

Low-voltage network topology and impedance identification using smart meter measurements

Amina Benzergera, Daniele Maruli, Antonio Sutera, Alireza Bahmanyar, Sébastien Mathieu, Damien Ernst

University of Liège

Liège, Belgium

{abenzergera, dmaruli, a.sutera, abahmanyar, smathieu, dernst}@uliege.be

Abstract—Distribution system operators have been upgrading their network over several decades, though not always keeping digital records of all changes. As a result, the operators do not always know exactly how their customers are connected to a network. Some of these customers are equipped with smart meters, providing voltage and current time-series. These measurements can be used to identify the network topology and the line impedances. This paper presents a method to identify radially operated low-voltage networks which can be applied with limited number of smart meters. The resulting identified model provides the map of the network and impedances of the inferred lines, allowing to perform subsequent analyses (e.g. power-flow). Simulation results on a case study with 128 nodes show an average error of 0.69% in computed voltages, while only 40% of the nodes are equipped with smart meters.

Index Terms—distribution system, impedance inference, low-voltage network, smart meters, topology identification.

I. INTRODUCTION

With the increasing use of distributed energy resources and electric vehicles, distribution system operators (DSOs) are encountering serious difficulties in guaranteeing the safety of their Low-Voltage (LV) network in the years to come. Being able to effectively integrate distributed energy resources and electric vehicles is a fundamental step to accelerate the energy transition process. At LV levels, distribution networks are mostly operated radially, and power is distributed through several feeders, i.e. main electrical lines carrying power from the substation to the customer. Most residential loads are connected to the feeder through a single phase and a neutral wire. The phase to which a load is connected may be selected arbitrarily. Power imbalances between phases are expected, leading to a reduced hosting capacity of the system. In order to implement effective preventive or corrective measures against voltage or congestion issues, DSOs need to be able to assess the system's response to various realistic scenarios. This analysis is usually performed through power-flow studies, but reliable solutions require accurate information about topology of the network and physical characteristics of the lines. DSOs do not always know how households, feeders and

other appliances are interconnected. Furthermore, LV networks topology can change over time because of faults, maintenance or reconfiguration, and existing databases can contain outdated or inaccurate information. DSOs can therefore lack possession of a reliable model of their LV network, which can hinder the efficient management and development of their system. Network identification is the mathematical process that allows to deduce this information. Effective network identification methods for LV networks are, thus, essential for the development of smarter grids [1]. This work presents a methodology to retrieve the topology and the cable parameters of an LV network from time-series measurements provided by a limited number of smart meters in the grid.

This paper is structured as follows. Section II reviews relevant literature. Section III states the problem, defining inputs and outputs. Section IV describes the methodology proposed to tackle the problem. Section V examines a specific case study to evaluate the potential of such methodology and discuss the results. Finally, Section VI summarises the conclusions and introduces potential future work.

II. LITERATURE REVIEW

In transmission systems, topology and cable parameter information is usually stored in appropriate databases [1]. Unfortunately, the same does not apply to LV distribution networks. With the increasing importance and complexity of distribution systems, the identification problem of LV networks has gained more attention and it has been recently tackled adopting different approaches. Recent literature, such as [1]–[3], focuses on identifying the topology of the network when limited information is available. Algorithms that aim, as this work does, to identify both network topology and lines parameters at the same time, are presented in [4]–[7]. Authors of [4] use the evaluation of voltage sensitivities with respect to active and reactive power injections and Prüfer sequences to identify the topology of small networks, assuming that only specific cables types and lengths are used for the lines. The identification problem in [5] takes the name of inverse power flow problem, where the system admittance matrix is found by solving an unconstrained least-squares problem. The case with non-measured nodes in the grid, also referred to as *hidden nodes*, is also tackled, both for meshed and also radial topology, with the assumption that these hidden nodes have zero net current injections. The inverse power-

This research is supported by the public service of Wallonia within the framework of the Silver project. AS is supported via the Energy Transition Funds project EPOC 2030-2050 organised by the FPS economy, S.M.E.s, Self-employed and Energy.

flow problem is extended to poly-phase systems in [6], with the full-observability assumption. Finally, an algorithm to jointly estimate both admittance and topology, assuming that the measurements for all non-zero power injecting nodes are available, is presented in [7]. A summary comparison table can be found in Table I. All the methods that have been examined share the assumption that every node in the network, or at least the power-injecting ones, has a meter attached to it. This paper presents a methodology to tackle the network identification problem even when some power-injecting nodes in the grid are not metered.

TABLE I: Literature review where \circ indicates that the article does not provide feature, \bullet partially provides feature, \bullet provides feature.

Features	papers	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[*]
Line parameters estimation		\circ	\circ	\circ	\bullet	\bullet	\bullet	\bullet	\bullet
Hidden non-injecting nodes		\circ	\circ	\circ	\bullet	\bullet	\circ	\bullet	\bullet
Hidden injecting nodes		\circ	\circ	\circ	\circ	\circ	\circ	\circ	\bullet
Required measurements		V	V	E	V, S	V, I	V, I	V, S	V, I
No assumptions on cables		\bullet	\bullet	\bullet	\circ	\bullet	\bullet	\bullet	\bullet
Unbalanced poly-phase		\circ	\circ	\circ	\circ	\circ	\bullet	\circ	\bullet

* Proposed methodology
V = Voltage; I = Current; S = Power; E = Energy

III. PROBLEM STATEMENT

Let us consider a three-phase four-wire LV radial distribution network, as represented in Figure 1. Let \mathcal{N} be the set of nodes in the network and let ψ_n be the two-element vector representing the geographical coordinates of node $n \in \mathcal{N}$. The three phases of the network are denoted by indices a, b and c and the set of phases by $\mathcal{P} = \{a, b, c\}$.

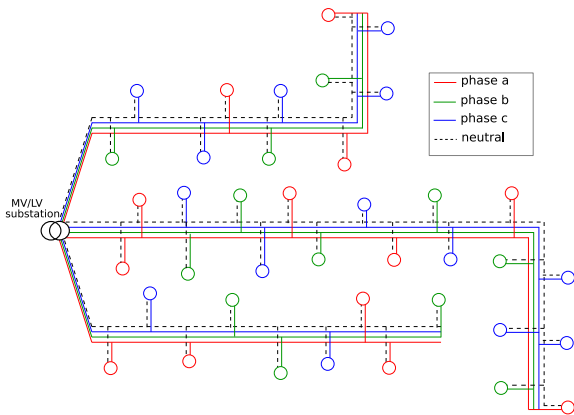


Fig. 1: Representation of a three-phase four-wire radial network with three feeders.

The network is observed over a finite time period \mathcal{T} discretised into time intervals of length Δt . The value of a variable in time-step $t \in \mathcal{T}$ is accessed by using an additional

subscript, i.e. $\bullet_{i,t}$ refers to the value of variable \bullet_i in time-step t . The absence of such a subscript denotes the entire time-series. Let $V_{n,t}^p$ and $I_{n,t}^p$ be the magnitudes of the voltage and the current injection in phase $p \in \mathcal{P}$ of node n , respectively. The set $\mathcal{M} \subseteq \mathcal{N}$ is the subset of nodes in the network that are equipped with smart meters. Information about $\psi_m, V_m^p, I_m^p, \forall m \in \mathcal{M}$ is considered as input of the problem. Smart meters are assumed to provide measurements over the entire observation period of voltage and current injection at phase p of the node $m \in \mathcal{M}$ to which they are connected. The method uses magnitude measurements since smart meters do not usually include a phasor measurement unit.

The aim of a network identification problem is to retrieve a model of the network describing both the network topology and edge impedances. Let $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ be a graph representing the network topology, with \mathcal{E} as set of edges. Let Z_e be the impedance associated to edge $e \in \mathcal{E}$ and \mathcal{Z} be the set of all the impedances. Let $\mathcal{H} = \mathcal{N} \setminus \mathcal{M}$ be the set of unmetered nodes in the network.

The network identification problem infers \mathcal{E}, \mathcal{Z} and $I_h^p, \forall h \in \mathcal{H}$, assuming that:

- the topology of the network does not change during the observation period;
- customers are connected to the main feeder through single-phase two-wire connection;
- the phase of the network to which metered customers are connected is provided;
- at least one node at every phase of every feeder of the network is equipped with a meter;
- each phase of the MV/LV substation is metered;
- the ratios between cable resistance and reactance are known, and denoted by γ .

The estimation of \mathcal{E}, \mathcal{Z} and $I_h^p, \forall h \in \mathcal{H}$ is valuable information to DSOs since it can be exploited, as an example, to perform power flow computations or to identify problematic network configurations.

IV. PROPOSED METHODOLOGY

The methodology is divided in two steps: the identification of three single-phase graphs and merging these into a three-phase one.

A. Single-phase identification algorithm

This step aims at modelling the phase $p \in \mathcal{P}$ of a network as a rooted tree $\hat{\mathcal{G}}^p$. The identification algorithm consists of three main parts, namely topology estimation, topology validation, and hidden node detection, that are applied to each phase $p \in \mathcal{P}$ independently. Let $\hat{\mathcal{N}}^p \subseteq \mathcal{N}^p$ be the pool of nodes made available to the identification algorithm. $\hat{\mathcal{N}}^p$ initially corresponds to the set of observed nodes \mathcal{M}^p .

Topology Estimation

The first step of the algorithm consists of the estimation of a rooted tree $\hat{\mathcal{G}}^p = (\hat{\mathcal{N}}^p, \hat{\mathcal{E}}^p)$ with a topology as close as possible to $\mathcal{G}^p = (\mathcal{N}^p, \mathcal{E}^p)$. This operation is carried out using correlation analysis on voltage measurements to infer

proximity between nodes. Correlation-based approaches have already been proven successful both in phase [8] and topology [1] identification methods. Load and production profiles at different nodes of the LV network can present similar patterns owing to comparable residential occupancy profiles and weather conditions. To be sure that this does not affect the results of the voltage correlation analysis, voltage time-series are pre-processed applying a high-pass filter, as suggested in [1]. Let w_{ij} be the Pearson Correlation Coefficient between the filtered voltage time-series of nodes i and j . Let \hat{G}_w^p be a weighted complete graph built on $\hat{\mathcal{N}}^p$, where the edge (i, j) weight is equal to w_{ij} . The estimated topology $\hat{G}^p = (\hat{\mathcal{N}}^p, \hat{\mathcal{E}}^p)$ is obtained computing the *maximum spanning tree* on \hat{G}_w^p .

Topology Validation

The second process checks each edge $e = (i, j) \in \hat{\mathcal{E}}^p$ and suggests where missing nodes and wrong connections, if any, are. This process is performed by evaluating the estimated impedance magnitude time-series $\hat{Z}_e^p = \{\hat{Z}_{e,1}^p, \dots, \hat{Z}_{e,T}^p\}$ for each edge $e = (i, j) \in \hat{\mathcal{E}}^p$ as, $\forall t \in \{1, \dots, T\}$:

$$\hat{Z}_{e,t}^p = \frac{V_{j,t}^p - V_{i,t}^p}{\hat{I}_{e,t}^p} \quad (1)$$

where the estimated current $\hat{I}_{e,t}^p$ flowing in e is given by

$$\hat{I}_{e,t}^p = \sum_{n \in \hat{\mathcal{D}}_e^p} \hat{I}_{n,t}^p \quad (2)$$

where $\hat{\mathcal{D}}_e^p$ is the set of downstream nodes with respect to edge e . Since impedances are constants, the values \hat{Z}_e^p should be close to the same constant value. Edge e is considered valid if the relative standard deviation of time-series \hat{Z}_e^p , $RSD(\hat{Z}_e^p)$, is less than an arbitrary threshold λ , e.g. determined by statistical tests for the largest accepted standard deviation.

Hidden Node Detection

Assume a hidden node $H \in \mathcal{H}^p$ exists in the path connecting it to the root node. Since the contribution of \hat{I}_H^p is missing in the computation of the topology validation step, the estimated current flowing through such path is not the correct one. This causes the rejection of the edges in that path, as shown in Figure 2. Note also that since node 5 is hidden and it is not a terminal node, the previous topology estimation step mistakenly detects an edge connecting node 3 to node 6. Let $A \in \hat{\mathcal{N}}^p$ be the node with the longest path to the root among all the nodes that are connected to a rejected edge in $\hat{\mathcal{U}}^p$ (node 6 in Figure 2). Let B and C be the parent and grandparent nodes of A , respectively. An unobserved node $H \in \mathcal{H}^p$ may be placed adjacent to A in three topological configurations, as shown in Figure 3.

Let $A \in \hat{\mathcal{N}}^p$ be the node, among all the nodes that are connected to a rejected edge, with the longest path to the root (node 6 in Figure 2). Let nodes B and C be the parent and grandparent node of A , respectively. Node $X \in \mathcal{H}^p$ is assumed to be adjacent to A . Three topological configurations can occur, as shown in Figure 3. The hidden node detection

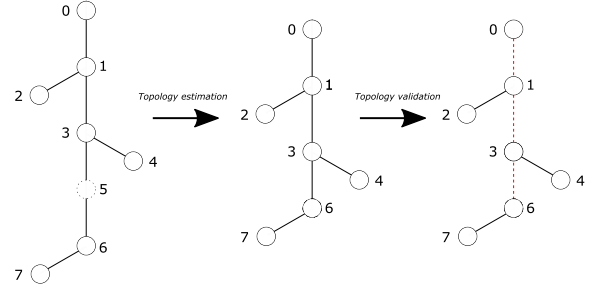


Fig. 2: First two steps of the identification algorithm with node 5 as hidden node. Dashed edges are rejected by the validation.

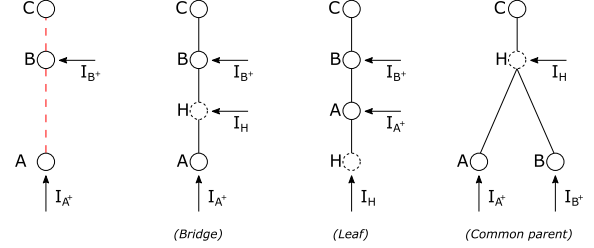


Fig. 3: Possible locations for a hidden node H adjacent to A .

step examines these configurations to assess which suits best. Let $\hat{I}_{A+}^p = \hat{I}_{AB}^p$ be the contribution to the estimated \hat{I}_{BC}^p flowing through BC of A and its descendant nodes, and

$$\hat{I}_{B+}^p = \hat{I}_{AB}^p - \hat{I}_{BC}^p \quad (3)$$

be the rest of the current. In order to detect the correct location of X , the algorithm solves three optimization problems:

Configuration “Bridge”:

$$\delta^b = \min \sum_{t \in \mathcal{T}} |V_{B,t} - V_{C,t} - \hat{Z}_{CB}^p (\hat{I}_{A+,t}^p + \hat{I}_{B+,t}^p + \hat{I}_{X,t}^p)| \quad (4a)$$

subject to, $\forall t \in \mathcal{T}$

$$V_{B,t} = V_{A,t} - \hat{Z}_{XA}^p \hat{I}_{A+,t}^p - \hat{Z}_{BX}^p (\hat{I}_{A+,t}^p + \hat{I}_{X,t}^p) \quad (4b)$$

with $(\hat{I}_X^p, \hat{Z}_{XA}^p, \hat{Z}_{BX}^p, \hat{Z}_{CB}^p) \in \mathbb{R}^4$.

Configuration “Leaf”:

$$\delta^l = \min \sum_{t \in \mathcal{T}} |V_{B,t} - V_{C,t} - \hat{Z}_{CB}^p (\hat{I}_{A+,t}^p + \hat{I}_{B+,t}^p + \hat{I}_{X,t}^p)| \quad (5a)$$

subject to, $\forall t \in \mathcal{T}$,

$$V_{B,t} = V_{A,t} - \hat{Z}_{BA}^p (\hat{I}_{A+,t}^p + \hat{I}_{X,t}^p) \quad (5b)$$

with $(\hat{I}_X^p, \hat{Z}_{BA}^p, \hat{Z}_{CB}^p) \in \mathbb{R}^3$.

Configuration “Common parent”:

$$\delta^{cp} = \min \sum_{t=0}^T |(V_{A,t}^p - \hat{Z}_{XA}^p \hat{I}_{A+,t}^p) - (V_{B,t}^p - \hat{Z}_{XB}^p \hat{I}_{B+,t}^p)| \quad (6)$$

with $(\hat{Z}_{XA}^p, \hat{Z}_{XB}^p) \in \mathbb{R}^2$.

Such problems are formulated by exploiting the current and voltage relationships occurring in each configuration. Once problems (4 - 6) have been solved, the algorithm selects the

configuration with the smallest $\delta \in \{\delta^b, \delta^l, \delta^{cp}\}$. Depending on the predicted location of X , the pool of nodes $\hat{\mathcal{N}}^p$ is updated accordingly. If the algorithm picks the bridge configuration, an additional node X is added to $\hat{\mathcal{N}}^p$. The estimated current injection magnitude \hat{I}_X^p is extracted from solution of (4), along with the values of \hat{Z}_{XA}^p and \hat{Z}_{BX}^p . The estimation of the voltage time-series V_X^p is computed as:

$$\hat{V}_{X,t}^p = V_{A,t}^p - \hat{Z}_{XA}^p \hat{I}_{A+,t}^p \quad \forall t \in \{1, \dots, T\} \quad (7)$$

If the leaf configuration is selected, the current injection \hat{I}_X^p is extracted from the solution of (5), while the voltage time-series of X cannot be estimated. To proceed with the identification process, node $A \in \hat{\mathcal{N}}^p$ is substituted by an auxiliary node X' with the same voltage as A and its current injection is given by $\hat{I}_{X'}^p = \hat{I}_A^p + \hat{I}_X^p$. This allows one to fix the current flowing in the path to the root of A for the next validation step and to continue the identification process. If the common parent configuration is selected, this implies that impedances \hat{Z}_{XA}^p and \hat{Z}_{XB}^p could be found, which leads to a common voltage \hat{V}_X^p . The voltage magnitude associated with \hat{V}_X^p is computed as the average voltage given by the two voltage drops:

$$\hat{V}_{X,t}^p = (V_{A,t}^p - \hat{Z}_{XA}^p \hat{I}_{A+,t}^p + V_{B,t}^p - \hat{Z}_{XB}^p \hat{I}_{B+,t}^p) / 2 \quad (8)$$

A node X is added to $\hat{\mathcal{N}}^p$ with voltage \hat{V}_X^p and a zero net current injection, since (6) voltages and currents in nodes A and B do not provide information to estimate any potentially missing current injection. If X is indeed a net zero-power injecting node, edge CX is accepted in the topology by the validation step. Otherwise, the correct \hat{I}_{CX}^p is obtained in the next validation iteration by solving problem (5), since node X is added in a leaf configuration.

Whenever more than one hidden node is located in the same area, solving problems (4) - (6) may not provide a meaningful estimation of X corresponding to an actual hidden node in the actual network. To check if the investigated area of the network has been correctly identified, the algorithm checks whether or not the addition of H would pass the topology validation step. The hidden node is accepted if the relative standard deviation of impedance time-series Z_{CX} for *bridge* and *leaf* configurations, or both Z_{AX} and Z_{BX} for the *common parent* one, are below the tolerance λ . If not, to proceed with the network identification, a new node Y is added to the the pool of nodes $\hat{\mathcal{N}}^p$ instead, obtained using a different approach. The algorithm arbitrarily sets the position of the hidden node Y midway between nodes A and B , connected to them by two branches of equal impedances, to remove one degree of freedom to the problem. The impedance associated to edges YA and BY , i.e., $\hat{Z}_{YA}^p = \hat{Z}_{BY}^p$, is evaluated by multiplying a per-unit-of-length default impedance value Z^* by the distance between the two nodes it connects. Voltage and current injection of Y are computed, following Ohm's law, as:

$$\hat{V}_{Y,t}^p = V_{A,t}^p - \hat{Z}_{YA}^p \hat{I}_{A+,t}^p \quad \forall t \in \mathcal{T} \quad (9)$$

$$\hat{I}_{Y,t}^p = \frac{V_{B,t}^p - V_{Y,t}^p}{\hat{Z}_{BY}^p} - \hat{I}_{A+,t}^p \quad \forall t \in \mathcal{T} \quad (10)$$

Once the pool of nodes $\hat{\mathcal{N}}^p$ has been updated, topology validation and estimation steps are processed again. The algorithm is performed until all the edges in $\hat{\mathcal{G}}^p$ are labelled as valid. Finally, the impedance magnitude of each validated single-phase edge is given by the mean value of \hat{Z}_e^p in equation (1).

B. Three-phase model identification

This step identifies a three-phase model of the network starting from the identified single-phase graphs. Let $\hat{\mathcal{N}}$ and $\hat{\mathcal{F}}$ be the sets of all the nodes and feeders in the single phase graphs $\hat{\mathcal{G}}^a$, $\hat{\mathcal{G}}^b$ and $\hat{\mathcal{G}}^c$ respectively.

Each inferred single-phase graph $\hat{\mathcal{G}}^p$ contains \hat{F}^p paths starting from the root node corresponding to one feeder each. These paths must be gathered into $\hat{F} = \max\{\hat{F}^p | \forall p \in \mathcal{P}\}$ groups containing one feeder from each phase. These groups are obtained by solving the following optimization problem:

$$\min_{x \in \mathbb{B}^{|\hat{\mathcal{F}}| \times |\hat{\mathcal{F}}|}} \sum_{f_1 \in \hat{\mathcal{F}}} \sum_{f_2 \in \hat{\mathcal{F}}} x_{f_1, f_2} \Delta_{f_1, f_2} \quad (11a)$$

subject to:

$$x_{f_1, f_2} = x_{f_2, f_1} \quad \forall (f_1, f_2) \in \hat{\mathcal{F}}^2 \quad (11b)$$

$$\sum_{f_2 \in \hat{\mathcal{F}}} x_{f_1, f_2} = 3 \quad \forall f_1 \in \hat{\mathcal{F}} \quad (11c)$$

$$x_{f_1, f_2} \geq x_{f_3, f_1} + x_{f_3, f_2} - 1 \quad \forall (f_1, f_2, f_3) \in \hat{\mathcal{F}}^3 \quad (11d)$$

where f_1 and f_2 are two single-phase feeders, x_{f_1, f_2} is a binary variable equal to 1 if f_1 and f_2 are in the same three-phase feeder. Symmetry of x is imposed by equality (11b). Equation (11c) ensures that every three-phase feeder contains one feeder from each single-phase graph. Constraint (11d) imposes that if single-phase feeder f_3 is associated to single-phase feeder f_1 and to single-phase feeder f_2 then f_1 and f_2 are also associated and the three of them belongs to the same three-phase feeder. The distance between two single-phase feeders, Δ_{f_1, f_2} , is defined by:

$$\Delta_{f_1, f_2} = \frac{\sum_{k \in f_1} \delta_{k, f_2}}{|f_1|} + \frac{\sum_{k \in f_2} \delta_{k, f_1}}{|f_2|} \quad (12)$$

where $|f_1|$ and $|f_2|$ are the number of nodes in f_1 and f_2 , respectively. The distance between a node k and a single-phase feeder f is defined as:

$$\delta_{k, f} = \min\{\|\psi_k - (\rho\psi_i + (1-\rho)\psi_j)\|, \forall (i, j) \in f, \rho \in [0, 1]\}, \quad (13)$$

where i and j are nodes of f . The right part of the subtraction is the coordinates of the closest point to node k on the part of the feeder delimited by these two nodes. This distance is illustrated in Figure 4 for two single-phase feeders.

Now that the single-phase feeders are grouped, the nodes belonging to feeders in the same group must be linked together to form the feeders of the three-phase graph. Optimization problem (14) builds these links by solving a minimum spanning tree problem involving node depths [9]. The solution to

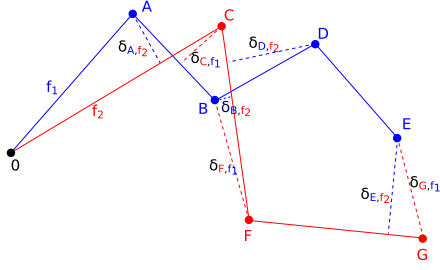


Fig. 4: Distance between two single-phase feeders (f_1 and f_2) from different phases.

this problem is an incidence matrix y , defining the edges $\hat{\mathcal{E}}$ of the graph $\hat{\mathcal{G}} = (\hat{\mathcal{N}}, \hat{\mathcal{E}})$.

$$\min_{y \in \mathbb{B}^{|\hat{\mathcal{N}}| \times |\hat{\mathcal{N}}|}, l \in \mathbb{N}^{|\hat{\mathcal{N}}|}} \sum_{i \in \hat{\mathcal{N}}} \sum_{j \in \hat{\mathcal{N}}} y_{i,j} W_{i,j} \quad (14a)$$

subject to:

$$\sum_{i \in \hat{\mathcal{N}}} y_{i,0} = 0 \quad (14b)$$

$$\sum_{i \in \hat{\mathcal{N}}} y_{0,i} = \hat{F} \quad (14c)$$

$$\sum_{j \in \hat{\mathcal{N}}} y_{j,i} = 1 \quad \forall i \in \hat{\mathcal{N}} \quad (14d)$$

$$l_0 = 0 \quad (14e)$$

$$l_j \geq l_i + y_{i,j} - |\hat{\mathcal{N}}|(1 - y_{i,j}) \quad \forall (i,j) \in \hat{\mathcal{N}}^2 \quad (14f)$$

$$l_j \leq l_i + 1 + |\hat{\mathcal{N}}|(1 - y_{i,j}) \quad \forall (i,j) \in \hat{\mathcal{N}}^2 \quad (14g)$$

$$l_j \geq l_i + 1 \quad \forall (i,j) \in \hat{\mathcal{E}}^p \quad (14h)$$

Objective function (14a) minimises the distance between the selected edges ($y_{i,j} = 1$). Distance $W_{i,j}$ is given by the Euclidean distance between node i and node j only if they belong to the same phase or the same feeder:

$$W_{i,j} = \begin{cases} \|\psi_i - \psi_j\| & \text{if } (i,j) \in \hat{\mathcal{E}}^p \text{ or if } x_{i,j} = 1 \\ +\infty, & \text{otherwise.} \end{cases} \quad (15)$$

Constraints (14b)-(14e) force the result to be a rooted tree that presents the substation as its root node 0 and the expected number of feeders in the topology. The node depth of node j , l_j , is defined by equations (14e)-(14g). The root node depth is set to zero by equality (14e). Equations (14f)-(14g) states that if node i is connected to node j , $l_i = l_j + 1$. Equation (14h) imposes that if, in the single-phase graph, node i is followed by node j , the depth of node i in the three-phase graph must be greater than the one for j .

Finally, impedances are assigned to the resulting three-phase edges. The path impedance of an edge $e^p \in \hat{\mathcal{E}}^p, \forall p \in \mathcal{P}$ is

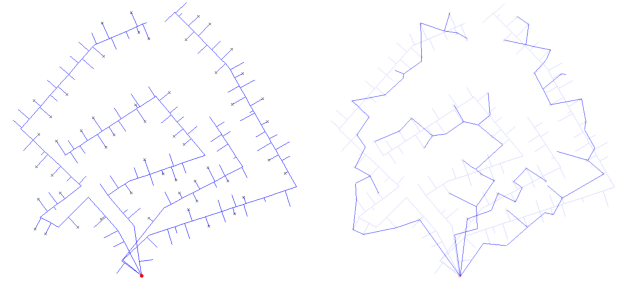
$$Z_e^p = \frac{Z_e^p}{d_{\hat{\Pi}(e^p)}} l_e \quad \forall b \in \hat{\Pi}(e^p) \quad (16)$$

where Z_e^p is the single-phase impedance estimated in Section IV-A and d_e the length of the set of consecutive edges, connecting i to j in the three-phase topology $\hat{\mathcal{G}}$.

Since this process is computed for all the single-phase edges, each edge in the three-phase model has three different impedance values associated to it. Assuming that an edge in the model has the same type of conductors in every phase, the self-impedance of each three-phase edge in the network model is set to the mean value.

V. SIMULATION RESULTS

The developed methodology is tested on a three-phase four-wire network with 128 single-phase customers similar to a Belgian LV network, depicted in Figure 5a. The voltage readings are obtained by simulating the network in OpenDSS over 30 days at a 30 minutes resolution, resulting in 1440 time-steps. For this power-flow analysis, two different cable types, whose self-impedance values are $\|Z_1\| = 0.0012 \Omega/m$ and $\|Z_2\| = 0.0009 \Omega/m$, are selected to model the main feeders and their laterals, respectively. Consumption profiles are obtained from residential consumption readings of the Low Carbon London Project [10]. Active and reactive power profiles are generated associating a different power factor ranging from 0.93 to 0.97 to each customer.

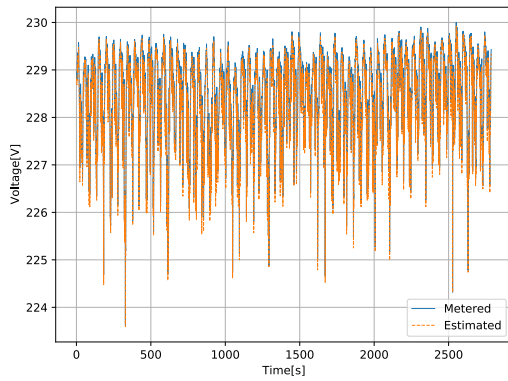


(a) Reference topology.

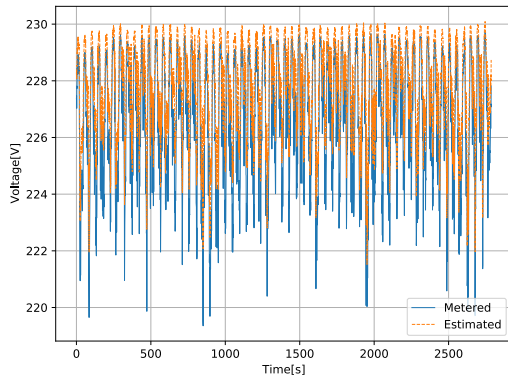
(b) Estimated topology.

Fig. 5: Network case study where each marker in Figure 5a corresponds to a metered customer.

Only 40% of customers' nodes outputted by the power flow analysis are randomly selected as smart meters nodes. The inputs of the identification algorithm are voltages and current injections of these nodes and their geographical data. To simulate common smart meter measurements, only magnitude values are used as input in the algorithm with a unit power factor. The hyper-parameter settings are the following: $\lambda = 0.1$, $\gamma = 0.1$ and $Z^* = 0.001 \Omega/m$. The estimated topology of the network is presented in Figure 5b. This figure shows that the proposed methodology accurately identifies the general shape of the network. The algorithm detects the right number of feeders and correctly associates the metered nodes. Estimated edge impedances modules have a mean value of $0.00106 \Omega/m$ with a standard deviation of $0.00098 \Omega/m$ which is well within the expected range $[0.0009, 0.0012]$. Note that 10 of the 84 estimated impedances are set using Z^* (see end of Section IV-A). Not considering these values still leads to impedances of average values of $0.00108 \Omega/m$ with a standard deviation of $0.00104 \Omega/m$.



(a) Node 180



(b) Node 215

Fig. 6: Metered and estimated voltage magnitude.

Another load-flow analysis is computed using the topology, the edge characteristics and the estimated current injections of the unmetered nodes. This load flow provides voltage time-series estimations which can be compared to the input readings. The minimum, median and maximum observed Root-Mean-Square Errors (RMSE) are 0.09%, 0.69% and 3.53% respectively. This suggests that the model, despite the few information provided to the algorithm, can lead to reliable voltage approximations. Figure 6 shows voltage magnitude of two other nodes. While Figure 6a depicts a low estimation error, the estimated voltage profile of Figure 6b presents several under-voltages corresponding to less severe cases in the correct time-series. This error is due to a localized inaccuracy in the topology estimation that led to under-estimation of the consumption around the meter.

produce a model whose topology reflects the structure of the

VI. CONCLUSION

This paper proposes an algorithm which identifies the topology of a low-voltage network when a subset of nodes, either customers or nodal points of connection, is not equipped with smart meters. This objective is achieved by analysing voltages, currents and geographical data of the metered nodes, without relying on additional information. The performances of the algorithm are evaluated on a case study with 128 customers, 51 of them equipped with smart meters. Despite the low observability of the system, the algorithm is able to

network. A load-flow analysis performed using the inferred model shows that the computed voltage time-series matches the correct values with average RSME of less than 1%.

Future work could focus on how to exploit the solution of additional load-flow analysis using the estimated model to understand which area in the model topology presents larger inaccuracy and to investigate how to improve it. Furthermore, it would be interesting to investigate what the minimum number of meters needed is to obtain relevant solutions and where they should be installed for an optimal estimation of the model. Additionally, further effort could be put into a more exhaustive estimation of the edge parameters, taking into account the evaluation of mutual and shunt impedances. Finally, the model provided by this algorithm could be used for further studies to maximise the integration of renewable generation and electric car connections within LV networks.

REFERENCES

- [1] S. Bolognani, N. Bof, D. Michelotti, R. Muraro, and L. Schenato, "Identification of power distribution network topology via voltage correlation analysis," *IEEE 52nd Conference on Decision and Control*, 2013.
- [2] Y. Weng, Y. Liao, and R. Rajagopal, "Distributed energy resources topology identification via graphical modeling," *IEEE Transactions on Power Systems*, 2016.
- [3] S. Pappu, N. Bhatt, R. Pasumarthy, and A. Rajeswaran, "Identifying topology of low voltage distribution networks based on smart meter data," *IEEE Transactions on Smart Grid*, 2017.
- [4] K. Soumalas, G. Messinis, and N. Hatziargyriou, "A data driven approach to distribution network topology identification," *2017 IEEE Manchester PowerTech*, 2017.
- [5] Y. Yuan, O. Ardakanian, S. Low, and C. Tomlin, "On the inverse power flow problem," *arXiv preprint arXiv:1610.06631*, 2016.
- [6] O. Ardakanian, Y. Yuan, V. Wong, R. Dobbe, S. Low, A. von Meier, and C. J. Tomlin, "On identification of distribution grids," *IEEE Transactions on Control of Network Systems*, 2019.
- [7] J. Yu, Y. Weng, and R. Rajagopal, "PaToPa: A data-driven parameter and topology joint estimation framework in distribution grids," *IEEE Transactions on Power Systems*, 2017.
- [8] F. Olivier, A. Sutera, P. Geurts, R. Fonteneau, and D. Ernst, "Phase identification of smart meters by clustering voltage measurements," *2018 Power Systems Computation Conference*, 2018.
- [9] T. F. Abdelmaguid, "An efficient mixed integer linear programming model for the minimum spanning tree problem," *Mathematics 2018*, 2018.
- [10] UK Power Networks, *SmartMeter Energy Consumption Data in London Households*, <https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households>.