Pacific Gas and Electric Company

EPIC Final Report

Program	***Electric Program Investment Charge (EPIC)***
Project	***EPIC 2.14 – Automatically Map Phasing Information***
Reference Name	***EPIC 2.14 – Phase ID***
Department	*Electric Asset Management: Grid Modernization*
Project Sponsor	Ferhaan Jawed
Project Business Lead	Anne-Lise Laurain
Contact Info	EPIC_Info@pge.com
Date	November 28, 2018
Version Type	Final
Version	

Table of Contents

List of Tables

List of Figures

Table of Acronyms

ADMS	Advanced Distribution Management System
AMI	Advanced Metering Infrastructure
CPUC	California Public Utilities Commission
DER	Distributed Energy Resources
DERMS	Distributed Energy Resource Management System
EPIC	Electric Program Investment Charge
ESFT	Electronic Secure File Transfer
EVC	Ensemble Voltage Cluster
ERVVCC	Ensemble Relative Virtual Voltage Constrained Cluster
FRC	Feature Reduced Cluster
GHG	Greenhouse Gas
GIS	Geographic Information System
IAGM	Impedance Adjusted Geographic Match
ICDM	Impedance Corrected Distance Mapping
IOU	Investor-owned utility
IPC	Iteration Pseudo-code
kV	Kilovolt
kWh	Kilowatt Hour
LDM	Logical Data Model
LFMC	Load Flow Monte Carlo
LiDAR	Light Detection and Ranging
MDB	Master Database
MIF	Metadata Inconsistency Flagging
MW	Megawatt
NDRCA	Neighborhood Dimension Reduced Cluster Assignment
µPMUs	Micro Synchro-phasor Measurement Units
PG&E	Pacific Gas & Electric Company
Phase ID	Phase Identification
RCC	Regression Corrected Correlation
RCKM	Reduced constrained k-Means
SCADA	Supervisory Control and Data Acquisition

SCE	Southern California Edison
SDG&E	San Diego Gas and Electric
SSD	Source Side Device
TD&D	Technology Demonstration and Deployment
t-SNE CHC	t-SNE constraint-driven hybrid clustering
V	Volt
VCDRCA	Voltage Clustered Dimension Reduced Cluster Assignment

Bibliography

Berrisford, 2013. *A tale of two transformers: An algorithm for estimating distribution secondary electric parameters using smart meter data..* s.l., s.n.

Mitra, R. et al., 2015. *Voltage Correlations in Smart Meter Data.* Sydney, NSW, Australia, ACM, pp. 1999-2008.

Pezeshki, H. & Wolfs, P., 2012. *Correlation based method for phase identification in a three phase LV distribution network.* Bali, Indonesia, Australasian Universities Power Engineering Conference (AUPEC).

Short, T. A., 2013. Advanced Metering for Phase Identification and Secondary Modelling. *IEEE Transactions on Smart Grid,* June, 4(2), pp. 651 - 658.

Wang, W. et al., 2016. *Phase Identification in Electric Power Distribution Systems by Clustering of Smart Meter Data.* Anaheim, CA, USA, IEEE.

Wang, W., Yu, N. & Lu, Z., 2017. *Advanced Metering Infrastructure Data Driven Phase Identification in Smart Grid.* Rome, Italy, s.n.

1 Executive Summary

Pacific Gas and Electric Company's (PG&E) Electric Program Investment Charge (EPIC) Project 2.14 *Automatically Map Phasing Information* primary goal was to discover, evaluate, select, develop, and validate specific methods to automatically identify meters' connectivity.

1.1 Objectives

EPIC 2.14 successfully evaluated the following project objectives:

- Compare various analytical methods and other potential alternatives to determine best project approach for phase identification (Phase ID) and meter-to-transformer mapping
- Develop analytical algorithms that use Supervisory Control and Data Acquisition (SCADA), Geographic Information System (GIS), and SmartMeter™ data for automated phase identification and meter-to-transformer mapping
- Determine market readiness for meter connectivity solutions in a real grid environment and compare with internal algorithms
- Identify gaps for full-scale deployment and make recommendations on how the methods would be deployed for operational use

1.2 Project Overview

As described in PG&E's EPIC 2 Application,[1] the distribution network model is central to multiple existing control systems, system analyses, and work processes. Utilities striving to modernize their distribution networks require improved visibility into both the physical state of their system and its real-time load flow conditions. As the load characteristics of the distribution network evolve such as with the growth of Distributed Energy Resources (DER), it is becoming more important to have accurate and up to date information to be able to actively manage the distribution system to ensure reliability for customers. This project focused on two important steps towards modernization.

Phase Identification

Accurate phasing information is increasingly necessary for cost effective operation of the modern distribution system, and is required for the safe and efficient integration of DER. For example, improved phase information supports the requirements for new grid modernization functions such as unbalanced power flow modelling and state estimation that exist in the advanced distribution management system (ADMS) and distributed energy resource management system (DERMS) platforms.

Meter-to-Transformer Connectivity

Although PG&E's databases contain information on which transformer each meter is connected to, records do not always reflect the field conditions. Accurate meter-to transformer-connectivity information is needed to ensure proper transformer loading levels. To address these needs, this project explored a variety of pre-commercial analytics and hardware options to automatically map three-phase electrical power information and meter-to-transformer connectivity to improve the

[1] PG&E EPIC Triennial Plan (2015-2017), May 1, 2014, Attachment 1, p. 44. https://www.pge.com/includes/docs/pdfs/about/environment/pge/epic/Attachment1.pdf.

distribution network model.[2] Normal means of improving the model included field surveys as outlined in California Public Utilities Commission (CPUC) General Order 165. Field surveys represent a significant expense, have a repeatability to cost factor (i.e. similar or increased costs for each repetition of the field survey), and may not collect the detailed information needed for distribution operations.

To determine the best path forward for this needed information, the project explored whether a data-driven approach has an advantage over providing results from independent specific events or other physical approach methods. Such an approach could allow utilities to:

- Perform phased load flow analysis and distribution network state estimation
- Correct meter to distribution transformer connectivity models
- Determine lateral phase connectivity

Demonstrating an analytics based approach to automatically map phasing information on PG&E's system is a first step to improving engineering practices for phase balancing that achieve greater reliability across PG&E's system in support of various initiatives such as a DERMS platform. Within this technology demonstration endeavor, the goal was to determine which methods or approaches may be successful, in degrees of success when compared to one another against a set of measured field data. Considering Phase ID analytics, many demonstrations in the literature are presented in smaller terms as compared to the 24,000 meter sampling utilized in this project, including both verification and validation datasets.

From a programmatic perspective, exploring automated phase mapping and meter-to-transformer connectivity aligns with the primary EPIC principle of greater reliability. Additionally, identifying phase and meter-to-transformer mapping methods capable of achieving satisfactory accuracy levels at a reasonable cost also aligns with the primary EPIC principle of lower costs. Such methods could help to improve electric service affordability for customers through improved electric operations efficiencies over time, after implementing a low cost, accurate, and repeatable mapping process.

In pursuit of these and other EPIC principles, the project assessed available technologies, such as the use of advanced metering infrastructure (AMI) data, and a range of analytical methods to achieve this automated capability. Preferred methods were selected based on the results of a prioritization analysis for proof-of-concept (POC) design, build out, and testing.

Two demonstrations were conducted during this project to evaluate the best methods to determine phasing and meter to distribution transformer (meter-to-transformer) connectivity.

POC Demonstration Scope:

- Three 4-wire (21 Kilovolt (kV)) circuits
- 4 external methods evaluated – 3 external participants
- Phasing data on the three circuits collected and shared with all the participants
- 5min AMI, GIS, SCADA data

Small-scale Demonstration Scope:

- Three 3-wire (12kV) circuits and an additional 4-wire (21kV) circuit; seven circuits in total, including the three 4-wire circuits from the POC demonstration
- 1 internal method developed and 4 external methods/participants evaluated

[2] Ibid., p. 43.

- Phasing data and meter-to-transformer connectivity collected
- Hourly AMI, GIS, SCADA data, Outage records data

1.3 Key Accomplishments

The following summarizes the key accomplishments of the project over its duration:

Novel field data collection approach. A novel wireless Phase ID hardware tool that does not require direct interaction with live wires was evaluated during the field verification. The use of this tool lowered the cost of the field data collection and reduced risk due to the contactless nature of the tool with hot wires.

Phase ID Algorithm testing on 4-wire circuits. PG&E collected field data on four 4-wire (21kV) circuits and evaluated eight methods from six different organizations (including PG&E and one university) via two demonstrations. In total, 3,700 meters were chosen to build the algorithms and about 14,000 meters were used to validate the methods built for this project. Several methods provided results with an accuracy level potentially high enough for implementation.

Phase ID Algorithm testing on 3-wire circuits. PG&E collected field data on three 3-wire (12 kV) circuits and evaluated five methods from four different organizations (including PG&E and one university). One 3-wire circuit dataset was utilized to calibrate each model (data corresponding to approximately 1,500 meter points), and the remaining two 3-wire circuits were used to validate each model approach (approximately 5,000 meter points). The model training process used field results from the first circuit to calibrate model parameters, while the field results from the testing set were not used until the final validation assessment. Results obtained had a lower level of accuracy than for 4-wire circuits, however the results are promising, and performance sufficient for implementation may be achievable with a few input data enhancements, such as receiving phase to phase substation voltage data.

Meter-to-transformer Algorithm testing. PG&E collected field data on three 3-wire (12 kV) and one 4-wire (21kV) circuits and evaluated five methods from five organizations (including PG&E and one university). Data from one of the 3-wire (12kV) circuits was used to create each model (corresponding to 1,500 meters) and the remaining three circuits (about 8,000 meters) were utilized to validate the models built. The results for fully automated meter-to-transformer re-assignment are not satisfactory for implementation at this time, because the reassignments would introduce more errors than they would solve. However, some algorithms do show skill in identifying incorrectly assigned meters, so they could be used to identify a subset of candidates for targeted field validation. The exploration of meter-to- transformer algorithms evaluated in this project will allow utilities and the broader industry to better understand a number of challenges linked to this issue.

1.4 Key Takeaways

The following are the key findings and lessons learned from this project:

Project Success Metric

Method effectiveness for Phase ID was determined based on the level of asset mapping accuracy, and was measured as the percent of meters in the common dataset where the method prediction matched the field-verified value.

For the POC demonstration, there were approximately 14,400 meters in the dataset, split across three 4-wire and 21 kilovolt (kV) distribution circuits. Developers of all methods were requested to use the field verification data from 4-wire circuit 1 as the training set to tune the models, and to reserve the data from 4-wire circuit 2 and 4-wire circuit 3 for the final analysis. Table 1 provides the results from four methods developed by external organizations.

Table 1. Phase ID Proof-of-Concept - Results by Circuit

Method	4-wire circuit 1	4-wire circuit 2	4-wire circuit 3	Average
Regression Corrected Correlation (RCC)	62.8%	69.5%	77.7%	**70.5%**
Ensemble Voltage Cluster (EVC)	94.5%	97.2%	94.7%	**95.7%**
Reduced constrained k-Means (RCKM)	94.2%	92.7%	93.4%	**93.3%**
Feature Reduced Cluster (FRC)	90.8%	94.0%	91.8%	**92.4%**

For the small-scale demonstration, hourly voltage data, which was more representative of what would be available in full deployment, was used instead of the 5-minute interval data that had been used in the POC demonstration. Results were thus expected to be slightly less accurate. All the new participants had access to hourly voltage AMI data, GIS and SCADA data for all the seven circuits (three 3-wire 12kV and four 4-wire 21kV). Although the EVC Method performed well in the POC demonstration for 4-wire circuits with 5-minute interval data, the method did not perform as well in the small-scale demonstration. Therefore, PG&E developed its own internal method to handle 3-wire circuits and less granular data. Table 2 provides the results from the new method from PG&E and the methods from other participants.

Table 2. Phase ID Small Scale Demo – Results by Circuit

	3 Wires - 12 kV			4 wires - 21kV			
Methods description	**3-wire circuit 1**	**3-wire circuit 2**	**3-wire circuit 3** [3]	**4-wire circuit 4**	**4-wire circuit 1** [4]	**4-wire circuit 2**	**4-wire circuit 3**
t-SNE constraint-driven hybrid clustering (t-SNE CHC)	32.0%	30.7%	73.2%	76.3%	Phase 1	Phase 1	Phase 1
Impedance Corrected Distance Mapping (ICDM)	50.2%	Not available	94.2%	Not available	81.8%	59.3%	67.8%
PG&E's method	64.8%	52.5%	78.2%	78.2%	83.6%	71.6%	75.7%
Load Flow Monte Carlo (LFMC)	36.5%	33.6%	55.4%	35.1%	Not available	30.4%	56.7%
EVC (revised)	68.0%	42.2%	72.9%	64.5%	89.8%	58.6%	63.6%

Depending on the type of circuit, establishing the phasing connectivity through computer-based analytic methods raises different challenges. In a 4-wire system the phases operate independently, resulting in identifiable metrics at the phase level. In a 3-wire system, where each phase affects the voltage on the remaining two, phase independent characteristics are much harder to decipher. In PG&E's current GIS database, the phasing information was defaulted to A, AB or ABC depending on the type of circuit. An accuracy level of 70% was used as the performance target for all circuits to assess model suitability for full scale implementation.

Finally, Table 3 provides the results for the meter-to-transformer assignment for each of the five methods evaluated.

[3] Field data shared with all participants to allow them to build their model with 3-wire circuit.

[4] Field data shared with all participants to allow them to build their model with 4-wire circuit.

Table 3: Meter-to-Transformer Methods Results for All Four Circuits

Methods description	Accuracy with only algorithm	Accuracy with Algorithm and Field Verification	# changes needed but not predicted	# reassignment suggested
Voltage Clustered Dimension Reduced Cluster Assignment (VCDRCA)	90.20%	96.48%	312	319
Impedance Adjusted Geographic Match (IAGM)	93.76%	95.78%	354	39
PG&E's Method	91.75%	96.32%	322	214
Metadata Inconsistency Flagging (MIF)	93.43%	95.55%	368	29
Neighborhood Dimension Reduced Cluster Assignment (NDRCA)	93.71%	**97.61%**	**244**	**268**
GIS	**93.80%**	93.80%	375	N/A

None of the methods provided accuracy above the current accuracy level of 93.8%. The algorithms generally struggled to identify the majority of incorrect assignments, and even when incorrect assignments were identified, none of the algorithms effectively identified the correct re-assignments.

Data Requirements

The project evaluated the sensitivity of the EVC Method, the method with the best results in the POC demonstration, to different levels of voltage data frequency and precision available from meters on the tested circuits. The three scenarios evaluated included: (1) trial high resolution data – same condition as for the POC demonstration; (2) a medium case achievable through a PG&E-driven firmware upgrade[5] to SmartMeter™ system; and (3) using the lower resolution currently available across the territory, i.e. voltage reads provided in whole volt increments every hour. These results informed the scoping of the small-scale demonstration and will inform the data requirements for a larger rollout of network connectivity analysis. The phase prediction results achievable with the firmware upgrade and hourly reads have accuracy only slightly below the 5-minute reads used for the POC demonstration. However, moving from one decimal place to the whole volt increments provided by much of the current SmartMeter™ system causes severe accuracy performance degradations, as depicted in Table 4.

[5] During the summer of 2017, Pacific Gas and Electric (PG&E) upgraded the firmware of all the SmartMeter™ system population, enabling meters to send voltage readings with a single decimal place of resolution as long as the meter hardware is capable of recording at this resolution.

Table 4. EVC Method Results by Data Source

Data Source	Max Voltage Decimals[6]	Sampling Time	4-wire circuit 1	4-wire circuit 2	4-wire circuit 3	Total
High Resolution	1	5 minutes	94.5%	97.2%	94.7%	**95.7%**
Medium Resolution	1	60 minutes	94.4%	89.2%	87.1%	**89.9%**
Low Resolution	0	60 minutes	33.8%	48.9%	30.3%	**38.8%**

Through this analysis, it was decided for the small-scale demonstration to use the hourly voltage information that was starting to become available as a result of the firmware upgrade. Voltage information is now available at PG&E for the entire territory with 15min or 60min interval usage data, depending on the rate of the customer.

1.5 Challenges and Resolutions

Multi-Vendor and Vintage Metering Equipment

The developed methods were found to be sensitive to the challenges of using real-world production data. Roughly 60% of meters' hardware enables them to collect voltage data in decivolt resolution. Changes were made to meter configurations though a firmware upgrade to enable the collection of data at this higher resolution. However, due to hardware limitations, roughly 40% of meters cannot precisely measure voltage and are limited to increments of whole volts, hiding small differences between meters.

Circuit Configurations

Physical properties of the distribution system were found to impact algorithm performance. Phase connection configuration where phase to phase and phase to neutral connections were mixed within a single circuit posed a challenge to all methods. A key source of information, to improve the ability of algorithms to map the phases in these scenarios, will be improved transformer high side configuration records, which are now being collected in a field-asset inventory effort.

Asset Data Quality

Data related to some asset attributes were found to have some inconsistencies that impacted the various methods. For example, some transformers which were recorded as three-phase transformers in the database were single-phase and vice versa. As part of this project, a series of consistency checks were developed which may be used to identify potential inaccuracies. In addition, the current field asset inventory effort will improve this data quality.

[6] The higher resolution cases enable the more precise voltage reads from meters, but do not guarantee them. Roughly 60% of PG&E meters territory-wide have the hardware capability for high resolution voltage reads, while ~40% are hardware-limited to whole-volt increments.

Measurement Data Quality

Production systems were found to have missing data, out-of-range data, varied time zones, and null values, and required robust data processing prior to execution of methods. SCADA data was found to have data quality problems, because of SCADA or Historian configuration problems, and measurement equipment failures. To address this for future deployment of Phase ID methods, systematic root cause analysis of SCADA historian data quality problems will need to be undertaken to correct and maintain these measurement archives.

Measurement Data Sources

There are measurements which are not currently available in the PG&E system which would be extremely beneficial to future deployment of Phase ID methods:

- Phase to phase bus voltage measurements for three-wire delta circuits
- Per phase RMS current at meters
- Per phase Power Factor at meters

Field Validation

Validation data collection for the project was challenging due to the sensitivity of the hardware utilized to verify the connectivity model. A wireless tool was used to determine phase connectivity, functioning by comparing a precisely timestamped local reading to that of a remote reference point with known phasing.

To reduce noise in tool readings, hardware was procured to provide a remote reference point within the PG&E service territory. A second, more robust round of field measurement was required to finalize reference measurements for the first experimentation.

1.6 Recommendations

This project has shown that there is significant merit in developing automated Phase ID solutions based on the methods explored in this project, and potentially other methods, using a scaled approach that would support distribution engineering applications and reduce expected expenditures related to boots on the ground approaches. An algorithm-based approach could also run at intervals to ensure the system model remains up-to-date with minimal marginal costs. Below are key recommendations to other utilities considering implementing similar Phase ID solutions which will help to make their implementations successful:

Employ Meters with Decimal Voltage Precision

Voltage data with decimal precision significantly improved algorithm performance. Currently, most of PG&E's meters do not have decivolt precision. However, as existing meters are replaced, their replacements will provide decivolt precision and over time the performance of PG&E's algorithms will continue to improve. Utilities pursuing similar Phase ID approaches should assess the performance sensitivity to their methods of voltage precision and consider planning to replace end of life meters with meters capable of providing decimal precision.

Collect Voltage Data in at Least 15-Minute Intervals

In 2017, PG&E upgraded SmartMeter™ firmware to collect voltage data either hourly or every 15 minutes depending on the customer's rate. However, due to the high proportion of hourly voltage data, only 60-minute interval usage could be used to run the algorithms. The ability to collect AMI voltage data with a 15-min or 5-min frequency is expected to improve the overall accuracy of the algorithms, and utilities pursuing similar approaches should consider enabling at least 15-minute interval collection.

Collect High Resolution SCADA Data

High resolution of the data in PG&E's SCADA and historian system is required to enable future deployment of PG&E's Phase ID methods. Most of these data resolution considerations can be addressed by updating the configuration of SCADA and historian change band and compression settings. Prior to this project, SCADA data was mostly used by PG&E's engineers to evaluate peak consumption and not typically used for applications that required this greater granularity. Utilities pursuing similar approaches should ensure that their SCADA systems are configured to provide high resolution SCADA data.

Ensure Accurate GIS Data

During this project, it was clear that some of the GIS data did not always reflect all of the conditions in the field. However, within PG&E a multi-year field asset inventory effort is being conducted that should verify transformer connectivity (Phase to phase or phase to neutral, three-phase or single-phase) and operating voltage. With this information, the performance of the algorithms could improve, and utilities pursuing similar Phase ID approaches should ensure that it is accurate in their systems. This information would also allow for 4-wire circuits to separate meters that are connected phase to phase with meters that are connected phase to neutral.

None of the methods evaluated in this project combined line to neutral with phase to phase solutions. 4-wire circuits (21kV) have a combination of line to neutral and phase to phase configuration, and due to inaccuracies in information on transformer connectivity configuration, only phase to neutral connections are predicted. When the proportion of phase to phase configuration is higher, this inaccuracy will have a larger impact on the performance of algorithms using this assumption.

Automate Phasing Record Updates

As part of the implementation process for automated Phase ID, utilities should implement a process to allow for automated bulk updates to their phasing records with the output of Phase ID algorithms, while retaining all of the constraints required for maintaining electrical connectivity. Automating this process will reduce the manual workload associated with any periodic record updates after the initial effort.

Establish Target Accuracy Level for Phase ID Algorithms

One of the challenges in this project was to determine the sufficient level of accuracy for Phase ID algorithms. Though the automated algorithms developed in this project show a clear improvement over the defaulted values currently populated in PG&E's GIS system, it is not known what PG&E's minimum requirement of accuracy is to achieve a stable solution of the unbalanced three phase model in the ADMS. It is recommended that at the onset of their analytical Phase ID efforts, utilities perform sensitivity analyses on unbalanced load flow models with varying levels of phase designation accuracy to understand their accuracy requirements. For PG&E, where DER penetration is projected to be high, the required Phase ID accuracy may be higher than for other utilities with lower projected DER penetration.

Establish Confidence Prediction for Phase ID

Though it is anticipated that the accuracy of a utility's phase identification algorithms will improve as the various input data improvements listed above are addressed, there will continue to be some errors. It would be beneficial to have an indication of the probability of error in phase identification to support targeted field validation. Utilities pursuing automated Phase ID methods should develop and evaluate algorithms to incorporate this confidence prediction as part of their solutions.

1.7 Conclusion

This project developed and analyzed a variety of pre-commercial analytics solutions to automatically identify meter phase and meter-to-transformer connectivity for seven circuits. Although the performance of the meter-to-transformer methods explored was not sufficient for deployment, two phase identification methods did yield very promising results that would improve upon the accuracy of meter phasing records if deployed with the input data currently available. PG&E has begun work on additional improvements to input data, to further improve algorithm performance, and on scaling the results of this project to production through the implementation of an automated phase identification solution. Implementing an automated approach at scale will be significantly less expensive than the conventional boots on the ground alternative.

This project was PG&E’s first project to fully utilize voltage data from the SmartMeters™ combined with voltage data from SCADA and asset data from GIS. Leveraging the SCADA and AMI data was more challenging than expected. Issues with data quality were revealed, and multiple work streams are now being created to address these issues before solution implementation.

Accurate meter phase mapping will be required before the implementation of ADMS and DERMS platforms. Other utilities will also be facing similar issues when working on the implementation of an ADMS in their system as the level of data quality required for this type of system is very high. This project showed that data analytics may be a viable long-term industry solution. In California especially, where the grid structures are similar to PG&E’s, this work will give other utilities a chance to faster define their strategy to tackle this issue.

2 Introduction

This report documents the EPIC 2.14 – Phase ID project achievements, highlights key learnings from the project that have industry-wide value, and identifies future opportunities for PG&E to leverage this project.

The CPUC passed two decisions that established the basis for this demonstration program. The CPUC initially issued D. 11-12-035, *Decision Establishing Interim Research, Development and Demonstrations and Renewables Program Funding Level*[7], which established the EPIC on December 15, 2011. Subsequently, on May 24, 2012, the CPUC issued D. 12-05-037, *Phase 2 Decision Establishing Purposes and Governance for Electric Program Investment Charge and Establishing Funding Collections for 2013-2020*[8]*,* which authorized funding in the areas of applied research and development, technology demonstration and deployment (TD&D), and market facilitation. In this later decision, CPUC defined TD&D as “the installation and operation of pre-commercial technologies or strategies at a scale sufficiently large and in conditions sufficiently reflective of anticipated actual operating environments to enable appraisal of the operational and performance characteristics and the financial risks associated with a given technology.”[9]

The decision also required the EPIC Program Administrators[10] to submit Triennial Investment Plans to cover three-year funding cycles for 2012-2014, 2015-2017, and 2018-2020. On November 1, 2012, in A.12-11-003, PG&E filed its first triennial EPIC Application with the CPUC, requesting $49,328,000 including funding for 26 TD&D Projects. On November 14, 2013, in D.13-11-025, the CPUC approved PG&E’s EPIC plan, including $49,328,000 for this program category. On May 1, 2014, PG&E filed its second triennial investment plan for the period of 2015-2017 in the EPIC 2 Application (A.14-05-003). CPUC approved this plan in D.15-04-020 on April 15, 2015, including $51,080,200 for 31 TD&D projects.[11]

Pursuant to PG&E’s approved 2015-2017 EPIC triennial plan, PG&E initiated, planned and implemented the following project: EPIC 2.14 – Phase ID. Through the annual reporting process, PG&E kept CPUC staff and stakeholder informed on the progress of the project. The following is PG&E’s final report on this project.

---

7 http://docs.cpuc.ca.gov/PublishedDocs/WORD_PDF/FINAL_DECISION/156050.PDF.

8 http://docs.cpuc.ca.gov/PublishedDocs/WORD_PDF/FINAL_DECISION/167664.PDF.

9 Decision 12-05-037 pg. 37.

10 PG&E, San Diego Gas & Electric Company (SDG&E), Southern California Edison Company (SCE), and the California Energy Commission (CEC).

11 In the EPIC 2 Plan Application (A.14-05-003), PG&E originally proposed 30 projects. Per CPUC D.15-04-020 to include an assessment of the use and impact of EV energy flow capabilities, Project 2.03 was split into two projects, resulting in a total of 31 projects.

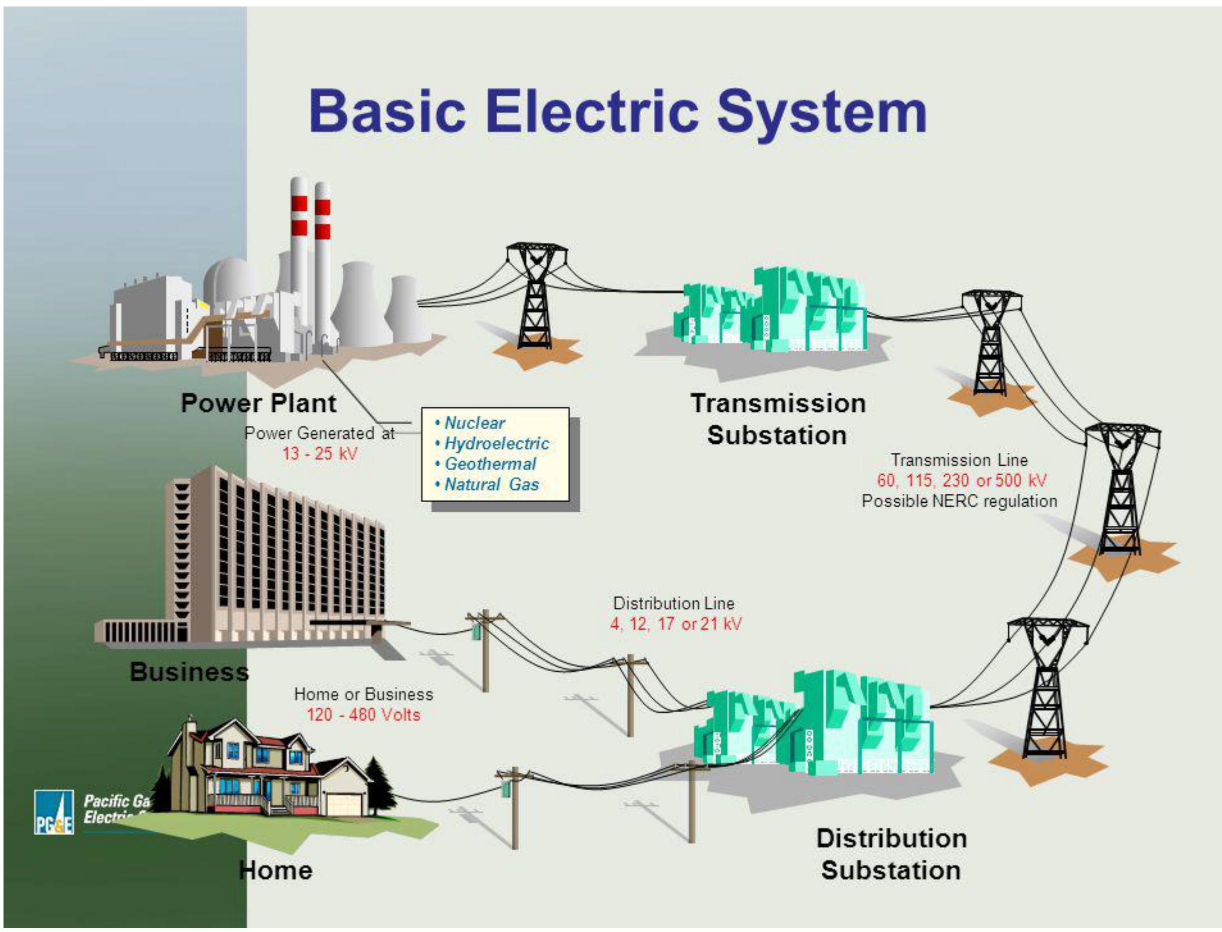
3 Project Summary

This project explored a wide range of methods for mapping meter phase and meter-to-transformer connectivity in PG&E's distribution network. This section summarizes the industry gap that the project addresses, as well as the project's objectives, the scope of work, and major tasks.

3.1 Issues Addressed

As described in PG&E's EPIC 2 application,[12] the distribution network model, represented in Figure 1, is central to many existing control systems, analyses, and processes.

Figure 1: Basic Electric System



Innovative utilities striving to modernize their distribution networks require improved visibility into both the physical state of their system and its load flow conditions. As the load characteristics of the distribution network evolve such as with the growth of DER, it is becoming more important to have

---

12 PG&E EPIC Triennial Plan (2015-2017), May 1, 2014, Attachment 1, p. 44, https://www.pge.com/includes/docs/pdfs/about/environment/pge/epic/Attachment1.pdf.

accurate and up to date information to be able actively manage the distribution system to ensure reliability for customers.

This project focused on the two following important steps towards modernization:

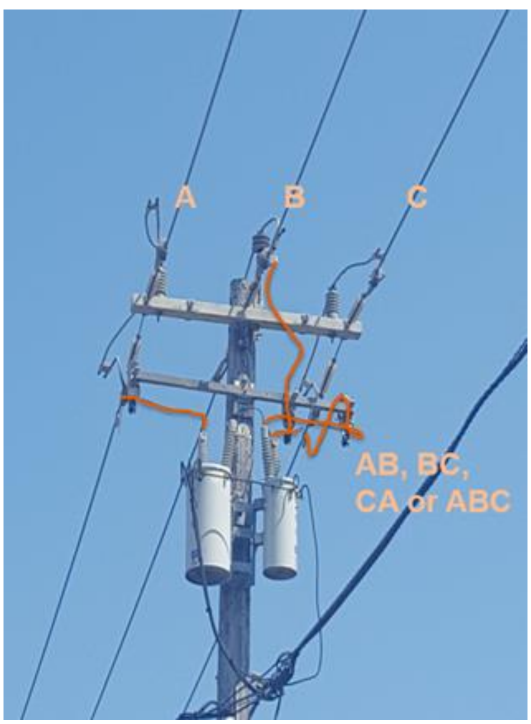
Phase Identification

PG&E has two main types of electric distribution circuits: 3-wire (typically 12kV) and 4-wire (typically 21kV) circuits. The primary configuration at PG&E is 3-wire. Depending on the configuration of the circuit, different phase connections are possible (see Table 5 below for more details).

Table 5: Possible Phasing Connection Based on the Type of Circuit

	4-wire circuit	3-wire circuit
single-phase tap with one neutral	A, B, or C	Not Applicable
two hot wires and no neutral	AB, BC, or AC	AB, BC, or AC
two hot wires with one neutral	A, B, C, AB, BC, or AC	Not Applicable
three hot wires with no neutral	AB, BC, AC, or ABC	AB, BC, AC, or ABC
three hot wires with one neutral	A, B, C, AB, BC, AC or ABC	Not Applicable

Figure 2: Phasing Options on a 3-Wire Circuit



In PG&E's current GIS database, the phasing information was defaulted to A, AB or ABC depending on the type of circuit.

The picture in was taken on one of PG&E's 3-wire (12kV) circuits during this project. There are three electrified (hot) wires, and in this case, the line on the right of the picture is connected to the C phase, the left one to the A phase, and the middle one to the B phase. Below, the transformer bank is connected to all those three phases; therefore, the secondary wire(s) below can have multiple phasing connectivity possibilities: AB, BC, CA or ABC. ABC are three phase transformers, and are connected to all 3 hot wires, and potentially the neutral if available. AB, BC and CA are phase to phase designations, where the transformer is connected between two of the hot lines. For a circuit with a neutral line (e.g. 4-wire), phase to neutral A, B or C designations are available, where a single hot line is connected to the neutral.

Accurate phasing information is increasingly necessary for the cost-effective operation of the modern distribution system, and is required for the safe and efficient integration of DER. Improved phase information is required for unbalanced power flow models and state estimation in an ADMS, and a DERMS platforms.

The normal means of improving a distribution model typically involve labor intensive field visits to each meter to see which phase they are connected to. Field surveys represent a significant expense, have a repeatability to cost factor (because return visits are necessary) and reliability issues because of their manual nature. Industry stakeholders frequently consider big data approaches as a tool that can drive innovation in enterprise utility operations and improve cost-effectiveness and efficiency across multiple processes that rely on the distribution model, including decision support for engineering, operations, and capital investment.

Meter-to-transformer connectivity

Although PG&E's databases contain information on which transformer each meter is connected to, records do not always reflect the field conditions. Accurate meter-to-transformer connectivity information is needed to ensure proper transformer loading levels.

To address these needs, this project explored a variety of pre-commercial analytics and hardware options to automatically map three-phase electrical power information and meter to transformer connectivity to improve the distribution network model.

3.2 Project Objectives

The primary goal of the EPIC 2.14 – Phase ID project was to determine the best solution for identifying phase and meter-to-transformer connectivity. In pursuit of this goal, the following objectives were established:

- Compare various analytical methods and other potential alternatives to determine best project approach for phase identification and meter-to-transformer mapping.
- Develop analytical algorithms that use SCADA, GIS, and SmartMeter™ data for automated phase identification and meter-to-transformer mapping.
- Determine market readiness for meter connectivity solutions in a real grid environment and compare with internal algorithms.
- Identify gaps for full-scale deployment and make recommendations on how the methods would be deployed for operational use.

3.3 Scope of Work and Project Tasks

The overall scope of this project was to evaluate, select, develop, and validate specific methods for cost-effectively identifying phase and meter-to-transformer connectivity. Listed below are brief descriptions of each major project task:

Literature Review

Evaluate methods and strategies for conducting phase identification and meter-to-transformer mapping. This includes the use of AMI data, Light Detection and Ranging (LiDAR) mapping technology, micro phasor measurement units (µPMUs), and hardware at the transformer to achieve this automated capability, as well as physical "boots-on-the-ground" methods.

Method Library

Develop a library of potential methods identified in the literature review, develop a robust set of criteria to assess their merits, score each method and select the subset to be explored further through POC demonstration.

Data Review

Identify the data associated with each of the selected methods, collect and review sample data. Document master data sources, data dictionary, and logical data models for each method.

Field Data Collection

Evaluate and select methods and tools for conducting the field validation. Collect field data to serve as the basis for model development and evaluation.

Proof-of-Concept Demonstration

Conduct preliminary evaluation of external methods for phase identification on three 4-wire circuits with the highest time resolution available for this project.

Small-Scale Demonstration

Continue evaluation of various phase identification methods and begin evaluating meter-to-transformer approaches. Apply analysis to more representative grid conditions by including 3-wire circuits, which are more representative of PG&E’s territory, and by using hourly AMI data, to see if models are viable with the current data coming from the SmartMeter™.

Develop Full Deployment Recommendations

Compile the final results of all the phase identification and meter-to-transformer approaches explored in the project, share results and lessons learned, collect feedback and develop full deployment recommendations.

4 Project Activities, Results, and Findings

4.1 Literature Review

As the first major project task, a literature review was performed to explore a wide range of candidate approaches for phase identification and meter-to-transformer mapping.

4.1.1 Technical Development and Methods

The literature review was structured to assess a landscape of possible methods to consider in the project, building on a perspective that a referenced method for phase identification may be a valid approach, and that there are advantages of some methods over others. The literature review was structured to show evidence of a method within four separate domains that either presented a method directly or supported the use of the method. These included:

- Resources from technical industry literature, meetings, and working groups
- Resources from patents
- Resources from conference proceedings
- Resources from analytical or other methods found from journals, papers, and vendors

The literature review identified the following methods to populate the Method Library:

1. *Correlation-based regression methods to contrast meter data, substation and line equipment data, and spatial data* – This method correlates voltage information for the utility's transformer and sensor system with voltage information from customer SmartMeter™.
2. *LiDAR mapping technology* – Arial-based LiDAR is used to render three-dimensional imagery and map connections to transformers in the distribution grid. Low flying drones, helicopters, and fixed-wing aircraft may be used with varying costs to deploy and operate. Results and error rate *may* vary based on difficulty with analysis of physical features.
3. *Photogrammetry drone mapping technology* – Arial based video is used to map *connections* to transformers in the distribution grid. Low flying drones may be used, with varying cost to deploy and operate. Results and error rate may vary based on difficulty with analysis.
4. *µPMUs and other hardware at the transformer and meter that could provide automated capability* – This method compares the voltage phase angle at different locations using measurements form a µPMU recorder. *Literature* suggests coupling with voltage correlation method.
5. *Discrete multi-variate modeling technology to identify correlation using information theory* - Information and graph theory is used to determine, to a highly specific degree, correlation between time series datasets. Notably, this has never been applied for use in Phase ID.
6. *Conservation of energy method* – A method that utilizes a combination of power measurements from SmartMeter™ data and transformers, adding up the energy in a combinatorial search to identify phasing.
7. Manual *field verification* - Using line-crew labor to physically verify connectivity.

4.1.2 Challenges

During the literature review effort, it was noted that many sources regard phase identification mapping as a potential outcome from the utilization of analytics, but few methods demonstrating the testing of such analytical approaches were available. A project requirement included a review of the technical approach detail where a method might be explored further for application and testing with the project-specific dataset.

4.1.3 Results and Observations

The project team conducted a review of approaches to phase identification and meter-to-transformer mapping, resulting in a broad set of candidates informed by technical literature, industry meetings, working group reports, patents, conference proceedings, journals, vendors, and white papers.

4.2 Method Library

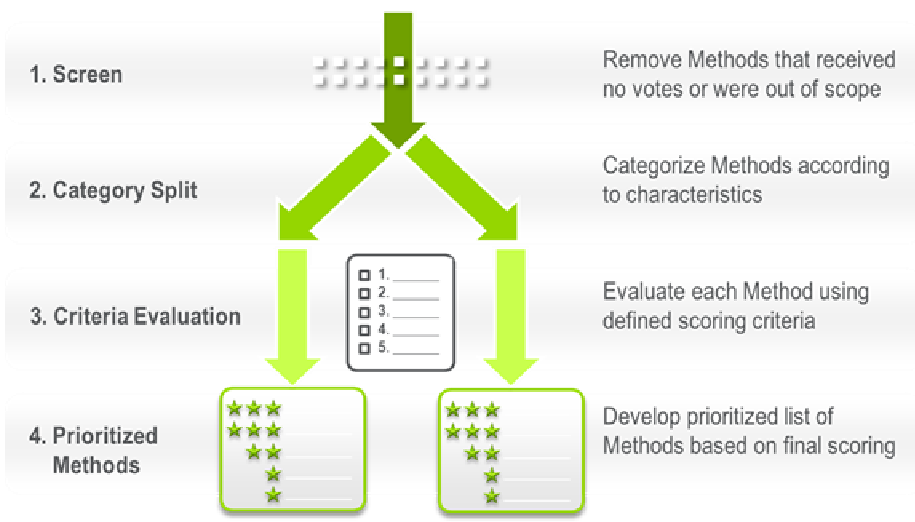
As the second major project task, the project developed a library of the candidate methods identified in the literature review, developed a robust set of criteria to assess their merits, scored each method and selected the subset to be explored further through the POC demonstration.

4.2.1 Technical Development and Methods

The prioritization approach considered criteria developed by the PG&E project team and reflected industry best practices for considering technology, people, financial, and process-related aspects based on available information.

Value assessment of methods was performed using the prioritization process steps outlined in Figure 3 by applying scoring criteria, evaluating results within subject matter experts, and prioritizing methods for selection to move to the POC phase.

Figure 3: Prioritization Process



4.2.1.2 Screening

Prioritization analysis was performed using the specific screening criteria illustrated in Table 6.

Table 6. Criteria Overview

Category	Sub-categories (as applicable)	Description	Scoring Criteria Scale
Data Requirements	• Physical • Analytical	Presence, accessibility of data in PG&E's system	Low to High
Data Quality	N/A	Performance relative to data accuracy, noise, etc.	Low to High
Cost	• Equipment • Training • Labor • Repeatability	Cost to implement, repeat	Millions (USD)
Method Duration	N/A	Duration to complete method	Days to Years
Implementation Characteristics	• Qualifications requirements • Level of Effort expected	• Staff qualifications • Staff quantity	• Skill Level, Low to High • Resource Count, Number of Resources

4.2.1.2 Weighting of the categories

To complete the analysis framework approach, each category was weighted to allow a normalized comparison of category fitness. The evaluation category weights are shown in Table 7.

Table 7. Evaluation Category Weighting

Assessment Category	Assigned Weight
Data Requirements	31%
Data Quality	19%
Cost	28%
Method Duration	11%
Implementation Characteristics	11%

4.2.1.3 Selection

To set a threshold limit for which methods would be appropriate for the POC demonstration, the project team re-evaluated the method evaluation scoring in relation to available resources and remaining time for the existing approved project. This review indicated that a threshold setting of 3.5 was appropriate. Only projects with an overall score of 3.5 or more were considered in the POC demonstration

4.2.2 Challenges

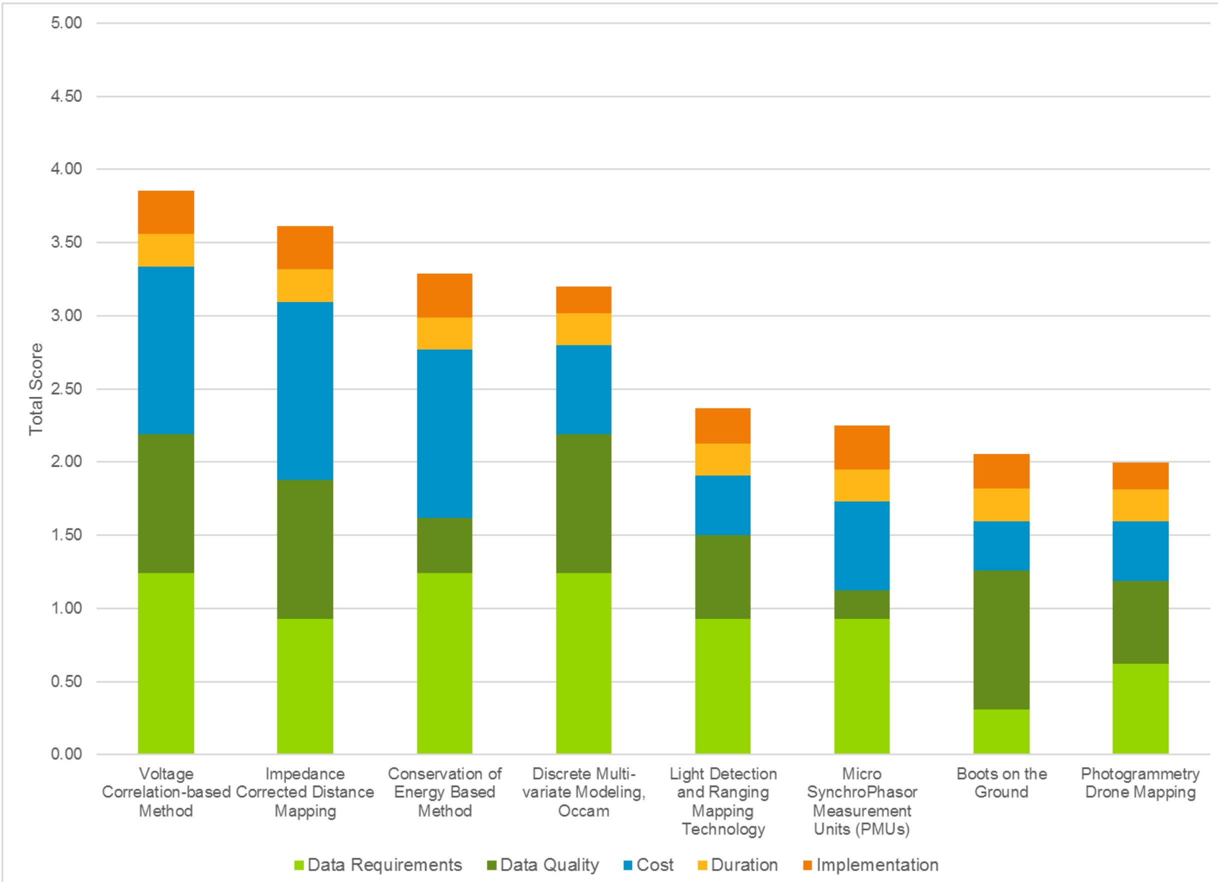
The scoring process had a limitation in that it was only a snapshot-in-time of the available data sources and data quality. If data quality was improved or if key data sources were to be made readily available within the existing enterprise computer environment in the future, scoring of the methods may have changed. The scoring process, therefore, was blind to potential forecasted improvements in data sources.

Additionally, cost estimates did not take into considerations how other data source development efforts could have been leveraged to share resource and cost burdens, specifically, in how it relates to making data sources available. Although it would generally be considered a potential improvement in scoring, changes in data access or quality could also have a negative impact on the scoring.

4.2.3 Results and Observations

This task prioritized the candidate approaches produced by the literature review, to inform the technical approach to the project and select methods for demonstration and evaluation. The final attribution of value based on the evaluation criteria is illustrated in Figure **4**.

Figure 4: Attribution of Evaluation Criteria Value



The screening results illustrated in Table 8 were thus explored further in the POC demonstration. These were the Voltage Correlation-based Method (found in two academic methods) and the Impedance Corrected Distance Mapping Method (developed by an external vendor).

Table 8. Method Screening Results

Total Score	Method Name	Description	Potential Method Sources
3.8575	Voltage Correlation-based Method	Correlates voltage information from the utility's transformer and sensor system with voltage information from customer SmartMeter™.	Correlation-Based Method for Phase ID in a Three-Phase LV Distribution Network (Pezeshki & Wolfs, 2012) Advanced Metering for Phase ID, Transformer Identification, and Secondary Modeling (Short, 2013) Voltage Correlations in Smart Meter Data (Mitra, et al., 2015)
3.615	Impedance Corrected Distance Mapping	Compares both voltage, circuit diagram and physical location to compute phase	Vendor Method

4.3 Data Review

In this task, the data needed for each of the selected methods was studied.

4.3.1 Technical Development and Methods

The approach to this task included:

- Reviewing data and formalizing data structures
    - Reviewing PG&E AMI, GIS, and SCADA data samples for all the methods that were evaluated
- Defining the data dictionary to document all data sources for all methods
- Setting forth processes to verify and validate the various methods studied
    - A dataset was needed to evaluate the development of the system, called training data,
    - Another separate dataset was needed to evaluate the success of the methods, called testing data

4.3.1.1 Input to all methods

- Meter Data Profiles
    - Time Resolution
        - POC: High-resolution trial with data reads every 5 minutes
        - Small Scale Demonstration: Mixture of hourly and 15-minute data
    - Time Duration
        - POC: One month of data was used, in a period spanning November and December 2016.

        - Small Scale Demonstration: Data was used from September 2017 through March 2018. However, SCADA data quality was improved after November 15th, so participants were encouraged to use data after that period.
    - Data Resolution: For some meters, decivolt precision was available, but in general, meter hardware was unable to provide reads more precise than single volt increments.
        - POC: For the circuits tested, between 64% and 71% of the meters had decivolt precision
        - Small Scale Demonstration: For the additional circuits tested during this demonstration, between 41% and 51% of the meters had decivolt precision
- SCADA Data Profiles
    - Voltage is recorded at the substation banks and the current for each circuit is recorded at circuit breakers. The values are pulled from SCADA Historian and interpolated in 5 minute measurements.
- Meter Metadata
    - The meter latitude and longitude coordinates were originally taken as a snapshot from the customer care and billing database
    - Utilized existing relationship between meters and transformers
- GIS Network Map
    - The PG&E network map is imported from a GIS export. Topology is provided down to the primary level.
- Outage Records
    - Records of equipment outages were exported from the PG&E outage log

4.3.1.2 Outputs

- Meter Phase Assignments
- Transformer Phase Assignments
- Meter-to-transformer Assignments[13]

4.3.1.3 Data Preparation

All methods required some level of data processing and quality control. For example, for the POC demonstration, 2.1% of the active meters in the study area had missing or incomplete voltage information in the study period. This could be caused by legacy meters, malfunctioning meters, service conditions, replaced meters, or other issues.

The following data quality steps were taken by all methods unless otherwise noted:

- Meter Voltage Profiles
    - Standardize timestamps to known time-zone
    - Filter meter voltage measurement periods with 0 values
    - In situations where no value is found for a meter channel, this will manifest in the database as an integer overflow value of 32767. These values are filtered out.
    - For clustering based methods, the different voltage levels recorded at the meters will be categorized as separate clusters, though they may be on the same phase. To avoid

[13] Small scale demonstration only.

this, raw voltage reads are scaled to a per unit common voltage level of 120 V, to allow comparison between different voltage levels.

- SCADA Voltage Profiles
  - Standardize timestamps to known time-zone
  - Filter 0 value Voltage time periods
  - Scaling to 120 V common voltage level
  - For the small-scale demonstration, automated data filtering to filter periods of stale data or periods where SCADA quality flags were active was implemented.

4.3.2 Challenges

The Data Review presented several challenges associated with providing the input data needed to run the project's various methods:

- GIS Data Quality
  - Some meters were assigned to the wrong transformer
  - Some transformers were assigned to the wrong circuit. Several transformers were moved from one circuit to another, which caused errors in many of the algorithms.
  - Errors in connectivity. E.g., Single phase lateral in GIS may in the field be composed of two separate single phase laterals.
- Polyphase Meters
  - The listed phase assignments for three-phase meters were not assumed consistent from meter to meter. That is, each three-phase meter reported voltage for Phase A, Phase B, and Phase C, but there was no guarantee that a Phase A reading on one meter is taken from the same phase as a Phase A reading on another.
- Meter Metadata
  - The latitude/longitude coordinates found in the customer database are not always aligned with the calculated latitude/longitude based on the customer service address. After consultations with the PG&E GIS team, the difference between the two options was not significant, with only around a dozen non-trivially impacted meters per circuit.
- SCADA Data Quality
  - Due to compression settings on the SCADA and the historian and other configuration problems, issues with measurement resolution were encountered. Work was required to apply quality filters to the SCADA data to ensure that data was not interpolated inappropriately. Additionally, one of the substations had a phase voltage reading become very inaccurate as the measurement equipment began to fail. Getting those issues fixed was challenging because different teams are responsible for SCADA and the historian.
  - Substation voltages are only recorded in the historian as line to neutral voltages. Phase to phase substation voltage needed to be calculated.
  - In some substations, the phase labels of the currents may be incorrect in SCADA, and care must be taken to address this for methods which use this data.
- Mapping AMI with GIS data
  - Sometimes meter metadata and transformer information did not align, making it look like the meter could not be connected to that transformer, but it was observed that the information stored in the database was not always accurate. Therefore, the transformer connectivity details were not used.

4.3.3 Results and Observations

Data review was a challenging task. Rarely before this project have AMI, SCADA and GIS data been combined to get a full understanding of the grid connectivity. Moreover, effort was required to understand the data available, the reliability of the data available and decide which data to use. Some of the data PG&E was hoping to use was not available, e.g. secondary network or phase to phase voltage measurement at the substation. Throughout the project, PG&E's team worked closely with all the participants of this project to make sure that the right data was provided when needed and all data quality issues were fixed as soon as possible.

4.4 Field Data Collection

This task collected the field data that would be needed for method development and validation. Field data collection was done multiple times: Twice for the three 4-wire circuits used for the POC demonstration, and one time for the additional 4-wire circuit and the three 3-wire circuits used for the small-scale demonstration. A summary of the circuits used in the two demonstrations is provided in Table 9.

Table 9: Circuit Configuration Description

Demonstration	Meter-to-transformer	Phase Identification	Configuration	Voltage Level	Circuit
Proof-of-concept & Small-Scale Demonstrations	Not field verified	Field verified	4 Wire - Wye	21 kV	4-wire circuit 1
					4-wire circuit 2
					4-wire circuit 3
Small-Scale Demonstration	Field verified				4-wire circuit 4
			3 Wire - Delta	12 kV	3-wire circuit 1
					3-wire circuit 2
					3-wire circuit 3

4.4.1 Technical Development and Methods

Currently, to check the phase of a line, PG&E requires line-crew staff with special certification to use a phase identification hot stick tool. This works well to check a few data points, but availability was limited and labor cost was prohibitive for the scope of field data collection required for this project. Therefore, another tool that did not require contact with energized lines was selected.

For the POC demonstration, a key component was choosing the specific field locations for the phase check with the selected tool. To minimize the number of locations to check in the field and reduce the cost, the team reviewed the maps of the selected circuits and partitioned them into groups where all transformers in each group would necessarily be the same phase, thus requiring only one to be measured from each group. For example, all the transformers on a section of the circuit with only one hot and one neutral wire were assumed to be the same phase. The approach was divided by physical characteristics:

- **Single-Phase Line with a Neutral**
  The project utilized PG&E GIS data to identify the single-phase taps for each demonstration circuit. For each single-phase tap, the source side device (SSD) nearest to the tap and all downstream SSDs were identified. For each transformer on a tap, the SSD nearest to the tap was selected as the field location to check for the phase, given that all transformers on that single-phase tap will have the same phase.
- **Two-Phase Lines**
  As in the single-phase line approach, the two-phase taps and all associated SSDs were identified. Two-phase transformers (21 kV phase to phase) were identified and grouped with the most upstream SSD; however, it was noted that single-phase transformers (12 kV line to neutral) on two-phase lines need to be checked individually. Open-delta (or open-wye) transformers were also identified, as well as each transformer supplying single-phase customers marked for separate phasing.
- **Three-Phase Lines**
  Single-phase, two-phase, and open-delta transformers on three-phase lines required a phase check, while other three-phase transformers would not need to be checked.

For the small-scale demonstration, as the meter-to-transformer associations were also verified, the phasing was collected at each meter for those connected on overhead transformers. The team in the field had the GIS data for the underground primary and secondary on tablet PCs for reference, and assumed that data to be correct.

Depending on the type of circuit and number of energized lines and presence of a neutral line, phasing connectivity can vary. Table 10 summarizes all the phasing possibilities by circuit type.

Table 10: Phasing Possibilities by Circuit Type

	4-wire circuit	3-wire circuit
single-phase tap with one neutral	A, B, or C	Not Applicable
two hot wires and no neutral	AB, BC, or AC	AB, BC, or AC
two hot wires with one neutral	A, B, C, AB, BC, or AC	Not Applicable
three hot wires with no neutral	AB, BC, AC, or ABC	AB, BC, AC, or ABC
three hot wires with one neutral	A, B, C, AB, BC, AC or ABC	Not Applicable

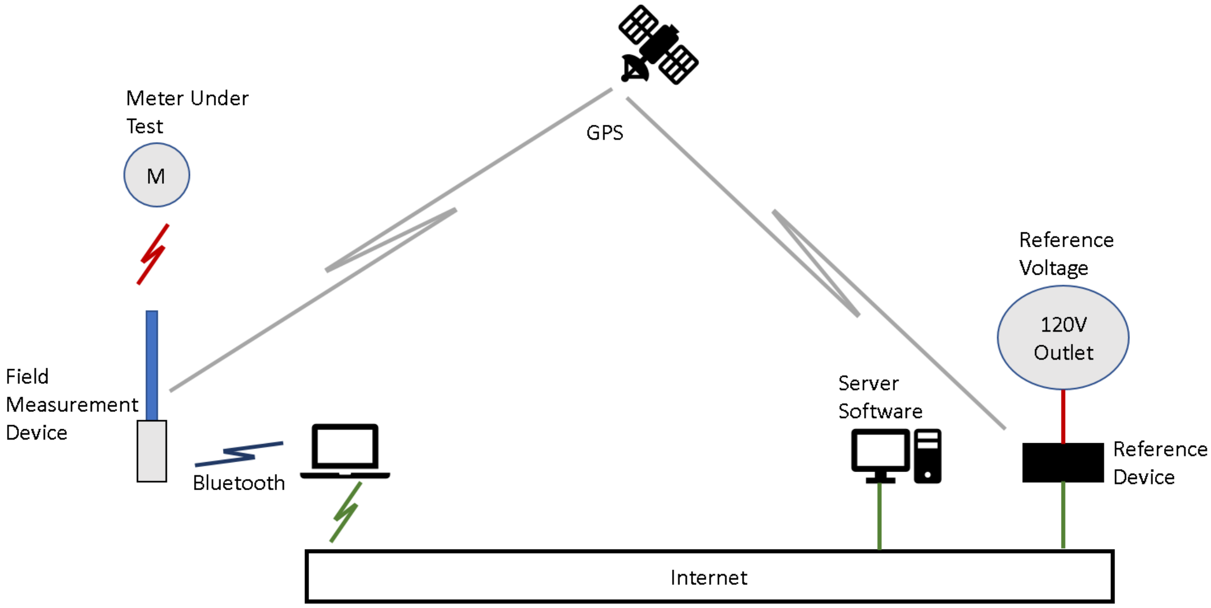
Wireless Phasing Tool Description

The Phase Identification Wireless Tool was selected as the tool for field phase verification because it does not directly interact with an energized wire, and is useable by a meter technician. The current process requires direct contact with the wire and is performed by a Trouble-man. This current process is depicted in Figure **5** below.

Figure 5**: Phase Identification Tracker Usually Used by PG&E Line-Crew Staff**

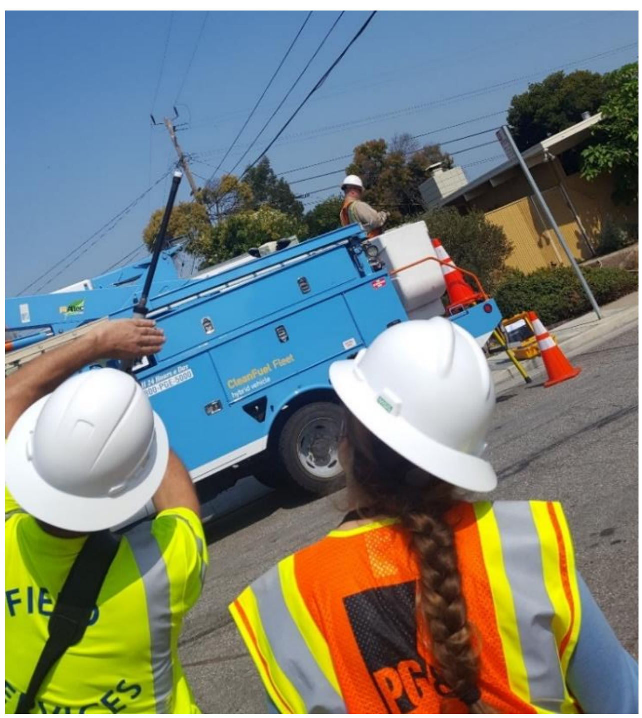
The wireless tool has 2 main components. The first component is the base station, which is composed of the server software and the reference device. The base station has a device that is internet connected, has GPS, and is monitoring a 120V AC line. The base station also consists of server software running on a computer. The second component is the field unit. The field unit has an electric field measurement device with GPS and Bluetooth. The field unit also has a laptop running the software that communicates with the base station. When a field measurement is taken, the electric field measurement device communicates with the laptop via Bluetooth. The laptop then communicates with the server software over the internet. The server software compares the phasing of the field and stationary measurements using the GPS time synchronization. Figure **6** below depicts the setup.

Figure 6: Phase ID Tool Diagram



Prior to taking a measurement, the field device needs to be calibrated (Figure **7**). For the calibration, PG&E lineman was needed to identify the phase, using the standard hot stick tool. With the wireless tool, a measurement of that phase is taken and the known phase is recorded in the laptop software to calibrate the wireless tool. When a measurement is taken, 2 samples are taken. A phase error for each measurement sample is reported by the wireless tool and should be very close to each other.

Figure 7: Calibration of the Field Device for Phase Identification



Proof-of-Concept Demonstration Field Data Collection

A PG&E engineer worked with a troubleman to test and calibrate the tool at the substation, checking the results against those of a tool that does apply directly to the wire and the labeled phase at the substation. Two single-phase points for the POC demonstration were checked and matched across both tools. Due to time constraints, a line-to-line point was not checked.

For the POC demonstration, each circuit was segmented into sections such that all transformers in each section would be the same phase, such as along a single-phase lateral. Segmentation was performed based on the data contained in GIS. One transformer was selected from each of these sections, and the meter technicians were given a list of service points associated with each of these transformers to verify. The meter technicians went down the list of service points for each transformer until they could gain access to the customer meter and get a clear reading of the phasing of that meter. This was assumed to be the phasing of all customers on that transformer and any other customers in that electrical segment.

The issues with the first attempt at field verification were discovered due to measurements reported as phase to phase from meters that were on single-phase laterals per the GIS database. An extended comparison of the phasing received from the field crew to an extract from PG&E's GIS database confirmed that a significant number of transformers were assigned AB, BC, and CA phases when the phase designation field was single-phase of both the transformer object and the attached line objects. Descriptions of the potential causes of incorrect readings follow.

4.4.1.1. Small-Scale Demonstration Field Data Collection

Whereas in the POC demonstration only the phasing was measured, in the small-scale demonstration, both phasing and meter-to-transformer measurements were made. The procedure for doing this involved calibrating and validating the tool with a lineman certified to use a hot-stick phase tester. Because the meter-to-transformer connectivity was being validated, all meter phases were recorded and measurements were taken at the meter. Since there is a 30-degree phase shift between the primary voltage and the secondary voltage on delta-to-wye connected transformers the tolerance range is reduced to 30 degrees when comparing the hot-stick tester results to the measurements at the meter.

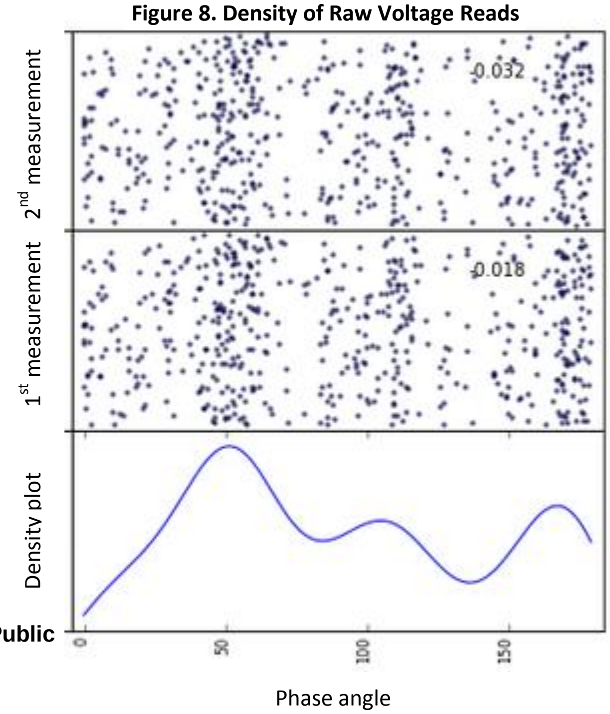
4.4.2 Challenges

The project had constrained line-crew availability and budget. Line-crew effort was considered cost-prohibitive for the verification of the field data. Following development of the training and testing dataset using the meter-centric tool, it was discovered that the tool was highly sensitive to calibration, which created problems the first time the team went on the field to collect the data.

Wireless Phasing Tool Challenge 1: Phase Offset Not Calibrated Correctly

The initial calibration used both the wireless and wired tool, and ensured they both gave the same reads on two single-phase laterals.

Figure 8. Density of Raw Voltage Reads

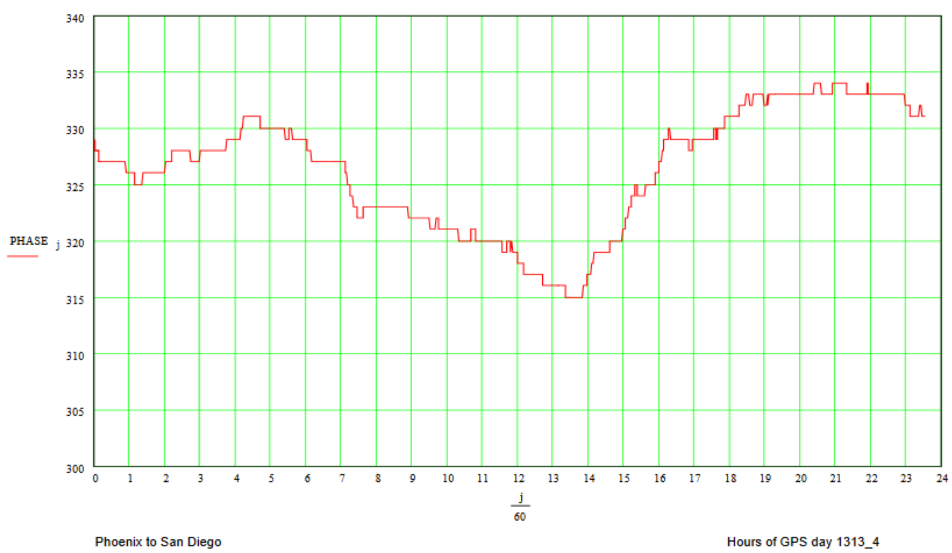


However, a noisy calculation of this phase offset could result in all the previous field readings being off by the same amount. One would expect raw field readings to cluster into groups, just uniformly translated by the offset error. In Figure 9, the raw phase angle of each reading is shown on the x-axis, and the points are spread uniformly across the y-axis for display purposes. The first and second panels from the top are the first and second sets of measurements taken, and the third panel is a density plot. While there were areas of concentrated readings 60 degrees apart as expected, many readings were also scattered between the peaks, not the +/- 5 degrees discrepancy that was expected by the tool manufacturer.

Wireless Phasing Tool Challenge 2: Phase Offset Not Static

This base station must be on the same electrical grid as the wireless tool (Western Interconnect, Texas, Eastern Interconnect), but it may also be necessary to be in the same service territory. For the initial reads, the base station was in Phoenix, Arizona. The tool manufacturer published a white paper that describes potential movement in this phase offset if there is a significant transmission flow between the two locations. In Figure 10 the phase offset between base stations in San Diego and Arizona could swing by up to 20 degrees within a 24-hour period. For the second round of analysis, a local base station was procured and used to control for this behavior.

Figure 10. Phase Offset Between San Diego and Phoenix



Wireless Phasing Tool Challenge 3: Electrical Interference on Wireless Reads

As the wireless tool works based on a capacitive connection, extraneous electrical fields can induce faulty measurements. This may have been the case at the substation with 115kV transmission lines overpowering the distribution voltages. The physical orientation of distribution wires on the 4-wire circuits (triangular rather than three horizontally) also makes it difficult to get clear readings on individual phases. It is recommended to find single-phase laterals without secondary wires to get the cleanest reads for calibration, and to directly measure at the meter to get phasing.

Overcoming Wireless Phasing Tool Challenges

To address the challenges, a subcontractor with experience using the wireless tool was selected to perform the second attempt for the POC demonstration of field verification as well as for the small-scale demonstration. The improved methodology aimed at addressing the challenges used multiple readings at each measurement site to control for noise in the measurements, and relied on the distribution system expertise of the field resources to flag questionable results for additional measurement. While marketed as a single button, point-and-shoot tool, the subject matter expertise required to obtain accurate readings fell somewhere between a meter technician and a distribution line technician. A key component of the improved method was establishing an accurate secondary offset value stored in the software on the field tablet PC, which resulted in more confidence in the accuracy of the phase readings.

4.4.3 Results and Observations

The field verification results summary is presented in Table 11 and shown in Figure 11. As expected, three wire systems (circuits 1105, 1109 and 1101) have only phase to phase designations. The methods developed in the POC and small-scale demonstrations were judged on the percentage of meters where the prediction matched field verification, excluding meters on 3 phase transformers.

There were three potential error categories recorded in the field survey:

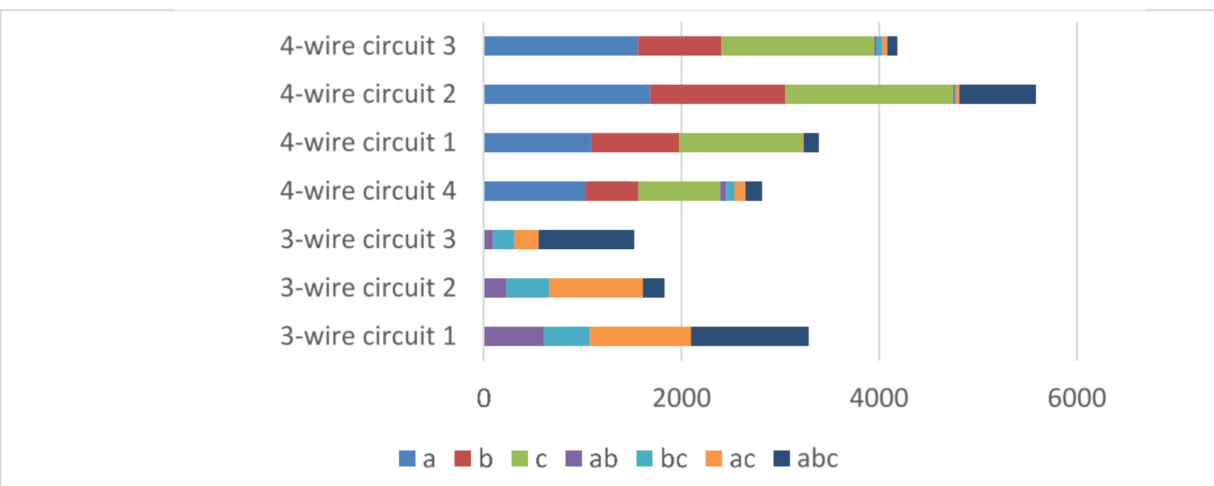
1 Continued, unexpected, or questionable readings from the wireless tool
2 Unable to take reading (no access to meters, transformer only feeds street lights, etc.)
3 Does not exist (transformer removed due to recent construction)

For the assessment of method accuracy, neither meters associated with these three error categories nor the meters associated with ABC were included. Those transformers were field validated to be three-phase on the secondary side, but the phasing of downstream meters was unable to be determined due to underground wires, locked meter rooms, etc.

Table 11: Validated Transformer Phase Designation of Meter by circuit

Phase Designation	A	B	C	AB	BC	AC	ABC	Does Not Exist	Questionable Reading	Unable to Access
3-wire circuit 1				609	464	1024	1190			
3-wire circuit 2				230	432	950	218			2
3-wire circuit 3				93	219	244	968			
4-wire circuit 4	1027	538	827	62	83	110	168			1
4-wire circuit 1	1088	889	1248			11	154	1	1	345
4-wire circuit 2	1685	1364	1699	12	18	32	774	23	3	224
4-wire circuit 3	1563	841	1548	19	59	52	101	2		426

Figure 11: Validated Transformer Phase Designation of Meter by Circuit



4.5 Proof-of-Concept Demonstration

For this demonstration, three 4-wire circuits from one substation were selected for phase evaluation, as the field verification could be used to support other EPIC projects, including EPIC 2.02 *Pilot Distributed Energy Management Systems (DERMS)*, in the study area. In addition, a trial for additional meter capabilities was purchased to collect 5 minute interval voltage data from the SmartMeters™ on those circuits.

The field-verified dataset was split into a training set of circuits used in model development, and a test set of circuits that was reserved until the final evaluation of the methods. All the method participants were asked to use 4-wire circuit 1 as the training set and the 4-wire circuit 2 and 4-wire circuit 3 as the test set for their methods, but this was not enforced as all field data was shared directly with all participants.

In the POC demonstration, four external approaches were evaluated against field-verified data:

- **Ensemble Voltage Cluster (EVC)**: Approach based upon an adaptation of methods described in "Voltage Correlations in Smart Meter Data," (Mitra, et al., 2015). In addition to this core approach, a weighted ensemble was implemented.
- **Regression Corrected Correlation (RCC)**: Implementation of methods described in "Advanced Metering for Phase Identification, Transformer Identification, and Secondary Modeling," (Short, 2013)
- **Reduced Constrained k-Means (RCKM):** External academic method described in "Phase Identification in Electric Power Distribution Systems by Clustering of Smart Meter Data" (Wang, et al., 2016).
- **Feature Reduced Cluster (FRC)**: The methods used by this participant were not shared in detail, but utilized feature reduction and clustering.

4.5.1 Ensemble Voltage Cluster (EVC)

4.5.1.1 Technical Development and Methods

Approach Overview

This method leveraged the algorithms described in the paper, *Voltage Correlations in Smart Meter Data*, by R. Mitra, R. Kota, S. Bandyopadhyay, and V. Arya (Mitra, et al., 2015). The method was adapted to the conditions of the study area and parameters like the distance metrics were selected through analysis of the 4-wire circuit 1 training dataset. In addition, an ensemble prediction across subsets of time was developed to generate a more accurate and stable prediction.

Algorithm Approach

First, the method performs some simple transformations of the meter voltage profiles in the study area. The paper proposes a raw transformation (no change), a delta (compute the difference between voltages in each time step), and a binary (compute whether voltage goes up or down each time step). During the development and calibration on 4-wire circuit 1, the delta (differential) transformation of the data performed the best compared to the verification dataset and was used for all other circuits as well.

Second, the method performs an ensemble k-means clustering on the transformations of daily voltage profiles. This is an iterative process that starts with cluster centroids based on SCADA data, and then iteratively assigns meters to each of those clusters, trying to minimize the sum of the distances between each centroid and the meters in that cluster. Several distance metrics were evaluated using the 4-wire circuit 1 circuit, and Euclidean was selected based on superior performance.

Third, a similarity matrix is constructed with the association between two meters equal to the percentage of samples that predicted those meters to be in the same cluster. A spectral clustering is used on this similarity matrix to determine final meter predictions. This differs from the Mitra approach and was implemented to address a lack of consistency in the predictions when run over different time periods.

Finally, the set of meters associated with each transformer in the given mapping are assigned the final prediction that occurs most often among that group of meters on that transformer.

Assumptions

- Voltage Data Cleaning
  - All voltage readings were normalized to 120 V for comparison purposes.
  - Any voltage readings greater than 10% from nominal were flagged as possibly erroneous and removed from the calculation.
  - Three phase meters were removed from the dataset.
  - Due to clock drift, some of the voltage reads took a small number of seconds away from the requested 5-minute interval. These reads are all assigned to that requested 5-minute interval.
- Clustering
  - Based on the results of data transform testing, the meter voltages were transformed by calculating the difference between each timestamp.
  - Euclidean distance was used to determine the similarity between voltage profiles.

Testing

- Data Transforms
  - Raw (no change), delta (difference between consecutive timestamps), and binary (whether change between timestamps was positive or negative) were tested. In contrast to the Mitra paper, delta significantly outperformed the others. The delta transform was used in subsequent analyses after initial comparisons to the raw and binary transforms.
- Distance
  - Euclidean distance, correlation, and cosine distance were tested as the measurement of similarity between voltage profiles. Euclidean distance was shown to produce the best results in the training set so it was used for subsequent analyses.
- Separation of Phase to phase and Phase to neutral Meters
  - As discussed in more detail in the following Challenges section, the proportion of phase to phase and phase to neutral meters in the study area influenced results. A test was performed separating the two configurations prior to analysis using field verified phasing, with improved results compared to the combined set, especially for the smaller phase to phase group. This result is promising for future development of the method on circuits with significantly mixed configurations. This method was not implemented in any of the methods used for this study because of the lack of reliable information about the transformer connectivity configuration, which would be required to determine whether connections were phase to phase or not.

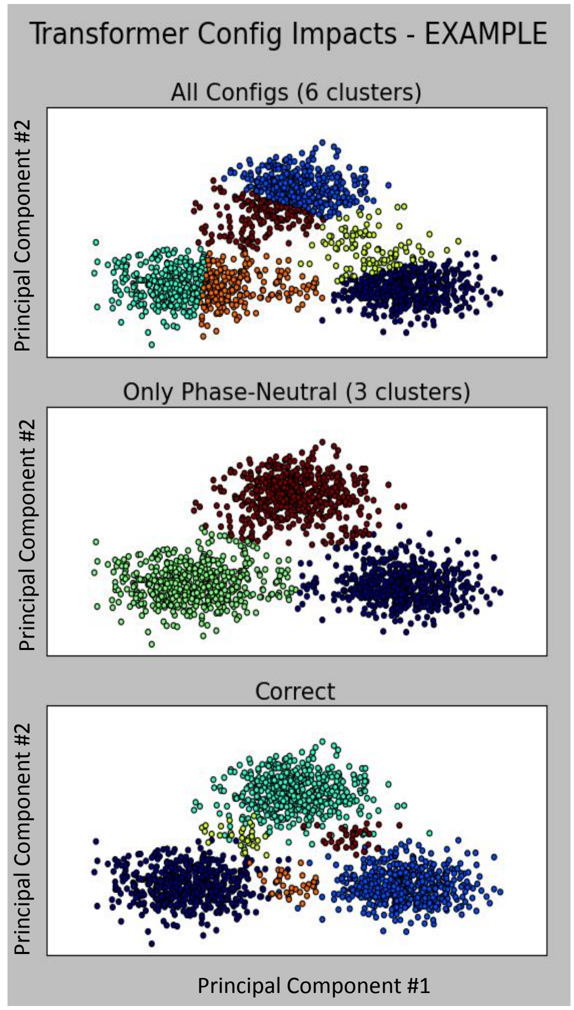
4.5.1.2 Challenges

Distribution Transformer Configuration

Phase to Phase transformers

Unsupervised clustering algorithms like those implemented in this method are found to be sensitive to the number of anticipated clusters (Phase values; A, B, C, AB, BC, AC, ABC), as well as to the size of each of those clusters. The method has difficulty when both phase to neutral and phase to phase connectivity are considered at the same time, especially considering that in general for 4-wire circuits, most of the connections are phase to neutral (A, B or C). Figure 12 provides a simulated example of the effects of small groups of intermediate data on the clustering results. When looking for all six phase assignments (all Configs part of the figure), the first panel displays the larger groups incorrectly cut in half. In the second panel, assuming there are only three total groups, the smaller three are ignored and consumed by the large ones. The third panel provides the contrived ground truth results, where there are three large groups and three smaller ones filling in the spaces between.

Figure 12. Example Transformer Data Groupings



The third panel is similar to the arrangement on 4-wire circuits, with large numbers of phase to neutral transformers and a much smaller number of phase to phase transformers that can cloud or alter the results. For this reason, it is recommended to draw out samples from the reference data where differentiation between meter types may be known to improve results. Using existing GIS data to determine transformer configuration to consider phase to phase and phase to neutral transformers separately has the potential to improve the method accuracy (see next section for more details).

Inaccurate Transformer Configuration

As mentioned above, to be able to determine whether a meter is connected phase to phase or phase to neutral, knowing the high side transformer connectivity will be required with an accurate meter-to-transformer connectivity. The transformer connectivity includes three-phase or single-phase as well as phase to neutral or phase to phase connectivity. However, with the field data, it was clear that not all data coming from GIS was accurate. Alternate methods that were attempted, such as assuming transformers with a three-phase meter on them should be three-phase transformers also did not align with data collected from the field, due to apparent meter-to-transformer miss-assignment. These inconsistencies cause difficulty in performing operations such as separating clustering steps by phase to phase or phase to neutral, and in determining whether transformers are single phase or three-phase. Because of this issue, no phase to phase predictions are made for the 4-wire circuit, and some single-phase transformers are inaccurately classified as three-phase transformers and vice versa.

Phase to Phase Substation Voltages

As described previously, no clusters were performed on phase to phase meters; therefore this challenge was not addressed by this method.

Solution Stability

Small Clusters in Ensemble k-Means

The k-means algorithm minimizes the sum of the distance from cluster centroids to the elements of each cluster. If there is an unexpectedly similar group of meters in a sample, they can be assigned their own cluster, merging two other clusters that should have been separate. From the validation testing, this can produce poor results for a significant number of samples in the ensemble, but the spectral clustering across the ensemble appears to be resilient against it and determines the correct assignments from the well-separated sample.

Voltage Outliers

When present, outlier readings such as zero voltage can be so far from the other voltages that they are assigned their own cluster. Any readings greater than 10% away from nominal were flagged as possibly erroneous and removed from the calculation. The method can still make predictions for these outlier meters based on values from other days in the ensemble or from other meters on the same transformer.

Low Resolution Meters

When meters were restricted to integer volt readings, groups of them could look artificially similar and could be perceived as closer together than they should be. This especially affected the delta and binary transformations. Through initial analysis of the 4-wire circuit 1 data, this may also be corrected for by the days in the ensemble where there are significant enough differences.

Meter Voltage Measurement Interval

Because this method uses the delta transformation, the voltage interval can have a significant impact on the accuracy. This relationship is explored further in Section 4.5.5.

Sensitivity to Connectivity Errors

After developing predictions for individual meters by phase, this method uses information about meter-to-transformer connections. Although there are some cases of miss-assignment, current data coming from GIS were used and overall results obtained by the algorithm were improved with this assumption. Similar tests were made to add constraints based on single-phase lateral connections but this reduced the accuracy. Because the lateral connectivity was observed to reduce the accuracy of the forecasts, they were not included in the final predictions.

4.5.1.3 Results and Observations

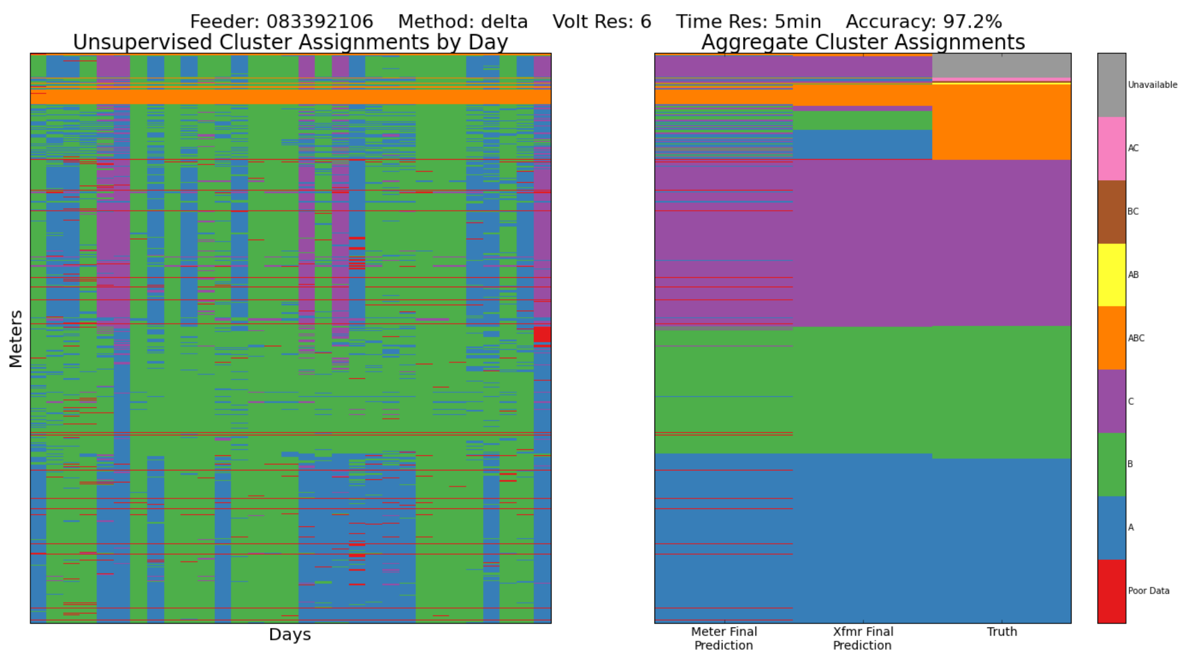
The results for the EVC method are shown in Table 12.

Table 12: EVC Results With High Resolution Data

	4-wire circuit 1	4-wire circuit 2	4-wire circuit 3	Total
Ensemble Voltage Cluster (EVC)	94.5%	97.2%	94.7%	**95.7%**

Figure 13 displays the results of the ensemble prediction. On the left panel, each column represents a day (30 days used in the sample), and each row represents each meter used in the study. The assigned cluster for any meter is not fixed from day-to-day. This is why each cell is colored by phase designation depending on the result of the algorithm. The right panel is divided into three columns: the left column presents the final phasing prediction for each meter as the result of the ensemble prediction, the middle one represents the phasing prediction for each meter constrained by the transformer assignment from GIS, and in the right column, the field-verified phasing. The ensemble prediction can smooth out outlier days from individual meters, resulting in a much less noisy panel on the left. In addition, the consideration of the predictions of neighboring meters allows inferences to be made about meters with missing or incomplete data that show up in red in the both panels.

Figure 13. EVC Clustering Results With High Resolution



The impact of the amount of data used was also tested. This method saw similar levels of accuracy with six weeks instead of four weeks of data, as well as only three weeks. However, less than three weeks of data caused the accuracy to drop sharply. As the data was collected automatically and the method operation time is linear with the number of days involved, varying the amount of data provided is not expected to have a significant impact on the process duration.

Multiple Circuit Configurations

The Mitra paper was written with the expectation of a 4-wire circuit, phase to neutral system configuration. This was also the case in the study area. However, most of the PG&E territory is composed of 3-wire circuit with phase to phase configurations. In the small-scale demonstration, this method in a phase to phase configuration would be tested. The fundamentals of the algorithm are sound; however, adjustments were required to adapt the method to the different configuration.

Ensemble Prediction

While there were many days in the study period that did not provide three clusters, the ensemble prediction was able to focus on the days that did provide three reasonably balanced clusters, making the method robust to poor or missing data in a significant portion of the period. There may be further improvements available by investigating the characteristics of the days providing poor separation.

4.5.2 Regression Corrected Correlation (RCC)

This method is an implementation of methods described in "Advanced Metering for Phase Identification, Transformer Identification, and Secondary Modeling," (Short, 2013). The approach uses a bottoms-up grouping approach to test the correlation of meters specifically pre-assigned to a transformer. Rather than test assumptions for connectivity in a search pattern based on correlation from a broad set of possible outcomes, the method works to validate connectivity by working from the presumed connectivity model, testing for error in the connectivity assumption. If the error level is low, then the assignment is considered correct, and the algorithm may progress to the next comparison. In this way, as the algorithm progresses through a circuit's presumed connected meters, phase and the inferred connectivity model for meter-to-transformer are produced at the same time.

4.5.2.1 Technical Development and Methods

Algorithm Approach

Meters are ordered ahead of the evaluation, by meter ID and grouped by assigned transformer in GIS (herein called set A). From set A, the first two meters, *i* and *j*, are evaluated in piecewise fashion. For each meter pair *i* - *j* in set A, the regression model in equation (1) below is solved. The input data into the regression model is the voltage time series value from each meter and the real component of current at each meter.

Method 1 then finds the meter pair *i* – *j* with the highest value for the coefficient of determination (R^2) and calculates voltage using equation (2) and current using equation (3) at a virtual upstream common point. This virtual upstream point is then added as a new metering point, *k*. Meters *i* and *j* are then removed from set A and the new metering point *k* is added. The process is repeated (*i.e.* combining *k* with the next metering point, and so forth) until all meters are paired or the R^2 value drops below a threshold. With all meters paired, voltage and power estimates are compared to each transformer secondary from the substation dataset. These values then identify phasing by solving the regression model in equation (4) where V_{sub}^{p}, W_{sub}^{p}, $V_{k,xfmr}$, and $W_{k,xfmr}$ are the substation voltage on phase *p*, substation power on phase p, voltage at transformer *k*, and power on transformer *k*, respectively. The phase with highest correlation in terms of R^2 is assigned to the transformer and all meters under the transformer. The general logic for the bottoms-up grouping and single-phase transformer Phase ID is described in Algorithm 1 in Table **13**.

Table 13. Method 1 Algorithm Description

Step A1	For each single-phase transformer performing a bottoms-up grouping for the set of meters (set A)		
		A1.1	Find meters on each single-phase transformer
		A1.2	For each meter *i*, solve the regression model in (1) paired with every other meter in set A
		A1.3	Select meter *j* that has the highest coefficient of determination (R^2) value in regression with meter *i*
		A1.4	For the selected meter pair *i – j*, store line parameters for each branch from the regression model. Also, find voltage using (2), current using (3), and calculate power at the common upstream point (this forms a meter pairing point *k*)
		A1.5	Remove meter pair *i – j* from set A and add new meter pairing point *k*
		A1.6	Repeat starting from A*1.2* until all meters in set A have been paired or the R^2 value drops below a threshold. This gives a virtual metering point at the secondary side of each transformer
Step A2	Identify phasing for each single-phase transformer		
		A2.1	With the virtual metering point for each transformer, separately run a regression using (4) to each of the three phases at the substation
		A2.2	Assign the transformer phasing based on the phase with the highest R^2 value
		A2.3	Assign meter phase based on transformer phasing

Referenced Equations

$$V_i = \beta_0 + \beta_1 V_j + R_j I_{j,R} + X_j I_{j,X} + R_i(-I_{i,R}) + X_i(-I_{i,X}) \quad (1)$$

$$V_k = (V_{k,estimate\_i} + V_{k,estimate\_j})/2 = [(V_i + R_i I_{i,R} + X_i I_{i,X}) + (V_j + R_j I_{j,R} + X_j I_{j,X})]/2 \quad (2)$$

$$I_k = I_i + I_j \quad (3)$$

$$V_{sub}^p = k_0 + k_1 V_{k,xfmr} + k_2 W_{k,xfmr} + k_3 W_{sub}^p \quad (4)$$

Assumptions

- Regression Model Simplification
  - The regression model in equation (1) is simplified to the real power-only regression model in equation (5) since reactive power to calculate reactive current is not provided by PG&E meter data.

$$V_i = \beta_0 + \beta_1 V_j + R_j I_{j,R} + R_i(-I_{i,R}) \quad (5)$$

- Connectivity Model from Available Data
    - Set A (meters under the same transformer) is obtained from the existing GIS record of meter-to-transformer connectivity. Each meter is associated to a transformer based on the mapping, and all meters under same transformer are grouped in one set.
- Unity power factor at substation
    - Power-per-phase at substation was calculated as the product of voltage and current assuming no reactive displacement.
- R^2 Threshold
    - The threshold selected was 99% as described in the academic paper.
- Bottoms-Up Grouping Approach
    - Transformers with more than two meters
        - The bottoms-up grouping approach is continued until the R^2 value drops below 0.99. Meters that are electrically close together correlate strongly.
    - Transformers with one meter
        - The bottoms-up approach cannot be performed and the meter measurement values are passed to the transformer to solve regression model in (2).
    - Meter pair $i - j$ with missing data
        - Observations from both meters were removed from their measurements' time series if one of the meters had missing data
- Removal of three phase and phase to phase meters
    - Three phase and phase to phase meters were removed from the calculations since and approach to including them is not described in the academic paper. This configuration difference may be a significant factor in the relatively low accuracy of Method 1.

Testing

All model development was performed using the 4-wire circuit 1 circuit.

A simulated test set was used to validate the functioning of the implemented algorithm under ideal conditions, and it was shown to function properly.

4.5.2.2. Challenges

Distribution Transformer Configuration

Phase to Phase and three-Phase transformers

The bottoms-up grouping analysis was performed only on transformers with single-phase meters. Phase to phase and three phase meters identified in the field results were removed from all datasets since the academic method did not describe how to account for these meters. This can be expected to deteriorate the performance of the algorithm because the bottoms-up approach relies on complete metering downstream of each grouping point to have accurate estimates at the upstream points, and excluding the three phase meters will have a large impact on those flows.

Inaccurate Transformer Configuration

Because this method excluded three-phase and phase to phase meters based on field designations, the evaluation of this method was not impacted by this challenge.

Phase to Phase Substation Voltages

The academic method does not address how to handle phase to phase connections. However, the step A2.1 would require phase to phase voltages from the substation to work correctly. This challenge was not addressed by this method.

Solution Stability

Unpaired Meters

As mentioned above, the pairing of meters using the bottoms-up approach depends on several factors such as the number of meters under the same transformer, the R^2 threshold value, and existing connectivity from meter data. In the analysis, several meters were not grouped. Having unpaired meters had an impact on the performance, since the bottoms-up approach relies on complete metering downstream of each grouping point to have accurate estimates at the upstream points. A possible explanation for the meters left out of the grouping is that they were matched to the wrong transformer in the meter metadata. Legacy meters without voltage profiles or with low resolution voltage measurements would also create discrepancies between the bottoms-up calculated voltage profile and the real profile.

Low Resolution Meters

This method is likely negatively impacted by low resolution meters in the regression between the meters. The mitigation for these meters is that they may be treated as unpaired due to the regression threshold. However, when there are many meters excluded, it can cause deterioration in the solution.

Meter Voltage Measurement Interval

No sensitivity to this was evaluated, but this method is likely not significantly impacted by the voltage interval, as no time-based calculations are applied to the data.

Sensitivity to Connectivity Errors

This method is impacted by meter-to-transformer mapping errors, as it impacts the regression method which creates the virtual voltages on the transformer. However, it addresses this by excluding meters which do not meet the regression threshold. Unfortunately, it does not include a mechanism to find proper homes for the meters, so if there are many meters impact, it can contribute to a deterioration of the solution. This method did not implement single phase lateral constraints.

4.5.2.3 Results and Observations

The results from the RCC approach are shown in Table 14. The RCC method had more difficulty producing accurate results than the other methods evaluated.

Table 14. RCC Results With High Resolution Data

	4-wire circuit 1	4-wire circuit 2	4-wire circuit 3	Total
Regression Corrected Correlation (RCC)	62.8%	69.5%	77.7%	**70.5%**

Likely causes of the poor performance of this algorithm are the lack of inclusion of the three-phase meters, as well as the dropping of single-phase meters due to the regression threshold. Low resolution voltage data of some of the meters and meter-to-transformer mapping inaccuracies may also contribute to poor correlation between meters. As the method steps up through the circuit topology, the impact of these errors has a cumulative effect. Missing transformer load in the bottoms-up grouping would result in a reduction of the correlation to upstream meters, which would in turn result in more dropped meters. Due to this initial lack of performance in comparison to other methods, this method was not pursued in the small-scale demonstration.

4.5.3 Reduced Constrained k-Means (RCKM)

RCKM is an external academic method described in “Phase Identification in Electric Power Distribution Systems by Clustering of Smart Meter Data” (Wang, et al., 2016). Like the EVC method, this algorithm uses k-means clustering to group meters into phase groups, however in RCKM the clusters are constrained by the electrical connectivity, and the final label assignments are done by comparing the principal component of the circuit breaker bus voltages to the centroids of the clusters.

4.5.3.1 Technical Development and Methods

This method is described in detail in the academic paper. The general approach is as follows:

1. Normalize and center AMI Voltage and extract top Principal Components (PC)
2. Generate cluster connectivity constraints from network model
    a. These “must-link” constraints require that a set of members in the same constraint group are to be in the same cluster
    b. Must-link constraint groups are implemented for meters at the same voltage level and on the same single-phase lateral
3. Use constrained k-means clustering to group meters
4. Identify the phase connectivity of each cluster
    a. This is achieved by minimizing the Euclidean distance between the pairing options of the principal components of the circuit breaker bus voltage and the principal component of the cluster centroids.

4.5.3.2 Challenges

Distribution Transformer Configuration

Like other methods, this implementation did not differentiate between phase to phase and phase to neutral transformers, and only predicted phase to neutral results. If more accurate transformer connectivity was available, this algorithm could be applied to the meters assigned to those transformers separately.

Phase to Phase Substation Voltages

Because SCADA is used to map the clusters to the voltage phases, phase to phase substation measurements would be required for this method to accurately predict for those circuits.

Solution Stability

This method is apparently stable for high frequency high resolution data. Though it does not use ensembles, the constraints in the clusters are apparently sufficient to prevent some of the unstable solutions observed in the EVC method.

Low Resolution Meters

As with other methods, this method was negatively impacted by low resolution meters. No resolution for this challenge is provided by this method.

Meter Voltage Measurement Interval

No sensitivity to this was evaluated, but this method is likely not significantly impacted by the voltage interval, as no time-based calculations are applied to the data.

Sensitivity to Connectivity Errors

Because of the use of meter-to-transformer and single-phase laterals as constraints in the k-means constrained cluster, errors in these constraints may introduce errors in the phasing. In addition, because the constraints are strict and included in the clustering steps, these errors may create instability in the clusters that could deteriorate the overall solution.

4.5.3.3. Results and Observations

The results for the RCKM method are shown in Table 15.

Table 15: RCKM Results With High Resolution Data

	4-wire circuit 1	4-wire circuit 2	4-wire circuit 3	Total
Reduced constrained k-Means: (RCKM)	94.2%	92.7%	93.4%	**93.3%**

4.5.4 Feature Reduced Cluster (FRC)

This method was implemented by an external participant who did not provide details about their implementation. In general, their method used some form of feature reduction and a clustering method to label phases.

4.5.4.1 Technical Development and Methods

As this method was external and proprietary, no detail was provided on technical development and methods.

4.5.4.2 Challenges

Distribution Transformer Configuration

This approach did not provide different forecasts for phase to phase transformers and only predicted for phase to neutral options.

Phase to Phase Substation Voltages

This was not addressed by the participant, but it is assumed that the lack of phase to phase voltages would impact the accuracy of this method on 3-wire systems.

Solution Stability

This method is apparently stable for high frequency high resolution data. Because the method is proprietary, it is not known what mechanism achieves this.

Low Resolution Meters

No sensitivity to this was evaluated, but this method is likely negatively impacted by low voltage resolution.

Meter Voltage Measurement Interval

The primary challenge raised by this participant was the importance of having high time frequency measurements to improve the capability of the algorithm.

Sensitivity to Connectivity Errors

Like other methods, this will be negatively impacted by connectivity errors when the results are constrained by connectivity. The participant did not specify whether constraints were applied within the algorithm.

4.5.4.3 Results and Observations

The results for the FRC method are shown in Table 16. This method performed comparably to the other best methods in the POC demonstration. However, due to external factors, this external participant chose not to participate into the small-scale demonstration.

Table 16: FRC Results With High Resolution Data

Approach	4-wire circuit 1	4-wire circuit 2	4-wire circuit 3	Total
Feature Reduced Cluster (FRC)	90.8%	94.0%	91.8%	**92.4%**

4.5.5 Measurement Frequency and Precision Sensitivity on the EVC method

The project evaluated the sensitivity of the EVC method to different levels of voltage data frequency and precision available from meters on the tested circuits.

Sensitivity to the time frequency

As the POC demonstration utilized 5 min interval voltage data, it was possible to evaluate the sensitivity of the method to different sampling intervals. The performance of the EVC method was tested based on (1) High frequency scenario using 5-minute interval voltage data and (2) Low frequency scenario using data every 60 minutes. Table 17 summarizes the results for each circuit.

Table 17: EVC Sensitivity to Voltage Interval Frequency

Scenario	Max Voltage Decimals	Sampling Time	4-wire circuit 1	4-wire circuit 2	4-wire circuit 3	Total
High Frequency	1	5 minutes	94.5%	97.2%	94.7%	**95.7%**
Low Frequency	1	60 minutes	94.4%	89.2%	87.1%	**89.9%**

These results can help to inform the data requirements for a larger rollout of network connectivity analysis. During the summer 2017, PG&E started a meter firmware upgrade to collect voltage information from the SmartMeter™. The time frequency was based on the rate schedule of the customer (15min for industrial and commercial customers and mostly 60min for residential customers). As only a low percentage of the meters on each circuit have 15-minute resolution, using data from the firmware upgrade would require running the model using the hourly reads only. The low frequency scenario was tested to see the impact of the performance of the algorithm with hourly data. The results achievable with the firmware upgrade have accuracy only slightly below the high-resolution trial implementation, therefore it was decided for the subsequent small-scale demonstration to use hourly voltage reads and not to spend additional resources on getting 5-minute interval reads.

Sensitivity to the voltage precision

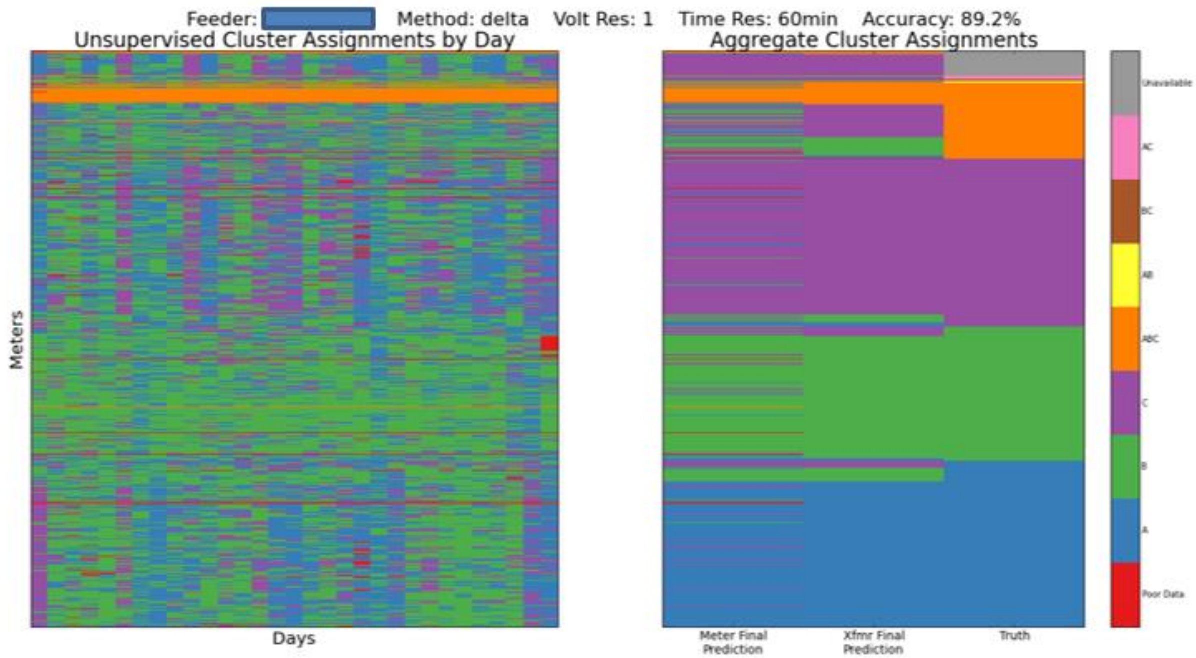
In PG&E's territory, about 60% of the meters have decivolt precision and 40% of the meters are incapable of providing this precision. A replacement of those meters would be necessary to get the decivolt level of precision. To determine the impact of this on the algorithm, a new scenario was tested, looking at the performance of the EVC method if we only had integer voltage precision. Table 18 presents these results. Moving from one decimal place to the whole volt increments available on the current SmartMeter™ system causes severe accuracy performance degradations.

Table 18: EVC Sensitivity to Voltage Resolution

Scenario	Max Voltage Decimals	Sampling Time	4-wire circuit 1	4-wire circuit 2	4-wire circuit 3	Total
Medium Resolution	1	60 minutes	94.4%	89.2%	87.1%	**89.9%**
Low resolution	0	60 minutes	33.8%	48.9%	30.3%	**38.8%**

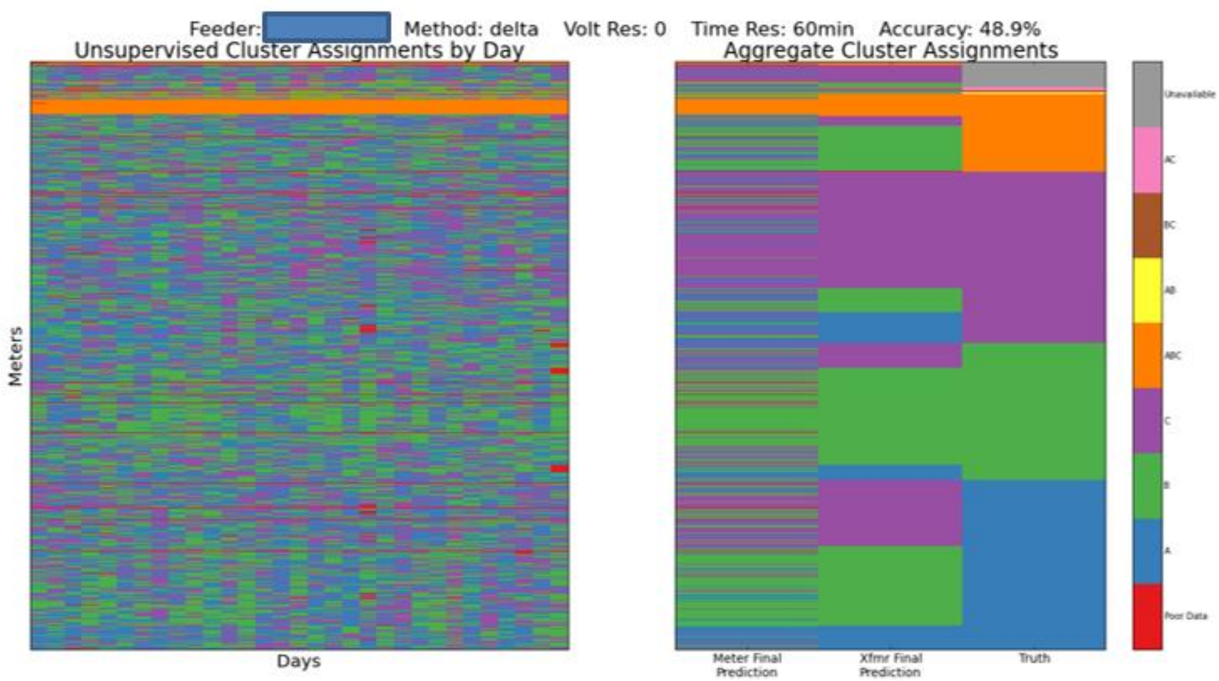
Figure 14 shows the heat map produced using voltage reads with decivolt voltage resolution, combined with 60-minute interval reads rather than 5-minute (shown in Figure 13 previously). The clustering for individual meters for individual days is much less clear, but when considered across the ensemble, much of the accuracy of the full resolution data case is recovered.

Figure 14: EVC Results – Medium Resolution



However, as seen in Figure 15, when the voltage read resolution drops to single volt increments, the predictive power of the method is almost completely lost.

Figure 15: EVC Results – Low Resolution



4.5.6 Results Summary

Table 19 provides a summary of all the results obtained for the POC demonstration for phase identification. As noted previously, all methods were supposed to use the field verification from 4-wire circuit 1 as the training set used to tune the models, and to reserve the data from 4-wire circuit 2 and 4-wire circuit 3 for the final analysis. The results are presented together for conciseness.

Table 19: Proof-of-Concept Results With High Resolution Data

Method	4-wire circuit 1	4-wire circuit 2	4-wire circuit 3	Average
Regression Corrected Correlation (RCC)	62.8%	69.5%	77.7%	**70.5%**
Ensemble Voltage Cluster (EVC)	94.5%	97.2%	94.7%	**95.7%**
Reduced constrained k-Means (RCKM)	94.2%	92.7%	93.4%	**93.3%**
Feature Reduced Cluster (FRC)	90.8%	94.0%	91.8%	**92.4%**

The EVC method is the best performing algorithm across all circuits. However, it is noted that this method did not use single phase lateral constraints, because they reduced the accuracy, while these constraints were included in the RCKM method. It is likely that these results would be closer if those implementation details were the same.

The POC demonstration proved that using data analytics to automatically identify the phase of meters is possible with accuracy greater than 90%. This demonstration was developed on 4-wire (21kV) circuits, which are the most common circuit configuration in the US. However, in California, the majority of circuits have a 3-wire configuration. For this reason, a small-scale demonstration was subsequently conducted to be able to provide results more representative of PG&E's distribution system.

4.6 Small-Scale Demonstration

For this demonstration, both phase identification and meter-to-transformer methods were assessed. Four additional circuits were selected and field validated: three 3-wire (12kV) circuits and another 4-wire (21kV) circuit. To build the algorithms, the field data from one 4-wire circuit (4-wire circuit 1, the same circuit used for the verification set in the POC) and one 3-wire circuit (3-wire circuit 3) were shared with all the participants. In between the two demonstrations, PG&E had upgraded the firmware of their SmartMeters™ to allow for collection of voltage information with at least hourly resolution. Considering the results on the impact of the time resolution on the EVC method from the POC demonstration, it was decided not to invest in additional meter capabilities to collect higher time resolution AMI data.

4.6.1 Phase Identification Methods

During this demonstration, PG&E first utilized the EVC algorithm developed in the POC demonstration, but due to the reduced accuracy with the stage 2 dataset and the 3-wire circuits, PG&E developed its own method and benchmarked results with four external solutions:

- **Ensemble Voltage Cluster (EVC) Revised:** This is the same approach implemented in the POC, but modifications were made to enable 3-wire circuits.

- **PG&E's method -Ensemble Relative Virtual Voltage Constrained Cluster (ERVVCC):** PG&E internally developed method derived from the EVC method which uses voltage clustering with several pre-processing and filtering steps (details below).
- **t-SNE constraint-driven hybrid clustering (t-SNE CHC):** External academic method. This approach is described in "Advanced Metering Infrastructure Data Driven Phase Identification in Smart Grid" (Wang, et al., 2017). This method uses t-SNE feature reduction of voltages prior to applying a constrained density based clustering.
- **Load Flow Monte Carlo (LFMC):** Method which uses detailed asset and connectivity information to develop a load flow model and uses this model to identify phasing.
- **Impedance Corrected Distance Mapping (ICDM):** This method utilized an impedance correction for the voltage measurements and then assigned phase labels based upon the minimum distance to the SCADA voltage.

4.6.2 EVC – Ensemble Voltage Cluster (Revised)

4.6.2.1 Technical Development and Methods

Approach Overview

This method was an implementation of the EVC as applied in the POC demonstration, but was modified to support phase to phase measurements by PG&E.

Phase to Phase Voltage Estimate

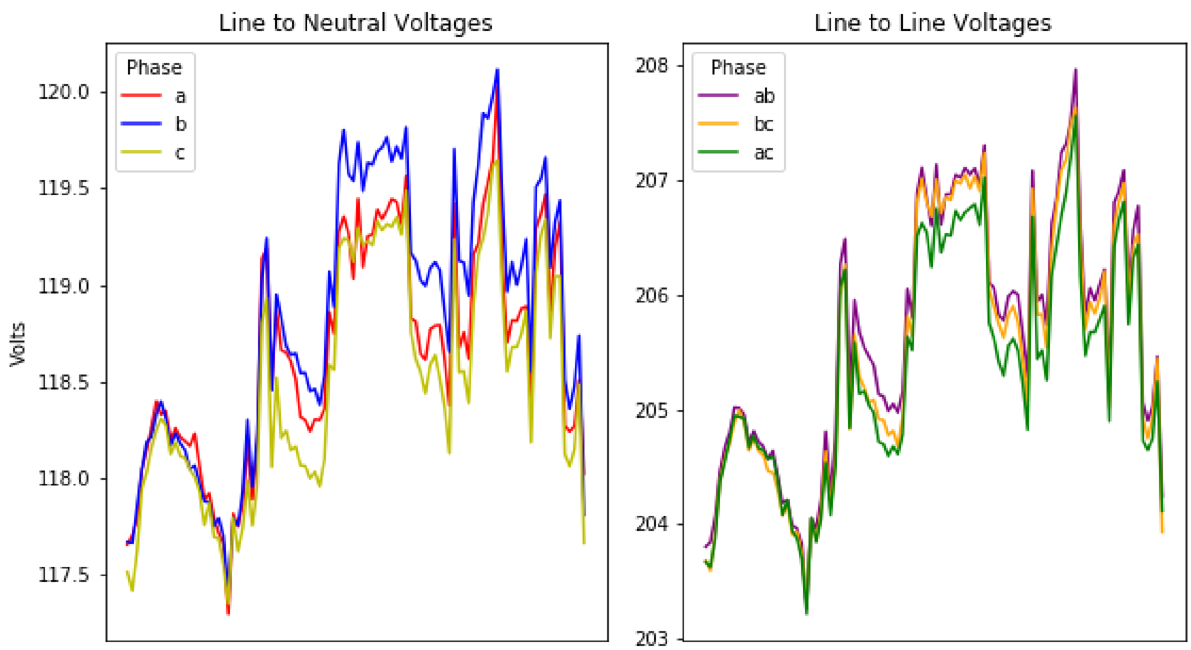
Because phase to phase voltage was not available at the substation, a proxy for this value was calculated from the phase to neutral measurements. This calculation assumes that the phase angles are balanced and then uses the relatively simple formula to estimate the phase to phase voltage.

For phase AB, the equation for the phase to phase voltage V_{ab} would be:

$$V_{ab} = \sqrt{(V_a - V_b \cos\theta)^2 + \ (V_b \sin\theta)^2}$$

Where V_a and V_b are the A and B phase to neutral voltages and theta is the phase angle, which must be 120 for this formula to be valid. An example of the results of this calculation is shown in Figure **16**.

Figure 16: Calculation of Approximate Phase to Phase Voltages from Phase to Neutral Measurements



4.6.2.2 Challenges

Distribution Transformer Configuration

As discussed previously, this implementation only predicted phase to neutral phase for the 4-wire circuits. For 3-wire circuits, phase to phase was the only option, so the model needed to be updated to assign the correct phase to phase label.

Phase to Phase Substation Voltages

Ideally for this project, phase to phase measurements would be available at the substation, but the current standard at PG&E is to record the line to neutral voltage at the substation bus. The Mitra algorithm uses the substation voltage for the centroid of the clusters to stabilize the solution and to align the cluster with the phase label. For 3-wire circuits or for phase to phase transformers on a 4-wire system, these need to be phase to phase voltages to work properly. In the POC demonstration, no solution was provided to this problem, because the circuits were all 4-wire and the phase to phase transformers on the 4-wire were not clustered separately as described in the section above.

Solution Stability

The ensemble spectral cluster helps to improve the stability of this method, but there appeared to be more variation in the ensemble cluster members in the solution for the 3-wire than there was for the 4-wire.

Low Resolution Meters

As shown in 4.5.5, this method is negatively impacted by low resolution meters.

Meter Voltage Measurement Interval

As shown in 4.5.5, this method is sensitive to the voltage measurement interval.

Sensitivity to Connectivity Errors

As discussed in 0, this method is sensitive to connectivity errors.

4.6.2.3 Results and Observations

The results for EVC in the small-scale demonstration are shown in Table 20.

Table 20: EVC Small Scale Demonstration Results

	3-wire circuit 1	3-wire circuit 2	3-wire circuit 3 *	4-wire circuit 4	4-wire circuit 1 *	4-wire circuit 2	4-wire circuit 3
EVC	68.0%	42.2%	72.9%	64.5%	89.8%	58.6%	63.6%

*Training circuits

This method performed worse for the circuits in this demonstration than in the POC demonstration. To a certain extent, this is expected because only hourly data was available for this demonstration, so based on the sensitivity results in Table 17, up to an 8%-point reduction in accuracy was anticipated. The results in 4-wire circuit 2 and 2107 seem to have deteriorated much more than would be anticipated from the POC demonstration results. This may be due to a lower number of stable cluster solutions in the period observed, as it can be seen from Figure 14 that there is much more variability in the solution for these circuits.

In addition, it was noted in other methods that 4-wire circuit 2 had a subset of meters which behaved markedly different from other meters on the circuit, because the field validation of this circuit was done in the previous year. It is possible that this might have been a circuit reassignment which was not captured in the metadata. This set of meters may have disrupted the clustering in this time period. Also of note, 4-wire circuit 4 had a measurement failure, so good data was only available on that circuit for about 3 weeks, as opposed to about 3 months for the other circuits.

4.6.3 PG&E Phase ID Method – Ensemble Relative Virtual Voltage Constrained Cluster

Technical Development and Methods

Approach Overview

The results from the POC demonstration were extremely promising, but in the small-scale demonstration, AMI data was only available at an hourly time resolution, and included 3-wire circuits which presented additional challenges. As shown in table 20, the EVC method did not perform well on 3-wire systems with hourly data. To address these issues, a new method was constructed.

Like EVC, this method uses a voltage clustering ensemble based approach. However, several pre-processing and feature engineering approaches were added to improve performance regardless of varying circuit characteristics.

Phase to Phase Voltage Estimate

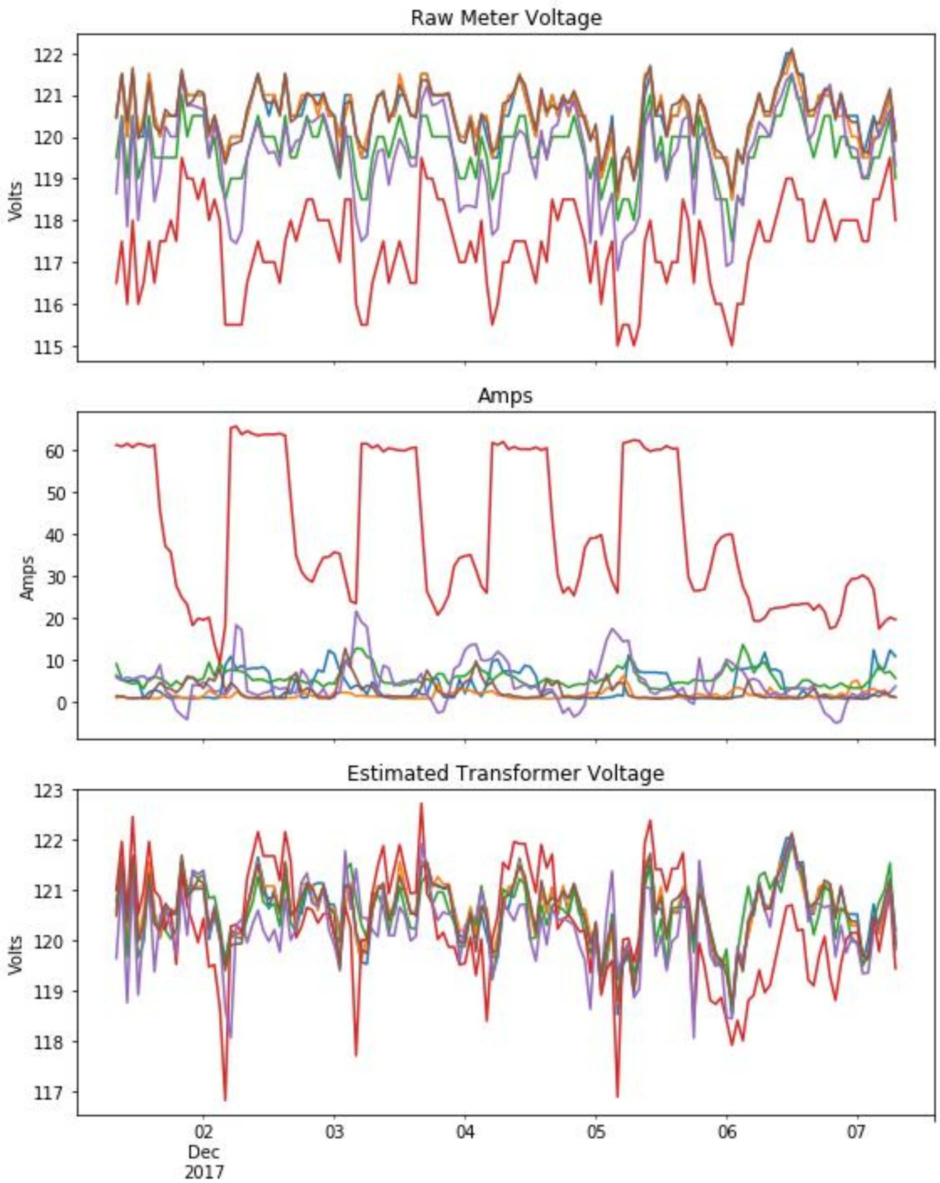
The phase to phase voltage estimate from Section 0 was used as the cluster initialization for three wire circuits.

Virtual Transformer Voltage

A key aspect of this new approach was the development of the “Virtual Transformer Voltage”. This method uses voltage and load measurements and an estimate of the impedance to estimate the voltage at the low side of the transformer. If multiple meters are available on at transformer, then an optimization can be used to improve the estimate of the impedance, by minimizing the difference over time of the various meters estimates of the transformer voltage. A similar approach is described in detail in (Berrisford, 2013), with the primary difference being that secondary connectivity was not available for this analysis, so radial connections were assumed on the secondary.

An example of applying this algorithm to a series of meters on a single service transformer is shown in Figure **17**. The first chart, titled Raw Meter Voltage, shows the raw AMI voltage readings of five meters. These five meters are all on a single-phase transformer, at the same location on the primary, but due to local load, they can vary substantially. The load can be seen in the second chart, where current has been estimated using the voltage and the integrated load measurements. The meter with the largest load has substantial impact to the raw voltage. The estimated transformer voltage is shown in the third chart, and it can be observed that the approach was able to successfully create a similar estimate of the transformer voltage from multiple AMI meters.

Figure 17**: Example of Raw AMI Voltage and Corresponding Virtual Transformer Voltage for Meters on a Single Transformer**



Phase Relative Voltage

A second modification to the Mitra approach is to generate new features to use for the clustering. When 5-minute data was available, the voltage "Delta" was a productive feature to use for clustering in the POC demonstration, but it was not as effective in this demonstration, in particular for the 3-wire delta systems. For this approach, a new feature called "phase relative voltage" was generated.

For each AMI voltage and each SCADA measurement, a calculation was made of the difference between that voltage and each phase of the SCADA voltage. To isolate relative changes, a series of standardizations was applied to the SCADA and AMI voltages prior to calculating these differences.

In an ideal world, the SCADA phase associated with a given meter measurement would have its phase relative voltage with that meter near zero, while the Relative phase against the other two SCADA phases for both the AMI and that SCADA phase would be equivalent.

The primary benefit of the phase relative voltage was that it made the SCADA voltages and AMI voltages more comparable, and improved the ability of the algorithm to map between the AMI clusters and the SCADA voltages. For three-wire delta systems, this approach was the only way found to achieve reasonable cluster grouping. In addition, it seemed to be more effective at isolating phase differences when there was less voltage imbalance on the bus.

The phase relative voltage features improved performance for the three phase circuits, but in certain situations, there were still challenges in getting appropriate separation between phases. To address this, attempts were made to isolate periods of time during which there was more distinction between the phases. Initially, a filter based on voltage imbalance was used, which showed some improvement. However, it was noted that on the three wire systems, when there was larger imbalance on one of the phases, the two other phases of the SCADA would become too close to each other to be easily differentiable. This caused problems both in the ability of the clustering algorithm to accurately separate the AMI into correct groups and to correctly assign the cluster groups to the appropriate phase. To address this, a new feature was generated from the SCADA voltage, and used to filter the times used for the analysis. This feature is simply the smallest difference between the 3 normalized and centered SCADA voltages. For several of the circuits, even setting a threshold of 0.2 volts drastically reduced the amount of time used for the analysis, and for some circuits filtered the entire time frame. However, when sufficient data was available for the algorithm after the filtering, the accuracy and stability of the predictions was improved.

Filtering

It was noted that there were some instances where the cluster solution did not find a good solution, or could not be mapped to the SCADA voltages effectively. It was noted that on 3-wire circuits there were periods of time where the target SCADA voltages would both be very close to two of the three clusters and one cluster would be relatively far from the SCADA voltages. It is possible that this is an artifact of the balanced phase assumption when estimating the phase to phase voltages, when one of the phases was unbalanced. To mitigate this, the data set was filtered for periods of time where there was a minimal difference between all three phases of the normalized SCADA voltages. For example, if the unit normalized phase AB and BC voltages did not differ from each other by a minimum threshold value, then that measurement was excluded from the analysis.

Algorithm Approach

1. Drop voltage outliers.
2. Convert AMI secondary voltage to normalized virtual transformer voltage.
3. Split meters into Phase to phase and Line to Neutral, and Single phase or poly phase groups.
4. For each group:
    a. If group is a Phase to phase AMI connection, transform SCADA measurements from line to neutral to phase to phase representation, assuming 120-degree phase difference.
    b. Filter times to isolate periods with sufficient voltage imbalance and sufficient separation between phases.
    c. Split the filtered times into subgroups, and then for each time group:
        i. Exclude meters which do not have a sufficient correlation to at least one of the SCADA phases. This excludes meters which have moved from one circuit to another, which are disruptive to the clustering process.

   ii. Create phase relative voltage features for both the SCADA and AMI voltages.
   iii. Filter for periods of time with more current imbalance, and a minimum difference between the voltages.
   iv. Prepare data for clustering by dropping timestamps and meters without sufficient data in the period.
   v. Apply must-link and cannot-link constraints
   vi. Constrained Spectral Cluster into 3 groups
   vii. Assign Group to SCADA phases which minimize the average Euclidean distance to all members of the group.
  d. Generate an ensemble forecast using weighted voting of the ensemble members. Weights are determined based meter distance from the cluster centroid. Imbalanced clusters are given reduced weights.
5. For each meter use the result from each round to vote for a phase label.
6. For each single-phase transformer, use the result from each meter to vote for a phase label.

Assumptions

- For virtual transformer voltage calculations, radial connection is assumed
- Assumptions used for reactive power are derived from customer type
- For substation phase to phase voltage calculation, 120 balanced phasing is assumed

Testing

All training and calibration was performed on the 4-wire circuit 1 and 3-wire circuit 3 dataset.

4.6.3.2 Challenges

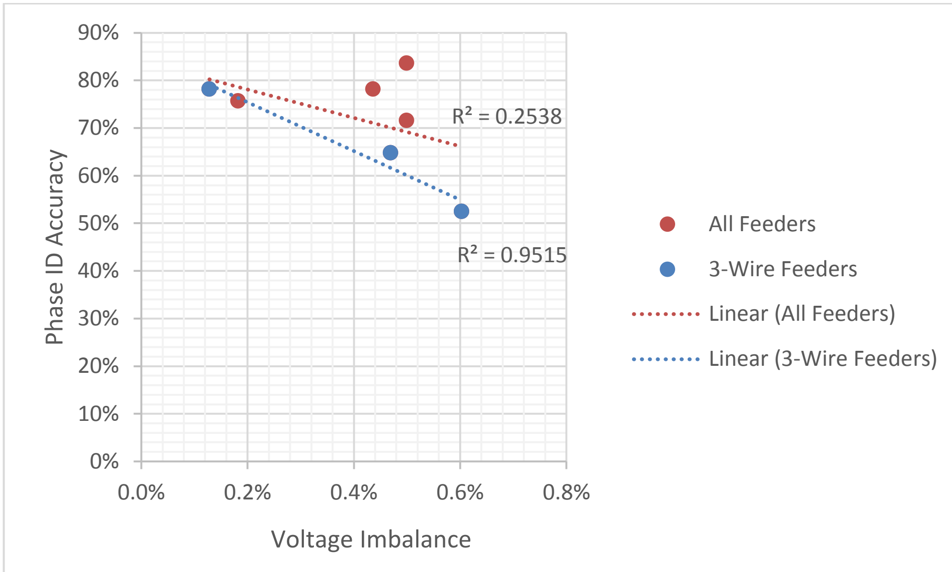
Distribution Transformer Configuration

This approach did not provide different forecasts for phase to phase transformers on 4-wire circuits and only predicted for phase to neutral options. However, the method was developed such that if more accurate transformer configuration was available, these transformers and associated meters could be clustered in a separate group to determine the meter phase labels. Three phase transformers were not included in the clustering, but the implementation was designed in such a way that they could be phase identified as well.

Phase to Phase Substation Voltages

Because no phase to phase measurements were available at the substation, a calculation (explained in 4.6.2.1) was used to estimate the phase to phase voltages from the phase to neutral measurement. It was noted that during certain periods, one phase would be very different from the others, and that in those periods, the other two phases would become very close, impacting the performance of the prediction. At these times, the cluster to voltage mapping was not aligning well. It is likely that this is an artifact of the balanced phase assumption in the phase to phase voltage estimate being inaccurate during high periods of imbalance. To mitigate this, periods where there was a small difference in two of the voltages were filtered from the dataset. This improved the results in most cases, but for some circuits drastically reduced the amount of data available for analysis. It would be much better to use the actual phase to phase voltage measurements if it was available. In Figure **18**, the accuracy of the PG&E method as a function of circuit voltage imbalance is shown. Though the sample size is very small, it appears that the 3-wire circuits are impacted negatively by the voltage imbalance. This is after the application of the voltage difference filtering described above, but even with the improvement associated with that, the impact of voltage increment can be seen to be a significant impact. It is anticipated that with phase to phase substation voltages, the performance for circuits with more voltage imbalance can be improved.

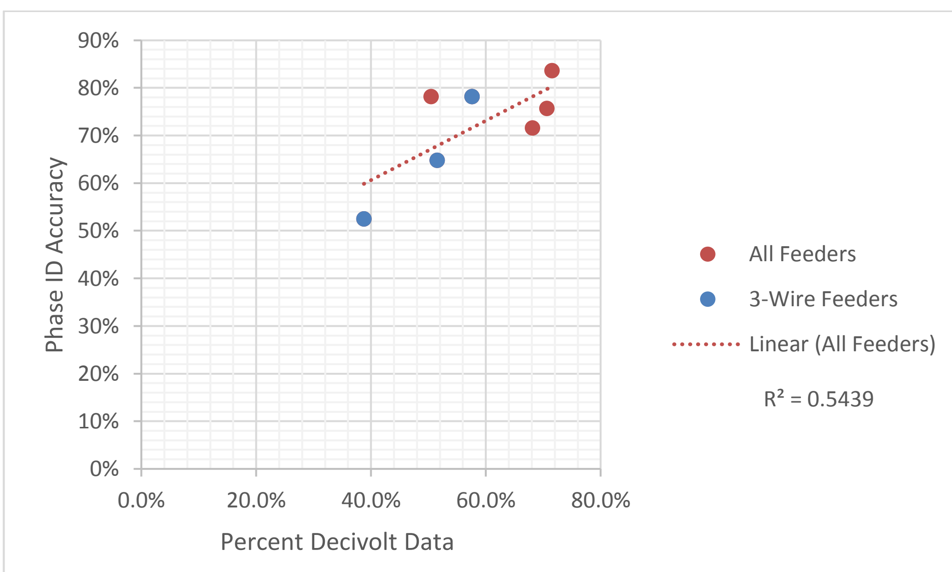
Figure 18: PG&E Phase ID Accuracy vs. Voltage Imbalance



Low Resolution Meters

Figure 19 shows the relationship of the circuit percentage of single phase meters with decivolt data on the PG&E phase ID accuracy. Though the sample size is small, there is an apparent positive correlation between the percent of decivolt meters and the accuracy of the method. This is consistent with the relationship seen in the EVC in the POC demonstration.

Figure 19: PG&E Phase ID Method Accuracy vs. Percent of Single Phase Meters With Decivolt Data on Circuit



Meter Voltage Measurement Interval

Because this method does not use a time delta as a feature, the measurement interval does not have a large impact on the performance.

Sensitivity to Connectivity Errors

Like other methods, this will be negatively impacted by connectivity errors when the results are constrained by connectivity. This method uses constrained spectral clustering to improve the stability of the clusters, but the constraints are not enforced in the cluster output so they can break the constraints if needed. In circuits where the solution is accurate, inconsistency between the predicted phase and the constrained phase can be used to help identify connectivity errors. However, if there are too many connectivity errors, the constraints may contribute to instability in the clustering solution.

4.6.3.3. Results and Observations

The results for the PG&E method are shown in Table 21.

Table 21: PG&E Phase ID Results

	3-wire circuit 1	3-wire circuit 2	3-wire circuit 3 *	4-wire circuit 4	4-wire circuit 1 *	4-wire circuit 2	4-wire circuit 3
PG&E	64.8%	52.5%	78.2%	78.2%	83.6%	71.6%	75.7%

*Training circuits

The results from the 4-wire circuits are in the 71% to 84% accuracy range. The results from the 3-wire circuits range from 52% to 78%. Based on the observations of the clustering solutions and trends related to voltage imbalance discussed above, it seems that the poor performance on some of the 3-wire circuits may be related to the lack of phase to phase measurements at the substation.

4.6.4 t-SNE constraint-driven hybrid clustering (t-SNE CHC)

This approach is described in “Advanced Metering Infrastructure Data Driven Phase Identification in Smart Grid” (Wang, et al., 2017)

4.6.4.1 Technical Development and Methods

The algorithm used for this method is summarized as follows:

- Voltage magnitude measurements are collected from the SmartMeters™.
- Each SmartMeter’s™ readings are centered and normalized by their standard deviation.
- Key features are extracted from the preprocessed voltage time series with a nonlinear dimensionality reduction method.
- The Constrained Hybrid Clustering (CHC) algorithm is leveraged to cluster the low-dimensional data points. This method uses the density-based spatial clustering of applications with noise (DBSCAN) algorithm to generate a dynamic number of clusters.
- Outliers from the DBSCAN clusters are reincorporated into the clusters using k-nearest neighbor (KNN)
- Must-link constraints are applied to cluster groups to enforce consistency with the network topology.

- The phases of the clusters generated are identified by performing field validations on a SmartMeters™ near the centroid of each cluster.

4.6.4.2 Challenges

Distribution Transformer Configuration

Like other methods, for 4-wire circuits, this implementation did not differentiate between phase to phase and phase to neutral transformers, and only predicted phase to neutral results. If more accurate transformer connectivity was available, this algorithm could be applied to those transformers separately.

Additionally, there is some benefit in the fact that this method can generate a dynamic number of clusters. One challenge in categorizing phase to phase connections in a 4-wire system is that it is possible that one of the phase to phase options is not present. If the number of clusters was assumed to be six, but only five categories were present, this could cause errors. In theory, this method may be more resilient in these cases.

Phase to Phase Substation Voltages

At the suggestion of PG&E, the phase to phase voltage calculation was used to estimate phase to phase voltages.

Solution Stability

In the original implementation of this approach, a set of clusters was developed, and the phases were assigned based upon field validations of the sub-clusters. However, that approach was not considered practical in this method since the goal was to not have field verification for all circuits in the future. Instead, the phases were mapped to the SCADA voltages, which were transformed to phase to phase representations by assuming balanced phases for the 3-wire systems. This mapping was detrimental to the performance of the algorithm, but additional field validations were not considered in this evaluation.

Sensitivity to Connectivity Errors

As is the case for EVC, this method is sensitive to connectivity errors. However, in this case, the connectivity is introduced as part of the constrained clustering algorithm, as opposed to a constraint applied at the end of the analysis.

Low Resolution Meters

As with other methods, this method was negatively impacted by low resolution meters. No solution for this challenge is provided by this method.

Meter Voltage Measurement Interval

No sensitivity to this was evaluated, but this method is likely not significantly impacted by the voltage interval, as no time-based calculations are applied to the data.

4.6.4.3 Results and Observations

The results from the t-SNE CHC are shown in Table 22.

Table 22: t-SNE CHC Small Scale Demonstration Phase ID Results

	3 Wires - 12 kV			4 wires - 21kV			
	3-wire circuit 1	3-wire circuit 2	3-wire circuit 3 *	4-wire circuit 4	4-wire circuit 1 *	4-wire circuit 2	4-wire circuit 3
t-SNE CHC	32.0%	30.7%	73.2%	76.3%	-	-	-

*Training circuits

This participant chose to utilize two different methods in the two phases of the project. They did not provide new results for this method for the circuits used in the small scale demonstration since they had provided those results using the RCKM method in the POC demonstration. Though this was not ideal for evaluating the performance of the two methods across the two stages of the project and comparing them, it was what was provided by the participant. The method performed well for 4-wire circuit and for 3-wire circuit 3, but poorly for the rest of the 3-wire circuits.

4.6.5 Load Flow Monte Carlo (LFMC)

This method uses detailed asset and connectivity information to develop a load flow model. Phase connectivity assumptions are systematically modified to develop connectivity solutions which match the observed phase currents. The method intended to use load and voltage measurements, but due to data and time limitations, the solution was only able to utilize load measurements.

4.6.5.1 Technical Development and Methods

This method utilizes detailed asset and GIS information as well as extensive consultation with engineering staff to develop a detailed 3-phase load flow model of the circuit. The load on the circuit is modelled in detail, and the assumptions about the phasing assignment of the load are modified and discarded if they do not conform to the current measured at the substation. Voltages are also modelled, and these voltages are compared to measurements in the field to further limit the possible assignments. A useful feature of this approach is that it also provides a confidence measure along with the phasing, as an output of the Monte-Carlo step.

4.6.5.2 Challenges

Distribution Transformer Configuration

This method generated separate phase to phase and phase to neutral phase predictions. However, transformer and meter metadata issues likely impacted the accuracy of these predictions.

Phase to Phase Substation Voltages

The phase to phase voltage calculation was used to estimate phase to phase voltages.

Solution Stability

The Monte Carlo method provides a probability associated with the phasing of each meter. However, it seemed that the probability was not accurate, as there were many meters with inaccurate phases predicted to have 100% probability of being correct. If this probability was more accurate, it could support mechanisms to prevent uncertain predictions from being used to update records in PG&E's database.

Low Resolution Meters

As with other methods, this method was negatively impacted by low resolution meters. No solution for this challenge is provided by this method.

Meter Voltage Measurement Interval

No sensitivity to this was evaluated, but this method is likely not significantly impacted by the voltage interval, as no time-based calculations are applied to the data.

Sensitivity to Connectivity Errors

A core part of this model is that it relies on asset and connectivity models to develop the model of possible outcomes. This model can be impacted by physical characteristics of the system which are currently not well documented or have inaccuracies. In the case of the circuits evaluated in this project, there may have been too many errors in the underlying data and assumptions to get the level of accuracy needed to isolate the phasing. This issue will likely be present system wide.

Additional Data Requirements

The method was unable to use the provided voltage measurements in the power flow approach. The solution used only load matching, which has too many possible solutions to succeed alone. To make use of the voltage measurements in this method, the following additional data would be required:

1. Average voltage measurements for the AMI 1-hour period, not a single measurement during the period
2. kW, kVAR, volts per phase at the circuit breaker, as opposed to current measurement of total across the phases
3. A longer period of sample data

4.6.5.3 Results and Observations

The results for the LFMC model are shown in Table 23.

Table 23: LFMC Phase ID Results

	3-wire circuit 1	3-wire circuit 2	3-wire circuit 3 *	4-wire circuit 4	4-wire circuit 1 *	4-wire circuit 2	4-wire circuit 3
LFMC	36.5%	33.6%	55.4%	35.1%	Not available	30.4%	56.7%

*Training Circuits

In the case of 4-wire circuit 1, the power flow model did not work, which may be due to the presence of a large actively managed battery resource on that circuit. Although the approach of this method looked promising, the results for this method were poor.

4.6.6 Impedance Corrected Distance Mapping (ICDM)

4.6.6.1 Technical Development and Methods

This method used GIS location data to correct for secondary voltage drop in the AMI. Then rather than use a clustering approach, this method assigned the phases to each meter by matching it to the labeled SCADA using the minimized distance.

4.6.6.2 Challenges

The participant using this method stated that integer voltages and hourly frequency were disruptive to the ability of the method to resolve the phases. In addition, the method was deemed unreliable by the participant for two of the circuits, so predictions were not provided for those circuits.

Distribution Transformer Configuration

Like other methods, for 4-wire circuits, this implementation did not differentiate between phase to phase and phase to neutral transformers, and only predicted phase to neutral results.

Phase to Phase Substation Voltages

It is not known how this method addressed this challenge.

Solution Stability

It is not known if any mechanisms are used to improve the stability of the solution. Using distance to the substation phases may be more stable than using clustering methods.

Sensitivity to Connectivity Errors

Meter-to-transformer connectivity errors may impact the impedance estimate. It is not known if lateral constraints were enforced in the solution.

Low Resolution Meters

As with other methods, this method was negatively impacted by low resolution meters. The participant called this out as the most significant impact on their accuracy. No solution for this challenge is provided by this method.

Meter Voltage Measurement Interval

No sensitivity to this was evaluated, but the participant claimed that this method was impacted by hourly frequency of the data.

4.6.6.3 Results and Observations

This method performed very well for 3-wire circuit 3 and 4-wire circuit 1, but it is noted that these are the circuits where results were made available to the participants. Also, the participant noted that they had used the data on those circuits to get better results. Results for 3-wire circuit 2 and 4-wire circuit 4 were not provided, as the vendor said there were problems with the source, data which resulted in unreliable solutions.

Table 24: ICDM Phase ID Results

	3-wire circuit 1	3-wire circuit 2	3-wire circuit 3 *	4-wire circuit 4	4-wire circuit 1 *	4-wire circuit 2	4-wire circuit 3
ICDM	50.2%	Not available	94.2%	Not available	81.8%	59.3%	67.8%

*Training Circuits

4.6.7 Phase ID Results Summary

Summary of Phase Identification Methods Results

Table 25summarizes the results of all methods employed in the small-scale demonstration. Overall, none of the methods consistently provided more than the 70% accuracy that had been targeted as the performance threshold to warrant full-scale deployment. However, two methods (ERVVCC and EVC) have the highest potential for future full-scale deployment.

Table 25: Summary of Phase ID Small Scale Demonstration Results

	3 Wires - 12 kV			4 wires - 21kV			
Methods	**3-wire circuit 1**	**3-wire circuit 2**	**3-wire circuit 3 ***	**4-wire circuit 4**	**4-wire circuit 1 ***	**4-wire circuit 2**	**4-wire circuit 3**
ICDM	32.0%	30.7%	73.2%	76.3%	Not Available	Not Available	Not Available
t-SNE CHC	50.2%	Not available	94.2%	Not available	81.8%	59.3%	67.8%
ERVVCC	64.8%	52.5%	78.2%	78.2%	83.6%	71.6%	75.7%
LFMC	36.5%	33.6%	55.4%	35.1%	Not available	30.4%	56.7%
EVC (revised)	68.0%	42.2%	72.9%	64.5%	89.8%	58.6%	63.6%

*Training Circuits

4.7 Meter-to-Transformer Methods

Another potential use of SmartMeter™ data, beyond Phase ID, is the automatic mapping of meters to their distribution transformers. Accurate meter-to-transformer connectivity information is needed to ensure proper transformer loading levels.

In PG&E’s customer database, records are kept for meter-to-transformer connectivity, but there can be several reasons for errors in those records:

- When customers are calling to start their service, there may be data entry errors in recording the correct transformer number.
- During an outage, as the priority of the crew on the field is to restore power safely and in a timely manner, customers could be connected to a different transformer than previously, but the new transformer association may not be updated.

- Due to increased load, a new transformer may be installed, and a subset of the meters moved to the new transformer, without the records being accurately updated. This results in one transformer which appears to be empty, and one which appears to be overloaded.

In the small-scale demonstration, as previously described, PG&E collected field data for three 3-wire and one 4-wire circuits. Both the phase and the meter-to-transformer association were collected.

Table 26 summarizes the comparison between the data collected in the field and the data currently found in GIS on overhead transformers. The overall accuracy is 93.8% for overhead transformers on the four circuits. 54 meters were identified to be the wrong circuit, and 321 meters were assigned to the wrong transformers. In total, 375 meters were not connected as expected.

Table 26: Field Data Versus GIS - Meter-to-Transformer Assignment

GIS vs Field	3-wire circuit 1	3-wire circuit 2	3-wire circuit 3	4-wire circuit 4	Total
Correct in GIS	2,040	1411	591	1,606	**5,648**
Incorrect in GIS	69	83	56	113	**321**
Meters on the wrong circuit	Not applicable	Not applicable	Not applicable	Not applicable	**54**
Accuracy	96.7%	94.4%	91.3%	93.4%	**93.8%**

Generically, the approach of most algorithms evaluated is to use a combination of geographical location and voltage readings to group meters, and then assign them to transformers. Those approaches should not only identify the 375 meters that are not connected where expected but also to re-assign the 321 meters to the right transformers. As the data shared are only on the four meters considered, a re-assignment of the 54 meters on the wrong circuits was not expected.

PG&E explored its own internally-developed method as well as several external methods for meter-to-transformer mapping:

PG&E Method

Candidates for reassignment were developed based upon inconsistent metadata or outlier voltage on a single transformer. Outliers and split transformers were determined using DBSCAN segmentation on the Voltage measurements. Reassignment was evaluated based on similarity of secondary voltage drop corrected voltage with meters on nearby transformers, or nearest transformer if it was empty.

Neighborhood Dimension Reduced Cluster Assignment (NDRCA) Method

A combination of geographic and electrical distance was used to reduce the set of potential transformer assignments. Within the potential transformer assignments, metrics were developed for feature-reduced voltage similarity and combined with the geographic and electrical distances to produce an aggregate likelihood of assignment.

Impedance Adjusted Geographic Match (IAGM) Method

This participant searched nearest transformers with impedance adjusted voltage to find the best match with nearby transformer voltages.

Voltage Clustered Dimension Reduced Cluster Assignment (VCDRCA) Method

Global customer clustering was performed based on voltage magnitude measurements. For each local area, meters were clustered based on nonlinear dimension reduction mapping. Clusters that shared common meter(s) were merged. Each cluster's centroid was calculated and the cluster was assigned to the nearest transformer.

Metadata Inconsistency Flagging (MIF) Method

Meter-to-transformer connectivity records were flagged as miss-assignment candidates based on meter and transformer locations as well as transformer and meter configuration records. In addition, overloaded transformers were flagged. No proposed assignments were provided by this method, but graphic tools could be provided to support customer to transformer reassignment.

4.7.1 PG&E Method

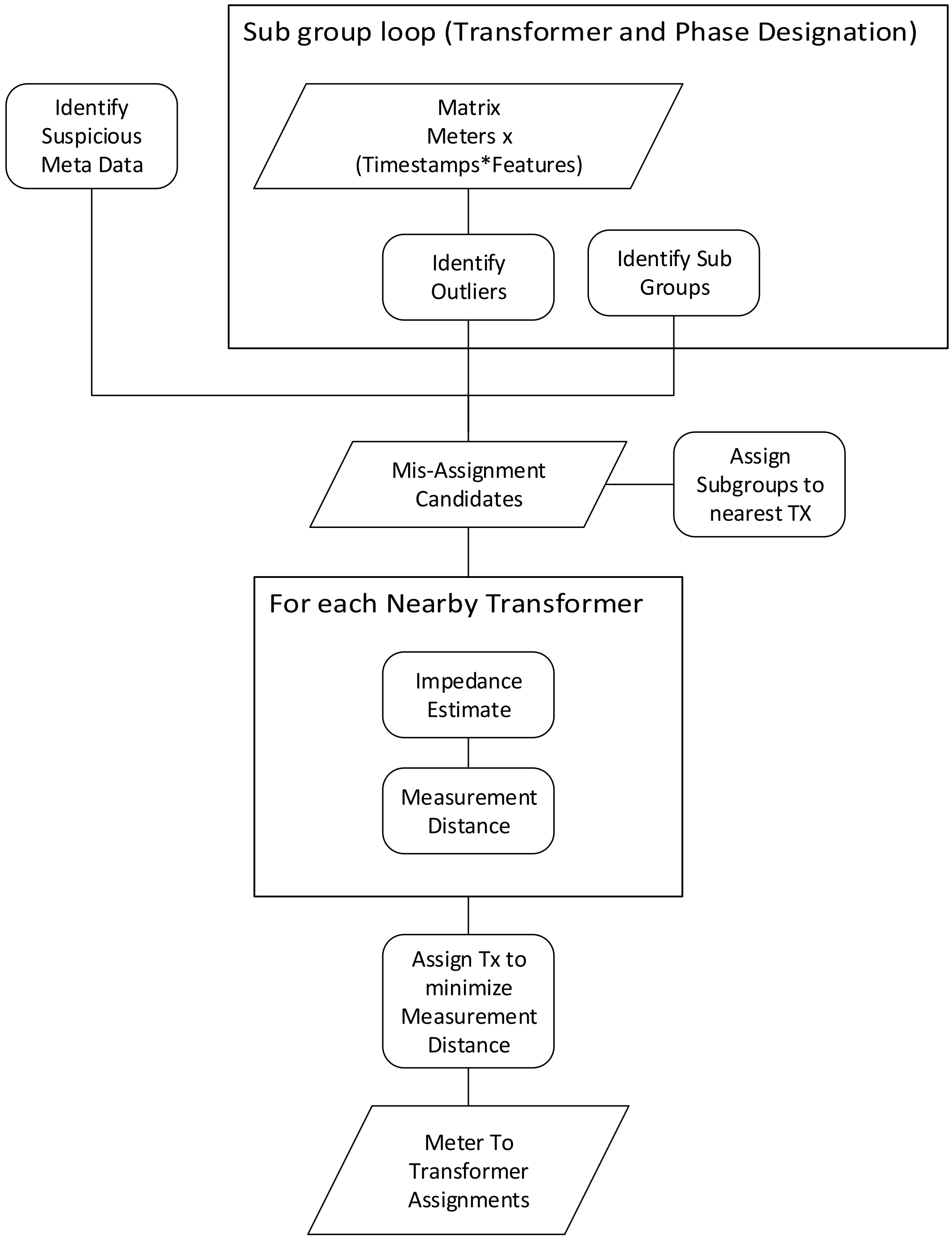
4.7.1.1 Technical Development and Methods

The approach taken by PG&E, represented in Figure 20, attempts to develop a subset of reassignment candidates and to only try to reassign those candidates. Candidates for reassignment were developed using two methods.

1. Identify suspicious assignments using transformer and meter metadata:
    a. Meter observed nominal voltages incompatible with transformer low side voltage. For example, a 120/240 Three Phase transformer cannot have a 120/208 meter on it.
2. Identify suspicious assignments using Voltage Data
    a. DBSCAN clustering was applied to voltages on a given transformer and phase, and was used to detect transformers with voltage outliers or with distinct subgroups which might indicate a transformer which had some subset of the meters reassigned to a new or alternate transformer to prevent overloading.

For each suspect assignment, a set of candidate transformers were identified based on either distance to the meter, or to the previously assigned transformer. If the nearest transformer was compatible and empty, then the assignment was made to that transformer. Otherwise, for each candidate, the transformer low side voltage was estimated given the load and distance between the meter and the transformer, and the average Euclidean distance from existing meters assigned to that transformer was evaluated. If a substantially better matching transformer was found, then the meter was re-assigned.

Figure 20: PG&E Method Conceptual Diagram



4.7.1.2 Challenges

Secondary Voltage Drop

The core of this problem is to try to find similarities between meters on the same service transformer. However, there are no measurements of the voltage on the transformer, and the measurements at the meters experience a voltage drop as a function of their local load. This secondary voltage drop can be a substantial impact, so ideally it could be corrected for. However, to do this accurately would ideally require the impedance and detailed topology of the secondary as well as power factor and per phase voltage and current measurements of each meter. Of these measurements, only the voltage and total interval energy is currently available.

For much of PG&E's system, there is not accurate, detailed mapping of the secondary connections between the service transformer and the customer meter. There are some records for underground secondary service, but for above-ground connections, records of the path of the secondary conductor and the conductor material are not systematic. This causes challenges in estimating the impedance between the customer and the transformer, and in the ability to correct for the voltage drop between the transformer and the meter.

This method addresses this change by utilizing the 'Virtual Transformer Voltage' method described previously, by estimating the voltage for the transformer currently assigned, and also for the nearby transformers considered for the assignment. When reassignments are considered, the Virtual Transformer Voltage is recalculated using that prospective secondary configuration.

Phasing

For three-phase transformers on an unbalanced system, the readings from the different phases on a transformer can be dissimilar if the phase voltages are different. The current phasing effort is focused on single phase transformers, so as phasing is determined this can be used to support the meter-to-transformer mapping, but this is not known for the three-phase meters. Because the phasing is not known for three phase transformers, this can cause challenges in knowing which voltages to compare in the meter-to-transformer mapping for three phase meters.

Inaccurate Meter and Transformer Locations

In most cases, the location of the service point is recorded at the meter location, so the distance between the meters and the transformer is a useful piece of information. However, in some cases the location of the meter can be inaccurate due to data entry errors. In some cases, these records appear to have been corrected by GIS analysts to use the location associated with the service address, but for service areas on large plots of land, the location of the address can be quite distant from the actual metering point.

This method addressed this problem by replacing meter locations which were very distant from their assigned transformer with the location of the transformer.

Three-phase Transformers

Because three-phase current or load measurements are not available, it is not currently possible to accurately correct three-phase meters for secondary voltage drop, as one phase may have a much larger load than the others. In addition, the three phases make it more difficult to match voltage measurements between meters on a transformer, because the phasing is not known.

This method addressed this by assuming that phase load on polyphase meters is proportional to voltage level. In addition, the Virtual Transformer Voltage optimization is modified to optimize for polyphase meters. This approach helps, but does not seem to be sufficient. It would be extremely helpful to this method to have per-phase instantaneous current or power.

Empty Transformers

When a transformer does not have any meters associated, it does not provide any reference voltages to compare against, and only proximity can be used. There can be a few different potential causes for this scenario; in some cases, the transformer serves unmetered load, like street lights, or PG&E's equipment. However, this can also occur when a new transformer has been added to address an overload condition without updating the meter-to-transformer records. In many cases the transformer is recorded as having unmetered load, but when these records are not accurate they cause challenges to algorithmic reassignment. This method attempted to address this by assigning meters to the nearest transformer if they were found to be outliers and had an empty transformer nearest to them. In general, this method will not perform as well as cluster based methods in these situations.

4.7.1.3 Results and Observations

The results for the PG&E method are shown in Table 27.

Table 27: PG&E Meter-to-Transformer Reassignment Results

	No Change Predicted	Change Predicted
Correctly Assigned in GIS	5487	161
Incorrectly Assigned in GIS	322	53
Correct change		37

322 of the 375 incorrectly assigned meters were not flagged by the algorithm for a change, which means that 85.9% of the changes needed were not identified. Of the 214 changes suggested, only 53 were correct and 37 of them were correctly changed. Overall, by only applying this algorithm, the accuracy of GIS will drop from 93.8% to 91.7%. However, if changes were field verified before implementation, and 53 incorrect records were updated, the overall accuracy in GIS would increase to 96.3%.

4.7.2 NDRCA Method

4.7.2.1 Technical Development and Methods

The NDRCA method was developed from a starting point of the method described in (Mitra, et al., 2015). The stages of the original algorithm are as follows:

1. The population of meters is split into "neighborhood" groups using the DBSCAN distance-based clustering algorithm and the geographic locations of the meters and transformers.
2. K-Means clustering is applied to the voltages of the meters in each neighborhood group to create n clusters, where n is the number of transformers in the group.
3. Each cluster is assigned to the transformer nearest to the geographical centroid of the cluster.

This method differs from the method described in Mitra in a few ways. In the first step, a combination of geographical distance and electrical distance is used to develop the neighborhood groups. Prior to the second step, a dimensional reduction algorithm is applied to the voltage. Finally, instead of reassigning automatically based upon the results of the cluster proximity, a metric is developed using the relative geographical and voltage distance of assignment options. This ratio is used as a threshold to determine whether to apply the reassignment.

4.7.2.2 Challenges

Secondary Voltage Drop

As with other methods, the secondary voltage drop can cause meters on the same transformer to look dissimilar. As this method was external, it is not known if this issue is addressed. It may be that the dimensional reduction method is effective in reducing the impact of this.

Inaccurate Meter and Transformer Locations

As with other methods, this approach is challenged by inaccurate geographic locations in the meter or transformer assets. However, the additional use of the ‘electrical’ distance in the neighborhood grouping seems to be effective in mitigating some of these challenges.

Three-phase Transformers

It is not documented how this method addresses the challenges of three phase transformers.

Empty Transformers

Because there is no dependency on existing assignments, this method is more effective than the PG&E method at finding meter-to-transformers assignments on transformers which have no existing meters assigned to them.

4.7.2.3 Results and Observations

The results for the NDRCA method are shown in Table 28.

Table 28: NDRCA Meter-to-Transformer Reassignment Results

	No Change predicted	Change Predicted
Correctly assigned in GIS	5511	137
Incorrectly Assigned in GIS	244	131
Correct change		**116**

244 of the 375 meters were not flagged by the algorithm for a change, which means that 65.1% of the changes needed were not identified. Of the 268 changes suggested, 131 were correct and 116 of them were correctly changed. Overall, by only applying this algorithm the accuracy of GIS will drop from 93.8% to 93.7%. However, if changes were field verified before implementation, and 131 incorrect records were updated, the overall accuracy in GIS would increase to 97.6%.

4.7.3 IAGM Method

4.7.3.1 Technical Development and Methods

In the IAGM method, meter voltage was first corrected to estimate the voltage at the low side of the transformer, using the distance from the meter to the transformer and an assumption about the resistivity of the secondary conductor. The algorithm then uses an undisclosed method to search the nearest transformers with impedance adjusted voltages to find best match with nearby transformer voltages.

4.7.3.2 Challenges

Secondary Voltage Drop

As with other methods, the secondary voltage drop can cause meters on the same transformer to look dissimilar. This method used a proprietary voltage correction to address this challenge; however, because this correction relies upon assumptions about the conductor to estimate this drop, there may be errors.

Inaccurate Meter and Transformer Locations

As with other methods, this approach is challenged by inaccurate geographic locations in the meter or transformer assets. It is not known how these issues were addressed by this method.

Three-phase Transformers

It is not documented how this method addresses the challenges of three phase transformers.

Empty Transformers

Like the PG&E method, this method likely struggles with assignments to transformers which do not have existing meters assigned to them. It is not known if the method has any mechanisms to address this.

4.7.3.3 Results and Observations

The results for the IAGM method are shown in Table 29.

Table 29: IAGM Meter-to-Transformer Reassignment Results

	No Change predicted	Change Predicted
Correctly assigned in GIS	5630	18
Incorrectly Assigned in GIS	354	21
Correct change		**17**

354 of the 375 meters were not flagged by the algorithm for a change which means that 94.4% of the changes needed were not identified. Of the 39 changes suggested, 21 were correct and 17 of them were correctly changed. Overall, by only applying this algorithm the accuracy of GIS will remain about the same since very few changes were suggested by this method. However, if changes were field verified before implementation, and 21 incorrect records were updated, the overall accuracy in GIS would increase to 95.8%.

4.7.4 VCDRCA Method

4.7.4.1 Technical Development and Methods

This method is composed of four separate steps:

1. Perform global customer clustering based on voltage magnitude measurements.
2. For each local area, cluster the meters based on nonlinear dimension reduction mapping.
3. Merge the clusters that share common meter(s).
4. Calculate each cluster's centroid and assign the cluster to the nearest transformer.

4.7.4.2 Challenges

Secondary Voltage Drop

As with other methods, the secondary voltage drop can cause meters on the same transformer to look dissimilar. It is not known whether this method had any mechanism to address this.

Inaccurate Meter and Transformer Locations

As with other methods, this approach is challenged by inaccurate geographic locations in the meter or transformer assets. The participant called this out as a key challenge to success in this algorithm, it was suggested that the utilization of address data to geolocate the service points and find geographical inconsistencies could improve this situation.

Three-phase Transformers

This method addresses three-phase transformers by using the step of clustering by voltage, and then merging clusters which have overlapping meters, because those meters must be polyphase and on three phase transformers.

Empty Transformers

Because this method uses clustering and geographic proximity, it should be able to map to empty transformers successfully.

4.7.4.3 Results and Observations

The results for the VCDRCA method are shown in Table 30.

Table 30: VCDRCA Meter-to-Transformer Reassignment Results

	No Change predicted	Change Predicted
Correctly assigned in GIS	5392	256
Incorrectly Assigned in GIS	312	63
Correct change		**35**

However, if changes were field verified before implementation, and 63 incorrect records were updated, the overall accuracy in GIS would increase to 96.5%.

4.7.5 MCIF Method

4.7.5.1 Technical Development and Methods

Meter-to-transformer connectivity records were flagged as miss-assignment candidates based on meter and transformer locations as well as transformer and meter configuration records. In addition, overloaded transformers were flagged. No proposed assignments are provided by this method, but graphic tools could be provided to support customer to transformer reassignment.

4.7.5.2 Challenges

Secondary Voltage Drop

Voltage is not used in this method, so it is not impacted by this.

Inaccurate Meter and Transformer Locations

This method flags meters and transformers which are unusually far from their transformers. The method may be challenged when the distance is inaccurate, but does not exceed the distance threshold. It may also inaccurately flag meter-to-transformer mapping where the mapping is correct, but one of the locations is inaccurate.

Three-phase Transformers

This method is not impacted by this challenge.

Empty Transformers

No reassignments are suggested, so this is not relevant.

4.7.5.3 Results and Observations

The results for the MCIF method are shown in Table 31.

Table 31: MCIF Meter-to-Transformer Reassignment Results

	No Change predicted	Change Predicted
Correctly assigned in GIS	5626	22
Incorrectly Assigned in GIS	368	7
Correct change		**N/A**

368 of the 375 meters were not flagged by the algorithm for a change which means that 98.1% of the changes needed were not identified. Of the 29 changes suggested, 7 were needed. If changes were field verified before implementation, and 7 incorrect records were updated, the overall accuracy in GIS would increase to 95.6%.

4.7.6 Meter-to-transformer Results Summary

Table 32 combines the results for all of the meter-to-transformer mapping methods. The table shows that NDRCA is the best performing method.

Table 32: Meter-to-Transformer Methods Results for All Four Circuits

Methods	Accuracy with only algorithm	Accuracy with Algorithm and Field Verification	# changes needed but not predicted	# reassignment suggested
VCDRCA	90.20%	96.48%	312	319
IAGM	93.76%	95.78%	354	39
PG&E	91.75%	96.32%	322	214
MIF	93.43%	95.55%	368	29
NDRCA	93.71%	**97.61%**	**244**	**268**
GIS	**93.80%**	93.80%	375	N/A

While NDRCA is the method with the highest potential, for implementation PG&E had expected to employ a method that would be able to identify almost all the changes needed.

None of the methods managed to identify all the changes needed. Meter-to-transformer mapping is a very challenging task, and with the lack of information at the transformer level, it is difficult to identify incorrect assignments when there are only one or two meters on one transformer. For the same reason, the reassignment of the meter to the correct transformer is also difficult, therefore if an algorithm was deployed to obtain a list of meters likely incorrectly assigned, corresponding field verification would be needed, to identify the meter's correct location. By running the algorithm and then conducting subsequent field verification, the NDRCA method would have improved the GIS accuracy from 93.8% to 97.6%.

5 Value Proposition

The purpose of EPIC funding is to support investments in TD&D projects that benefit the electricity customers of PG&E, SDG&E, and SCE. In California, all the utilities are facing the same meter connectivity issues and have a similar grid configuration. EPIC 2.14 – *Phase ID* has demonstrated voltage data from SmartMeters™ can be utilized effectively by multiple algorithms to map the phases of meters and transformers, thereby improving the reliability and safety of the grid at a reduced cost, relative to industry standard practices. It also demonstrated that voltage data is potentially useful for mapping meter-to-transformer connectivity, and although the reassignment of the meters is not conclusive, identifying incorrectly mapped meters alone is an improvement.

5.1 Primary Principles

The primary principles of EPIC are to invest in technologies and approaches that provide benefits to electric ratepayers by promoting greater reliability, lower costs, and increased safety. This EPIC project contributes to these primary principles in the following ways:

Greater Reliability

This project contributes to this principle by demonstrating, at a POC level, that automated phase asset mapping through algorithmic analysis can be achieved at a satisfactory accuracy level. This capability is critical to improving engineering practices for phase balancing that achieves greater reliability across PG&E's system. Improved phase balancing on the distribution network, and the resultant reliability improvements across the network, support the requirements for unbalanced power flow and state estimation in the proposed ADMS, and DERMS platforms.

Lower Costs

This project contributes to this principle because the cost of automated phase asset mapping through algorithmic analysis using data currently available is more economic than using physical resources and tools on the ground—notably avoiding a recurring high cost activity. This project recommends further full demonstration testing that will help determine whether the analytical methods for automated phase asset mapping can provide PG&E with improved efficiency over time for electric operations across the service territory, and to maintain and improve electric service affordability for customers.

Increased Safety

This project contributes to this principle by proving that granular visibility into distribution network phasing over time using SmartMeter™ voltage data is possible. When combined with other tools and techniques into a targeted operational use case, this capability could enable detection of broken wires with greater accuracy and support the restoration of power faster when an outage occurs.

5.2 Secondary Principles

EPIC also has a set of complementary secondary principles. This EPIC project contributes to the following four secondary principles: societal benefits, greenhouse gas (GHG) emissions reduction, economic development; and efficient use of ratepayer funds.

Societal Benefits:
Improved phase balancing achieves a lower cost of service through a reduction in losses and outages, lessening the overall cost of electric service to all customers, and extended to society through shared economic interests for efficiency improvement.

GHG Emissions Reduction:
Improved phase balancing could be achieved by future deployment of an automated phase asset mapping method that has been tested and proven on PG&E's full system. Better Phase ID can help accommodate the interconnection of renewable DER. These non-fossil fuel-based power generation technologies can facilitate GHG emissions reduction by offsetting fossil fuel-based generation on PG&E's system.

Economic Development:
Improved phase balancing achieved through automated phase mapping algorithms that rely on data from the SmartMeter™ Network have the potential to influence new markets for more efficient devices, new sensing and communications capabilities, and culminated in PG&E's vision for the Grid of Things™. Enabling DER adoption by reducing the impacts of those devices on the distribution system through phase balancing improves their value proposition and market potential.

Efficient Use of Ratepayer Funds:
This project demonstrated that the existing SmartMeter™ Network can be used to support automated phase asset mapping through algorithmic analysis. This lesson indicates that PG&E might not need to deploy field resources or rely on more costly communications solutions to achieve phase mapping. Through the full system method testing recommended by this project, PG&E will determine whether there are compelling cost advantages for the company and its customers by deploying an automated phase asset mapping method tested in this project. By carefully identifying, testing, and recommending cost-effective methods for phase mapping that leverage the SmartMeter™ Network, this project has taken a productive step toward potential savings for ratepayers.

5.3 Accomplishments and Recommendations

5.3.1 Key Accomplishments

The following summarizes the key accomplishments of the project over its duration:

Evaluated Potential Solutions to Solve Meter Connectivity Issues:
This project evaluated a set of potential approaches to Phase ID and meter-to-transformer mapping based on a literature review informed by technical literature, industry meetings and working group reports, patents, conference proceedings, journals, vendors, and white papers. This project considered LiDAR mapping technology, µPMUs, and hardware at the transformer. The methods were evaluated using a set of criteria developed by the project team to guide the project approach, keeping in mind the greatest benefit to PG&E, its customers, and other utility operations today. It was established by PG&E that the best possible long-term solutions would be driven by data analytics.

Tested new Technology for Field Data Collection:

A novel wireless Phase ID hardware tool that does not require direct interaction with live wires was evaluated during the field verification. The use of this tool lowered the cost of the field data collection and reduced risk due to the contactless nature of the tool with hot wires.

Tested Phase Identification Algorithms on 4-Wire Circuits:

PG&E collected field data on four 4-wire (21kV) circuits and evaluated eight methods from six different organizations (including PG&E and one university) via two demonstrations. In total, 3,700 meters were chosen to build the algorithm and about 14,000 meters were used to validate the methods built for this project. Several methods provided results with an accuracy level potentially high enough for implementation.

Tested Phase Identification Algorithms on 3-Wire Circuits:

PG&E collected field data on three 3-wire (12 kV) circuits and evaluated five methods from four different organizations (including PG&E and one university). One 3-wire circuit dataset was utilized to calibrate each model (data corresponding to approximately 1,500 meter points), and the remaining two 3-wire circuits were used to validate each model approach (approximately 5,000 meter points). The model training process used field results from the first circuit to calibrate model parameters, while the field results from the testing set were not used until the final validation assessment. Results obtained had a lower level of accuracy than for 4-wire circuits, however the results are promising and performance sufficient for implementation could be achieved with a few input data enhancements, such as receiving phase to phase Substation voltage data.

Tested Meter-to-Transformer Algorithms:

PG&E collected field data on three 3-wire (12 kV) and one 4-wire (21kV) circuits and evaluated five methods from five organizations (including PG&E and one university). Data from one of the 3-wire (12kV) circuits was used to create each model (corresponding to 1,500 meters) and the remaining three circuits (about 8,000 meters) were utilized to validate the models built. Although none of the methods evaluated demonstrated sufficient performance in identifying incorrect assignments, this allowed PG&E to better understand the challenges linked to this issue.

Identified Data Challenges:

This project was one of the first to combine AMI, SCADA and GIS data, and it revealed numerous data challenges, and also helped to identify a path forward to solve some of those challenges in support of deploying automated methods like those explored in this project.

Better Defined Circuit Imbalance Challenges:

One characteristic which may impact the performance of the phase identification algorithms is current and voltage imbalance. In (Wang, et al., 2017), it has been shown that there is evidence of some relationship between the level of current imbalance on a circuit and the performance of phase ID algorithms. Wang shows that circuits with a lower current imbalance have a reduced accuracy when using the phase identification algorithm in that paper. In their study, there is a roughly linear reduction in accuracy for circuits with current imbalance below 10%.

The formula for calculating the current imbalance at a given measurement time is:

$$100\% \, \frac{\Delta I_{max}}{I_{mean}}$$

where ΔI_{max} is the largest absolute current deviation from the mean out of the 3 phase currents, and I_{mean} is the average current magnitude of the three phases. A similar formulation is used for voltage. In the circuits used in the study, it is possible to see a circuit which has high current imbalance, but low voltage imbalance, or vice versa. This is because there can be multiple circuits on a bus, and though the currents are separate for each bus, the voltage is shared. Several of the algorithmic approaches use voltage imbalance and variation to map the phases, and balanced voltages seem to cause some challenges to those methods, which can be counterintuitive when the circuit current is balanced. The current imbalance of the circuits is shown in Figure 21, and the voltage imbalance is shown in Figure 23. The circuit current imbalance vs the PG&E phase ID accuracy is shown in Figure 22. There are not many circuits and the correlation is not strong, so it is not conclusive, but there does appear to be some relationship. This could be an interesting relationship, because it could be that a 3-phase unbalanced load flow model might be less impacted by phase errors when the circuit is balanced.

Figure 21: Circuit Current Imbalance Summary

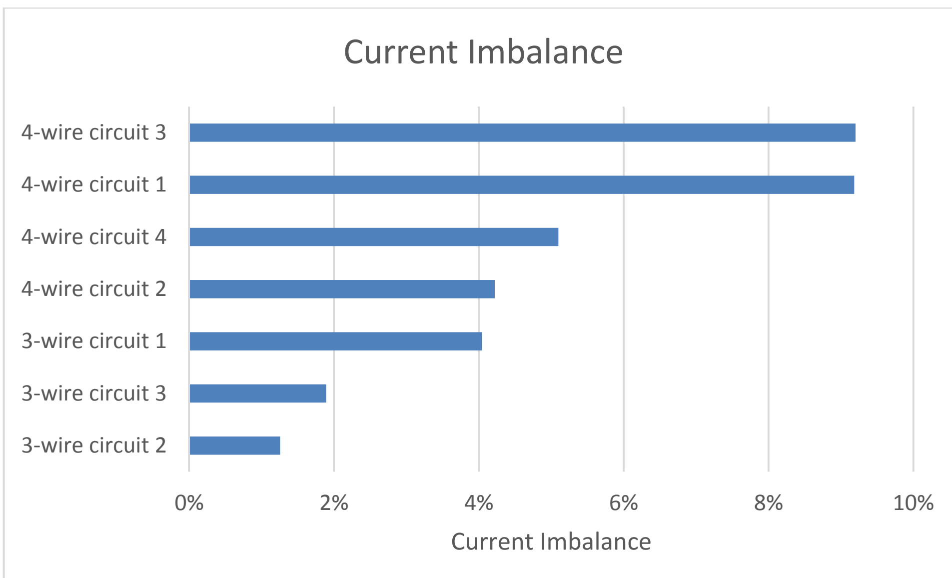


Figure 22: Current Imbalance Versus Accuracy

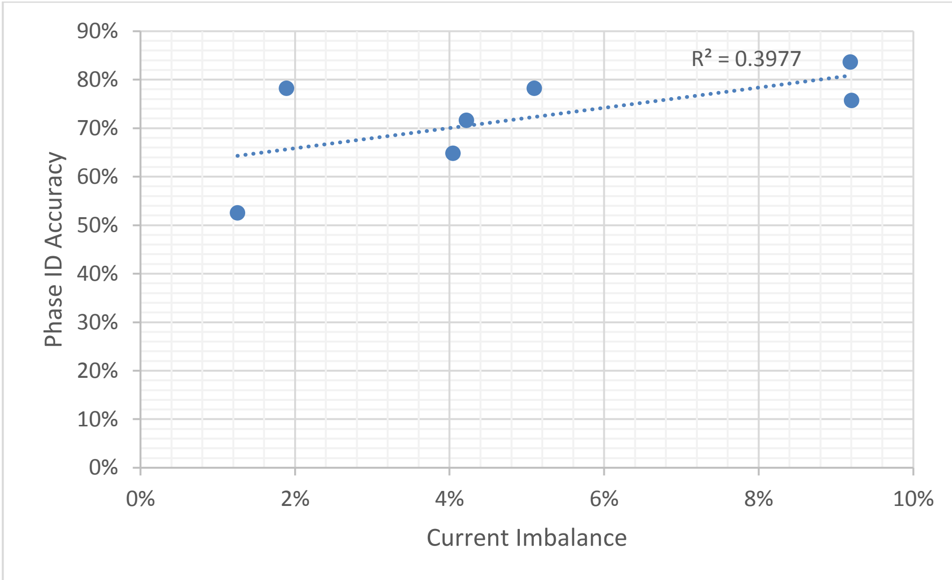
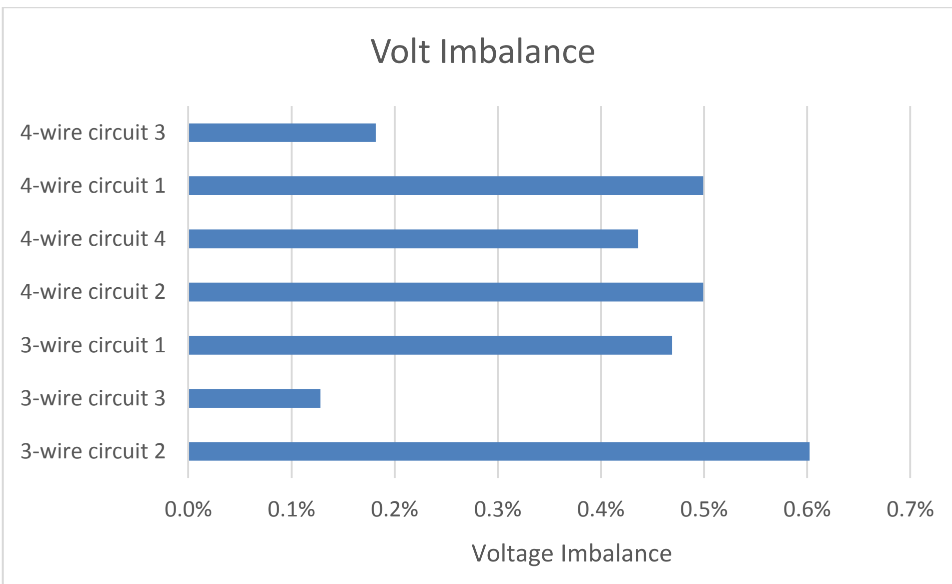


Figure 23: Circuit Voltage Imbalance Summary



Identified Full-Scale Deployment Gaps and Made Deployment Recommendations:

Two of the methods, the EVC and ERVVCC, are promising for implementation but further testing using additional data should improve the performance of those algorithms and allow PG&E to decide which algorithm to implement for full-scale deployment.

Evaluated Impact of Data Resolution on Phase ID Method Effectiveness:

The sensitivity of method accuracy to the frequency and precision of input data was evaluated. This informs the requirements of a wider-scale implementation of these methods. It should be noted that only some of the meter hardware types can be upgraded to collect higher precision voltage data, and that the proportions of each hardware type vary across the service territory.

5.3.2 Key Recommendations

This project has shown that there is significant merit in developing automated Phase ID solutions based on the methods explored in this project, and potentially other methods, using a scaled approach that would support distribution engineering applications and reduce expected expenditures related to boots on the ground approaches. An algorithm-based approach could also run at intervals to ensure the system model remains up-to-date with minimal marginal costs. Below are key recommendations to other utilities considering implementing similar Phase ID solutions which will help to make their implementations successful:

Employ Meters with Decimal Voltage Precision:

Section 4.5.4.3 demonstrated that voltage data with decimal precision significantly improved algorithm performance. Figure 24 shows the distribution of PG&E circuits based on their percentage of decivolt versus integer voltage data. Figure 25 shows this distribution for single phase meters only. Only single-phase meters were used for the evaluation of the phase ID methods. Most of PG&E’s 3-wire circuits have a very low proportion of decivolt meters compared to the 4-wire circuits. This contributed to the lower level of accuracy obtained with the algorithms for the 3-wire circuits.

Figure 24: Distribution of Circuit Deci-Volt Resolution

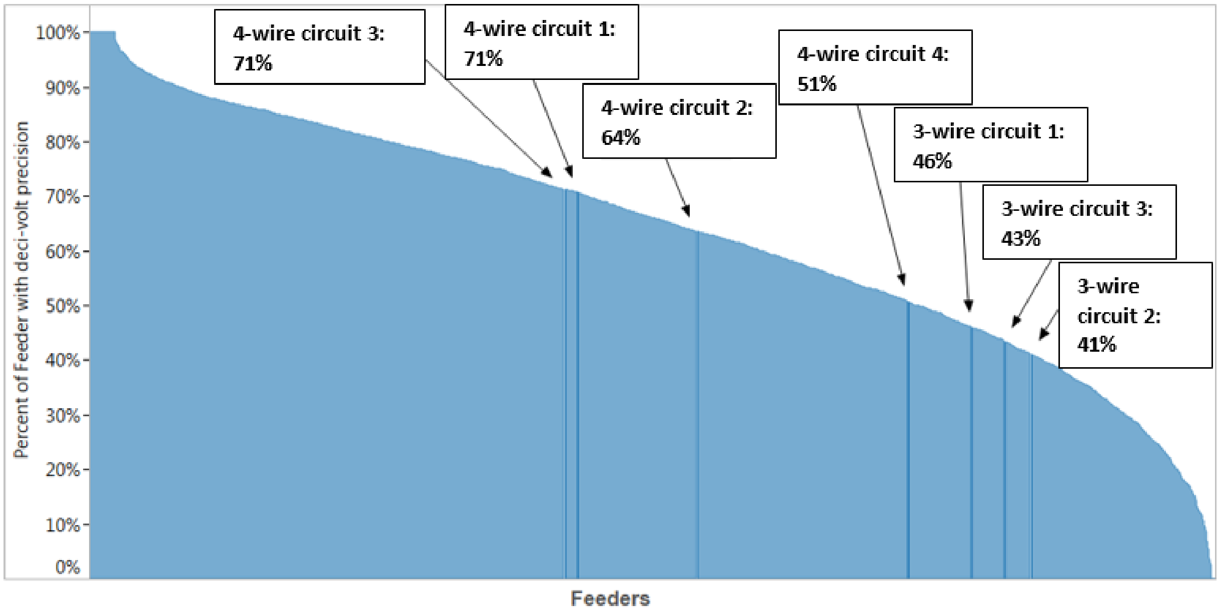
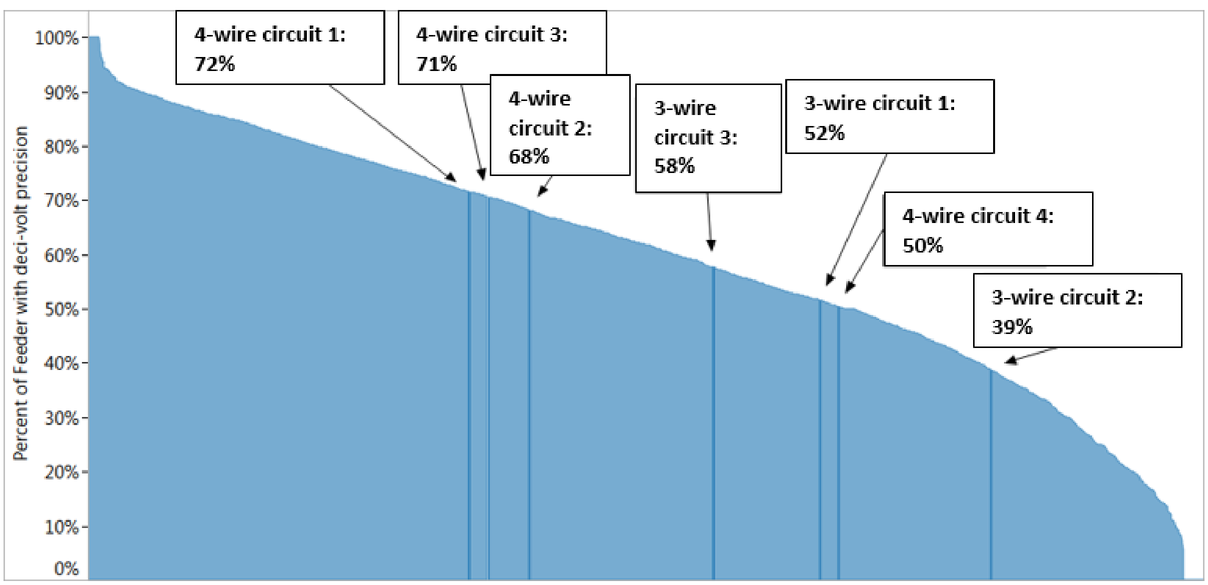


Figure 25: Distribution of Circuit Deci-Volt Resolution for Single Phase Meters Only



Most of the PG&E meters that do not have the capacity to provide decivolt precision are near their expected end of life, and are planned to be replaced over the next few years, but for the time being they continue to represent a significant portion of the fleet. However, over time all those meters will be replaced and the performance of PG&E's Phase ID algorithms will continue to improve.

Utilities pursuing similar Phase ID approaches should assess the performance sensitivity to their methods of voltage precision and consider planning to replace end of life meters with meters capable of providing decimal precision.

Collect Voltage Data in at Least 15-Minute Intervals:

In 2017, PG&E upgraded SmartMeter™ firmware to collect voltage data either hourly or every 15 minutes depending on the customer's rate. Figure 26 shows the distribution of circuits based on their proportion of 15-min vs 60-min interval data. The subset of single phase meters which have 15-minute data are shown in Figure 27. Due to the high proportion of hourly voltage data, only 60-minute interval usage could be used to run the algorithms. However, during this project it was proven that some methods, especially EVC, were impacted by the time resolution. The ability to collect AMI voltage data with a 15-min or 5-min frequency is expected to improve the overall accuracy of the algorithms, and utilities pursuing similar approaches should consider enabling at least 15-minute interval collection.

Figure 26: Distribution of Circuit 15-Minute Interval Meters as of 08/2018

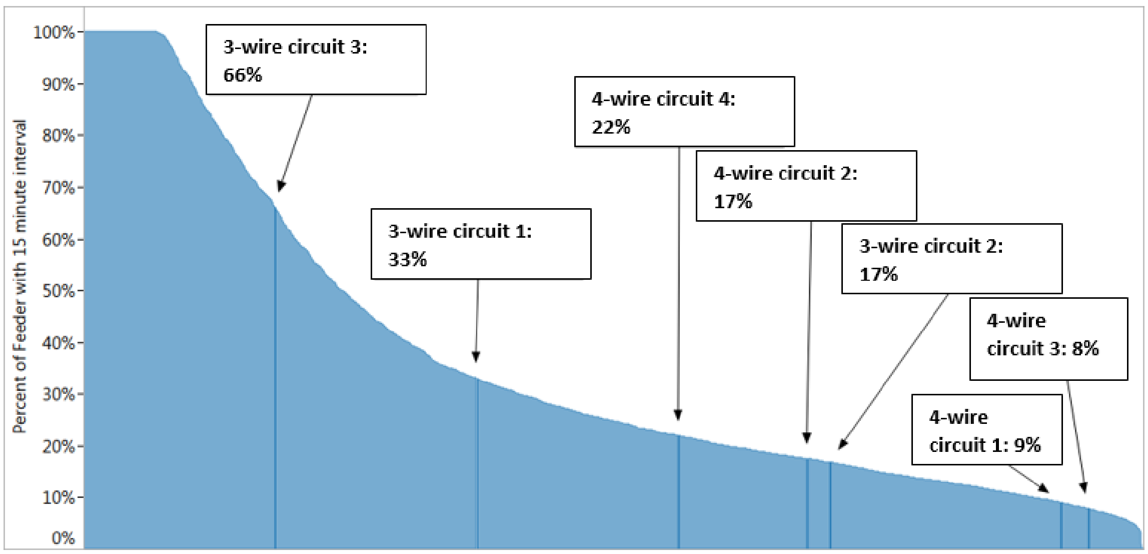
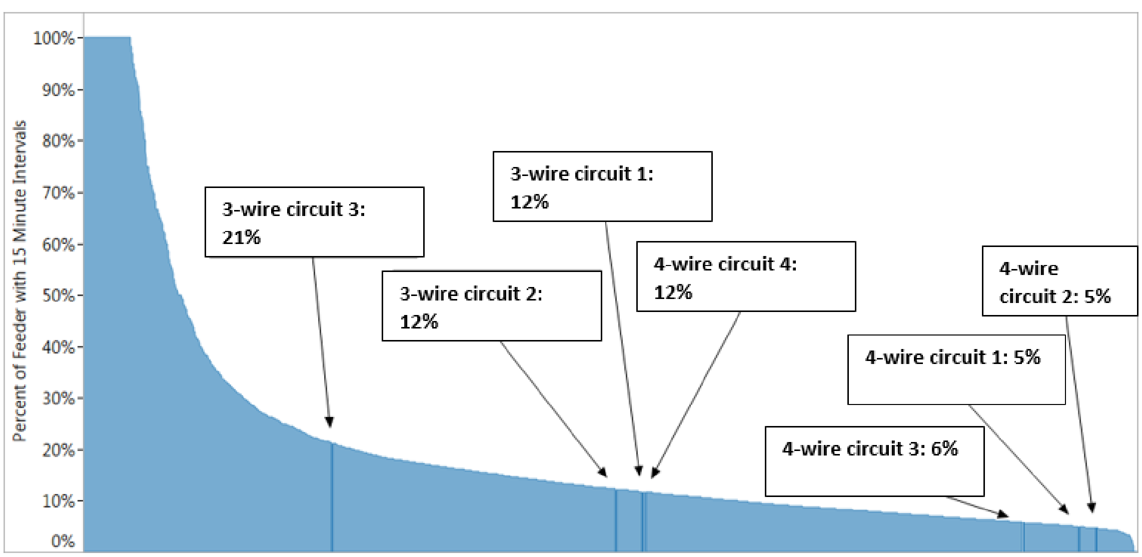


Figure 27: Distribution of Circuit 15-Minute Interval Data for Single Phase Meters as of 08/2018



Collect High Resolution SCADA Data:

As discussed in section 4.3.2, high resolution SCADA and historian system data is required to enable future deployment of Phase ID methods. It has been determined that most of these data resolution considerations can be addressed by updating the configuration of SCADA and historian change band and compression settings. Prior to this project, SCADA data was mostly used by PG&E's engineers to evaluate peak consumption and not typically used for applications that required this greater granularity. Utilities pursuing similar approaches should ensure that their SCADA systems are configured to provide high resolution SCADA data.

Ensure Accurate GIS Data:

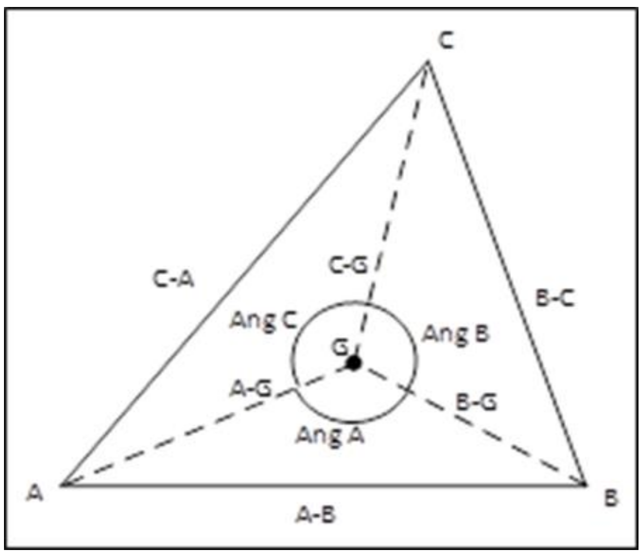
During this project, it was clear that some of the GIS data did not always reflect all of the conditions in the field. However, within PG&E a multi-year field asset inventory effort is being conducted that should verify transformer connectivity (Phase to phase or phase to neutral, three-phase or single-phase) and operating voltage. With this information, the performance of the algorithms could improve, and utilities pursuing similar Phase ID approaches should ensure that it is accurate in their systems. This information would also allow for 4-wire circuits to separate meters that are connected phase to phase with meters that are connected phase to neutral.

None of the methods evaluated in this project combined line to neutral with phase to phase solutions. As explained previously, 4-wire circuits (21kV) have a combination of line to neutral and phase to phase configuration, and due to inaccuracies in information on transformer connectivity configuration, only phase to neutral connections are predicted. When the proportion of phase to phase configuration is higher, this inaccuracy will have a larger impact on the performance of algorithms using this assumption.

Collect Phase to Phase Voltage Measurement at the Substation:

Algorithm performance for 3-wire circuits was significantly impacted by the fact that PG&E does not collect phase to phase voltage measurement at the majority of its distribution substations. As explained in section 4.6.2.1, the phase to phase voltage was calculated using the line to neutral voltage measurement. The assumption for this calculation is that all the phases are balanced, but in general that is not the case. Therefore, the phase angle is not equal to the 120 degrees used in the calculation. Unbalanced phase angle is depicted in Figure 28. For 4-wire circuits which have phase to phase measurements, this calculation will also be necessary to enable identification of phases for phase to phase configuration. PG&E does not collect this information for the majority of its distribution substations; however, some substations are equipped to collect this information. PG&E will conduct further tests subsequent to this project on some of these circuits with phase to phase voltage data to quantify the resulting improvements to the algorithms. Utilities with substations equipped to collect this information should also plan to apply it and assess its impact on their algorithms.

Figure 28: Unbalanced Phase Angles Drawing



Automate Phasing Record Updates:

As part of the implementation process for automated Phase ID, utilities should implement a process to allow for automated bulk updates to their phasing records with the output of Phase ID algorithms, while retaining all of the constraints required for maintaining electrical connectivity. Automating this process will reduce the manual workload associated with any periodic record updates after the initial effort.

Establish Target Accuracy Level for Phase ID Algorithms:

One of the challenges in this project was to determine the sufficient level of accuracy for Phase ID algorithms. Though the automated phase identification algorithms developed in this project show a clear improvement over the defaulted values currently populated in PG&E's GIS system, it is not known what PG&E's minimum requirement of accuracy is to achieve a stable solution of the unbalanced three phase model in the ADMS. It is recommended that at the onset of their analytical Phase ID efforts, utilities perform sensitivity analyses on unbalanced load flow models with varying levels of phase designation accuracy to understand their accuracy requirements. For PG&E, where DER penetration is projected to be high, the required Phase ID accuracy may be higher than for other utilities with lower projected DER penetration.

Establish Confidence Prediction for Phase ID:

Though it is anticipated that the accuracy of a utility's phase identification algorithms will improve as the various input data improvements listed above are addressed, there will continue to be some errors. It would be beneficial to have an indication of the probability of error in phase identification to support targeted field validation. Utilities pursuing automated Phase ID methods should develop and evaluate algorithms to incorporate this confidence prediction as part of their solutions.

5.4 Technology transfer plan

5.4.1 Path to Production

The planned end product is a cloud-based platform to run the Phase ID algorithm at scale across PG&E's service territory, and an associated interface to GIS to allow for the output of the algorithm to automatically update GIS' meter phasing records. To move towards this end product, the following activities would need to be conducted:

1. **Additional Algorithm Work:** Conduct evaluation of two Phase ID algorithms on additional circuits
2. **Load Flow Analysis:** Conduct load flow analysis to determine what level of meter phase assignment accuracy is needed for ADMS
3. **Input Data Improvements:** Implement input data improvements, including higher resolution Substation SCADA data and phase to phase voltage measurement at Substations. Two Phase ID methods did yield very promising results that would improve upon the accuracy of meter phasing records if deployed with the input data currently available, but additional improvements to input data are expected to further improve algorithm performance.
4. **Platform:** Build cloud-based platform for running Phase ID algorithm at scale in production
5. **GIS Interface:** Build interface to GIS to enable automated updating of records with algorithm's phase assignment output

There is no planned path to production for Meter-to-Transformer algorithms. The algorithms generally struggled to identify the majority of incorrect assignments, and even when incorrect assignments were identified, none of the algorithms effectively identified the correct re-assignments.

5.4.2 IOU's Technology Transfer Plans

A primary benefit of the EPIC program is the technology and knowledge sharing that occurs both internally within PG&E, and across the other IOUs, the CEC and the industry. In order to facilitate this knowledge sharing, PG&E will share the results of this project in industry workshops and through public reports published on the PG&E website. Specifically, below are information sharing forums where the results and lessons learned from this EPIC project were presented or plan to be presented:

Information Sharing Forums Held

2017 EPIC symposium

San Diego, California | October 2017

EPRI Grid Analytics and Power Quality conference 2018

Phoenix, Arizona | June 2018

Information Sharing Forums Planned

DistribuTECH 2019 Power Delivery Conference & Exp006F

New Orleans, Louisiana | February 2019

5.4.3 Adaptability to other Utilities and Industry

Utilities striving to modernize their distribution networks require improved visibility into both the physical state of their system and its real-time load flow conditions. As the load characteristics of the distribution network evolve such as with the growth of DER, it is becoming more important to have accurate and up to date information to be able to actively manage the distribution system to ensure reliability for customers.

Automated asset phase mapping through algorithmic analysis is a viable approach that other utilities should consider exploring for their systems. At project commencement, PG&E was not aware of any utilities evaluating automated phase mapping through algorithmic analysis at scale. Other utilities continue to explore similar phase mapping problems, with reports of some testing other physical approaches or specific analytical approaches unique to their systems; however, none had successfully demonstrated the optimal way to implement analytic Phase ID across all utilities. While additional work will be required after this EPIC project, it has demonstrated promising results for the future. However, the process is unique to each utility and requires a fitted approach in each case. PG&E first considered solutions that existed in the market, and none of them performed well enough, so PG&E created its own approach that ultimately performed the best. Consequently, the transferability of these results is limited to consideration by other utilities for testing on their unique systems subject to similar data requirements and other screening performed in the early phases of this project.

5.5 Data access

Upon request, PG&E will provide access to data collected that is consistent with the CPUC's data access requirements for EPIC data and results.

6 Metrics

The following metrics were identified for this project and included in PG&E's EPIC Annual Report as potential metrics to measure project benefits at full scale.[14] Given the POC nature of this EPIC project, these metrics are forward looking.

D.13-11-025, Attachment 4. List of Proposed Metrics and Potential Areas of Measurement (as applicable to a specific project or investment area)	Section
1. Potential energy and cost savings	
b. Total electricity deliveries from grid-connected distributed generation facilities Identify the phase on which a customer is connected should help PG&E to perform better load balancing.	3.1
3. Economic benefits	
a. Maintain/Reduce operations and maintenance costs Determining that one circuit is less or more balanced through an automated process that is repeatable and economical can result in more efficient engineering planning and practices. If implemented effectively the system of more balanced assets would require less operations and maintenance costs.	3.1
b. Maintain/Reduce capital costs Automated phase mapping means that PG&E might not need to build/expand other communication paths.	3.1
c. Reduction in electrical losses in the transmission and distribution system Should subsequent tests of an automated mapping process prove successful and inform regular engineering practices for rebalancing phases, the resulting rebalanced phases could result in fewer losses.	3.1
5. Safety, Power Quality, and Reliability (Equipment, Electricity System)	
a. Outage number, frequency, and duration reductions Automated Phase ID can support analytics for outage management through DERMS integration to enterprise systems. Better understanding PG&E system will allow employees to restore power more effectively	3.1

D.13-11-025, Attachment 4. List of Proposed Metrics and Potential Areas of Measurement (as applicable to a specific project or investment area)	Section

14 2015 PG&E EPIC Annual Report. Feb 29, 2016. http://www.pge.com/includes/docs/pdfs/about/environment/epic/EPICAnnualReportAttachmentA.pdf.

D.13-11-025, Attachment 4. List of Proposed Metrics and Potential Areas of Measurement (as applicable to a specific project or investment area)	Section
c. Forecast accuracy improvement If implemented, meter-to-transformer and phasing mapping enables ability to accurately forecast distribution loads on each phase in short-term scenarios. Use cases include capability for more flexible switching that can lead to increased safety, power quality and reliability.	3.1
d. Public safety improvement and hazard exposure reduction Improved phase balancing using low cost analysis can support avoided transformer overload failures and provide better loading data to mitigate unsafe loading conditions that could result in hazardous public exposure to exploding equipment or downed wires.	3.1
e. Utility worker safety improvement and hazard exposure reduction Automated asset phase mapping greatly reduces the need for manual field data collection on asset phasing, thereby reducing potential PG&E employee contact with live wires when collecting phase readings manually.	3.1
7. Identification of barriers or issues resolved that prevented widespread deployment of technology or strategy	
a. Description of the issues, project(s), and the results or outcomes Investigation provided insights into data flat form requirements that inform how far current data streams can serve business needs, while avoiding additional data costs, thereby optimizing existing grid resources using existing data collection capabilities.	3.1
b. Increased use of cost-effective digital information and control technology to improve reliability, security, and efficiency of the electric grid (PU Code § 8360) This project demonstrated the ability to leverage the existing SmartMeter™ infrastructure investment to support additional data transmission beyond day-to-day metering operations. The ability to identify the phase to which a customer is connected using SmartMeter™ data may lead to accurate connectivity, which is foundational to high levels of automation; success in this area would also support higher levels of distributed generation connectivity as well as voltage regulation.	4.3
c. Dynamic optimization of grid operations and resources, including appropriate consideration for asset management and utilization of related grid operations and resources, with cost-effective full cybersecurity (PU Code § 8360) This project defined the requirement for future ADMS applications	3.1

D.13-11-025, Attachment 4. List of Proposed Metrics and Potential Areas of Measurement (as applicable to a specific project or investment area)	Section
8. Effectiveness of information dissemination	
d. Number of information sharing forums held Twice or three times if internal Tech forum is counted	5.4
10. Reduced ratepayer project costs through external funding or contributions for EPIC-funded research on technologies or strategies	
a. Description or documentation of funding or contributions committed by others Vendors provided results from their methods funded outside of the project for benchmarking against the results from methods developed within the project producing a costs savings for the project.	4.5 4.6

7 Conclusion

This project developed and analyzed a variety of pre-commercial analytics solutions to automatically identify meter phase and meter-to-transformer connectivity for seven circuits. Although the performance of the meter-to-transformer methods explored was not sufficient for deployment, two phase identification methods did yield very promising results that would improve upon the accuracy of meter phasing records if deployed with the input data currently available. PG&E has begun work on additional improvements to input data, to further improve algorithm performance, and on scaling the results of this project to production through the implementation of an automated phase identification solution. Implementing an automated approach at scale will be significantly less expensive than the conventional boots on the ground alternative.

This project was PG&E's first project to fully utilize voltage data from the SmartMeters™ combined with voltage data from SCADA and asset data from GIS. Leveraging the SCADA and AMI data was more challenging than expected. Issues with data quality were revealed, and multiple work streams are now being created to address these issues before solution implementation.

Accurate meter phase mapping will be required before the implementation of ADMS and DERMS platforms. Other utilities will also be facing similar issues when working on the implementation of an ADMS in their system as the level of data quality required for this type of system is very high. This project showed that data analytics may be a viable long-term industry solution. In California especially, where the grid structures are similar to PG&E's, this work will give other utilities a chance to faster define their strategy to tackle this issue.