Advanced Metering for Phase Identification, Transformer Identification, and Secondary Modeling

T. A. Short, *Senior Member, IEEE*

***Abstract*—Advanced metering infrastructure (AMI) offers utilities new ways to model and analyze distribution circuits. Results from two circuits introduce a new method to identify phasing of transformers and single-phase taps using voltage and kilowatt-hour measurements from AMI. In addition to phase identification, we show how to use the same approach to create or check meter-to-transformer mappings. These algorithms are based on linear regression and basic voltage drop relationships. With this approach, secondary connectivity and impedance models can be auto generated. In addition, detection of unmetered load appears possible. Also demonstrated is use of AMI to estimate primary-side voltage profiles.**

***Index Terms*—power distribution, advanced metering infrastructure (AMI), phase identification, distribution secondary.**

I. Introduction

UTILITY models of distribution circuits have become more complex with widespread use of geographic information systems (GIS). These detailed models allow utilities to better plan and operate distribution circuits. Data accuracy is important for best use of models. Phasing errors are common with GIS data. In this paper, we introduce methods to check phasing of transformers and single-phase taps using AMI measurements. These methods rely on voltages measured at AMI meters augmented with kilowatt-hour measurements.

Dilek [1] describes a phase identification algorithm that uses a searching algorithm for identifying phasing of taps in a circuit model based on loading profiles. Arya et al. [2] present a method for phase identification that matches kilowatt-hour measurements from meters to kilowatt-hour measurements at the substation using integer programming. The advantage of using meter voltage measurements as described here is that the mathematics are simpler, and the approach does not require complete coverage of AMI to determine phasing. Only the area of interest needs to be considered. Approaches have been proposed to identify theft from metering measurements [3, 4].

Taylor et al. [5] show examples where AMI can be used to verify secondary-side voltage drops and to verify modeling of losses and voltage drop on secondary systems.

The methods described here for phase identification and other features are based on relatively simple fundamentals. Voltage measurements at meters that are electrically close together correlate strongly. Voltage drops through distribution components are a function of the real and reactive components of line current, and these can be readily estimated.

T. A. Short is with EPRI, Burnt Hills, NY, 12027 USA (518-596-8833; email: tshort@epri.com, t.short@ieee.org).

II. Voltage Correlations

Even on a substation with a load tap changing (LTC) transformer, the three voltages on each phase differ somewhat with time. At a substation with independent voltage regulators on each phase, the voltages generally differ significantly. We can use these differences for phase identification. How much the three phases differ will likely determine how much data we need for accurate phase identification. Fig. 1 shows correlations in terms of the coefficient of determination (R^2) between phases at several distribution substations buses. For phase identification, LTC's with tightly coupled voltages (R^2 above 0.9) will be the most challenging.

The features described here rely on time-varying voltages. Fig. 2 shows graphs of substation voltages plotted against voltages measured at two meters at 15-min intervals for a 30-day interval. This substation is a 24-kV substation regulated

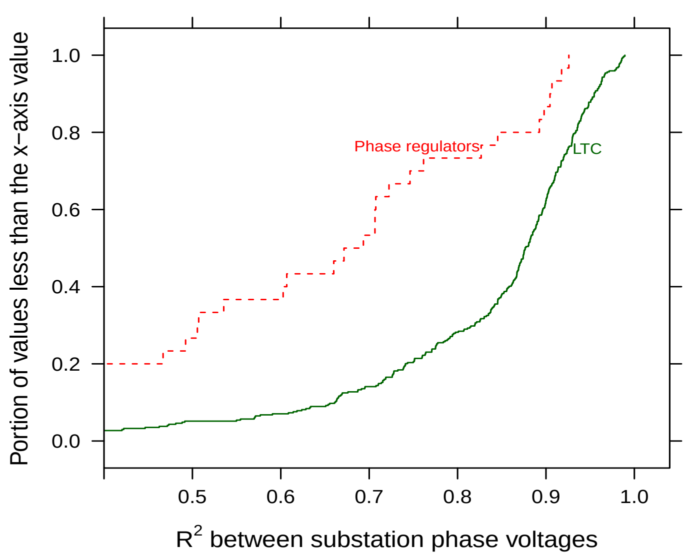

Fig. 1. Cummulative distribution of coefficient of determination (R^2) between phase voltages at 141 substations.

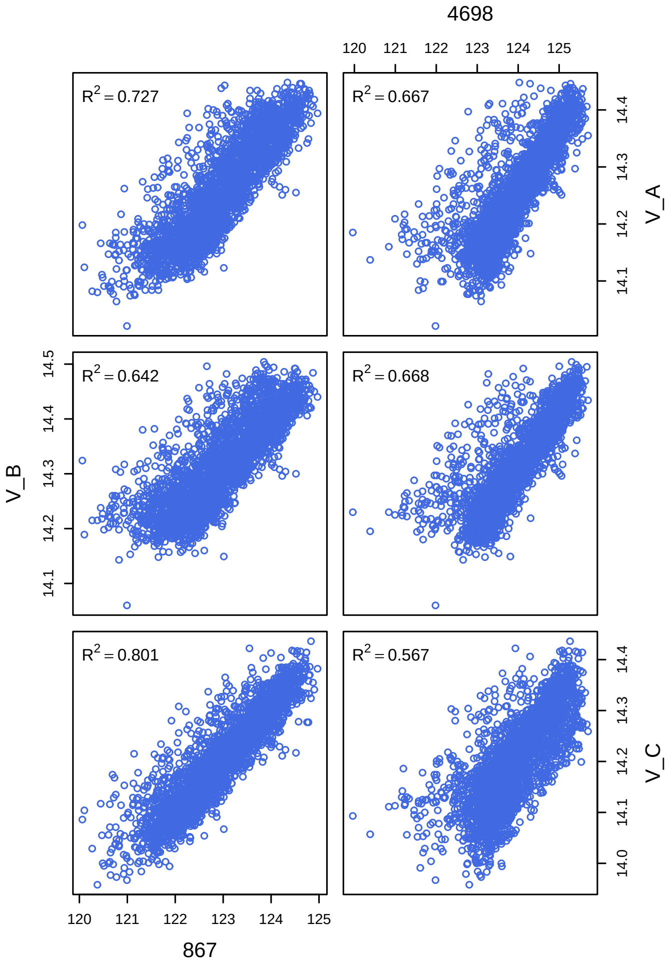

Fig. 2. Substation voltages (kV) compared to customer meter voltages (120-V base).

by an LTC. For this circuit, the customer metering captured an instantaneous voltage measurement. The graphs are arranged in panels. Each meter location is plotted against each of the three substation voltages. The overall coefficient of determination (R^2, the square of the correlation coefficient) for each comparison pair is given at the top left of each panel. Meter 867 (the left column of panels) correlates most strongly with phase C at the substation. Meter 4698 (right column) has the strongest correlations with phases A and B. At this substation, the correlations between phases at the substation were: $R^2_{\text{A-B}} = 0.88$, $R^2_{\text{B-C}} = 0.71$, and $R^2_{\text{A-C}} = 0.78$.

The following linear regression improves the phase distinction:

$$V_n = k_0 + k_1 V_m + k_2 W_m + k_3 W_n \qquad (1)$$

Where

k_i = regression coefficients

V_n = substation voltage on phase n

W_n = substation average power on phase n

V_m = voltage at meter m

W_m = average power on meter m

Including the substation average power (W_n) helps account for the voltage drop across the primary. Including the average power drawn by the customer (W_m) helps account for secondary-side voltage drop.

After correcting for load, the correct phasing for meter 4698 became more apparent. The R^2 values for phases A, B, and C were 0.775, 0.785, and 0.638 respectively, showing a stronger correlation to phase B.

Voltage correlations are even stronger for meters on the same secondary. Fig. 3 compares voltages of four meters on the same single-phase tap for a 30-day time period. Three of the meters are on the same phase. The AMI on this circuit records an average voltage at hourly intervals. Each panel in Fig. 3 shows how the voltages compare between two meters. For the combination of meter pairs shown, meters 1601 and 1613 have the highest correlation, so we expect them to be closely connected.

When load is accounted for as in eq. (1), the R^2 values show more dramatically the connection between meter pairs. Fig. 4 shows that meters 1601 and 1613 are very closely connected.

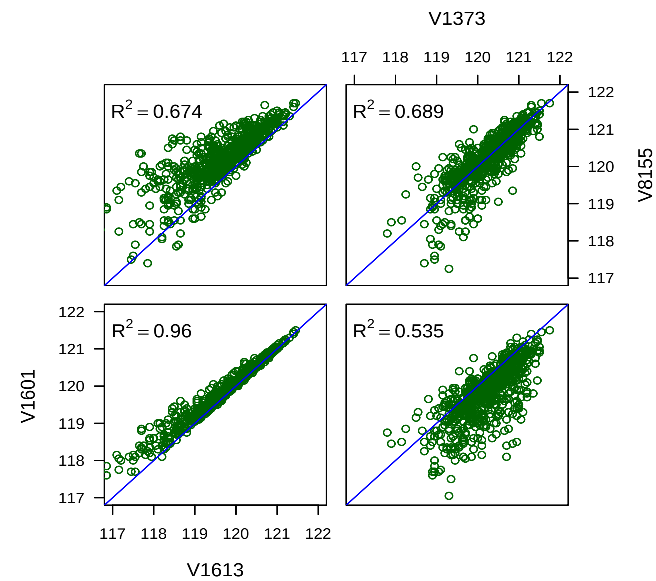

Fig. 3. Comparison of pairs of customer meter voltages (120-V base).

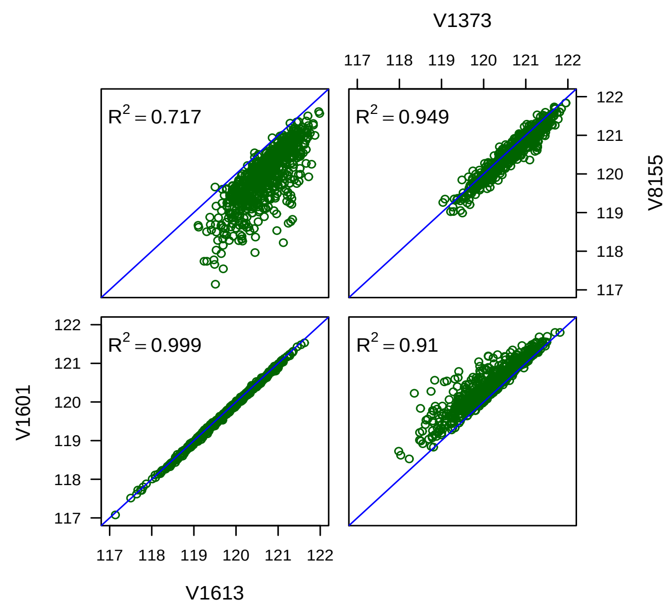

Fig. 4. Pairs of customer meter voltages corrected for load (120-V base).

III. Linear Regressions for Phase Identification

Correlation of meter voltages from AMI to substation voltages is a simple way to estimate meter phasing. For each meter, find the correlation coefficient between the meter voltage records and each of the three substation voltages. Assign the meter phasing based on which phase had the highest correlation coefficient. With this approach, phasing estimates were at least 80% accurate on four circuits where this algorithm was applied.

Using linear regression improves the accuracy of phase identification relative to using just voltage correlations. The algorithm becomes:

1) For each meter (V_m, W_m), separately run a regression using eq. (1) to each of the three phases at the substation (V_n, W_n).

2) Assign the meter phasing based on the phase with the highest R^2.

Fig. 5 shows results of phase identification on one circuit. This is a 24-kV suburban circuit with an LTC-regulated bus. The AMI captures kilowatt-hour and instantaneous voltage values every 15 minutes. In Fig. 5, the single-phase taps are color-coded according to the phasing in the utility's GIS model. Each meter is shown as a small dot. A colored square is overlayed on top of a meter if the phase identification differed from the GIS phasing. The color of the square shows the predicted phasing. The three single-phase taps labeled L1, L2, and L3 were all field checked. The phasing errors were confirmed, and the predicted phasing was correct.

Fig. 6 shows results of the phase identification using the regression model on a circuit at a different utility. This is a 24-kV rural circuit that includes single-phase taps with step-down transformers to 4-kV circuit sections. As with Fig. 5, phasing errors are highlighted with a color-coded rectangle. On this circuit, the AMI captures average voltage, kilowatt-hour, and kilovar-hour readings at hourly intervals. The regression algorithm is not perfect. There are several incorrect phase predictions scattered around. Several of them are on the long single-phase tap on the green phase just outside the substation. This lateral is a 4-kV section.

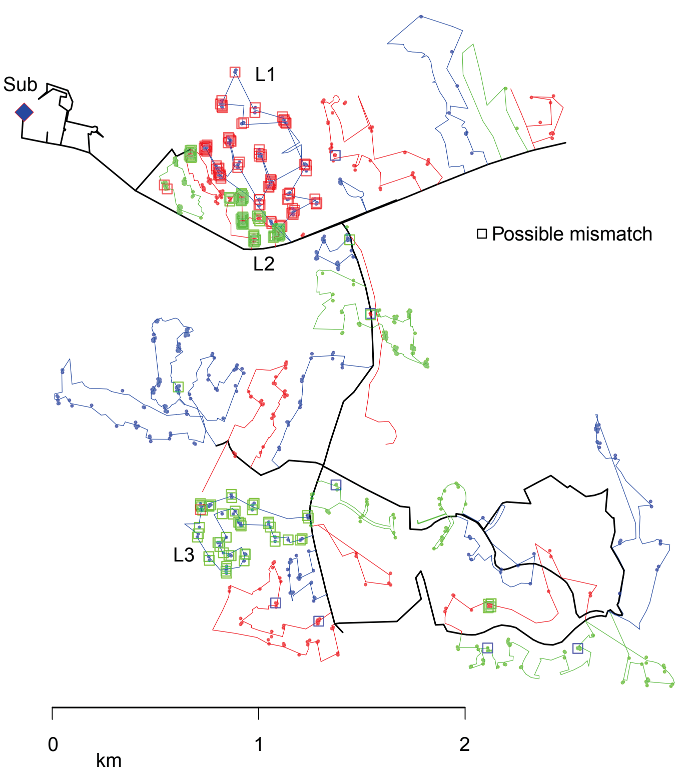

Fig. 5. Phase identification estimates on circuit A.

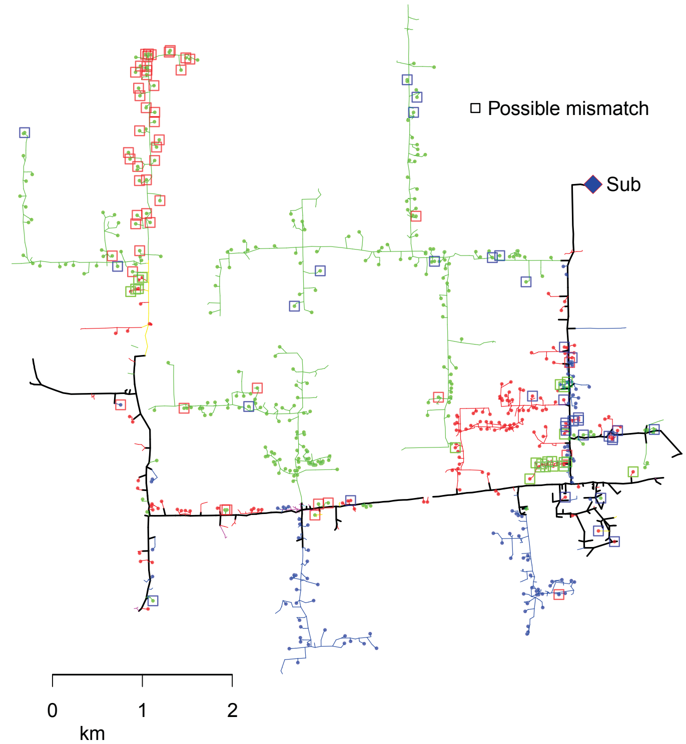

Fig. 6. Phase identification estimates on circuit B.

The regression approach was generally at least 95% accurate on the four circuits. This accuracy estimate is uncertain because of model uncertainties. Refining accuracy would require more detailed field checking.

Another way to reduce the duration of data needed for phase identification with this approach is to intentionally create a change in voltage on one phase. With individually regulated phases, this is easy, quick, and almost foolproof. The phasing on the circuit in Fig. 6 was verified using this method.

IV. Bottoms-Up Grouping

As we saw in Fig. 3 and Fig. 4, meters that are closely connected are highly correlated. We will use that to build up a mock circuit based on AMI data. Consider two meters that share an upstream common point as shown in Fig. 7. We will use linear regression to estimate R_1, X_1, R_2, X_2, and V_0 given I_1, I_2, V_1, and V_2. The linear regressions will rely on the common approximation to voltage drop (Short [6] and many other references):

$$V_{\text{drop}} \approx R \cdot I_R + X \cdot I_X$$

Using this we have the following:

$$V_0 = V_1 + R_1 \cdot I_{1,R} + X_1 \cdot I_{1,X} = V_2 + R_2 \cdot I_{2,R} + X_2 \cdot I_{2,X}$$

As a linear regression formula, this is:

$$V_1 = \beta_0 + \beta_0 \cdot V_2 + R_2 \cdot I_{2,R} + X_2 \cdot I_{2,X} + R_1 \cdot \left(-I_{1,R}\right) + X_1 \cdot \left(-I_{1,X}\right) \quad (2)$$

The regression coefficients to be found are β_0, β_1, R_1, X_1, R_2, and X_2. The coefficients β_0 and β_1 may not be strictly necessary, but including these helps to improve R^2 values, possibly because they account for metering errors. The input data into the regression model is the voltage time series from each meter (V_1 and V_2) and the real and reactive components of current at each meter ($I_{1,R}$, $I_{2,R}$, $I_{1,X}$, and $I_{2,X}$). The real and reactive components of current at meter i can be found as:

$$I_{i,R} = \frac{P_i}{V_i} \qquad I_{i,X} = \frac{Q_i}{V_i}$$

Where

V_i = Average voltage at meter i, volts (at nominal voltage, normally 240 V)

P_i = Average real power on meter i, watts

Q_i = Average reactive power on meter i, watts

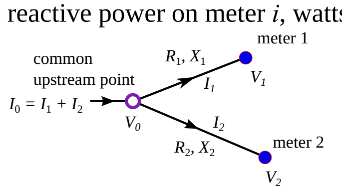

Fig. 7. Two meters sharing a common upstream point.

With the line resistances and reactances found, the common-point voltage V_0 can be estimated as:

$$V_{0,\text{estimate1}} = V_1 + R_1 \cdot I_{1,R} + X_1 \cdot I_{1,X}$$

$$V_{0,\text{estimate2}} = V_2 + R_2 \cdot I_{2,R} + X_2 \cdot I_{2,X}$$

$$V_0 = \left(V_{0,\text{estimate1}} + V_{0,\text{estimate2}}\right)/2 \quad (3)$$

With this regression model, we can build up a mock circuit based on AMI data with the following algorithm:

1) Start with a set of meters to be grouped (set A), each meter having series of voltage, watt, and var averages captured simultaneously.

2) For each meter i, solve the regression model in eq. (2) paired with every other meter in set A.

3) Pick the meter j that has the highest R^2 value in regressions with meter i.

4) For the new meter pair of closely coupled meters i and j, store the line resistances and reactances for each branch from the regression model. Also find the voltage [from eq. (3)] and real and reactive power at the common point. This forms a new metering point k.

5) Remove meters i and j from set A. Add the new meter point k to set A.

6) Repeat starting with step two until all meters have been paired together.

Fig. 8 shows results of this algorithm applied to a set of meters from two nearby secondaries on the same phase on circuit B based on 30 days of data. The pairing started with meters 1601 and 1613, the two meters that were most closely coupled. These are the same two meters highlighted in Fig. 4. The results in Fig. 8 are presented as a tree. For each meter pairing point, the R^2 values are shown. The length of each branch is based on the line resistance estimated from the regression model. Both of these secondaries are fed by 25-kVA transformers. The top secondary grouping (meters 8142 and 8155) have a length of secondary between the transformer and the first branching point.

This algorithm can become time consuming for large numbers of meters (hundreds or thousands) because computations increase as the number of meters squared. On a 2.2-Ghz Intel i7 processor, the grouping algorithm for 39 meters using four weeks of data takes 40 sec. Increasing this to 10,000 meters would require approximately 18 hours of computation time. To reduce computations, we can modify step 2 of the algorithm by using voltage correlations to reduce the set of meters for comparison:

2) For each meter i, solve the regression model in eq. (2) for the meters in set A that have the top N correlations between meter voltages.

Finding a voltage correlation matrix between all sets of meters is not computationally taxing.

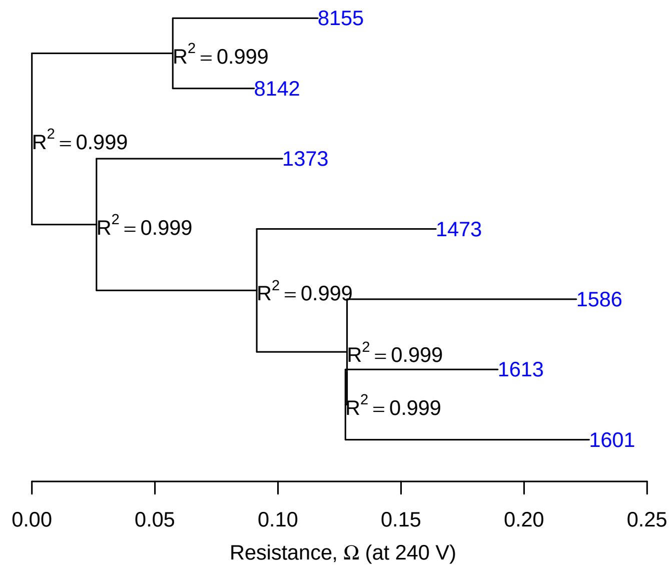

Fig. 8. Bottoms-up grouping of meters on two secondaries.

For circuit B, the utility has impedance models for secondaries from GIS data. Fig. 9 compares the connectivity and resistances of the GIS model with the model created with the bottoms-up grouping algorithm using AMI data. The connectivity of the AMI-based model is equivalent to the GIS model; the meters are grouped with their correct transformer, and the connectivity of the secondary lines matches. The main difference is that the GIS model has less resistance from each meter to the first upstream common point. A field check by the utility found that the services from the road to the meter were smaller than that given in the GIS model. In the model, most secondaries were 3/0 aluminum. In the field, the services were estimated to be #2 aluminum (2.5 times the resistance of 3/0) or #4 aluminum (4 times the resistance of 3/0). This matches findings from the model produced from AMI data.

Another issue is customer load balancing. The AMI voltages are measured at across both hot legs (240 V nominal) of the 120/240-V secondary, but customers have both 120-V and 240-V load. If the 120-V loadings are unbalanced, the unbalanced current will contribute to more voltage drop and higher effective impedance.

This example suggests that this approach could be used to auto-generate impedance models for distribution secondaries.

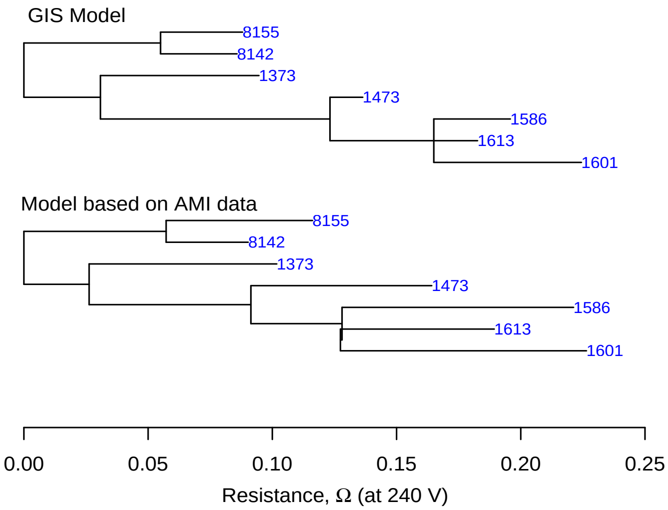

Fig. 9. Comparison of the GIS model to the model based on bottoms-up grouping.

The regression model in eq. (2) includes both metered real and reactive power. Many residential AMI systems do not record reactive power. For distribution secondaries, this is not a significant handicap. Fig. 10 compares results of the full regression with real and reactive power to a similar algorithm that only uses real power. The connectivity of both models is the same, and the resistive impedances are close. The regression model still works well because *X*/*R* ratios of secondary systems are low. Both of the secondaries used in this example are predominantly 3/0 aluminum triplex which has *X*/*R* = 0.28. With a power factor of 0.95, the resistive component of voltage drop is more than ten times the reactive component of voltage drop. Having reactive power makes more difference for the connectivity and impedance through the distribution transformer because *X*/*R* ratios are higher. Results were not too different in Fig. 10, but other cases have shown more difference.

AMI systems can have different types of voltage measurements. Those given in the examples above were based on average voltage. Results from another circuit with instantaneous voltage measurements gave lower R^2 values. The instantaneous voltage measurements have more noise. The regression model in eq. (2) applies if all values are averaged over the same time period. If real power is an averaged number (kWh), and the voltage is instantaneous, then eq. (2) is only an approximation.

V. Virtual Voltage Metering

With the bottoms-up grouping, we obtain an estimate of voltage at each upstream common point. We can use this feature to use AMI data as a virtual voltage meter. Circuit B had AMI meters installed on the secondary side of several transformers. The transformer on the five-meter secondary system shown on the bottom of Fig. 8 had a meter on the transformer secondary. Fig. 11 compares measurements of the voltages on the transformer secondary to estimates from bottoms-up grouping. The correlation coefficient between the measurements and estimates was 0.997.

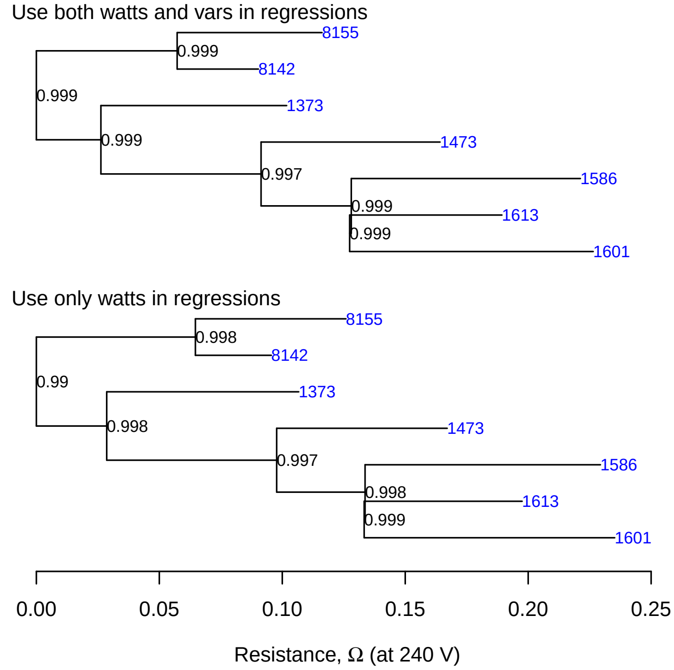

Fig. 10. Comparison of using watts and vars to watt-only regression.

With AMI from two transformers that are electrically close, we can estimate the primary voltage. Fig. 12 compares primary-side voltages based on two estimates. On estimate was from the two transformer secondary system in Fig. 8. The second estimate was from another pair of transformers close to the first two. The correlation coefficient between the two estimates was 0.998.

Using the AMI as a virtual meter could be used in several ways: (1) identification of open points on loops,

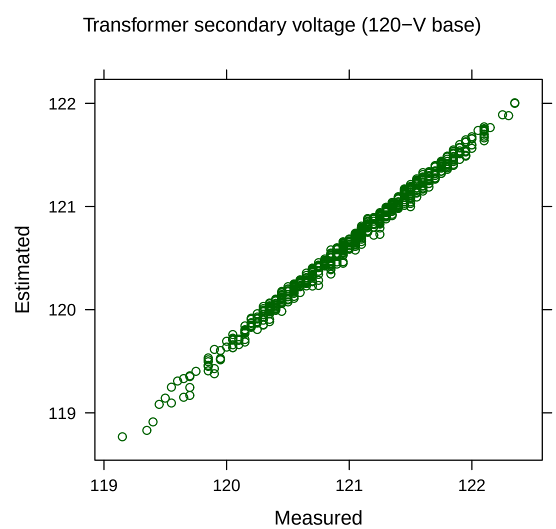

Fig. 11. Measurement of transformer secondary voltage versus an estimate from bottoms-up grouping (120-V base).

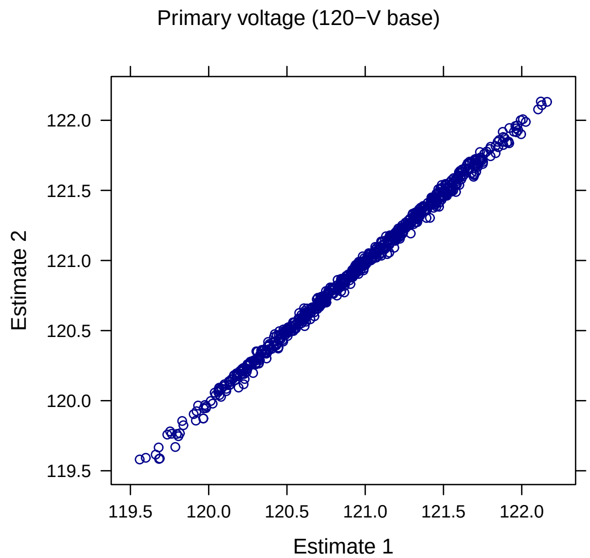

Fig. 12. Two estimates of primary-side voltage (120-V base).

(2) identifying transformers with off-nominal taps or ratios, (3) circuit model verification, (4) identification of switched capacitors with blown fuses, and (5) identification of misoperating voltage regulators.

Fig. 13 shows estimates of primary-side voltages along circuit B using the bottoms-up grouping. This voltage profile was developed with the following algorithm:

1) Perform a bottoms-up grouping for each transformer on the system. This gives a virtual metering point at the secondary side of each transformer.

2) For each transformer, find the other two transformer virtual metering points which combine with the highest R^2.

3) Find the primary voltage based on grouping of these three transformers using the same bottoms-up regression.

Each data point in Fig. 13 is color coded by phasing (using corrected phasing discussed later). The size of the data point symbol is based on the R^2 of the final grouping, so smaller symbols have less confidence. The portion of the circuit on the green phase (diamonds) at the end of the circuit has a step-down transformer to 4 kV that effectively boosts the voltage.

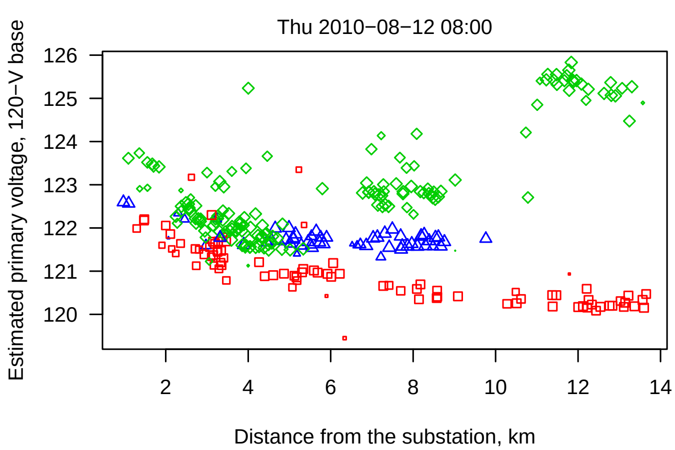

Fig. 13. Estimates of primary-side voltages along the circuit.

VI. Detection of Theft and Unmetered Load

The bottoms-up grouping relies on complete metering downstream of each grouping point. If there is missing load, the grouping becomes less accurate. We may be able to use this feature to identify unmetered load, including theft of service. Fig. 14 shows a four-transformer secondary with missing load. The R^2 values dropped, and the impedances were negative, indicating that there was not enough load on that secondary to account for the voltage difference.

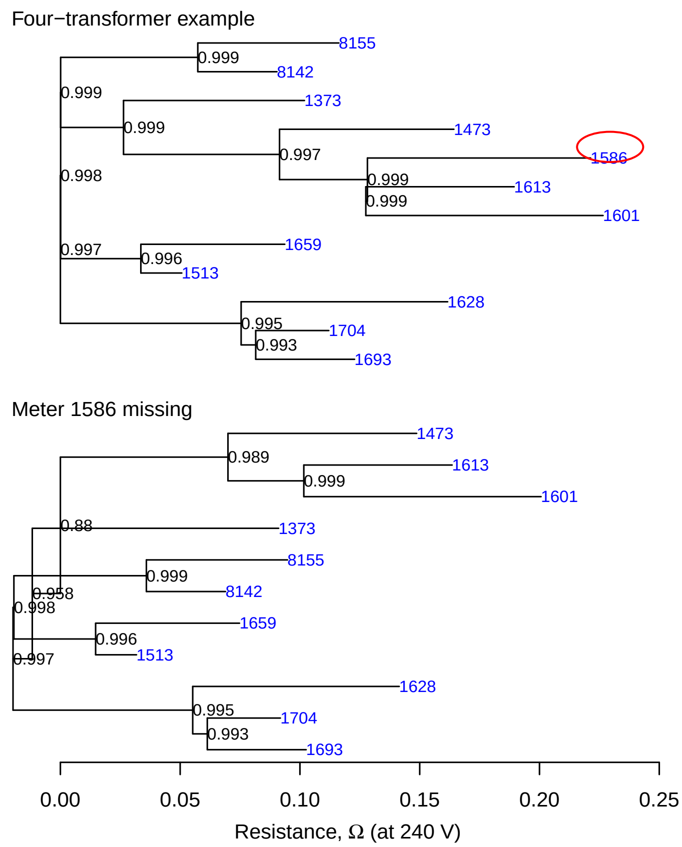

Fig. 14. Example of bottoms-up grouping with missing load.

This example suggests the following algorithm for detecting unmetered load:

1) Perform a bottoms-up grouping for each transformer on the system. This gives a virtual metering point at the secondary side of each transformer. If the R^2 value for a grouping is below a given threshold, unmetered load is likely.

2) For each transformer, find the transformer that matches best with that transformer (in terms of R^2). If the R^2 value for this grouping is below a given threshold, unmetered load is likely.

A set of 36 transformers on circuit B with metering at the transformer was used to test this approach for detection of unmetered loads. Of these, five transformer meters had more than 10% of the downstream load that was unmetered. One transformer had less load metered at the transformer than the sum of the meters indicating that one or more customer meters were really on a different transformer. For this set of meters, the following thresholds were good indicators of unmetered load:

1) Transformer secondary R^2 threshold = 0.985.
2) Primary-level R^2 threshold = 0.97.

The 30 meters without unmetered load had R^2 values exceeding both of these thresholds, and four of the transformers with unmetered load had one or both R^2 values below the given thresholds. One meter did show unmetered load that was not detected by these thresholds. On this meter, there was a constant 500 W of load that was unmetered, possibly a street light that was always on. This suggests that for unmetered load to be detected, it must change with time. The one meter with too much load also had R^2 values below these thresholds.

Both of these R^2 values should be checked. The transformer-level threshold is most effective at picking up unmetered load that is downstream of branching points beyond the transformer. The primary-side threshold is most effective at picking up unmetered load that is directly paralleled at the transformer secondary.

VII. Application on a Single-Phase Tap

Fig. 15 shows the bottoms-up grouping applied to a single-phase tap. The GIS model of this tap is shown on the bottom-right part of the figure. In the GIS model, each meter is color-coded and symbol-coded according to the transformer designated in the GIS model. In Fig. 15, the bottoms-up grouping was continued until the R^2 value dropped below 0.98. The top tree shown includes several secondaries attached to the primary. The remaining ungrouped meters and secondaries are shown below that. With the bottoms-up grouping, we can identify several GIS and metering errors. Several transformer groupings are highlighted. These indicate the following:

T1 – *Unmetered load* – The R^2 for the grouping on transformer T1 dropped below 0.98. The GIS model for this secondary system shows branches that do not go to AMI meters, suggesting locations for possible unmetered load. A field visit confirmed two missing meters.

T2 – *Potential phasing error* – Meter 1834, one of the meters left out of the grouping, had a low R^2 value, indicating a phasing error. A field visit found that this was an open-wye – open-delta transformer with AMI only on the lighting leg. This likely explains why this did not correlate.

T3 and T4 – *Mismatched meters and transformers* – The bottoms-up algorithm found that meters 1184 and 1264 were on transformer T4 rather than T3 as designated in the GIS model. Also, these meters are shown in the wrong location; meter 1184 appears to be closer to meter 1241, and meter

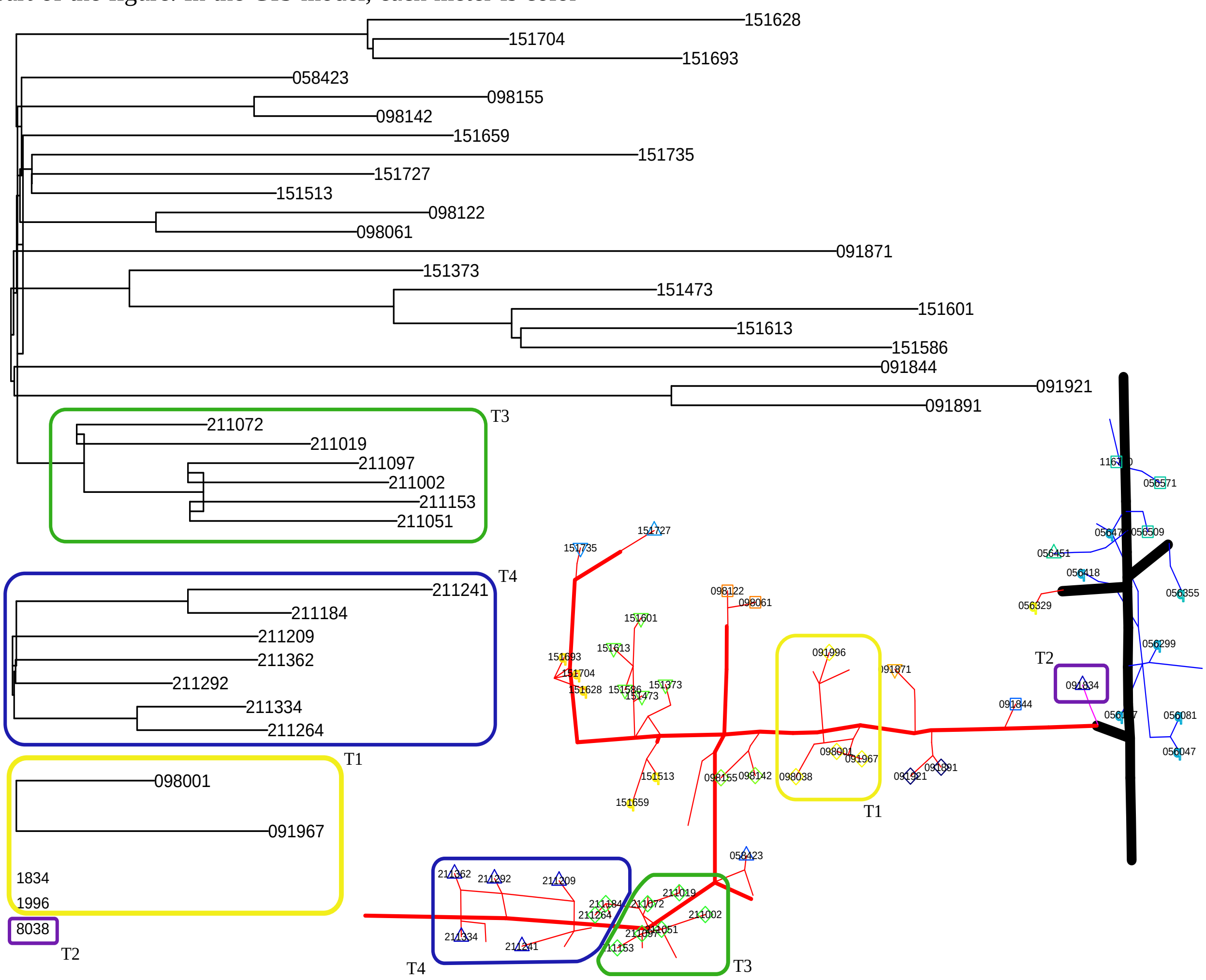

Fig. 15. Application of grouping on a single-phase tap.

1264 appears to be closer to meter 1334. Also transformer T4 has missing load indicated. A field visit found that the connectivity indicated by AMI is likely to be correct, but the utility could not identify connections completely, because these secondaries were underground.

Not all matches were perfect. Meters 1513 and 1659 did not group correctly. These meters have low load, and depending on the time period, they may or may not group. They did group correctly in Fig. 14. With a larger number of meters, the grouping was less consistent with time period. With smaller groupings (Fig. 8 and Fig. 14), groupings and estimated resistances were more consistent. More work is needed for groupings of larger numbers of meters.

VIII. Phase Identification using Grouping

The bottoms-up approach can also be used at the circuit level for phase identification. Fig. 16 shows phase identification results for circuit B using the following algorithm:

1) Find the meters on each single-phase tap. Each tap can be a single-phase primary lateral, or it can be a single-phase transformer connection.

2) Perform a bottoms-up grouping for each single-phase tap. This gives a virtual metering point at the upstream common point, on the primary side if there are more than one transformer on the tap and on the secondary side if there is just one transformer on the tap.

3) With the virtual metering point at each single-phase tap, use the regression approach using eq. (1) to identify phasing for the single-phase tap.

The phasing errors in Fig. 16 (seven single-phase transformers and two single-phase taps) have been verified by the utility. The phasing errors are highlighted at the meter level, but the algorithm results actually give phasing by tap.

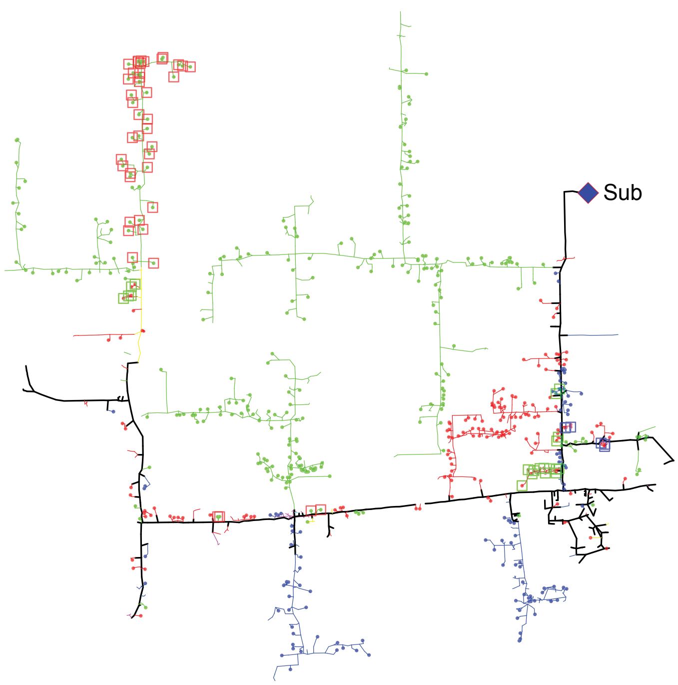

Fig. 16. Phase identification estimates on circuit B using grouping of single-phase taps.

Doing phase identification for each single-phase tap has given reasonable results; but on other circuits, it might not completely account for meters matched to the wrong transformer. To account for that, the bottoms-up grouping could be applied to every meter on the system.

IX. Future Directions

Results look promising in many areas for using AMI data for phase identification, transformer identification, theft detection, auto-generation of secondary circuits, and using AMI as a virtual metering system. Additional work is needed in several areas. We have limited data on accuracy because of the small number of circuits evaluated. More field verification is needed along with trials on additional circuits to better evaluate the accuracy of these methods. These algorithms need to be applied on more circuits to gather more data on how well they work and to identify problem areas.

For phase identification, the most challenging circuits are likely to be short, LTC-regulated circuits where the voltages on the three phases track each other well. Delta-connected step-down banks may also cause issues, although a bottoms-up approach should account for that.

More work is needed to find out what capabilities are needed from AMI to make the best use of these algorithms and to identify the limitations of these approaches. For example, having average voltages appears to give better results than gathering instantaneous voltages. We need more work to determine the optimal duration for data collection (length and determining if season makes a difference). Once algorithms are fine-tuned, work will be needed to find the best ways to integrate the algorithms into a utility’s information technology systems.

X. Acknowledgment

Southern Company was the main driver behind this project, suggesting the need for phase identification and ways for improving models based on GIS data. Don Parker of Alabama Power (Southern Company) and Jared Green (formerly with Alabama Power and now with EPRI) supplied AMI data and circuit data and performed field checking. Brian Bushey of Consumers Energy provided AMI data, circuit data, and field checking. Ron Belvin of Duke Energy also provided AMI data, circuit data, and field checking. Thanks to Jared Green (the project manager) and Brian Seal of EPRI for suggestions and support.

XI. References

[1] M. Dilek, "Integrated Design of Electrical Distribution Systems: Phase Balancing and Phase Prediction Case Studies," Virginia Polytechnic Institute, 2001. http://scholar.lib.vt.edu/theses/available/etd-11132001-222401/

[2] V. Arya, D. Seetharam, S. Kalyanaraman, K. Dontas, C. Pavlovski, S. Hoy, and J. R. Kalagnanam, "Phase Identification in Smart Grids," IEEE International Conference on Smart Grid Communications (SmartGridComm), 2011.

[3] S. S. Depuru, L. Wang, and V. Devabhaktuni, "Support vector machine based data classification for detection of electricity theft," IEEE/PES Power Systems Conference and Exposition (PSCE), 2011.

[4] C. J. Bandim, J. E. R. Alves, Jr., A. V. Pinto, Jr., F. C. Souza, M. R. B.

Loureiro, C. A. Magalhaes, and F. Galvez-Durand, "Identification of energy theft and tampered meters using a central observer meter: a mathematical approach," IEEE PES Transmission and Distribution Conference and Exposition, 2003.
[5] J. A. Taylor, T. A. Short, and B. Bushey, "Efficiency Impacts of Distribution Secondaries," IEEE Transmission and Distribution Conference, Orlando, FL, 2012.
[6] T. A. Short, *Electric Power Distribution Handbook*. Boca Raton, FL: CRC Press, 2004.

XII. Biographies

Tom A. Short (M'90 – SM'98) is with EPRI at an office in Burnt Hills, NY. Before joining EPRI in 2000, he worked for Power Technologies, Inc. for ten years. Mr. Short has a Master's of Science degree in Electrical Engineering from Montana State University (1990). Mr. Short authored the *Electric Power Distribution Handbook* (CRC Press, 2004). In addition, he led the development of IEEE Std. 1410-1997, *Improving the Lightning Performance of Electric Power Overhead Distribution Lines* as the working group chair.