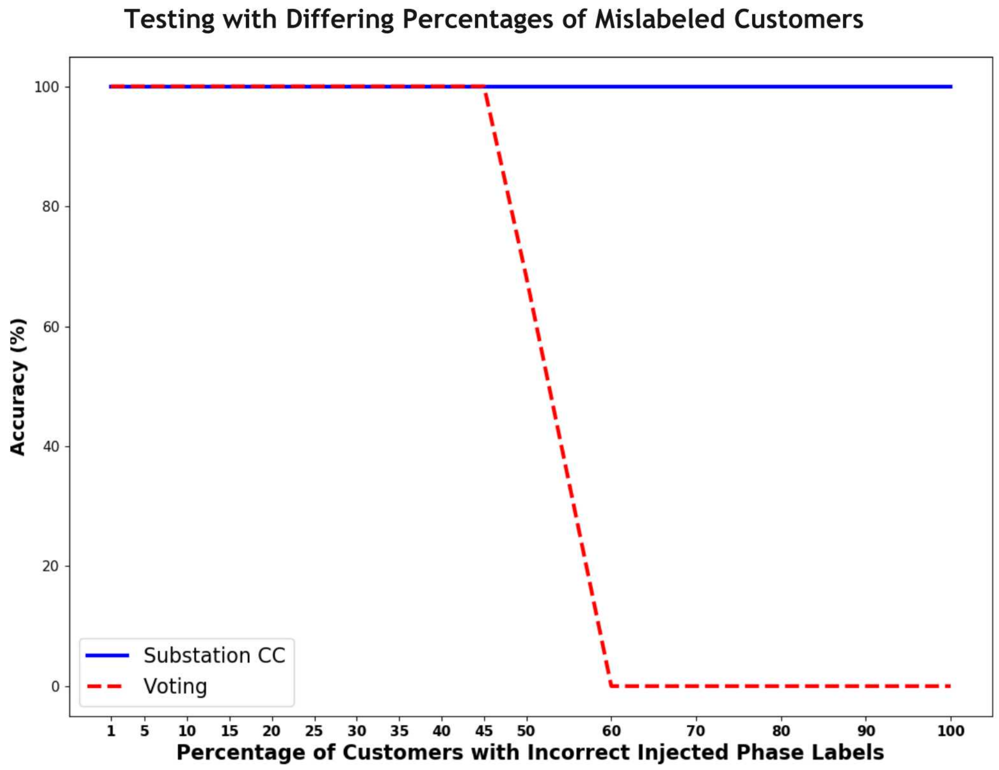




**Advanced Metering Infrastructure (AMI) meters provide timeseries voltage measurements at each residence**

# Phase Identification - Methodology

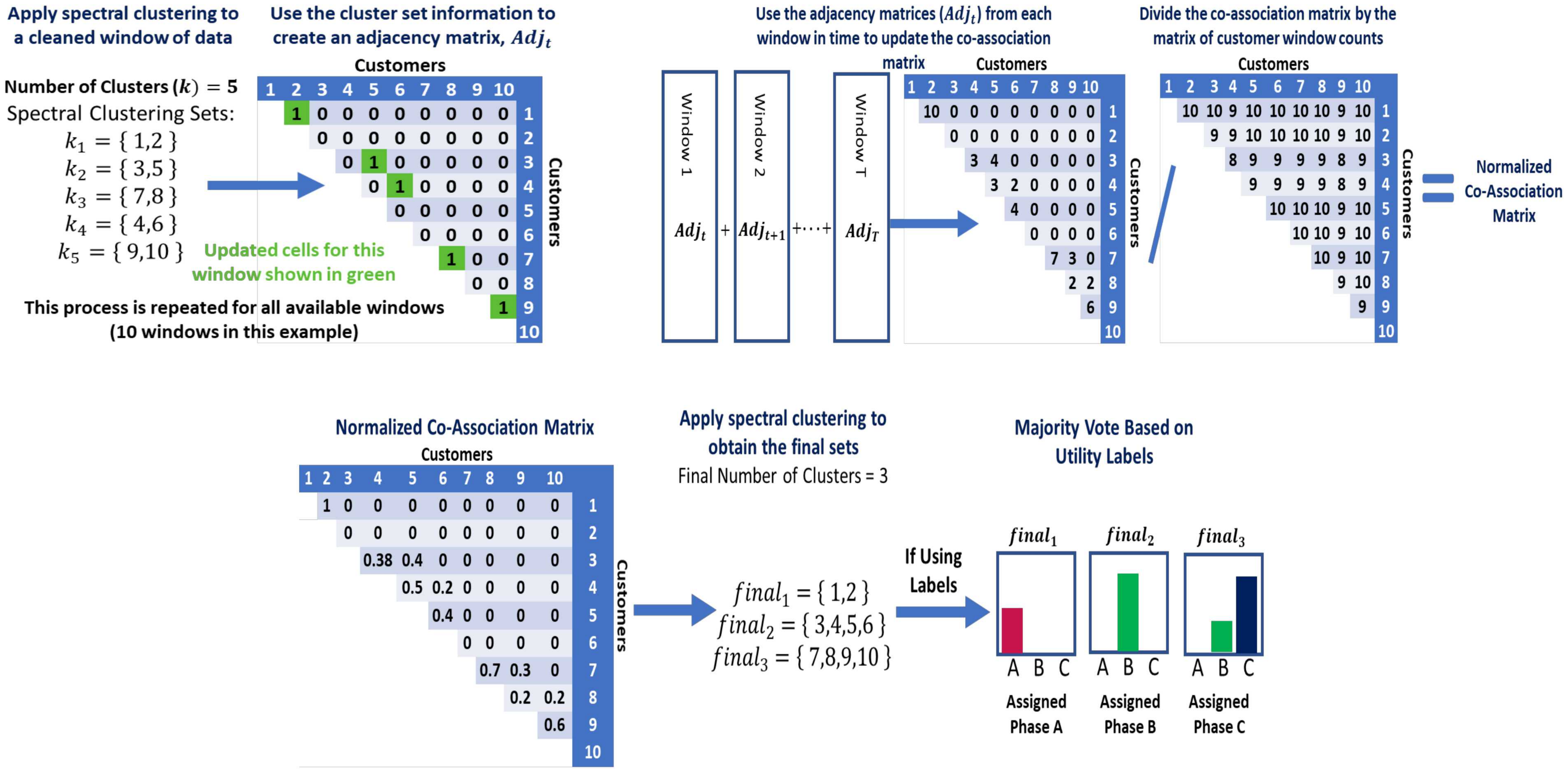



**An ensemble algorithm using spectral clustering is used to create a co-association matrix which is used to group customers by phase**




Choose how to assign final predicted phases based on the availability (and trust) in existing utility phase labels

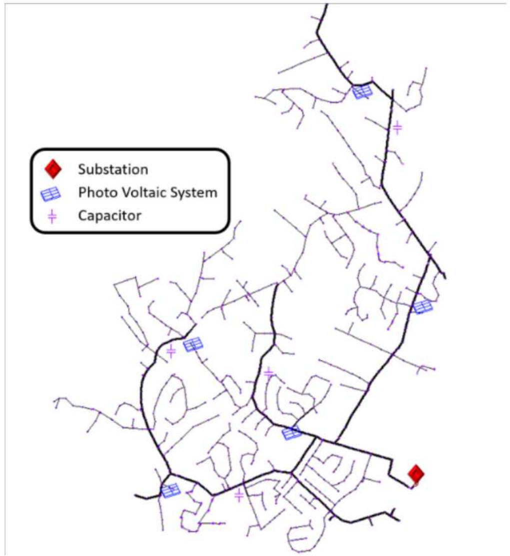
# Phase Identification - Methodology



# Phase Identification - Results

**Testing Circuits - Sandia:**

- Synthetic Data:
  - EPRI Ckt 5 – 1379 Residential Meters
  - Achieved 100% accuracy under rigorous testing scenarios
- Utility Data:
  - 1055 Residential Meters from a utility in the Northeastern US.
  - Predicted 143 customers (~13%) to be on a different phase from the original utility model




*EPRI Ckt 5*[1]

**Testing Circuits - CYME:**

- Synthetic Data:
  - Achieved >99% accuracy on all systems
- Utility Data:
  - 209 Residential Meters from a utility in the Northeastern US.
  - Predicted 35 customers to be on a different phase from the original utility model, 29 from one lateral and 6 from individual transformers

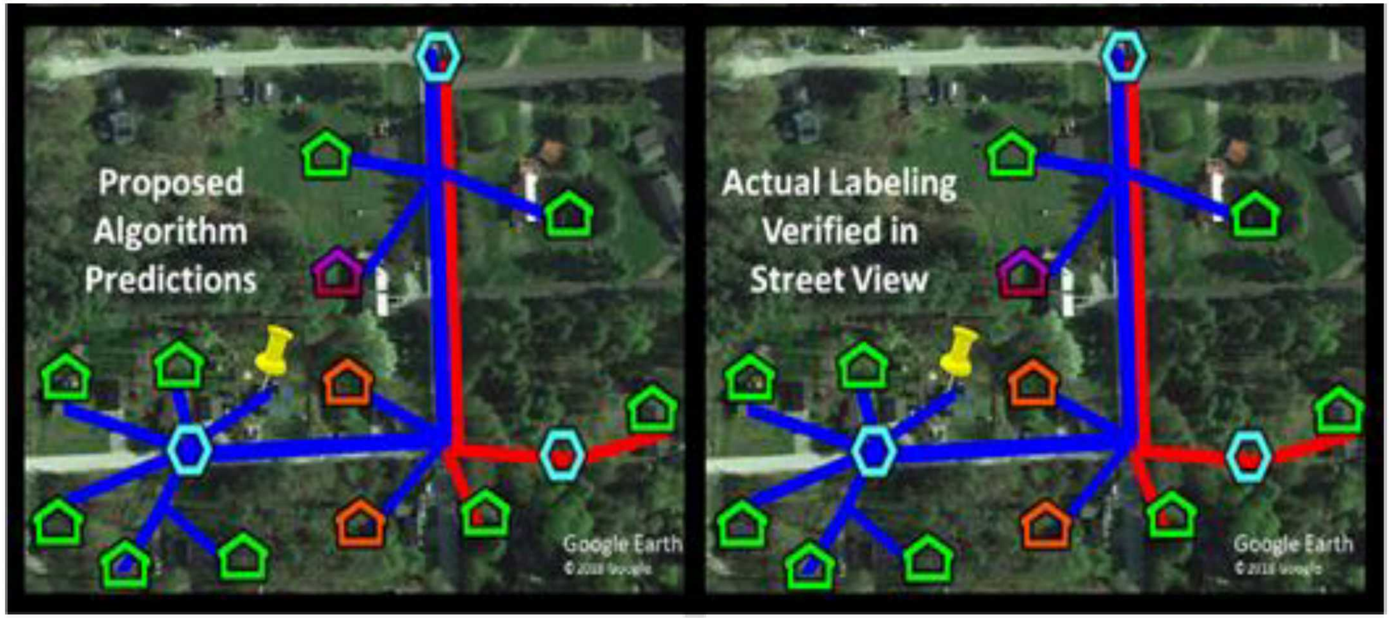
| Network | Nodes | AMI Meters | Substation Regulators | Inline Regulators |
|---|---|---|---|---|
| EPRI's CKT5 | 3003 | 1373 | 0 | 0 |
| North #1 | 2369 | 615 | 1 | 3 |
| North #2 | 4065 | 963 | 1 | 6 |
| South | 1778 | 447 | 0 | 1 |

Figure Credit: Francis Therrien – CYME | Eaton

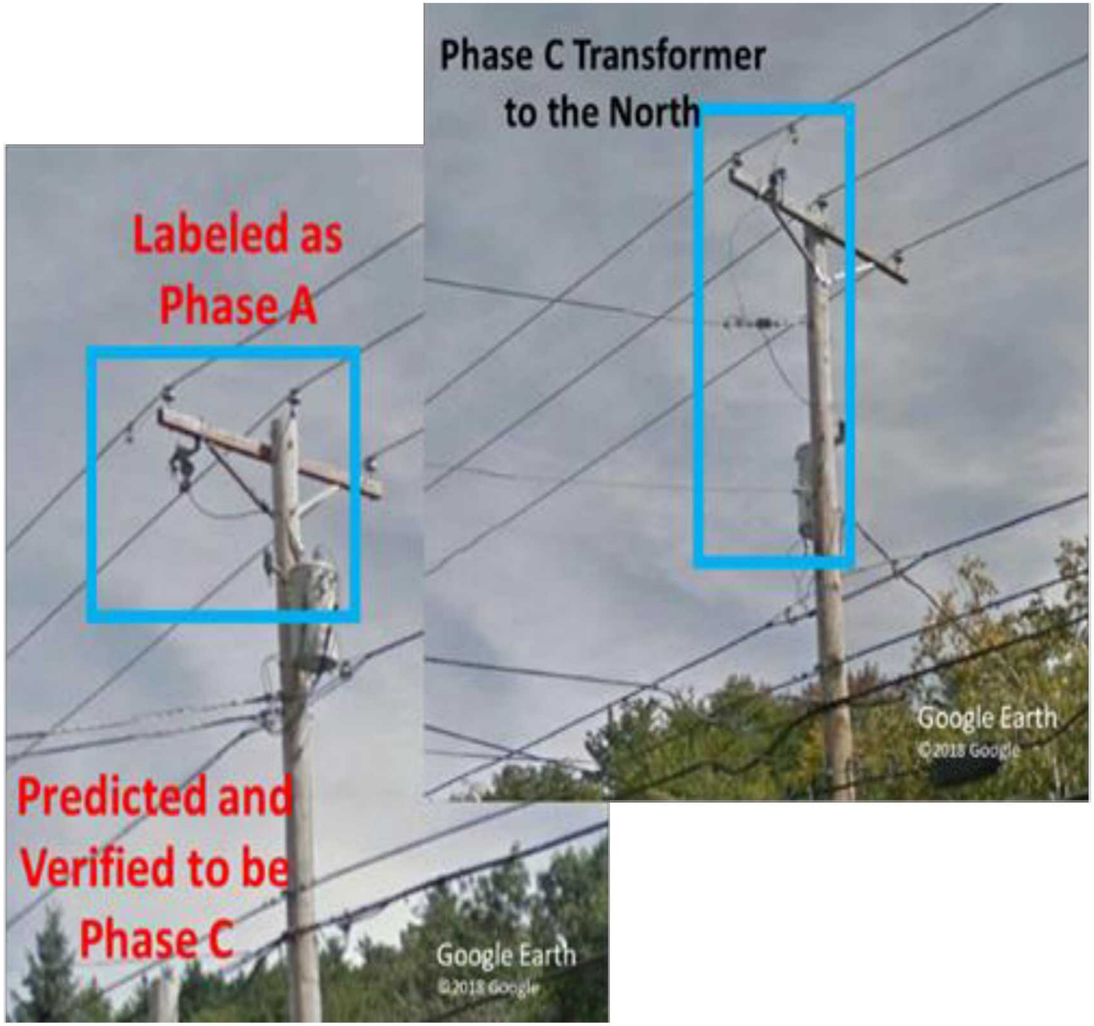
1: Figure Citation: J. Fuller, W. Kersting, R. Dugan, and S. C. Jr., "Distribution Test Feeders," *IEEE PES AMPS DSAS Test Feeder Working Group*, 2013. [Online]. Available: http://sites.ieee.org/pes-testfeeders/.

# Phase Identification - Results


Purple and Orange marked meters are incorrectly labeled in the original utility model



# Phase Identification - Results




Left-hand transformer labeled in the utility model as Phase A, but predicted to be Phase C by our phase identification algorithm

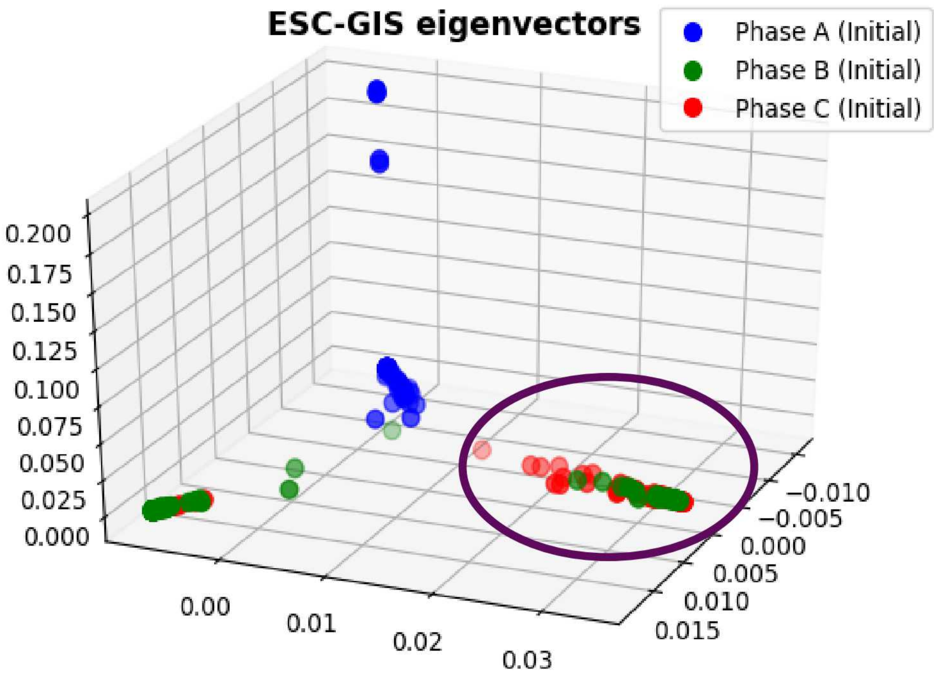
Right-hand transformer is the next Phase C transformer to the North (labeled and predicted to be C) is clearly connected to the same wire as the left-hand transformer

Not shown, the next Phase C transformer to the south is also connected to the same wire

# Phase Identification - Results

Note the group of Phase B (green) customers clustered with the majority Phase C (red) customers in the right-hand cluster

This lateral was identified by the phase identification algorithm as incorrect and the predicted labeling of Phase C was verified by the utility.

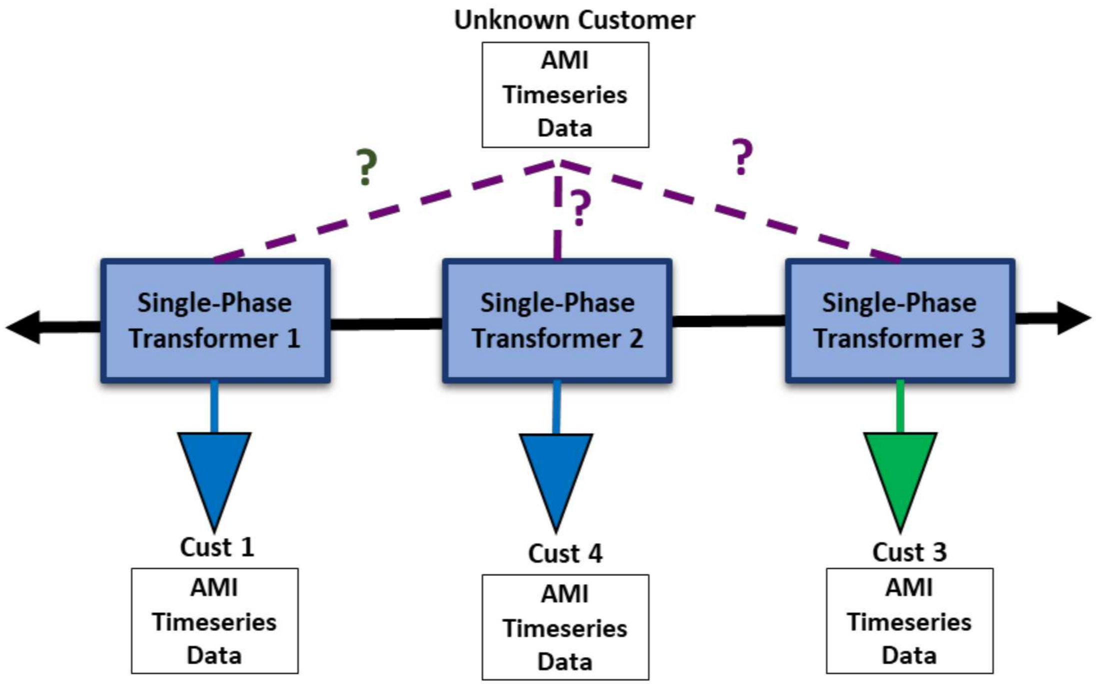


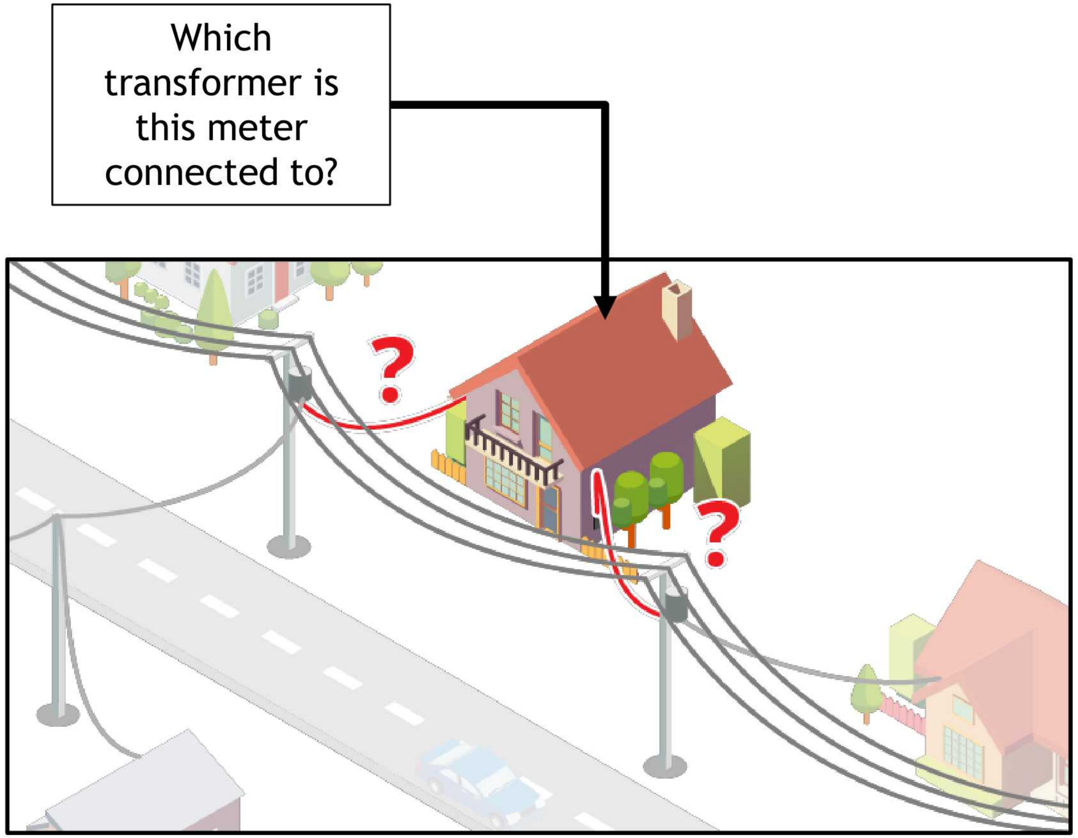

The utility reported that they had planned to move the lateral to Phase B from Phase C and changed the labeling in the model but did not physically move the lateral. Thus the 29 customers on that lateral remain on Phase C as predicted by the phase identification algorithm and shown in the figure.

Figure Credit: Francis Therrien – CYME | Eaton

# Meter to Transformer Task - Introduction

Problem Statement: Given AMI data from a set of customers, group the customers by transformer



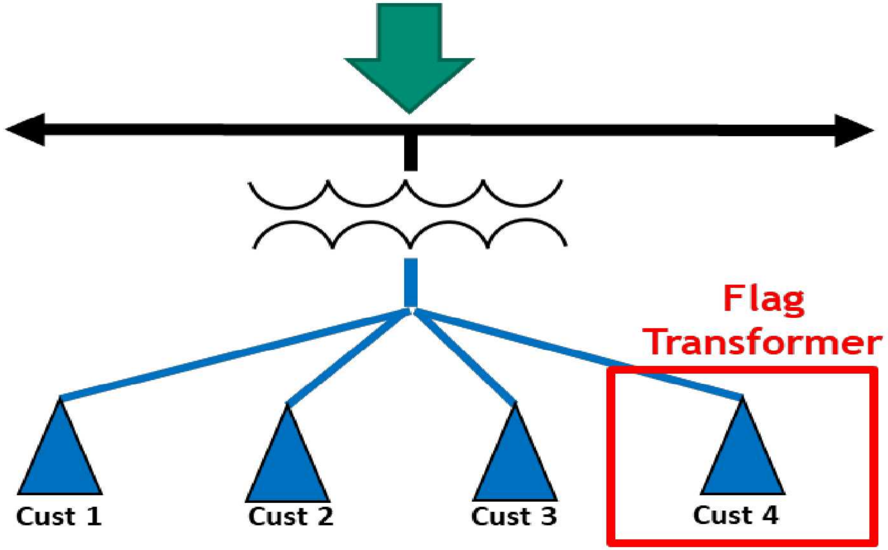
# Meter to Transformer Task – Error Flagging Methodology

**Correlation Coefficient analysis is used to flag transformer with customer connection labeling errors**

1. Calculate pairwise correlation coefficients using the 'window' method
2. Inspect the pairwise correlation coefficients for customers labeled on the same transformer
3. Flag transformers containing customer(s) with correlation coefficients below the threshold

Repeat for each Transformer

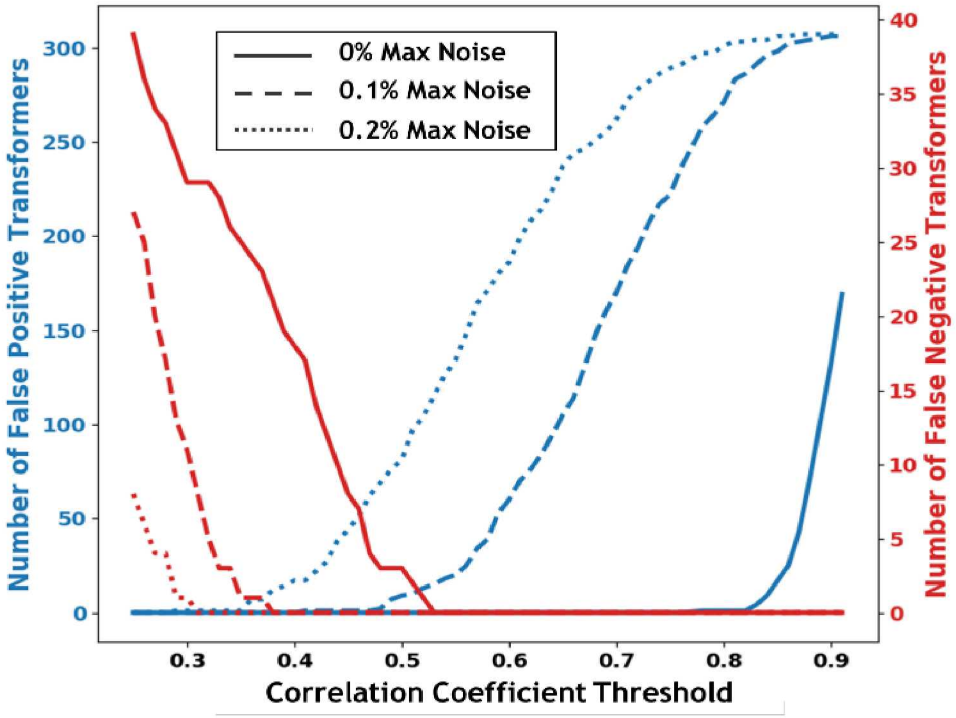
| | Cust #1 | Cust #2 | Cust #3 | Cust #4 |
|---|---|---|---|---|
| Cust #1 | 1 | 0.9 | 0.9 | 0.2 |
| Cust #2 | 0.9 | 1 | 0.9 | 0.2 |
| Cust #3 | 0.9 | 0.9 | 1 | 0.2 |
| Cust #4 | 0.2 | 0.2 | 0.2 | 1 |



# Meter to Transformer Task – Error Flagging Results

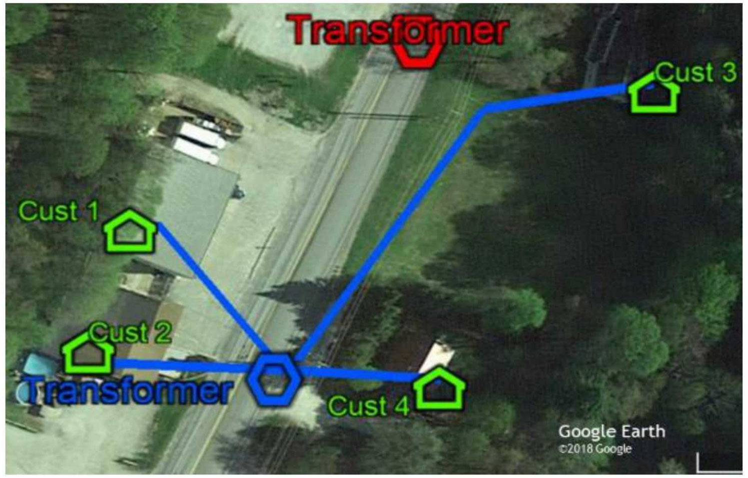
A significant concern with distribution system model validation is the false positive case, where a methodology injects new errors into a model

## Synthetic Data Testing

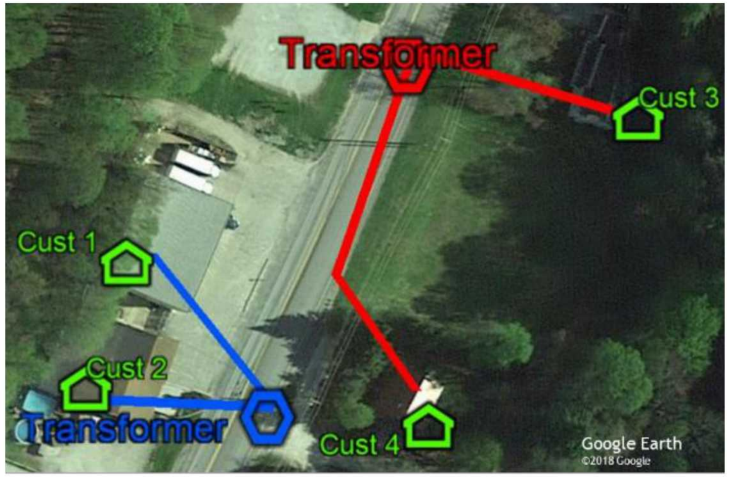



## Utility Data Testing

**Original Utility Labeling**




**Actual labeling verified using Google Earth imagery**

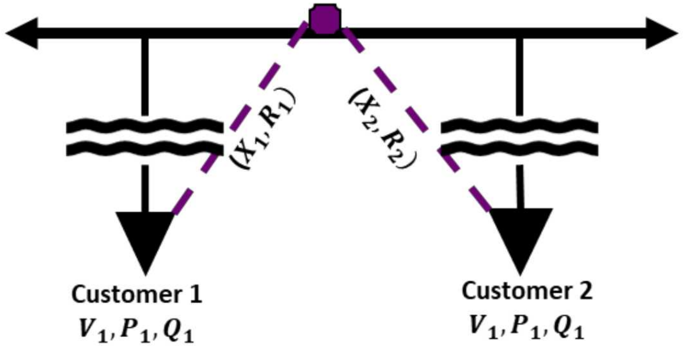


# Meter to Transformer Task - Pairing Methodology

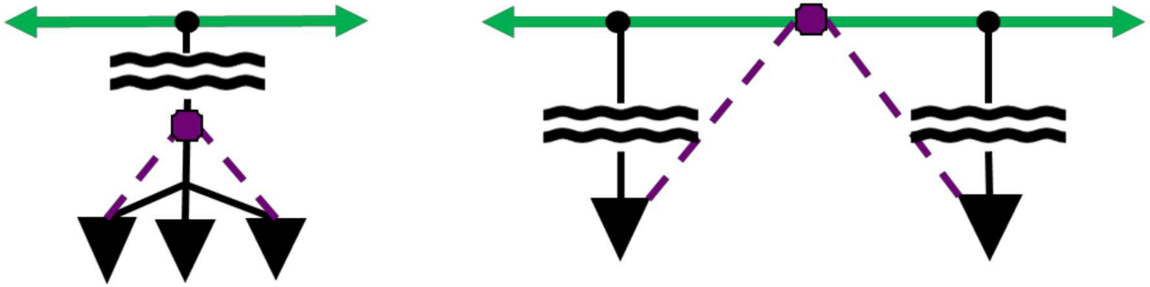
**A pairwise linear regression model is fit between customers to estimate the goodness of fit, reactance distance, and resistance distance for the customers**

$$V_1 - V_2 = P_1R_1 + Q_1X_1 - P_2R_2 - Q_2X_2$$

Fit regression model to calculate
- measure of fit, **MSE** or $R^2$
- Reactance Distance $(\boldsymbol{X_1} + \boldsymbol{X_2})$
- Resistance Distance $(\boldsymbol{R_1} + \boldsymbol{R_2})$
- $\boldsymbol{X}$ ***and*** $\boldsymbol{R}$ represent the 'distance' from the AMI meters to the nearest point the two meters are electrically connected ( )

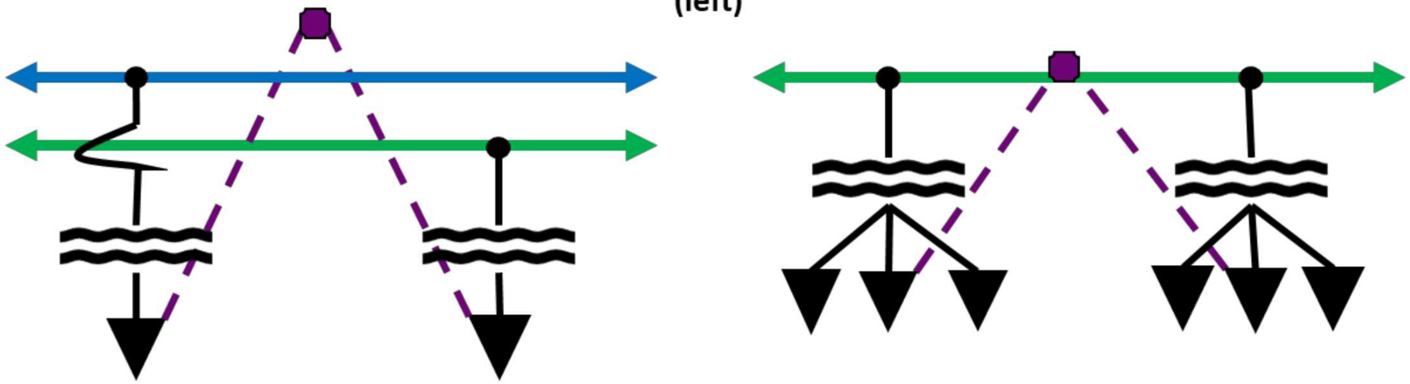



**When the regression fit ( ) is good there are two options: Customers on the same transformer (left), single customers on transformers (right)**

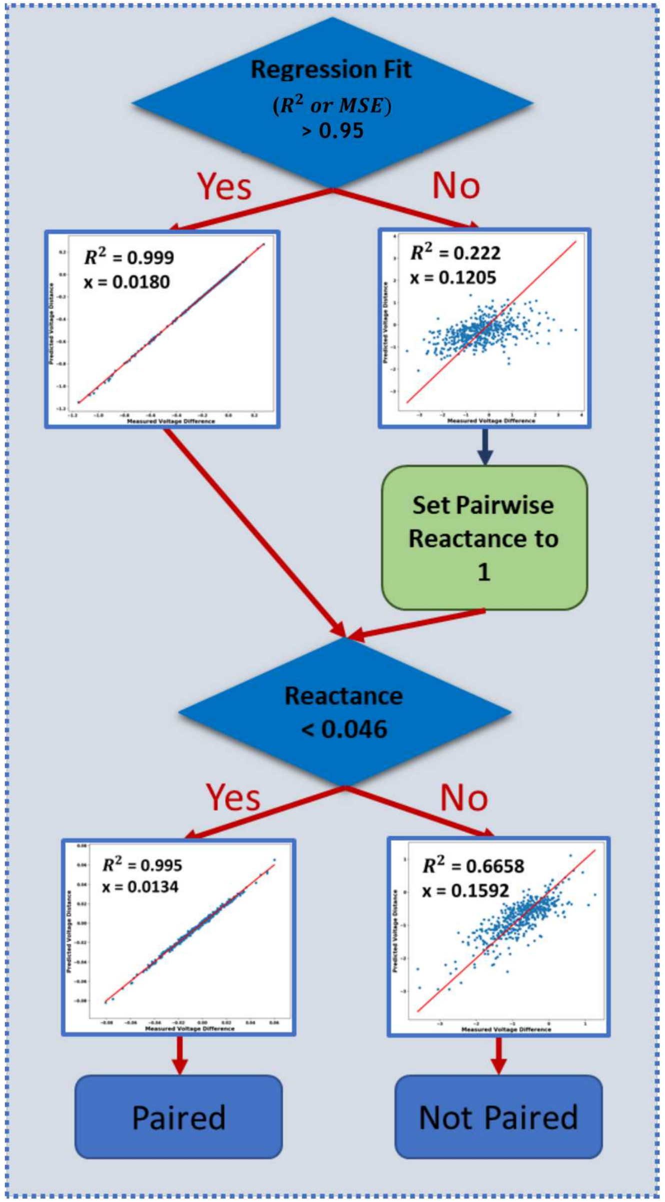


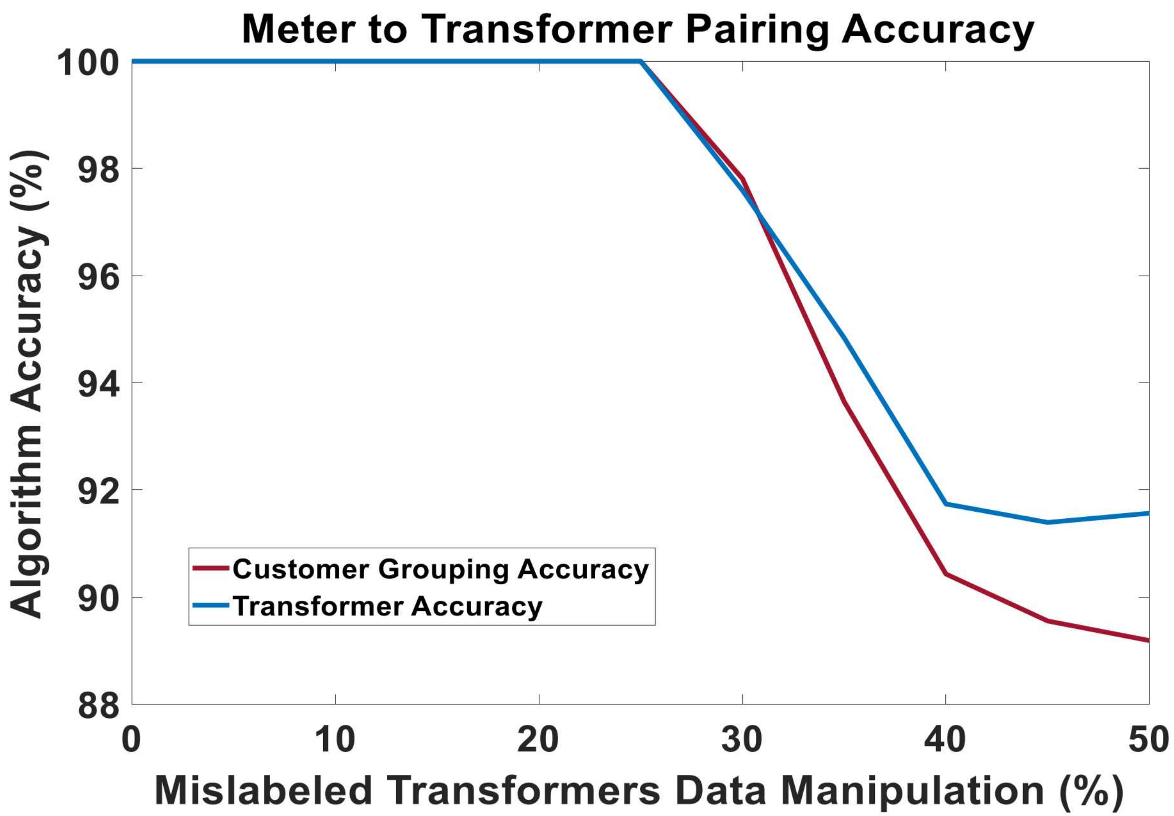
**The difference in these two cases is evident in the reactance distance $(X_1 + X_2)$**

**The regression fit ( ) is poor if there is physics, like voltage drop due to other customers on the transformer (right), or topology, like across phases (left)**



# Meter to Transformer Task - Pairing Methodology and Results




This method works well in the base case (unmanipulated data) however an ongoing research challenge is adapting this method to work in the under the various data issues mentioned in previous slides

# Conclusions and Future Work Directions

## Many innovations in AI and machine learning have not yet been applied to the power systems domain

As improvements and breakthroughs happen in other domains, those concepts can be adjusted and applied to solve power systems problems

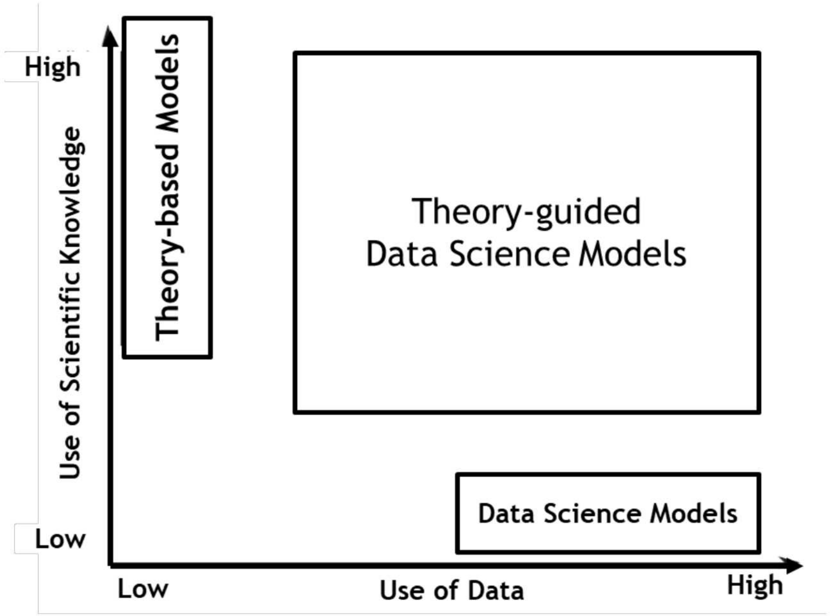
Similarly, lessons learned from other domains can be used to avoid similar situations

## Integration of Physics-based Constraints into AI

- Leverage existing knowledge (physical laws, power flow, etc) in AI-based algorithms
- Achieve more accurate results and faster training

- *SNL Project – "REDLY: Resilience Enhancements through Deep Learning Yields"*
- *loss = 'model-based' loss + 'physics-based' loss*
- $loss = loss_{data}(X_b, y_c) + (\mu loss_{KCL}(X_c) + \alpha loss_{KVL}(X_c))$

(Kirchoff's Current Law) (Kirchoff's Voltage Law)




A. Karpatne *et al.*, "Theory-guided Data Science: A New Paradigm for Scientific Discovery from Data," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 10, pp. 2318–2331, 2017.

## Explainable AI and Uncertainty Quantification

- Understand why a particular prediction/decision was given
- Understand the error bounds on predictions/decisions

- *SNL Project - "Opening the 'Black Box': An Experimentally-Validated Explainable Machine Learning Framework"*

# Conclusions and Future Work Directions

## Distributed, AI-based Controls using Fog Computing

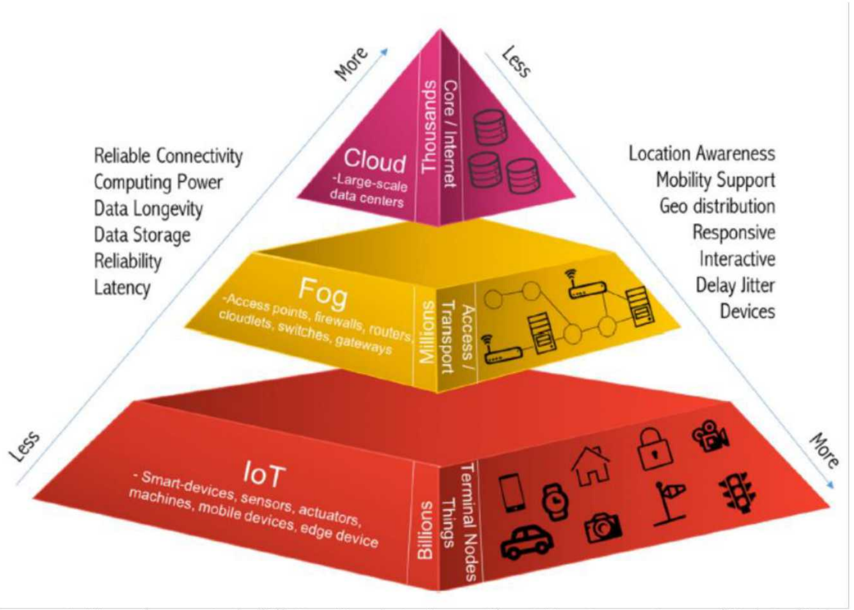
- Create resilient systems in the event of communication loss
- Accelerate systems with low latency because processing happens physically close to sensors

- *SNL Project – "HEDGES: High-Security Edge Computing for Smart Sensor Systems"*

## Semi-Supervised, Few-Shot Learning, or Synthetically-Generated Training Data

- Learn with few or no examples of critical events
- Generate realistic new data from existing samples

- *SNL Project – "Semi-Supervised Bayesian Low-Shot Learning for Explosive Device Characterization"*




A. Yousefpour *et al.*, "All One Needs to Know About Fog Computing and Related Edge Computing Paradigms: A Complete Survey," *J. Syst. Archit.*, vol. 98, pp. 289–330, Sep. 2019.

# Conclusion

- There are many promising applications of Big Data and Machine Learning in power systems.
  - It is an exciting time to be at this intersection – new algorithms, large datasets, computing power
- There are many challenging problems yet to be solved with some fascinating future research directions in big data and machine learning:
  - Integration of Physics-based Constraints into machine learning algorithms
  - Explainable AI and Uncertainty Quantification
  - Distributed AI-based Controls using Fog Computing
  - Semi-supervised, Few-shot learning, or Synthetically Generated Training Data
- Best results require integration between Big Data experts and power system experts

# Questions?

Sandia National Laboratories

Electric Power Systems Research Group

Matthew Reno
R&D S&E, Electrical Engineering
mjreno@sandia.gov

Logan Blakely
R&D S&E, Computer Science
lblakel@sandia.gov

# References


- K. Ashok, M. J. Reno, D. Divan, "Secondary Network Parameter Estimation for Distribution Transformers," IEEE Innovative Smart Grid Technologies (ISGT), 2020.
- K. Ashok, M. J. Reno, D. Divan, and L. Blakely, "Systematic Study of Data Requirements and AMI Capabilities for Smart Meter Connectivity Analytics," IEEE Smart Energy Grid Engineering (SEGE), 2019.
- M. Lave, M. J. Reno, and J. Peppanen, "Distribution System Parameter and Topology Estimation Applied to Resolve Low-Voltage Circuits on Three Real Distribution Feeders," IEEE Transactions on Sustainable Energy, 2019.
- M. Lave, M. J. Reno, R. J. Broderick, and J. Peppanen, "Full-Scale Demonstration of Distribution System Parameter Estimation to Improve Low-Voltage Circuit Models," IEEE Photovoltaic Specialists Conference (PVSC), 2017.
- J. Peppanen, M. J. Reno, R. J. Broderick, and S. Grijalva, "Distribution System Model Calibration with Big Data from AMI and PV Inverters," in *IEEE Transactions on Smart Grid*, 2016.
- J. Peppanen, M. J. Reno, R. J. Broderick, and S. Grijalva, "Secondary Circuit Model Creation and Validation with AMI and Transformer Measurements," North American Power Symposium, 2016.
- J. Peppanen, M. J. Reno, R. J. Broderick, and S. Grijalva, "Secondary Circuit Model Generation Using Limited PV Measurements and Parameter Estimation," IEEE PES General Meeting, 2016.
- J. Peppanen, M. J. Reno, R. J. Broderick, and S. Grijalva, "Distribution System Low-Voltage Circuit Topology Estimation using Smart Metering Data," IEEE PES Transmission & Distribution Conference & Exposition, Dallas, TX, 2016.
- J. Peppanen, M. J. Reno, R. J. Broderick, and S. Grijalva, "Distribution System Secondary Circuit Parameter Estimation for Model Calibration," Sandia National Laboratories, SAND2015-7477, 2015.
- L. Blakely, M.J. Reno, and J. Peppanen, "Identifying Common Errors in Distribution System Models," in Proc. 46th IEEE Photovolt. Spec. Conf. (PVSC), 2019.
- L. Blakely, and M.J. Reno, "Phase Identification Using Co-Association Matrix Ensemble Clustering", IET Smart Grid, June 2020.
- L. Blakely, M.J. Reno, and W. Feng, "Spectral Clustering for Customer Phase Identification Using AMI Voltage Timeseries", in Proc. Power Energy Conf. Illinois (PECI), Feb 2019.
- L. Blakely, M.J. Reno, and K. Ashok, "AMI Data Quality and Collection Method Considerations for Improving the Accuracy of Distribution System Models", Proc. 46th IEEE Photovolt. Spec. Conf. (PVSC), 2019.
- L. Blakely, M.J. Reno, "Identifying Errors in Service Transformer Connections", IEEE PES General Meeting, Aug 2020.