# Distribution System Low-Voltage Circuit Topology Estimation using Smart Metering Data

Jouni Peppanen, Santiago Grijalva
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA, USA

Matthew J. Reno, Robert J. Broderick
Sandia National Laboratories
Albuquerque, NM, USA

***Abstract*— Operating distribution systems with a growing number of distributed energy resources requires accurate feeder models down to the point of interconnection. Many of the new resources are located in the secondary low-voltage distribution circuits that typically are not modeled or modeled with low level of detail. This paper presents a practical and computational efficient approach for estimating the secondary circuit topologies from historical voltage and power measurement data provided by smart meters and distributed energy resource sensors. The accuracy of the algorithm is demonstrated on a 66-node test circuit utilizing real AMI data. The algorithm is also utilized to estimate the secondary circuit topologies of the Georgia Tech distribution system. Challenges and practical implementation approaches of the algorithm are discussed. The paper demonstrates the computational infeasibility of exhaustive secondary circuit topology estimation approaches and presents an efficient algorithm for verifying whether two radial secondary circuits have identical topologies.**

***Index Terms*—Load Modeling, Power Distribution, Power System Measurements, Smart Grids**

## I. Introduction

More accurate distribution system models are required to analyze and coordinate distribution system operation with rapidly increasing amount of PV and other distributed energy resources (DER). In particular, distribution system secondary (low-voltage) circuits down to the point of common coupling for DER are needed to achieve a comprehensive model of the distribution system. Distribution models are fairly complex involving a large number of components, settings, and connection details. Maintaining these models is not a trivial task due to the complexity and variety of changes that occur in a system over time, such as expansion, re-conductoring, phase balancing, changes in locations of poles and topologies, etc. Inaccuracies are usually present due to human errors, inaccurate manufacturing data, unrecorded network changes, incorrect tap information, etc. [1]. Incorrect component parameters and component connectivity, i.e., system topology are among the most common errors in Geographic Information Systems (GIS) [2]. Next to the overall modeling uncertainty, distribution system secondary circuits are typically either not modeled at all, or they are modeled with a low level of detail. However, it is becoming particularly important to accurately model the low voltage circuit sections where a large portion of the DERs are located. Moreover, since the per unit impedances are higher on these low-voltage circuits, any component errors are intensified in the voltage profile.

The extensive roll-out of smart meters and PV microinverters is rapidly increasing the available measurement data and expanding distribution system situational awareness [3], which in the past, has been limited downstream of the substation. This new data can be leveraged to calibrate existing utility feeder models [4]. Automated approaches are needed to achieve this in a cost-effective way [2].

This paper describes a practical and computationally efficient algorithm for radial distribution system secondary circuit topology and parameter estimation. This paper has the following structure. Section II discusses the relevant power system topology and parameter estimation literature. Section III introduces the distribution system secondary circuit joint topology and parameter estimation problem, discusses its complexity, and presents a novel algorithm capable of solving it. Section IV presents an algorithm for validating that the estimated secondary circuit topology is electrically identical to the true topology. Section V presents the method validation results for a 66-node three-phase test circuit. Section VI presents the results for selected Georgia Tech distribution system secondary circuits. Finally, section VII provides various conclusions of this work.

## II. Background

In transmission systems, topology estimation (TE) has been studied since 1970s whereas distribution system topology estimation (DSTE) is a rather new research area that is motivated by the increasing penetrations of DER and enabled by modern distribution system measurement sources. In transmission systems, where the substations topology types are typically known, topology estimation typically focuses on detecting topology errors that can be broadly categorized as errors in the status of switching devices and substation configuration errors [1]. Many of the conventional topology error detection approaches either require a residual vector from an existing state estimator or involve some modifications to the existing state estimator algorithm [1], [5], [6]. Due to the limited deployment of state estimators in distribution systems, these methods are not readily available. Moreover, these approaches are not intended for estimating entire circuit topologies. References [7]–[10] propose topology (switch status) detection algorithms, which do not require existing state estimator, but the methods are not intended for estimating entire circuit topologies. Additionally, the methods proposed in [7], [8], [10] require pervasive micro-PMU measurements, which are currently rare in distribution systems.

Distribution system parameter and topology estimation has recently received increasing attention. A linear optimization-based method for topology error detection and parameter estimation is proposed in [4]. In this work, the authors do not estimate the circuit reactances or utilize reactive power measurements. An algorithm for detecting incorrect smart meter placement in a GIS is presented in [2], [11]. The authors utilize historical smart meter data to detect neighboring meters based on their voltage correlations and meter depths in the circuit tree based on the voltage magnitudes. However, instead of estimating network component connectivity (and parameters) of entire circuit models, this approach is mainly intended for detecting errors in existing portion of the utility model. Topology detection algorithm for radial balanced 3-phase feeders is discussed in [12]. The algorithm, which is based on an approximation of the node voltages, relies on a rather restrictive assumptions that all lines (and transformers) have similar X/R ratios and that all system nodes are monitored. In practice, service drop impedances may vary significantly and typically only the leaf nodes of the radial circuit trees have smart meters and/or DER sensors. Practical methods for meter phase identification, meter-to-transformer mapping, and joint parameter and topology estimation are shown in [13].

This paper extends our earlier parameter estimation work in [14], [15] to the case of unknown circuit topologies. Our work further develops the method shown in [13] by allowing the estimation of any radial circuit topology. Our method, which is based on linearized voltage drop approximation and linear regression, is computationally efficient and can easily leverage large measurement samples generating an estimated circuit for a practical-sized secondary circuit. The computational time is within seconds even when thousands of measurement samples are utilized.

## III. Distribution System Secondary Circuit Parameter and Topology Estimation

### A. *Problem Formulation*

The objective of the distribution system secondary circuit topology (and parameter) estimation algorithm (DSTE) is to simultaneously identify topology (component connectivity) and component series impedance parameters of a given secondary circuit. Mathematically, secondary circuits are rooted trees, i.e., connected, directed graphs without cycles that have a designated root node (the secondary circuit transformer low-voltage node). Secondary circuit trees may be multifurcating, i.e., a given node may have more than two child nodes. Because the topology is unknown, the total number of nodes is also unknown. The algorithm requires that each secondary circuit tree leaf node has a smart meter or DER sensor measuring voltage and active and reactive power (or current and power factor) shown in blue in Fig. 1. Some of the internal nodes of the tree may also be metered. Utilizing this data, the objective is to identify the circuit topology and the component series impedance parameters shown in red in Fig. 1. It should be emphasized that the secondary circuit topology and component parameters are assumed to be completely unknown.

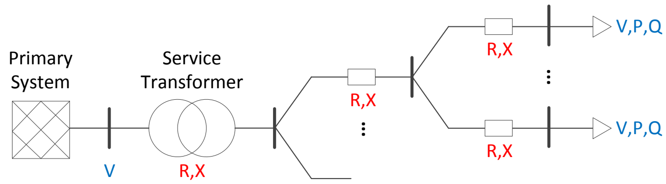

Fig. 1 Secondary circuit topology and parameter estimation problem

### B. *Infeasibility of Exhaustive Topology Search*

Theoretically, the topology of a secondary circuit can be estimated by performing an exhaustive search of all possible topologies. This could be achieved by first estimating the parameters for all possible topologies with the approach in [15] and then selecting the topology that results in the best accuracy, e.g., in terms of the mean squared error (MSE) of the simulated voltages compared to the measured voltages. This approach can quickly become infeasible for a secondary circuit with $N$ meters, due to the number of possible topologies to consider. The number of possible secondary circuit topologies is given by the number of rooted potentially multifurcating trees with $N$ labelled nodes some of which may be internal nodes. The number of such trees is of interest in evolutionary biology and has been calculated with a recursive relation in [16]. Table 1 lists the number of such trees for $N \in \{1, \ldots, 10\}$. Clearly, to evaluate all the alternative topologies becomes impractical even with 5 or 6 meters and is practically infeasible with 7 or more meters. Thus, an exhaustive comparison of all possible topologies would be a computationally demanding task. The next subsection presents a computationally efficient greedy-type joint parameter and topology estimation approach.

TABLE I. The Number of Rooted Trees with N Labelled Nodes, Allowing Multifurcations, and Allowing Some of The Internal Nodes to be Labeled [16]

| **# Leafs** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
|---|---|---|---|---|---|---|---|---|---|---|
| **# Trees** | 1 | 3 | 22 | 262 | 4,336 | 91,984 | 2.38e6 | 72.8e6 | 2.57e9 | 1.03e11 |

### C. *DSTE Algorithm*

In this section we introduce the distribution system secondary circuit parameter and topology estimation algorithm. The algorithm utilizes the well-known linear approximation of voltage drop ($V_{drop} = |V_1| - |V_2|$) over a series impedance $R + jX$ (on the right in Fig. 2):

$$V_{drop} = |V_1| - |V_2| \approx (RP + XQ)/V_2 = RI_R + XI_X, \quad (1)$$

where $P$, $Q$, $I_R$ and $I_X$ are the active power, reactive power, real current ($I_R = I(PF)$), and reactive current ($I_R = I\sqrt{1-(PF)^2}$) flowing over the branch, respectively [17].

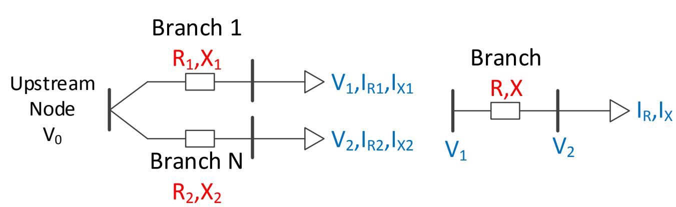

Fig. 2 Two meters connected in parallel (left) and in series (right)

The algorithm processes one secondary circuit at a time initialized with the list of all the meters of the secondary circuit and an empty mock circuit. For each meter pair at each iteration, the algorithm solves a linear regression problem for the parallel circuit type (on the left in Fig. 2)

$$\boldsymbol{V}_1 - \boldsymbol{V}_2 = \boldsymbol{I}_{R1}R_1 + \boldsymbol{I}_{X1}X_1 + \boldsymbol{I}_{R2}R_2 + \boldsymbol{I}_{X2}X_2 + \boldsymbol{\epsilon} \quad (2)$$

and a linear regression problem for the series circuit type (on the right in Fig. 2)

$$\boldsymbol{V}_1 - \boldsymbol{V}_2 = \boldsymbol{I}_R R + \boldsymbol{I}_X X + \boldsymbol{\epsilon}. \quad (3)$$

The order of meters 1 and 2 is irrelevant in regression model (2). If the secondary circuit does not have distributed generation causing reverse power flows, regression model (3) is solved only for the meter order with positive average voltage drop $\sum_{t=1}^{T}(V_{1,t} - V_{2,t}) > 0$. A wrong meter order simply results in negative estimated parameters. The complete DSTE algorithm is listed in Algorithm 1.

**Algorithm 1**: DSTE Algorithm

**Input**: Meter samples of $\boldsymbol{V}, \boldsymbol{I}_R, \boldsymbol{I}_X$

**Output**: Mock circuit (list of branches, $\mathcal{B}$, with the fields: from node names, to node names, impedances $R$ and $X$

**1.** Initialize the list of active meters, $\mathcal{L}$, as the list of all meters in the secondary circuit. Set, $\mathcal{B}$ empty.

**2.** If $\mathcal{L}$ has only one meter, **STOP**.

**3.** For all the meter pairs in $\mathcal{L}$, estimate the impedance parameters $R$ and $X$ with (2) and (3).

**4.** Select the linear regression model with the best fit in terms of the root mean squared error of the residuals. Denote the selected model $M_{12}$ and the corresponding meters $m_1$ and $m_2$.

**IF** model $M_{12}$ is of the parallel type (2)

**5.** Add two new branches with the impedance parameters obtained from model $M_{12}$ to $\mathcal{B}$.

**6.** Add a new virtual upstream node with the sum of meter $m_1$ and $m_2$ currents to $\mathcal{L}$.

**7.** Remove meters $m_1$ and $m_2$ from $\mathcal{L}$.

**ELSEIF** model $M_{12}$ is of the series type (3)

**8.** Add one new branch with the impedance parameters obtained from $M_{12}$ to $\mathcal{B}$.

**9.** Add downstream meter $m_2$ currents to the upstream meter $m_1$ currents.

**10.** Remove meter $m_2$ from $\mathcal{L}$.

**ENDIF**

**11.** Go to Step **2.**

When the list of active meters has only one meter left, Algorithm 1 stops and returns the mock circuit consisting of a list of branches, $\mathcal{B}$, with the fields: from node names, to node names, and impedances $R$ and $X$. The mock circuit includes all the meters in the original secondary circuit. The approach shown in [13] adds artificial close-to-zero-impedance branches in the common cases of three or more parallel meters and two or more meters in series. It can be challenging to set an impedance threshold to correctly remove the artificial branches while still preserving all the true branches. By utilizing the two circuit types shown in Fig. 2, Algorithm 1 does not create any artificial close-to-zero-impedance branches provided that at each iteration, the algorithm selects the correct meter pair and regression model type.

### *D. Practical Utility Implementation*

The principle of practical utility implementation of the DSTE algorithm is illustrated in Fig. 3. The algorithm is executed offline without the need of modifying any existing information systems. As inputs, the algorithm requires historical power (or current) and voltage measurements from the advanced metering infrastructure (AMI) and/or DER sensors. The meter and sensor locations as well as the available information of secondary circuit components and connectivity are received from the GIS database. Based on measurement data and known meter locations, the DSTE algorithm estimates the secondary circuit topologies (and parameters) that are verified against the original information (if any) before storing into the GIS.

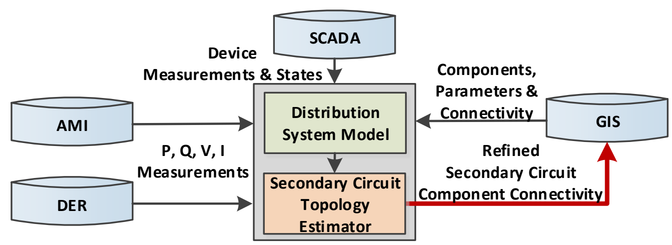


Fig. 3 Practical utility implementation

## IV. Topology Estimation Result Validation

In order to avoid cumbersome manual validations, an automated method was developed to compare an estimated circuit topology with a true circuit topology. We first clarify the requirements for two secondary circuit topologies to have electrically identical topologies (EIT) and then, present an automated algorithm to compare the topologies of two circuits.

*Definition 1 (Electrically identical topologies)* Two secondary circuits have electrically identical topologies if for all the nodes at all the depths of one of the circuits, in the other circuit at the same depth there is a node with equal sets of meters at the node and downstream of the node. This definition allows for different child node ordering between the two trees. The definition is illustrated in Fig. 4, which compares an original (true) topology with an electrically identical and non-identical topologies. The middle circuit has the internal order of the node pairs (1,2), (3,4), and (5,6) switched, but has EIT with the original circuit. In the right circuit, nodes 1 and 3 are siblings instead of nodes 1 and 2 being siblings, so these two circuits do not have EIT.

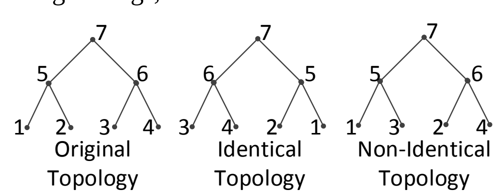


Fig. 4 Example of identical and non-identical secondary circuit topologies

Two circuit topologies cannot be directly compared utilizing node-edge incidence matrices because the internal nodes of the estimated circuit tree are unlabeled, i.e., there is no straightforward correspondence between the internal nodes of the two trees. Mathematically, two graphs are defined to be completely structural equivalent if the graphs are related by *isomorphism*, i.e., if there is a structure-preserving vertex bijection between the two graphs. Definition 1 and the definition of isomorphism (of rooted trees) are similar but not identical because rooted tree isomorphism allows graphs with arbitrarily permuted leaf nodes to be related through isomorphism. Isomorphism of two rooted trees can be verified with the AHU (Aho, Hopcroft, and Ullman) algorithm, which has a linear complexity with respect to the number of nodes $N$, i.e., $\mathcal{O}(N)$ [18]. Algorithm 2 was modified from the AHU algorithm to verify if two trees have EITs. Algorithm 2 correctly identified the tested various secondary circuit pairs, part of which have EITs and others do not have.

**Algorithm 2**: Check if Two Trees Have EIT

**Input**: Rooted trees T1 and T2
**Output**: TRUE / FALSE (do the trees have EIT)

**1.** If the trees do not have the same number of nodes N, return FALSE.
**2.** For all nodes of T1 and T2, determine $L_i$, node $i$ distance from the tree root node. Find the maximum node distance from the root node $\overline{L} = \max_{i \in \{1,\ldots,N\}} L_i$ and calculate the node depth: $D_i = \overline{L} - L_i$.
**3.** If maximum node distances $\overline{L}$ of T1 and T2 are not equal, return FALSE
**4.** Label nodes that have meters with the meter name and nodes without meters with an empty label.

**FOR** all nodes of T1 and T2 at each depth $i = 1, \ldots, \overline{L}$
**a)** Assign each node a label consisting of the sorted child node labels, e.g., if child node labels of a node are "1", "2", "7", then the node is labeled: "((1),(2),(7))".
**b)** If the list of the sorted labels of the nodes at depth $i$ are equal for T1 and T2, return FALSE.
**ENDFOR**
**5.** Return TRUE.

## V. Method Validation on 66-Node Test Circuit

### A. Circuit Overview

The DSTE algorithm was first analyzed on a 66-node test circuit with 10 commonly encountered secondary circuit topologies. Detailed description of the circuit is in [15] and the circuit topology is shown in Fig. 5.

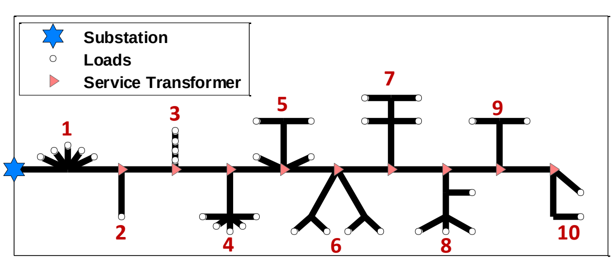


Fig. 5 66-node test circuit topology (secondary circuit numbers in red)

### B. Topology and Parameter Estimation Results

The DSTE algorithm was utilized to estimate the 10 secondary circuit topologies and parameters. The true and the estimated topologies and component parameters of secondary circuits 5 and 8 are compared in Fig. 6. The true and the estimated topologies match perfectly and the estimated impedances are very close to the true impedances.

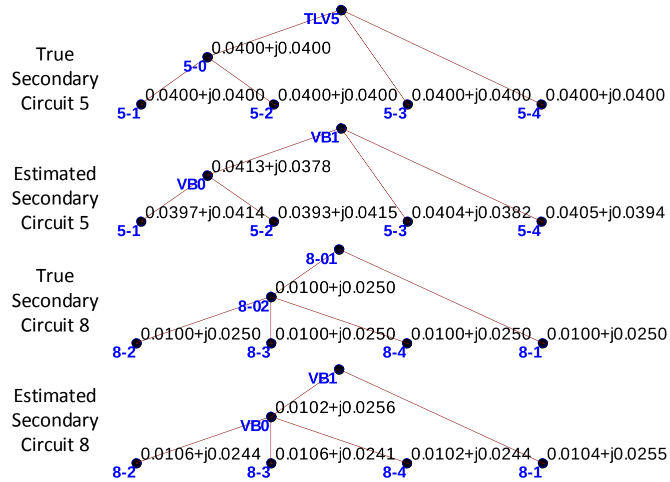


Fig. 6 True and estimated topologies of secondary circuits 5 and 8 (node names in bold blue, node upstream branch impedances, $R + jX$, in black)

Table II lists the average and maximum errors of the $R$ and $X$ parameters estimated with 8760 measurement samples both without and with measurement error. Without measurement error, most parameters are accurately estimated and even the worst-case accuracy is well within acceptable level. With the practical level: 1% P, 1% Q, and 0.2% V measurement error, most parameters are still estimated with a good accuracy and even the worst-case parameter estimation error of around 18% is an enormous improvement from having no information of the secondary system topology and parameters.

TABLE II. The Average Relative Errors of the Estimated R and X

| Meas. Error? | $R_{err,avg}$ [%] | $X_{err,avg}$ [%] | $R_{err,max}$ [%] | $X_{err,max}$ [%] |
|---|---|---|---|---|
| **No** | 0.45 | 0.36 | 2.80 | 1.44 |
| **Yes** | 3.31 | 3.84 | 17.57 | 11.66 |

### C. Discussion

The most challenging part of DSTE in Algorithm 1 is step 4, i.e., correctly selecting the linear regression model that provides the best fit. This led to various challenges with high measurement errors. In certain topologies, if the incorrect meter pair was selected even once, the final estimated topology would be wrong. Thus, it is crucial to pair correct meters at each iteration of the DSTE algorithm. Several regression model selection criteria were analyzed including R-squared and root mean squared error (RMSE). R-squared, which is a metric measuring to which degree a regression model describes the variation of the data, seems to prefer selecting meter pairs with larger voltage drop (larger response variable of the regression model) and thus, lower level of relative measurement noise. This resulted in an incorrect topology in many cases. RMSE turned out to be the best model selection criteria resulting in correct estimation of all the secondary circuit topologies in the 66-node test circuit.

## VI. Georgia Tech Feeder Results

The DSTPE algorithm was utilized to estimate the secondary circuit topologies of the Georgia Tech distribution system, which has been modeled in OpenDSS [19], [20]. Fig. 7 shows two examples of the original secondary circuit model and the estimated secondary circuit. For the Georgia Tech system, the secondary system topology is known from building diagrams, but the impedances, cable types, and lengths are not well known. The algorithm is not perfect, and in some cases when there is significant measurement error, the topologies are not correctly estimated. Like any other statistical estimation methods, the topology estimation results are strongly data-driven and thus, good quality data is an imperative to receive good results.

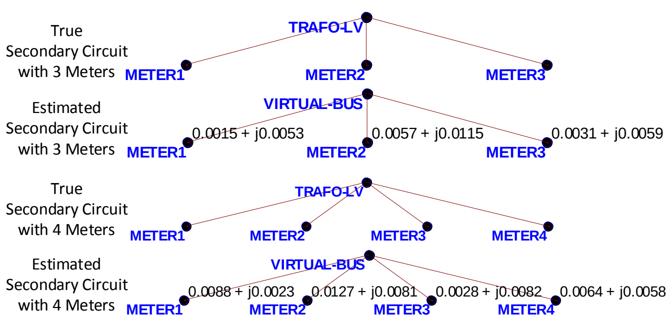


Fig. 7 Original and estimated topology of a secondary circuits with three meters and a secondary circuits with four meters

## VII. Conclusions

This paper presents a practical and computationally efficient algorithm for radial distribution system secondary circuit topology and parameter estimation. The paper also demonstrates the computational infeasibility of exhaustive secondary circuit topology estimation approaches and presents an efficient algorithm for verifying whether two radial secondary circuits have identical topologies.

The topology estimation algorithm correctly estimates all the ten secondary circuit topologies in a 66-node three-phase test circuit both without and with practical levels of measurement noise. The algorithm also correctly estimated approximately 9 out of 10 of the Georgia Tech distribution system secondary circuits.

The major challenge of the topology estimation algorithm is to pair the meters in the correct sequence. In certain topologies, if the incorrect meter pair was selected even once, the final estimated topology would be wrong. Future work will address the load and other conditions under which the algorithm is unable to pair the meters in a correct sequence.

## References

[1] A. Abur, *Power system state estimation: theory and implementation*. New York, NY: Marcel Dekker, 2004.

[2] W. Luan, J. Peng, M. Maras, and J. Lo, “Distribution network topology error correction using smart meter data analytics,” in *IEEE Power and Energy Society General Meeting*, Vancouver, BC, Canada, 2013.

[3] J. S. John, “Can Microinverters Stabilize Hawaii’s Shaky Grid? : Greentech Media,” 02-Feb-2015. [Online]. Available: http://www.greentechmedia.com/articles/read/enphase-to-help-hawaii-ride-its-solar-energy-wave.

[4] A. J. Berrisford, “A tale of two transformers: An algorithm for estimating distribution secondary electric parameters using smart meter data,” in *Annual IEEE Canadian Conference on Electrical and Computer Engineering*, Regina, SK, Canada, 2013.

[5] M. R. M. Castillo, J. B. A. London, and N. G. Bretas, “Network branch parameter validation based on a decoupled State/Parameter Estimator and historical data,” in *PowerTech, 2009 IEEE Bucharest*, 2009, pp. 1–7.

[6] M. R. M. Castillo, J. B. A. London, and N. G. Bretas, “An approach to power system branch parameter estimation,” in *IEEE Electric Power Conference*, Vancouver, BC, Canada, 2008, pp. 1–5.

[7] G. Cavraro, R. Arghandeh, G. Barchi, and A. von Meier, “Distribution network topology detection with time-series measurements,” in *Innovative Smart Grid Technologies Conference (ISGT), 2015 IEEE Power Energy Society*, 2015, pp. 1–5.

[8] G. Cavraro, R. Arghandeh, and A. von Meier, “Distribution Network Topology Detection with Time Series Measurement Data Analysis,” *ArXiv150405926 Cs*, Apr. 2015.

[9] A. S. Raffi and R. Rajagopal, “Feeder Topology Identification,” Mar. 2015.

[10] G. Cavraro, R. Arghandeh, K. Poolla, and A. von Meier, “Data-Driven Approach for Distribution Network Topology Detection,” Apr. 2015.

[11] W. Luan, J. Peng, M. Maras, J. Lo, and B. Harapnuk, “Smart Meter Data Analytics for Distribution Network Connectivity Verification,” *IEEE Trans. Smart Grid*, vol. PP, no. 99, pp. 1–1, 2015.

[12] S. Bolognani, N. Bof, D. Michelotti, R. Muraro, and L. Schenato, “Identification of power distribution network topology via voltage correlation analysis,” in *2013 IEEE 52nd Annual Conference on Decision and Control (CDC)*, 2013, pp. 1659–1664.

[13] T. A. Short, “Advanced metering for phase identification, transformer identification, and secondary modeling,” *IEEE Trans. Smart Grid*, vol. 4, no. 2, pp. 651–658, 2013.

[14] J. Peppanen, M. J. Reno, M. Thakkar, S. Grijalva, and R. G. Harley, “Leveraging AMI Data for Distribution System Model Calibration and Situational Awareness,” *IEEE Trans. Smart Grid*, 2015.

[15] J. Peppanen, M. J. Reno, R. Broderick, and S. Grijalva, “Distribution System Secondary Circuit Parameter Estimation for Model Calibration,” Sandia National Laboratories, SAND2015, Aug. 2015.

[16] J. Felsenstein, “The Number of Evolutionary Trees,” *Syst. Zool.*, vol. 27, no. 1, pp. 27–33, Mar. 1978.

[17] T. Gönen, *Electric power distribution system engineering*, 2nd ed. Boca Raton: CRC Press, 2008.

[18] A. V. Aho, J. E. Hopcroft, and J. D. Ullman, *The design and analysis of computer algorithms*. Reading, Mass: Addison-Wesley Pub. Co, 1974.

[19] J. Peppanen, J. Grimaldo, M. J. Reno, S. Grijalva, and R. G. Harley, “Increasing distribution system model accuracy with extensive deployment of smart meters,” in *IEEE Power and Energy Society General Meeting*, Washington D.C., 2014.

[20] M. J. Reno and K. Coogan, “Grid Integrated Distributed PV (GridPV) Version 2,” Sandia National Laboratories, SAND2014-20141, 2014.

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy’s National Nuclear Security Administration under contract DE-AC04-94AL85000.