Author/s:
Fang, L;Pengwah, AB;Andrew, LLH;Razzaghi, R;Munoz, MA

Title:
Three-phase voltage sensitivity estimation and its application to topology identification in low-voltage distribution networks

Date:
2024-07

Citation:
Fang, L., Pengwah, A. B., Andrew, L. L. H., Razzaghi, R. & Munoz, M. A. (2024). Three-phase voltage sensitivity estimation and its application to topology identification in low-voltage distribution networks. International Journal of Electrical Power and Energy Systems, 158, https://doi.org/10.1016/j.ijepes.2024.109949.

Persistent Link:
https://hdl.handle.net/11343/345151

Electrical Power and Energy Systems 158 (2024) 109949

Contents lists available at ScienceDirect

International Journal of Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes

Three-phase voltage sensitivity estimation and its application to topology identification in low-voltage distribution networks

Luxin Fang [a], Abu Bakr Pengwah [b], Lachlan L.H. Andrew [a,c,*], Reza Razzaghi [b], Mario Andrés Muñoz [a,c]

[a] Faculty of Engineering and Information Technology, The University of Melbourne, Parkville, 3010, VIC, Australia
[b] Department of Electrical and Computer Systems Engineering, Monash University, Clayton, 3800, VIC, Australia
[c] ARC Centre in Optimisation Technologies, Integrated Methodologies, and Applications, 700 Swanston Street, Carlton, 3053, VIC, Australia

ARTICLE INFO

Keywords:
Low-voltage distribution networks
Voltage sensitivity estimation
Network topology estimation
Smart meters

ABSTRACT

This paper aims to estimate the three-phase voltage sensitivity matrix and the network topology of a low-voltage distribution network from smart meter data, which measures voltage magnitude, current magnitude, and power factor with a lead/lag indicator. The targeted networks are three-phase networks with single-phase loads. Understanding network sensitivity and topology is crucial for fault detection, unmetered load identification, and addressing Dynamic Operating Envelope problems. The problem is formulated as a constrained optimization problem to estimate both the three-phase voltage sensitivity and the low-voltage transformer voltage. The estimated voltage sensitivity is used to further identify the network topology, by implementing an enhanced Recursive-Grouping and Backtracking algorithm, as well as a candidate topology selection technique. The proposed method is tested on the 55-node European feeder and several synthetic networks. Compared to the state-of-the-art, the results show a four-fold improvement in the accuracy of voltage sensitivity estimation and substantially fewer mistakes in the topology estimates. The result underscores the efficacy of a three-phase network model and voltage angle approximation in enhancing estimation accuracy.

1. Introduction

Thanks to advancements in computer science and information technology, a wide range of applications for power distribution networks has emerged in recent years. To monitor and control the network, demand-response [1] and distribution management systems (DMS)[2] have been developed, which require accurate knowledge of the network structure. This includes knowledge of both the network topology and line impedances. Such knowledge is also invaluable for applications including transmission line fault detection and classification [3], detection of unmetered loads and electricity theft [4], microgrid protection [5], and optimal power dispatch strategy [6]. Although the topology might have been well-documented at the initial rollout, changes can go undocumented or may not be communicated immediately to all relevant parties, often due to the ongoing addition of new customers. In addition, it is costly to update records through manual inspection. Consequently, data-driven approaches have emerged as a means to acquire knowledge of the network topology. These methods typically rely on data collected from devices within the network, such as line current sensors [7] or Distribution Phasor Measurement Units (D-PMUs)[8], which are expensive and uncommon measurement equipment. An alternative involves using data collected by smart meters that, installed primarily for billing purposes, are widely deployed across Europe, the U.S.[9], and Australia [10].

Deriving the topology from smart meter data presents significant challenges. A primary issue is the lack of voltage phase information at the customer's end from smart meters, which only provide magnitudes of current and voltage, along with the power factor and a lead/lag indicator. Second, residential distribution networks are primarily three-phase networks with single-phase loads, introducing non-negligible phase coupling through the neutral line[1]. A third challenge arises from the fact that smart meters are not typically installed at substations, leaving the secondary side voltage of low-voltage (LV) transformers unknown. Finally, measurement noise—whether in the data values themselves or in the synchronization of measurement times—poses a significant hurdle. While this can be mitigated by collecting a large number of samples, the typical five-minute sampling period [11] necessitates efficient use of each sample. This work specifically addresses the first three challenges, within the limitations imposed by the last one.

* Corresponding author.
E-mail address: llandrew@unimelb.edu.au (L.L.H. Andrew).

[1] Three-phase customers are less common outside North America since most of the rest of the world uses 220–240 V distribution.

https://doi.org/10.1016/j.ijepes.2024.109949
Received 26 September 2023; Received in revised form 27 January 2024; Accepted 13 March 2024
Available online 30 March 2024

Nomenclature

$1-\phi$	Single-phase
$3-\phi$	Three-phase
N	Total number of leaf nodes
N_a	Number of leaf nodes in phase A
N_b	Number of leaf nodes in phase B
N_c	Number of leaf nodes in phase C
L	Number of samples of smart meter measurements
$\mathfrak{R}$	Real part of
$\mathfrak{I}$	Imaginary part of
$\mathbf{j}$	Imaginary unit
Δ	Difference operator
a	Complex number multiplier, equal to $e^{-\mathbf{j}2\pi/3}$
$V_s \| V_{s,i}$	Complex LV transformer voltage, relative to phase A, or 0° $[V]$
$V \| V_i$	Complex voltage | Complex voltage on node i $[V]$
$\|V\| \Big\| \|V_i\|$	Voltage magnitude | the voltage on node i measured by smart meters $[V]$
$I \| I_i$	Complex current | The current injection on node i $[A]$
$\|I\| \Big\| \|I_i\|$	Current magnitude | the current on node i measured by smart meters $[A]$
pf	Power factor with a lead/lag indicator, measured by smart meters
θ_v	Voltage angle, relative to phase A [Rad]
$\tilde{\theta}_v$	Voltage angle, relative to the phase of the leaf node [Rad]
θ_i	Current angle, relative to phase A [Rad]
$\tilde{\theta}_i$	Current angle, relative to the phase of the leaf node [Rad]
θ_p	Power factor angle, the difference between the voltage and current angle [Rad]
$Z \| Z_{ij}$	Impedance matrix, a symmetrical complex matrix, explained in text
$S \| S_{ij}$	Sensitivity matrix, a symmetrical complex matrix, explained in text
S_I	In-phase component of the sensitivity matrix
S_Q	Quadrature component of the sensitivity matrix
S_S	Same-phase component of the sensitivity matrix, the voltage sensitivity on current of the same phase.
S_O	Other-phase component of the sensitivity matrix, the voltage sensitivity on current of the different phase.
ν	Vectorization operator, converts $B = [b_{ij}] \in \mathbb{C}^{N\times M}$ to $c \in \mathbb{C}^{1\times NM}$, where $c_{(i-1)M+j} = b_{ij}$
$\dagger$	Pseudo-inverse operator
G	Topology graph, a Steiner tree whose root represents the transformer node of the physical network, leaf nodes represent customers, internal nodes represent interconnection buses
$D \| D_{ij}$	Distance matrix, a complex symmetrical matrix | The complex electrical distance between node i and j $[\Omega]$
$\mathcal{A} \| \mathcal{A}_{ij}$	Adjacency matrix, a binary symmetrical matrix | Node i and j are directly connected in G if $\mathcal{A}_{ij} = 1$, they are not directly connected if $\mathcal{A}_{ij} = 0$
$\mathcal{K}$	Set of active nodes, nodes are eligible for decision-making
SNR	Signal-to-noise ratio
DCS	DeltaCon score, a score between 0 and 1 measures the similarity between two topology graphs, a higher score indicates two topology graph are more similar, 1 means they are identical.

1.1. Related literature

Work on topology estimation has focused mostly on single-phase (1-ϕ) networks, as demonstrated by the survey by Deka et al. [12] of over fifty papers, where only nine considered three-phase (3-ϕ) networks. Several of the latter use D-PMUs since they measure complex voltage and current at a high sampling rate with low synchronization error. Hence, they produce the ideal data to estimate the topology. For example, Moffat et al. [13] used D-PMU data to estimate the impedance matrix for metered nodes, then implemented a recursive grouping (RG) algorithm to estimate the network topology. Chauhan and Sodhi [14] developed an Active Distribution Network State Estimation (ADNSE) framework, which uses D-PMUs to detect changes in topology dynamically.

Examples of 1-ϕ networks topology estimation using smart meter data are the work of Peppanen et al. [15], who estimated 5–6 bus subnetworks within a 66-node network. Soumalas et al. [16] used voltage sensitivity to estimate the topology assuming that all unknown line segments between nodes are equal or integer multiples of a specific length. Liao et al. [17] recovered large-scaled network topology aided by partial network knowledge. Park et al. [18] employed recursive grouping to reconstruct the complete topology using exclusively smart meter measurements as information.

According to [12], 3-ϕ methods use D-PMUs exclusively, smart meter data exclusively or a combination of the two. Approaches using D-PMU data only (Adaptive LASSO [19], PaToPa [20] PaToPaEM [21]) achieve accurate results, but D-PMUs are expensive, and not widely deployed. Of the methods partially using smart meter data, Bariya et al. [22] proposed a method to estimate the grid topology using voltage magnitude, but this approach has large error for loads having realistically high correlation. Deka et al. [23] rely on the voltage angle information, which is not typically provided by smart meters, and also require a large sample size (more than 4×10^5) to get accurate results. Gandluru et al. [24] recover the topology of a large-scale 1069-node power system accurately with noisy measurements, but their method requires line meters on some transmission lines for power measurement. Liao et al. [25] proposed a method for finding the network topology directly by applying a maximum a-posteriori probability (MAP) model on the time differences of voltages, without estimating the line parameters. Studies also differ in how they treat the access lines leading from feeders to individual customers. Although these are short, they often have substantial impedances.

Of the studies solely using the smart meter data, Li et al. [26] considered three-phase balanced loads, which reduces the same problem as a single-phase network. Pengwah et al. [27] and Flynn et al. [28] both used a single-phase algorithm; for the simulation studies, they

decoupled the three-phase network by adding ideal ground stakes at each load which decouples the phase, although results for real measurements showed acceptable accuracy despite the coupling. Garcia et al. [29] identifies the phase only, which is a prerequisite of topology estimation. Shi et al. [30] estimated the three-phase line parameters and topology, and also recovered the unknown phase information from the smart meter. Both [26] and [30] modeling the topology as a Spanning tree of customers, which assumes the access lines have no impedance. A more realistic model [31] explicitly models access lines, leading to a "Steiner" tree model, in which nodes are either customers (leaves) or "internal" nodes at which lines join with no customer. We adopt the more accurate Steiner tree approach, also used in [27] and [28]. Our proposal is the first algorithm to study a full three-phase network of single-phase loads, while modeling the access line, which is a more common scenario than [26] and [30] and more accurate than decoupling phases, especially when measurement noise is low.

The innovation in [27] and [28] was to improve the accuracy of sensitivity estimation by imposing physically-based constraints, yielding a quadratic optimization problem. They extend the RG algorithm to find multiple candidate topologies, including known leaf nodes and concealed internal nodes. Finally, several goodness-of-fit measures are combined to choose the best candidate topology. This method has higher accuracy than [13] and [18], which similarly uses the RG algorithm. Moreover, Flynn et al. [28] extended the optimization problem in [27] to estimate the LV transformer voltage along with the voltage sensitivity on the 1-ϕ network and further improved the performance of the RG algorithm. However, these two methods have limitations. Because the sensitivity estimation considers one phase at a time, the coupling between phases dominates the effective signal to noise ratio (SNR). Moreover, during testing, an ideal grounding stake was placed at each node, which masks the coupling through the neutral line. While this simplification is justified by Kron reduction, its accuracy is due to the cancellation of neutral currents by the phases. This does not reduce the sensitivity of the voltage at one node to the current at another, which is the required property for estimating line impedances and hence, topologies.

Other approaches try to exploit other sources of information. Liu et al. [32] discussed the precision limit to the topology estimation problem and accordingly proposed a conservative progressive self-adaptive (CPS) algorithm to estimate the topology. Talkington et al. [33] seek to circumvent the unknown voltage phase angle problem by using a first-order approximation. However, their algorithms require a known scaled version of the sensitivity matrix itself, which may limit its applicability. Another approach is to use information about the geographic locations of the meters [34].

1.2. Contribution of this paper

This paper proposes a novel method to estimate 3-ϕ voltage sensitivity coefficients for a network of single phase loads purely from smart meter measurements. From that, it identifies the topology of a radial 3-ϕ power network consisting of single-phase loads with a single transformer as the root, as is typical of low-voltage distribution networks. The contributions are:

- A method to estimate the voltage sensitivity to current and transformer voltage in low-voltage three-phase distribution network of single phase loads, using smart meter data only. The proposed method models the network as a Steiner tree, which considers the impedance on the access line, explicitly works around the unknown voltage angle and considers the coupling effect between customers on different phases through the neutral line, in Section 2 which reduces the estimation error by four-fold compared to the state to the art in in-phase sensitivity elements, and, unlike existing algorithms, estimates the inter-phase sensitivity elements;
- An enhanced recursive-grouping and backtracking algorithm to recover the network topology with a more targeted backtracking procedure and faster candidate topology selection technique, in Section 3, which outperforms the state-to-the-art in both accuracy and speed when estimating the individual phase topologies. The proposed algorithm is also capable of recovering the full three-phase topologies of single phase nodes, unlike prior work.

The performance of the proposed topology estimation algorithm is studied in Sections 5 and 6 provides the concluding remarks.

2. Sensitivity matrix estimation

The first step for estimating the topology is to find the electrical distances between nodes. This is done by estimating the (complex) sensitivity of the voltage at each meter to the current at each other meter, known as the 3-ϕ voltage sensitivity matrix. The problem is first formulated with complex voltage and current, from which an approximation using only smart meter measurements is derived.

Fundamentally, a 3-ϕ power distribution network can be modeled as a *weighted unrooted tree*, $\hat{G} = (\mathcal{N}_T, \mathcal{E}, D)$, with $\mathcal{N}_T$ being a set of nodes, total number of nodes being N_T, $\mathcal{E}$ being a set of edges, representing power transmission lines, with complex weights $D : \mathcal{E} \to \mathbb{C}^{N_T \times N_T}$ representing the complex electrical distance between pairs of nodes. We write D_{ij} to denote the impedance of the path between two nodes (i, j). An *unweighted unrooted tree*, $G = (\mathcal{N}_T, \mathcal{E})$, represents the network topology. In other words, the topology G do not contain the distance or impedance information. Smart meters measure voltage magnitude, current magnitude, and power factor including whether it is leading or lagging and whether real power is absorbed or injected.[2] Voltage and current are sampled for L time slots, giving $N \times L$ matrices V and I, with smart meters recording $|V|$, $|I|$, pf, and whether $|V|$ leads or lags I, where $|\cdot|$ denotes element-wise absolute value. Complex quantities are sometimes decomposed as real and imaginary parts, denoted by subscripts r and x, or as in-phase and quadrature parts denoted by subscripts $\mathcal{I}$ and $\mathcal{Q}$. The root node of G corresponds to the LV transformer, whose voltage is $V_s \in \mathbb{R}$.

2.1. Voltage sensitivity and transformer voltage estimation

In this study, following [27] and [28], the voltage sensitivity S refers to the changes in a customer node's voltage due to the changes in node currents (rather than powers as is often used). Consider first the 1-ϕ case. Let nodes $1, \dots, N+1$ be metered nodes, and nodes $N+2, \dots N_T$ be the unmetered "nodes" (location where lines join, such as where an access line connects to a feeder). Then in the 3-ϕ network, node 1 is the transformer node, nodes $2, \dots, N_a + 1$ are on phase A, nodes $N_a + 2 \dots, N_a + N_b + 1$ are on phase B and $N_a + N_b + 2 \dots, N + 1$ on phase C. Let $\mathcal{L}_{ii}$ be the (unique) path from the transformer node s to node i, and $\mathcal{L}_{ij}$ denote the section of $\mathcal{L}_i$ that is common to $\mathcal{L}_j$. Let Z_{ii} and Z_{ij} be the impedances of these paths[3]. By Ohm's law,

$$V_s - V_i = \sum_{j=1}^{N_T} Z_{ij} I_j = \sum_{j=1}^{N} Z_{ij} I_j \tag{1}$$

provided there are no nonlinear elements, such as phase-changing transformers, because there is no current injected at unmetered nodes $j > N$. Thus the sensitivity of (minus) V_i to I_j is simply Z_{ij}. The sensitivity matrix S is a submatrix of the standard impedance matrix Z (inverse of the admittance matrix, Y) containing the rows and columns of the Z matrix that correspond to metered nodes (i.e., customers);

[2] Note that the lead/lag and absorb/inject data is vital, and was ignored in [16] and [18], and miscalculated in [27] and [28].

[3] Note that Z_{ij} is not the impedance of a line between node i and j, it is the i, jth component of the inverse of the admittance matrix Y.

that is, it is the Nth leading principal minor. The Z matrix was first used for topology estimation of urban power networks in [17], and further used in [13] and [26] for single-phase distribution network topology estimation. If measurement were available at all nodes, then it would be preferable to estimate Y and exploit its sparsity by methods such as LASSO [17], but that is not possible in this more realistic setting; indeed, the admittance between each pair of measured nodes is 0. Moreover, using Z was reported in [13] to be more robust. To generalize this to three phases, let $a = e^{-\mathbf{j}2\pi/3}$, N_a, N_b and N_c be the numbers of nodes on each phase. The sensitivity of voltage to *power*, rather than current, was first introduced in [35], and used in some of the most recent works, such as sensitivity estimation given an incomplete matrix in [33] and representing localized power system structure as a sensitivity matrix in [34]. However, this is a non-linear relationship, and so we use the S defined above, to improve robustness to voltage sag. Let $E_a = E_{N_a,1}$, $E_b = aE_{N_b,1}$ and $E_c = a^2E_{N_c,1}$, where $E_{n,1} \in \mathbb{R}^{n\times 1}$ is the all-ones matrix. Finally let $E_N = [E_a', E_b', E_c']'$. Then concatenating (1) for all leaf nodes $i = 1, 2, \ldots, N$ gives

$$E_N V_s - V = SI, \tag{2}$$

where $V_s \in \mathbb{R}^{1\times L}$ is the time series of LV transformer voltages as a row vector, and $E_N V_s \in \mathbb{C}^{N\times L}$ is time series replicated to N rows and rotated to the three phases, L is the sample size and $V, I \in \mathbb{C}^{N\times L}$ are the voltage and current at the customer's nodes. As (2) is linear, it also holds for pairwise differences ΔV_s, ΔV, and ΔI:

$$E_N \Delta V_s - \Delta V = S\Delta I. \tag{3}$$

Unlike when using the power sensitivity [35], this is not needed to reduce the linearization error. However, it does improve robustness to uncertainties in V_s. The subsequent equations in this section also apply to pairwise differences.

Let $\theta_v, \theta_i \in R^{N\times L}$ be the matrix of unknown voltage and current angles relative to 0°, and $\tilde{\theta}_v, \tilde{\theta}_i \in R^{N\times L}$ be the angles relative to their corresponding phase. A least-squares estimation of S can easily be obtained if V_s and θ_v are known, they are not. Instead, (2) can be written in terms of the smart meter measurements $|V|$, $|I|$, the phase angle $\theta_p = \theta_i - \theta_v = \tilde{\theta}_i - \tilde{\theta}_v \in [-\pi\ \pi]$ between them, and the unknown voltage angles θ_v and voltages V_s as

$$E_N V_s - |V| \odot e^{\mathbf{j}\theta_v} = S(|I| \odot e^{\mathbf{j}(\theta_v+\theta_p)}), \tag{4}$$

where $\odot$ is element-wise multiplication, $e^{(\cdot)}$ is element-wise exponentiation. Its i, j th element can be derived from two smart meter channels: C_1, whose magnitude is $|I|$ and which is negative when the customer is injecting power, and C_2, whose magnitude is the power factor and which is positive when it is lagging (the voltage leads the current). Note that when the voltage leads the current, angle $\tilde{\theta}_i < \tilde{\theta}_v$, thus when both $C_{1,ij}$ and $C_{2,ij}$ are positive, θ_p is negative and its absolute value equals $\arccos(C_{2,ij})$ by the definition of power factor. Consider the load is still absorbing power ($C_{1,ij} > 0$) but $C_{2,ij} < 0$, then the current leads the voltage and $\theta_p = \arccos(|C_{2,ij}|)$. When the customer is injecting power $C_{1,ij} < 0$, then if $C_{2,ij} > 0$ (the voltage leads the current), θ_p is still $\arccos(C_{2,ij})$ relative to the negative real axis ($-\pi$), but it is instead in the 3rd quadrant. Similarly, if $C_{2,ij} < 0$ (the voltage lags the current), θ_p is in the 2nd quadrant. Therefore, the relationship between θ_p and C_1, C_2 is:

$$\theta_p = \begin{cases} -\arccos(C_{2,ij}), & \text{if } C_{1,ij} > 0 \text{ and } C_{2,ij} > 0 \\ \arccos(|C_{2,ij}|), & \text{if } C_{1,ij} > 0 \text{ and } C_{2,ij} < 0 \\ -\pi + \arccos(C_{2,ij}), & \text{if } C_{1,ij} < 0 \text{ and } C_{2,ij} > 0 \\ \pi - \arccos(|C_{2,ij}|), & \text{if } C_{1,ij} < 0 \text{ and } C_{2,ij} < 0, \end{cases}$$

or more concisely,

$$\theta_p = -\mathrm{sgn}(C_{2,ij})(\pi/2 - \mathrm{sgn}(C_{1,ij})(\pi/2 - \arccos(|C_{2,ij}|))). \tag{5}$$

To estimate the voltage sensitivity matrix S from (4), the LV transformer voltage V_s and voltage angle θ_v are required. Works like [32] and [33] specifically pointed out the impact of lacking the phasor information θ_v. It is straightforward to estimate V_s along with S as it occurs linearly, but a different approach is required for θ_v. We consider two approaches for approximating θ_v and evaluate their impact on S.

Let V_A be a $N_a \times L$ matrix of the voltages of devices on phase A, and V_B and V_C are defined analogously.

1. **Ignore**: this approach approximates θ_v by the nominal angle, which are a^0 (phase A), a^1 (phase B) or a^2 (phase C), (where $a = e^{-\mathbf{j}2\pi/3}$), obtained using phase identification as in [27]. Then, S is approximated as the least-squares solution to (4) subject to the constraints described below. This is the approach implicitly used when sensitivity to real and reactive power is used.
2. **Proj**: using the estimates of θ_v from **Ignore**, solve for S by minimizing the squares of only the in-phase components of (4); that is, quantities are projected onto the nominal phases before taking the least square. This exploits the fact that the component of V in the direction of the nominal phase is not very sensitive to θ_v, which mainly affects the quadrature components. This will be explained in more detail below.
3. **Oracle** uses the actual θ_v instead of estimating it from the measurements, for benchmarking purposes.

The **Proj.** seeks to minimize only $\|V_{s,I} - V_I - (SI)_I\|_F^2$, rather than considering the mismatch $\|V_{s,I} - V_I - (SI)_I\|_F^2 + \|V_{s,Q} - V_Q - (SI)_Q\|_F^2$.

To see why we should ignore the quadrature component, let $\tilde{\theta}$ be the voltage angle relative to its phase and note that the voltage drop on the transmission line is relatively small, whence $|V| \gg |SI|$, and $\tilde{\theta}_v \ll 1$. Thus $\cos\tilde{\theta}_v \approx 1 - \tilde{\theta}_v^2/2$, and $\sin\tilde{\theta}_v \approx \tilde{\theta}_v$. Therefore, $\cos\tilde{\theta}_v$ is less sensitive to error in $\tilde{\theta}_v$, making the right hand side of

$$\begin{aligned} E_N V_{s,I} - V_I - (SI)_I = |V_s| - |V| \odot \cos\tilde{\theta}_v \\ - S_I|I| \odot \cos(\theta_p + \tilde{\theta}_v) + S_Q|I| \odot \sin(\theta_p + \tilde{\theta}_v) \end{aligned} \tag{6}$$

more accurate than that of

$$\begin{aligned} E_N V_{s,Q} - V_Q - (SI)_Q = -|V| \odot \sin\tilde{\theta}_v \\ - S_I|I| \odot \sin(\theta_p + \tilde{\theta}_v) - S_Q|I| \odot \cos(\theta_p + \tilde{\theta}_v). \end{aligned} \tag{7}$$

Both **Ignore** and **Proj** involve solving quadratic programs, constructed as follows. In order to simplify taking the in-phase component later, (4) is rewritten as

$$(E_N V_s - |V| \odot e^{j\theta_v}) \odot R = (S(|I| \odot e^{\mathbf{j}(\theta_v+\theta_p)})) \odot R, \tag{8}$$

where $R_{it} = a^0$ if node i is on phase A, a^2 for phase B and a^1 for phase C ($a = e^{-\mathbf{j}2\pi/3}$); that is, its columns are the element-wise reciprocals of those of E_N. This allows the action of taking the in-phase component to be replaced by taking the real part.

Hence, the approximation $(e^{j\theta_v})_{nt} = 1/R_{nt}$ gives

$$|V| = (E_N V_s - S(|I| \odot e^{\mathbf{j}(\theta_v+\theta_p)})) \odot R, \tag{9}$$

or canonically

$$\nu(|V|) = xH, \tag{10}$$

where ν is the "vectorization" operator that maps $B = [b_{ij}] \in \mathbb{C}^{N\times M}$ to $c \in \mathbb{C}^{1\times NM}$ with $c_{(i-1)M+j} = b_{ij}$, and $x = [\nu(S)\ V_s] \in \mathbb{C}^{1\times(N^2+L)}$ contains the unknowns S_{ij} and $(V_s)_t$, for some $H \in \mathbb{C}^{(N^2+L)\times NL}$, since the right hand side of (9) is linear in x. The next step is to express H explicitly.

We again substitute $1/R_{nt}$ for $(e^{j\theta_v})_{nt}$ into (9), so that the (i,t) th component in $|V|$, or $(i-1)L + t$ th in $\nu(|V|)$, is

$$\begin{aligned} |V|_{it} &= \left((E_N V_s)_{it} - \sum_{k=1}^{N} S_{ik}|I|_{kt} e^{\mathbf{j}(\theta_p)_{kt}} (R_{kt})^{-1}\right) R_{it} \\ &= \left(a_{it} x_{N^2+t} - \sum_{k=1}^{N} x_{(i-1)N+k}|I|_{kt} e^{\mathbf{j}(\theta_p)_{kt}} (R_{kt})^{-1}\right) R_{it}. \end{aligned} \tag{11}$$

Hence the $((i-1)N+k,(i-1)L+t)$th element of H (dependence of $|V|_{it}$ on I_{kt}, $k \in [1,N]$) is

$$-|I|_{kt}e^{j(\theta_p)_{kt}}R_{it}/R_{kt} \tag{12}$$

and the $(N^2+t,(i-1)L+t)$th element (dependence of $|V|_{it}$ on $(V_s)_t$) is $a_{it}R_{it}=1$ ($a_{it}=a^0,a^1,a^2$ respectively for i being phase A,B,C node), and all other elements are 0.

2.2. Optimization problem

A least squares solution of **Ignore** can be obtained by post-multiplying (10) by the pseudoinverse $H^\dagger$ of H. A better solution can be obtained by minimizing the sum of squares of the mismatch subject to physicality constraints outlined below.

A real optimization is obtained by representing every complex number in (10) as

$$u+jv \equiv \begin{bmatrix} u & -v \\ v & u \end{bmatrix}. \tag{13}$$

In this representation, $x \in \mathbb{R}^{2\times(2N^2+L)}$ but only has $2N^2+L$ degrees of freedom, which will cause trouble for the solver. To avoid this, the bottom row of x and $v(|V|)$ due to the $[v,u]$ can be omitted, as it contains the same information as the top row due to $[u,-v]$, yielding x^- and $v^-(|V|)$. Furthermore, the columns of x and rows of H corresponding to the imaginary parts of V_s can be omitted to enforce the constraint that V_s is real, yielding H_1.

The following constraints apply to x^-, for all $i,j \in [1,N]$. First, as all elements of S, including same-phase (refers to the voltage sensitivity to current of the same phase) and other-phase (refers to the voltage sensitivity to current of the different phase) elements, are impedance, the real (resistive) part of S must be non-negative:

$$x^-_{2i-1} \geq 0. \tag{14}$$

Furthermore, as S_{ij} is the impedance on the common path between the transformer s to i and s to j, we have $S_{ij}=S_{ji}$:

$$\begin{aligned} x^-_{(i-1)N+2j-1} &= x^-_{(j-1)N+2i-1} \\ x^-_{(i-1)N+2j} &= x^-_{(j-1)N+2i}. \end{aligned} \tag{15}$$

As S_{ij} is a subpath of S_{ii}, and their resistances are non-negative, hence we have $\Re(S_{ij}) \leq \Re(S_{ii})$. Thus,

$$x^-_{(i-1)N+2j-1} \leq x^-_{(i-1)N+2i-1}. \tag{16}$$

Let k be the most downstream node common to the paths from s to i and from s to j, and

$$D_{ij}=(D_{si}-D_{sk})+(D_{sj}-D_{sk})=S_{ii}+S_{jj}-2S_{ij}, \tag{17}$$

denote the impedance on the path from i to j. Again, its resistive part must be non-zero, yielding

$$x^-_{(2i-1)N+2i-1}+x^-_{(2j-1)N+2j-1}-2x^-_{(2i-1)N+2j-1} \geq 0. \tag{18}$$

Finally, as the LV transformer voltage variation is within $\pm\alpha\%$, for $i \in [N^2+1,N^2+L]$,

$$(1-\alpha)\sqrt{2}V_{tr} \leq x^-_{2i-1} \leq (1+\alpha)\sqrt{2}V_{tr}, \tag{19}$$

Where V_{tr} is the nominal voltage in the secondary side of the transformer, in Australia, it is 230 V. Method **Ignore** then finds S and V_s as the solution to

$$\begin{aligned} \min_{x^-} \quad & \|v^-(|V|)-x^-H_1\|_2^2 \\ s.t. \quad & (14)\text{–}(19). \end{aligned} \tag{20}$$

The next step is to convert this to a problem to solve **Proj**. Minimizing the sum of squares of only the in-phase components corresponds to minimizing the sum of the squares of the real parts of the mismatch in (10). This is achieved by simply omitting every second column of $v^-(|V|)$, which contain the imaginary parts, and correspondingly every second column of H_1, yielding $v^=(|V|)$ and H_2. Then **Proj.** finds them as the solution to

$$\begin{aligned} \min_{x^-} \quad & \|v^=(|V|)-x^-H_2\|_2^2 \\ s.t. \quad & (14)\text{–}(19). \end{aligned} \tag{21}$$

The numerical results will show that this modification reduces the error in the estimated sensitivity matrix by around 22-fold. Note that this method can also find the sensitivity for non-radial networks. This is useful for applications such as operating envelope calculations. However, for such networks there is no simple relationship between the sensitivities and the individual line impedances.

3. Topology estimation by recursive grouping

From the sensitivity matrix of the leaf nodes, it is possible to estimate the locations of the (unmetered) internal nodes in the network, and thus find the complete topology. Here, we present a variant of the recursive grouping algorithm first introduced in [36], first applied in topology estimation in [13], and also used in [18,26,27], and [28]. However, unlike [27] and [28], who used only the real part of sensitivities, and [13], who used complex values but estimated the imaginary part poorly by approximating the phase angle by the nominal angle, i.e., applying the **Ignore** approach, we are able to use the full complex value. We also improve the backtracking decisions to select more likely candidates, and propose new features that allow the optimal candidate to be chosen more quickly.

Recursive grouping treats each customer (leaf) node as a degenerate tree. Then, it repeatedly joins pairs of trees as either siblings of a new node or with one as a child of the root of the other. To explain the basic recursive-grouping process, we first introduce an active node set $\mathcal{K}$, which contains the root of each sub-tree formed so far, initially the leaf nodes. For each node pair (i,j) in $\mathcal{K}$, there are three possible relationships – parent–child, siblings, and "others" – and we want to make decisions of joining pairs of the first two types. Parent–child and sibling pairs (i,j) can be distinguished from "other" pairs because their relative distances, $D_{ik}-D_{jk}$, to all other nodes $k \in \mathcal{K}$ is constant, as will be explained in Section 3.1. Fig. 1 shows recursive grouping of a 5-node network. To begin with, as shown in (a), the set $\mathcal{K}$ initially contains the LV transformer node (purple, labeled 1) and all leaf nodes (green, labeled 2–6). Pale nodes show nodes (other than the transformer) not in $\mathcal{K}$. Steps (b), (c) and (d) show nodes being joined as siblings, step (e) shows node 8 being joined as a child of node 9, and step (f) shows the special case of the transformer joining at the end. Note that after each decision step, the number of elements in $\mathcal{K}$ is decrease by 1.

The proposed recursive-grouping (RG) algorithm has two extra steps. The four steps are: Pair Selection, Node Join, Backtracking, and Topology Selection. To explain the last two, we introduce the concept of a decision tree (or single-player game tree), which is distinct from the topology tree discussed above. In a decision tree, the root represents the initial state of the RG process (Fig. 1(a)). Each internal node represents a decision to be made by the algorithm, for example, the state in Fig. 1(b) is a child of the root in the decision tree. If we join node pair (3,4) instead of (2,3) in Fig. 1(b), this is another child of the root and another internal node in the decision tree. Each edge in the decision tree represents a decision to merge two subtrees and how they are merged (siblings or parent/child). Finally, the leaf nodes in the decision tree are the final solutions of the RG, which in our case are candidate topologies. The state of Fig. 1(f) is one of the leaf node in the decision tree. The RG algorithm of Fig. 1 only considers a single path to a single leaf, but the backtracking introduced in [27] and used in this paper considers multiple children of selected nodes to discover multiple leaves. The final topology selection step selects one of the leaves of this decision tree.

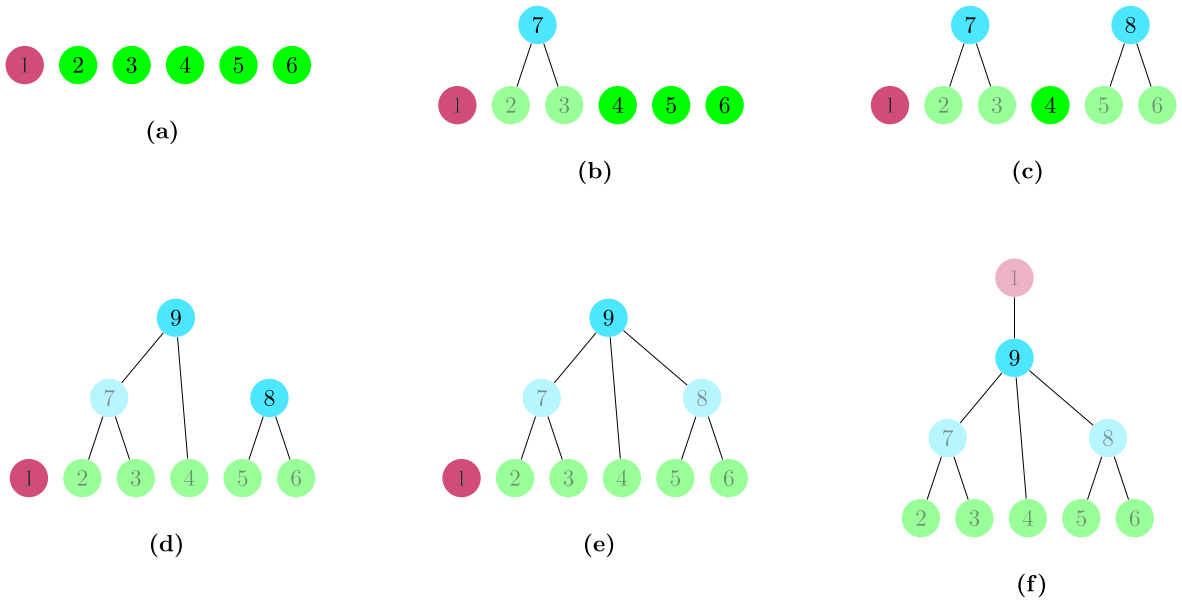


Fig. 1. Example of Recursive-Grouping. Step (e) is a parent–child join; the others are sibling joins.

3.1. Pair selection

For each pair $(i, j) \in \mathcal{K}^2$, let

$$\Phi_{ijk} = D_{ik} - D_{jk}, \quad k \in \mathcal{K} \setminus (i, j). \tag{22}$$

Then Φ_{ijk} is constant for all k if and only if i and j are either siblings or parent and child [36]. For sibling pair (i, j), $\Phi_{ijk} = \pm(D_{ih} - D_{jh})$ where k is the parent of i and j, whereas $\Phi_{ijk} = \pm D_{ij}$ if (i, j) are parent and child; the latter can be considered as h coinciding with either i or j, leading to either D_{ih} or D_{jh} being zero. However, due to estimation error in the sensitivity matrix, the value of ϕ_{ijk} varies slightly with k even if i, j are siblings or parent–child. Let

$$E_{\mathrm{s},ij} = \max_{k \in \mathcal{K}\setminus\{i,j\}} |\Phi_{ijk}| - \min_{k \in \mathcal{K}\setminus\{i,j\}} |\Phi_{ijk}|, \tag{23}$$

be an estimate of the measurement noise that is reflected on ϕ values, and

$$E_{\mathrm{p},ij} = |D_{ij}| - \frac{1}{|\mathcal{K}|} \sum_{k \in \mathcal{K}} |\Phi_{ijk}|, \quad k \in \mathcal{K} \setminus \{i, j\} \tag{24}$$

measure the mismatch to the parent/child condition.

These are used both to select which pair to join, and whether to join them as siblings or as parent and child, as follows.

Let $\mathcal{C} = (\{\mathrm{s}, \mathrm{p}\} \times \mathcal{K}^2) \setminus \mathcal{I}$, where $\mathcal{I}$ is a set of "invalid" combinations in which a leaf is a parent, or the LV transformer node is a child. We calculate E_{s} and E_{p} for all decisions in $\mathcal{C}$, and find the smallest value $E_{\min}$ amongst all E_{s} and E_{p}. If $E_{\min}$ corresponds to (23), our first guess is that this pair are siblings, otherwise, we guess parent–child.

However, due to errors in ϕ_{ijk}, this may yield an incorrect decision. To combat this, we record possible decisions for which the relative error,

$$\tilde{E}_{tij} = E_{tij} \Big/ \min_{(t',i',j') \in \mathcal{C}} E_{t'i'j'} \tag{25}$$

for $t \in \{\mathrm{s}, \mathrm{p}\}$, is less than a threshold T_c, and the algorithm later back-tracks to resume at this point to explore these other decisions. This differs from [27] and [28] which backtrack over at most one parent–child pair and at most one sibling pair.

Alternative decisions are only recorded if they are inconsistent with the first guess; that is, if they share at least one node to be joined. By contrast, [27] reserves every pair-relationship if their $\tilde{E} \le T_c$, which introduces unnecessary backtracking. Note that the depth of the decision tree is $N - 1$, regardless of the decisions.

3.2. Node join

After each pair is decided to join in the decision-making step, the distance matrix D and graph G is updated accordingly depending on the relationship of the pair. We first introduce a 0/1 symmetric adjacency matrix $\mathcal{A}$. A "1" in $\mathcal{A}$ indicates that the corresponding pair of nodes are directly connected in graph G, and a "0" means two nodes are not directly connected. The graph G can be uniquely determined by $\mathcal{A}$. This part of the algorithm connects the identified pair (i, j) into the network and update the adjacency matrix $\mathcal{A}$ and distance matrix D.

If a pair (i, j) is identified as siblings, a common parent node h is created and added to the active node set $\mathcal{K}$, then pair (h, i) and (h, j) are connected and nodes i and j are removed from the active node set $\mathcal{K}$. Consequently, we update the adjacency matrix $\mathcal{A}$ by creating a zero row and a zero column representing h and setting $\mathcal{A}_{hi} = \mathcal{A}_{ih} = \mathcal{A}_{hj} = \mathcal{A}_{jh} = 1$. For the distance matrix D, we first create a new row and a new column in D with distances

$$D_{ih} = D_{hi} = \frac{1}{2}(D_{ij} + \frac{1}{|\mathcal{K}|} \sum_{k \in \mathcal{K}} \Phi_{ijk}), \tag{26}$$

$$D_{jh} = D_{hj} = D_{ij} - D_{ih}, \tag{27}$$

$$D_{hk} = D_{kh} = \frac{1}{2}(D_{ik} - D_{ih} + D_{jk} - D_{jh}) \quad k \in \mathcal{K} \setminus \{i, j\} \tag{28}$$

where $\mathcal{N} \subset \mathcal{N}_T$ is the set containing all discovered node of the current status of graph G.

If a pair (i, j) is identified as parent–child, i is the child if the average over k of Φ_{ijk} is negative, and the parent otherwise. We connect (i, j) and set $\mathcal{A}_{ij} = \mathcal{A}_{ji} = 1$. Then we remove the child node from the active node set $\mathcal{K}$. As no new node is discovered, there is no need for updating D.

3.3. Backtracking

The original "greedy" algorithm of [36] will traverse a single path of the decision tree from the root to a leaf. Instead, we follow [27] and explore the tree more thoroughly to find multiple candidates, and then select among these by assessing how plausible each is. Once the first candidate topology is reconstructed, we first find the last decision node that has reserved candidate decision(s), then follow the candidate with the lowest relative score except for those that have already been followed. We join the corresponding pair and continue the RG algorithm to create a new candidate topology.

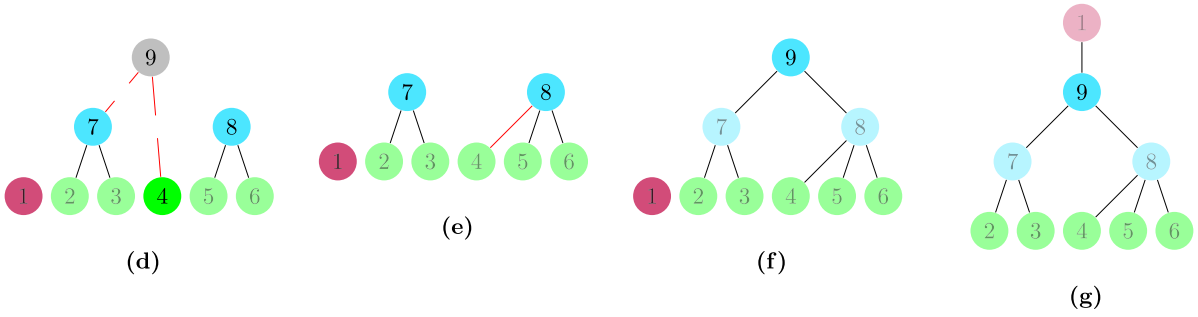


Fig. 2. Example of Backtracking, constructing an alternative topology of (d) in Fig. 1.

Fig. 2 shows the backtracking process. In this example, the decision of joining pair (4,7) as siblings in Fig. 1(d) has a reserved candidate decision of joining 4 as a child of 8. The initial decision is revoked by disconnecting edges between 4,9 and 7,9, removing 9 from the active node set $\mathcal{K}$ and returning its children 4,7 to $\mathcal{K}$, as shown in Fig. 2(d). Then according to the reserved candidate decision, internal node 8 is a parent of leaf node 4, hence we connect (4,8) and remove leaf node 4 from $\mathcal{K}$, as shown in Fig. 2(e). We keep making decisions until we obtain a new candidate topology shown in Fig. 2(g), which is different from the candidate topology in Fig. 1(f). If sufficient reserve candidates are stored, this could be continued until all possible trees had been evaluated. However, this would be too computationally expensive, and so we further set a maximum number of candidate topology N_c. Once N_c candidate topologies have been considered, the backtracking stops despite there being unvisited decision tree nodes. The full algorithm is presented in Algorithm 1. Notation $\Leftarrow$ indicates that the right-hand side is appended to the list on the left-hand side, and $\bar{\mathrm{s}} = \mathrm{p}$, $\bar{\mathrm{p}} = \mathrm{s}$ swaps between sibling and parent/child.

3.4. Topology selection

After obtaining the candidate topology library from the backtracking algorithm, we need an additional step to choose the best topology to be the final estimate, or to rank the candidates. This is done by defining several figures of merit for each candidate C_n that measure different aspects of the suitability of the estimate, which are then combined by a regression procedure to form a final score.

Since the RG algorithm seeks (suboptimally) to minimize $\hat{E}$ given in (25), the first figure of merit is the mean of accumulated squared error of E_s and E_p of all decisions:

$$F_\mathrm{s}(C_n) = \frac{1}{N-1}\sum_{i=1}^{N-1}(E_{\{\mathrm{s,p}\},n})^2\,, \tag{29}$$

where $E_{\{\mathrm{s,p}\}}$ includes E_s and E_p depends on the corresponding pair-relationship; N, the total number of leaf nodes, is also the number of decisions to complete the topology. If this were the only figure of merit, then backtracking and selecting the topology to minimize $F_\mathrm{s}(\cdot)$ would work together as a global optimization procedure for $F_\mathrm{s}(\cdot)$. However, there is an opportunity to use additional engineering insights that are not easy to incorporate into the objective $\hat{E}$ of the grouping procedure, because they cannot be calculated until the whole candidate topology is known.

The first ancillary figure of merit reflects the empirical observation that, when the topology is mis-estimated, it is common for some line lengths to be very large or very small. This causes the selection of the wrong topology to be positively correlated with having a large variance of line lengths. This suggests the figure of merit

$$F_\mathrm{RI}(C_n) = \sum_{i,j\in B_n}\left(\Re S_{ij} - \frac{1}{n}\sum_{k,l\in B_n}\Re S_{kl}\right)^2\,, \tag{30}$$

where B_n be the set of internal edges of candidate network C_n.

Finally, due to constraints on geometry and material, the ratio of reactance $\Im S_{ij}$ to resistance $\Re S_{ij}$ (X/R ratio) is unlikely to vary wildly between different lines, and unreasonably large variance in this ratio can again indicate incorrect topology estimates. This suggests, the second ancillary feature used is the variance of the X/R ratio of all edges in each candidate network:

$$F_\mathrm{XR}(C_n) = \mathrm{Var}\left[\frac{\Im S_{ij}}{\Re S_{ij}}\right], (i,j)\in\mathcal{T}\,, \tag{31}$$

where Var refers to the sample variance. Note that we are not claiming that either F_{RI} or F_{XR} should be zero for a real network, in contrast to F_s which should be. It is merely that they have been observed to be large for many of the mis-estimated topologies.

It remains to combine these features (figures of merit) into a single score. The simplest method is to take a linear combination of the features, which turns out to be adequate. Optimal coefficients are found by linear regression, even though it is not expected that there is any underlying linear relationship. Logistic regression was also tested, but found not to improve performance. The topology is then selected as

$$\arg\min_{C_n}\ \beta_\mathrm{s}F_\mathrm{s}(C_n) + \beta_\mathrm{RI}F_\mathrm{RI}(C_n) + \beta_\mathrm{XR}F_\mathrm{XR}(C_n), \tag{32}$$

where the β are the regression coefficients. The accuracy and computational speed comparison between the proposed topology selection method and existing methods are discussed in Section 5.1.2.

3.5. Non-radial networks

The proposed method is only applicable in radial distribution networks, which the vast majority of LV networks are. However, in High-Voltage (HV) and Medium-Voltage (MV) networks, meshed networks are common. In such networks, the recursive-grouping algorithm cannot be applied. Moreover, any non-radial network measured only at leaf nodes can be transformed to an "equivalent circuit" (in terms of steady-state V and I at the leaves) that is radial, which means that the discrete measurements of the voltage and current at the customer premises are not enough to determine a non-radial topology, even if we know both their magnitudes and phase angles. Deka et al. [37] estimated the topology of meshed networks under the condition that there is no cycle of unmetered nodes. However in our setting, due to all measurements being at leaf nodes, determining the topology is not possible. This can be viewed as a disadvantage of modelling the access line, but there are some Steiner points in the LV network even if we ignore the access line, which still causes ambiguity.

4. Evaluation of sensitivity estimation

The numerical evaluation of the proposed techniques will be in two parts. This section considers only the accuracy of evaluating the sensitivity matrix, and the following section considers the accuracy of the algorithm to estimate the topology from the sensitivity.

Algorithm 1: Topology Estimation Algorithm

Input:
Sensitivity matrix S
Number of leaf nodes N
Cut-off threshold T_c
Maximum number of candidates N_c
Output:
Library of candidate topologies C
Adjacency and distance matrices $\mathcal{A}_f$, $\mathcal{D}_f$
Initialization:
An array P indexed by pairs in the decision tree storing priority queues,
An empty decision index n,
An empty adjacency matrix $\mathcal{A}$,
An empty active node set $\mathcal{K}$,
A discovered node set $\mathcal{N} \leftarrow [1, 2, \ldots, N+1]$.
Calculate D from S using (17)
$\mathcal{K} \leftarrow \mathcal{N}$,$\mathcal{N}_o \leftarrow \mathcal{N}$
while $|C| \le N_c$ **do**
  if $|\mathcal{K}| > 2$ **then**
    Calculate $\tilde{E}$ using (25)
    $n \leftarrow |\mathcal{N}_o| - |\mathcal{K}| + 1$ (* depth in decision tree *)
    $(t, i, j) = \arg\min_{(t,i,j)\in\{\mathrm{s},\mathrm{p}\}\times\mathcal{K}^2} \tilde{E}_{t,ij}$
    $P_n \Leftarrow (\tilde{E}_{t,ij}, t, i, j)$
    if $\tilde{E}_{\bar{t},ij} \le T_c$ **then**
      $P_n \Leftarrow (\tilde{E}_{\bar{t},ij}, \bar{t}, i, j)$
    end
    if $t = \mathrm{s}$ **then**
      for $(i', j') \in \{i, j\} \times \mathcal{K} \setminus \{i, j\}$ *s.t.* $\tilde{E}_{\mathrm{s},i'j'} \le T_c$ **do**
        $P_n \Leftarrow (\tilde{E}_{\mathrm{s},i'j'}, \mathrm{s}, i', j')$
      end
    end
    $(\mathcal{A}, D, \mathcal{K}, \mathcal{N}) \Leftarrow \mathrm{NodeJoin}((t, \{i, j\}), \mathcal{A}, D, \mathcal{K}, \mathcal{N})$
  else
    $ID \leftarrow (\mathrm{p}, \mathcal{K})$ (* Last two parent/child *)
    $C \Leftarrow \mathrm{NodeJoin}(ID, \mathcal{A}, D, \mathcal{K}, \mathcal{N})$
    if *any* $|P| \ge 2$ *(* backtrack *)* **then**
      Find last decision point m where $|P_m| \ge 2$
      $\mathcal{K}_r \leftarrow \mathcal{K}$ at m
      Remove the current lowest $\tilde{E}$ from P_m
      Empty P_n for all $n > m$
      (* Build tree with new lowest $\tilde{E}$ in P_m *)
      $\mathcal{K} \leftarrow \mathcal{N}_o$
      for $n = 1$ *to* m **do**
        $ID \leftarrow$ lowest-$\tilde{E}$ decision in P_n
        $(\mathcal{A}, D, \mathcal{K}, \mathcal{N}) \Leftarrow \mathrm{NodeJoin}(ID, \mathcal{A}, D, \mathcal{K}, \mathcal{N})$
      end
    else
      return C
    end
  end
end
return C

4.1. Voltage sensitivity and LV transformer voltage estimation

In this study, all experiments are conducted in MATLAB, and the optimization problems are solved using YALMIP with the MOSEK solver. This experiment uses the IEEE PES 55-node European LV feeder [38] (EU feeder), and other synthetic networks with 21, 30, 42, 55 customers and fifty different networks per size (200 synthetic networks in total). Of the synthetic networks, the physical parameters of the transmission line such as reactance to resistance (X/R) ratio, average length, and impedance per unit length are chosen in the range of those in the EU feeder. Note that in this experiment, the network is a residential 3-ϕ network, which consists of 1-ϕ loads from all three different phases. The network does not include any 3-phase loads. This design is informed by the fact that in regions such as Europe, Australia, New Zealand, China, and much of other Asian countries, the standard nominal voltage of V_{tr} = 220-240 V is adequate for the majority of residential applications.

First, we implement the **Proj**. estimation in (21) to evaluate the accuracy of the estimated voltage sensitivity coefficients and the LV transformer voltage. The active power consists of two components. The first is representing the power consumption of customers, noted as P^L, given for $i \in [1, N], t \in [1, L]$ by

$$P^L_{i,t} = \lambda \hat{P}^L_{i,t-1} + (1-\lambda)\hat{P}^L_{i,t}, \tag{33}$$

where λ is the auto-correlation, $\hat{P}^L_{i,t}$ are i.i.d.úniform on $[P_a, 9P_a]$ for an adjustable power P_a. The second component noted as P^S, represents solar photovoltaic (PV) installed on 20% of nodes selected at random. Note that the PV system generates primarily active power, and does not inherently generate reactive power. Consequently, no reactive component is considered in PV generation.[4]

$$P^S_{i,t} = \begin{cases} -\max(0,\ 30P_a \sin(2\pi t/T)) & \text{if } i \in \mathcal{P}; \\ 0 & \text{otherwise} \end{cases}, \tag{34}$$

where T is the number of samples per day. Smart meters typically measure once per five minutes, or $T = 288$ times per day. Set $\mathcal{P}$ is the subset of customers that have solar PV.

For the customer's consumption, the power factor pf^L are randomly generated within the range of $[0.9, 1] \wedge [-1, -0.9]$. Lower power factors can occur in the measured power when there is cancellation of the active consumption by PV generation. The corresponding reactive power consumption is:

$$Q^L_{i,t} = \mathrm{sign}(pf^L_{i,t})\sqrt{1-(pf^L)^2_{i,t}}\,P^L_{i,t}, \tag{35}$$

which leads to the total active power $P = P^L + P^S$, total reactive power $Q = Q^L$. Note that the active power factor pf measured by smart meters differ from pf^L, which is

$$pf_{i,t} = \begin{cases} \arctan(|Q_{i,t}/P_{i,t}|) & \text{if } P_{i,t} > 0, Q_{i,t} > 0; \\ -\arctan(|Q_{i,t}/P_{i,t}|) & \text{if } P_{i,t} > 0, Q_{i,t} < 0; \\ \pi - \arctan(|Q_{i,t}/P_{i,t}|) & \text{if } P_{i,t} < 0, Q_{i,t} > 0; \\ -\pi + \arctan(|Q_{i,t}/P_{i,t}|) & \text{if } P_{i,t} < 0, Q_{i,t} < 0. \end{cases} \tag{36}$$

Lastly, the average transformer voltage varies during the day, with a minimum during peak solar generation, between the morning and evening consumption peaks. It is modeled as a perturbation of a sinusoidal voltage with mean $\sqrt{2}E_n$ volts, amplitude $0.1\sqrt{2}E_n$ volts, and period one-day($T = 288$ samples). The perturbation is i.i.d. uniformly distributed multiplicative noise $r_t \in [0.9\ 1.1]$ to simulate the sag caused by disturbance, faults, etc., giving

$$V_{s,t} = \sqrt{2}(1 + 0.1\sin(2\pi t/T))r_t E_n. \tag{37}$$

In this study, we use the relative average absolute error (abbreviated to relative error) to quantify the accuracy of the voltage sensitivity coefficients. Let S be the true sensitivity, and $\hat{S}$ be the estimated value, then the relative error between S and $\hat{S}$ be

$$E_r = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{S_{ij} - \hat{S}_{ij}}{S_{ij}}\right| = \mathbf{M}(|S - \hat{S}| \oslash |S|), \tag{38}$$

where $\oslash$ denotes element-wise division, $|\cdot|$ denotes the element-wise absolute value, and $\mathbf{M}(\cdot)$ denotes the average value of elements in a matrix.

Fig. 3 plots the relative average error of the absolute value of the sensitivity coefficients $|S|$ and the LV transformer voltage V_s against the *SNR* on current for different network size. The SNR of current

[4] This assumption is not crucial, as the algorithm itself does not distinguish between power from PV or the load; we only separate them to get a suitable joint distribution of the total real and reactive powers.

Table 1
Comparison of the $|S|$ and V_s estimation among [27], [28] and the proposed method in different sized network and SNR. ALG 1: modified [28], prop.: the proposed method.

Size	SNR	$\|S\|$ % error			V_s % error	
		[27]	[28]	prop.	[28]	prop.
30	25	16.1	5.9	3.8	2.30	0.32
	30	16.0	5.6	2.6	2.25	0.18
	35	16.0	5.2	2.0	2.23	0.12
	40	16.0	5.2	1.7	2.23	0.09
42	25	16.0	6.6	4.0	2.67	0.30
	30	16.0	6.4	2.7	2.62	0.21
	35	15.9	6.1	2.0	2.62	0.16
	40	15.9	6.1	1.8	2.62	0.14
55	25	17.7	10.8	4.2	3.50	0.44
	30	17.6	10.5	2.8	3.48	0.38
	35	17.5	9.9	2.5	3.48	0.36
	40	17.5	9.7	2.3	3.47	0.35

varies from 10 dB to 40 dB, and the SNR on voltage is always 6 dB higher. This is in contrast to [13,18,27,28] which had no noise on the voltage. The adjustable power P_a is 400 W, the number of samples used in the simulation is 40,000, and the auto-correlation coefficient $\lambda = 0.8$. Note that the actual average active power consumption is $5P_a$. The results are averaged over all 50 networks in each size. Fig. 3(a) shows a distinct downward trend of relative estimation error in both same-phase and other-phase coefficients of $|S|$ as SNR increases. The other-phase coefficients exhibit approximately double the relative error of the same-phase coefficients. This disparity arises because the coupling of voltage on current across different phases is smaller by a approximate factor of four compare to the current of the same phase. A factor of 1/2 is due to nodes in different phases are solely coupled via the neutral wire, compared with the two-wire coupling between the same-phase nodes. Another factor of 1/2 is from the different phase elements are out of phase by $\|\cos 2\pi/3\|$. Note however that the other-phase coupling remains significant, showing that for applications such as dynamic operating envelope calculations, the other-phase interactions should not be ignored as is current practice. Within the conventional range of smart meter measurement noise ($SNR_I = 40$ dB, $SNR_V = 46$ dB), the relative errors for same-phase and other-phase are around 1.5%–3% and 3-5% respectively. As seen in Fig. 3(b), the estimation of V_s tends to be more precise in larger networks at lower SNR values, but this pattern reverses as SNR increases. In smaller networks, the estimation accuracy is more susceptible to individual large measurement errors. Contrarily, larger systems exhibit an improved capacity to average out these errors, thereby enhancing overall accuracy. Fig. 4 shows the same y axis as in Fig. 3 against the adjustable power P_a, regarding different network size. Here, P_a ranges from 100 W to 800 W (actual average active power from 500 W to 4 kW), while SNR_I and SNR_V are kept constant at 40/46 dB, respectively. This experiment also uses a sample size of 40,000 and an auto-correlation coefficient (λ) of 0.8.

As shown in Fig. 4, the relative errors in the estimated $|S|$ is initially decrease then increase as progressively heavier loads are applied. The reason is that the measurement noise on $V_s - |V|$ is roughly a constant, and $V_s - |V|$ is proportional to the load. Consequently, as the load increases, the relative noise decreases, leading to smaller errors in voltage sensitivity S. However, as I further increases beyond a certain threshold, the discrepancy between the complex voltage and its same-phase component starts to become dominant (approach **Proj**. only uses the same-phase component), causing the error in sensitivity to increase. For the V_s estimation, the primary determinant of accuracy is the error in $|V|$, which grows with load power, making V_s error strictly rise with P_a. This underscores the need for accurate θ_v estimation in future research for improved sensitivity in high-load situations. Fig. 5 shows the impact of the number of time samples, varying from 100 to 40,000. Other parameters are: P_a=400 W (average load power XYZ); SNR_I=40 dB; SNR_V=46 dB; and λ=0.8. As seen in Fig. 5(a), at sample size = 10,000 of more, the relative error on same-phase element of S is typically less than 5%, and different-phase error is less than 10%.

Table 1 presents a comparison of the relative error in estimated $|S|$ and V_s between the proposed method and prior work. Note that [26] and [30] is not included as it finds a spanning tree for topology representation, contrasting our use of a Steiner tree. Moreover, their approach does not assess accuracy against the topology's ground truth, presenting a challenge for a comprehensive performance comparison. We applied the 1-ϕ sensitivity estimation method in [27,28] on each separate phase of 3-ϕ network. We then took the weighted average of the relative error to compare with the proposed 3-ϕ method. Note that only same-phase estimation results are compared as the 1-ϕ methods cannot estimate the other-phase sensitivity coefficients. The proposed method outperforms [27] by a factor of 5 to 10 in the accuracy of $|S|$. In [27], which approximates the LV transformer voltage as constant, the error in estimating S_x is almost 100%, while the error in the estimate of S_r is around 3 times larger than that of the proposed method. In [28], the LV transformer voltage was estimated using a regularization term $\alpha(|V_{s,t+1} - V_{s,t}|)$ in the objective to prevent excess fluctuations in estimates of V_s. We choose $\alpha = 0.01$. Although that method accurately estimates both $|S|$ and V_s in a single-phase network, applying it to one phase of a 3-ϕ network yields errors 2 to 4 times as large in $|S|$, and 8 to 20 times as large in V_s, as shown in Table 1.

4.2. Sensitivity estimation error source discussion

There are four sources of error when estimating the voltage sensitivity: applying a 1-ϕ method on a 3-ϕ network; approximations made about the unknown voltage phase angle θ_v; the unknown LV transformer voltage; and smart meter measurement noise. In this experiment, the error in $|S|$ caused by each of these sources is evaluated.

The tested network is the European 55-node feeder and fifty 55-node synthetic networks. Note that $|S|$ consists of two parts: the block-diagonal (same-phase) elements coupled by the neutral and line conductors; and the off-block-diagonal (cross-phase) elements only coupled by neutral conductors. The 1-ϕ method estimates the block-diagonal elements only.

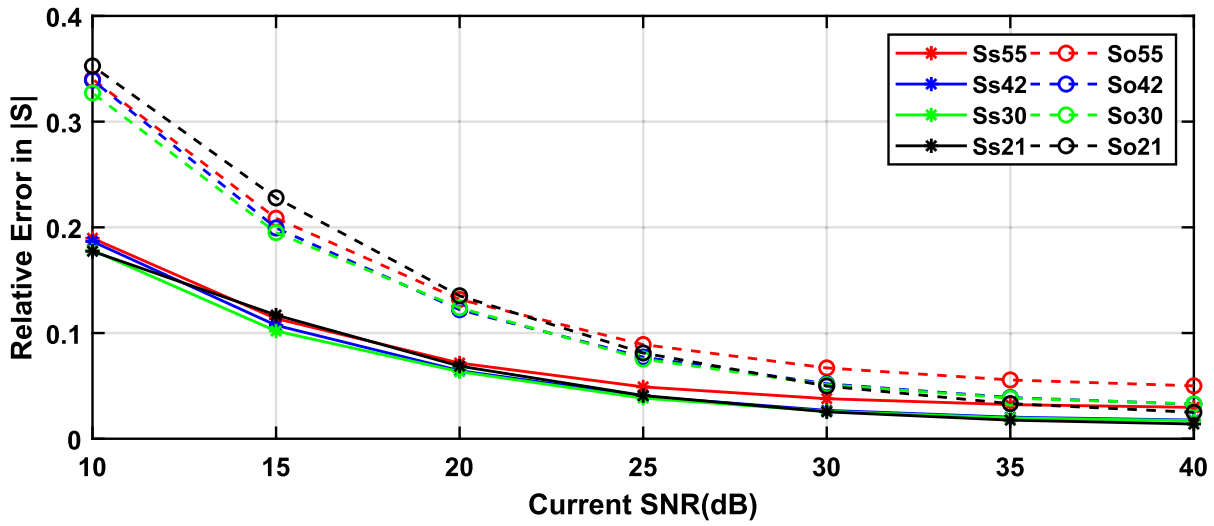
In the control group, the voltage sensitivity is estimated using the 3-ϕ method with known θ_v (**Oracle**) and V_s, and typical measurement noise ($SNR_I = 40$ dB, $SNR_V = 46$ dB). We apply a range of pertubations from the conditions in control group to repeat the $|S|$ estimation, as shown in the first column of 2. Here "Estimated V_s" indicates that the LV transformer voltage V_s is estimated via the proposed optimization method alongside with S, "Constant V_s" is when V_s is set to 230 V. The 1-ϕ estimation stands for applying the proposed sensitivity estimation method using the single-phase measurements only. Both methods [27,28] use 1-ϕ estimation, considering current noise only. They also project the complex voltage onto the real axis, a procedure akin to **Proj**.. The transformer voltage is estimated in [28], and treated as a known constant in [27].[5]

We evaluate the "ratio of relative error in $|S|$" after each change, and the errors in the diagonal and off-diagonal relative to those of the control group are shown separately in the second and third columns of Table 2. This leads to the following conclusions:

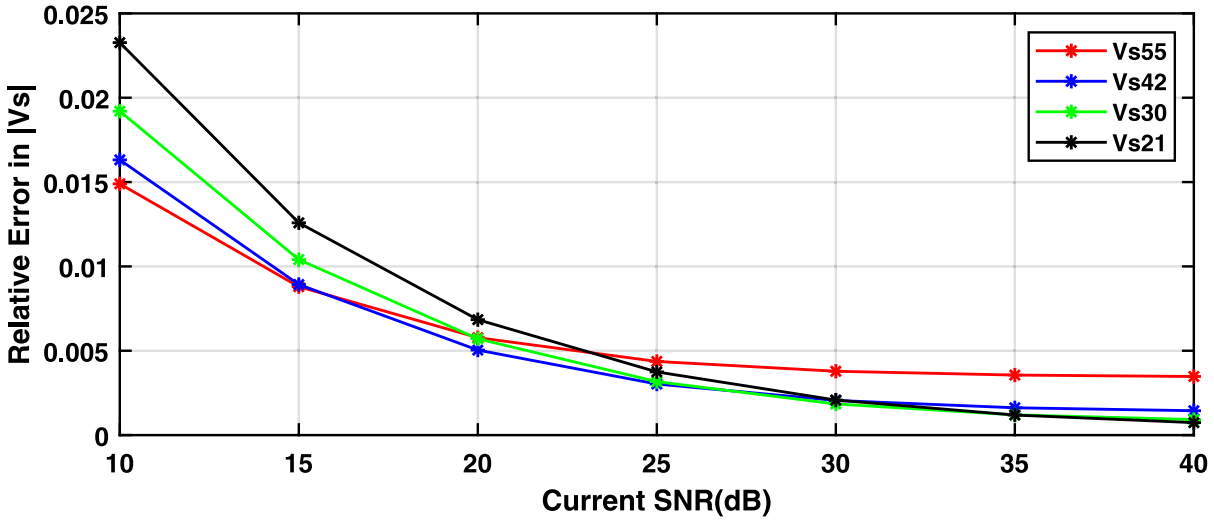
Table 2 demonstrates that applying the 1-ϕ method to each phase of a 3-ϕ network, as done in [17,18,27,28], ignoring the coupling effect between phases via the shared neutral line, leads to a threefold error in the same-phase component of $|S|$, and fails to estimate the other-phase component entirely.

Of the ways of handling unknown θ_v, ignoring both the phase shift and the quadrature component yields much less error than ignoring the phase shift but considering the full complex form of (4). As seen

[5] [28] described their method as complex equation, similar to **Ignore**. However, they actually used **Proj.** in implementation.



(a) $|S|$



(b) V_s

Fig. 3. Average relative error of (a) the estimated sensitivity matrix and (b) transformer voltage against current SNR, with voltage SNR 6 dB higher, for adjustable power $P_a = 400$ W, auto-correlation coefficient $\lambda = 0.8$. In the legend, S_S refers to same-phase sensitivities; S_O to other-phase sensitivities, V_s to transformer voltage, followed by the number of customers.

Table 2
Comparison of how major factors affect the average estimation error in diagonal and off-diagonal sensitivity elements. "Ratio" refers to the ratio of the average error in $|S|$ for the condition after the → to the average error for the condition before the →.

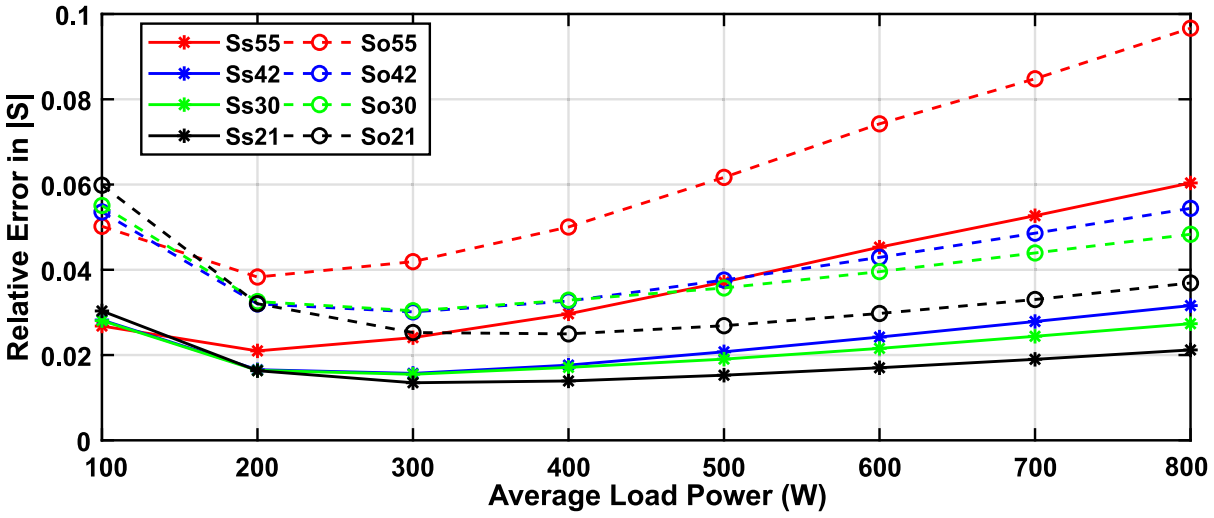
Changes to the Control Group	Diag ratio	OffDiag ratio
3-ϕ estimation → 1-ϕ estimation	3.18	N.A.
Oracle $\theta_v \rightarrow \theta_v$ approach "**Ignore**"	27.6	26.5
Oracle $\theta_v \rightarrow \theta_v$ approach "**Proj.**"	1.23	1.17
Known V_s → Estimated V_s	1.27	1.42
Known V_s → Constant V_s	8.20	8.41
$SNR_I = 40 \rightarrow SNR_I = 10$	2.01	1.69
$SNR_V = 46 \rightarrow SNR_V = 16$	10.7	11.0
Control group → Proposed method	1.70	1.76
Control group → [27]	12.69	N.A.
Control group → [28]	7.02	N.A.
Proposed method → [27]	7.46	N.A.
Proposed method → [28]	4.13	N.A.

in Table 2, the former only has an error in $|S|$ around 1.3 times the control group, whereas the latter has error more than 25 times as large. As the $\tilde{\theta}_v$ is typically smaller than 0.5°, it fits the scenario described in the paragraph before (6), which explains why adding the quadrature component to the optimization problem is detrimental to the accuracy of $|S|$.

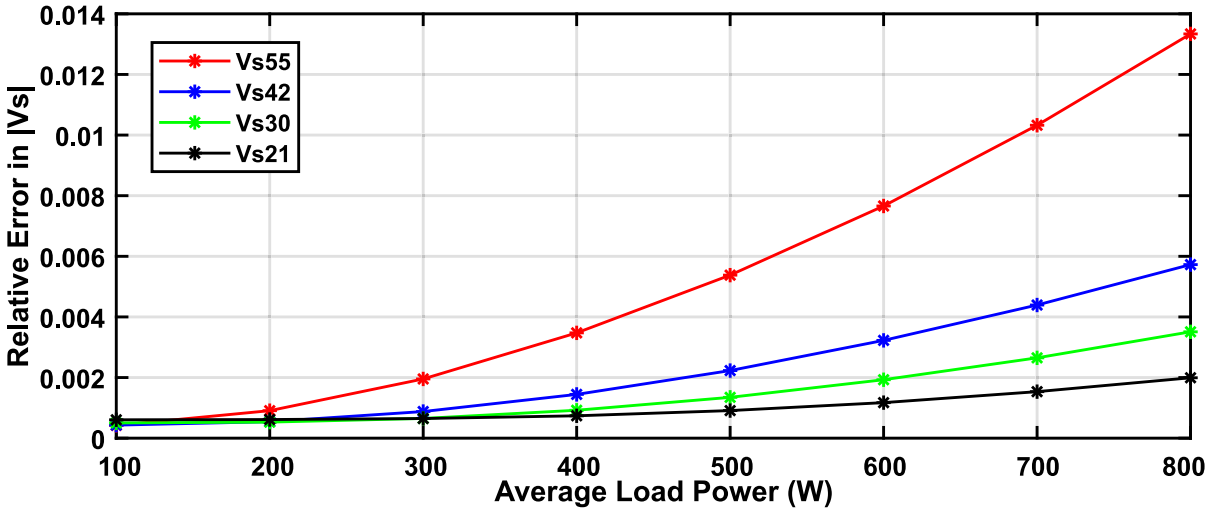
The unknown LV transformer voltage is another source of error [28]. In our setting, ignoring variations in V_s causes more than 8 times as large an error in $|S|$ as using the known V_s. This is because the difference $V_s - V$ is the small voltage drop across the transmission line in a local network, which is often small compared with the fluctuations of V_s. By contrast, an estimation of V_s by the proposed method can reduce the error in $|S|$ to less than 1.5 times as large as known $|S|$, which is a 6.5 fold improvement over assuming constant V_s.

Our proposed method demonstrates robustness against current noise. As shown in Table 2, substantial current noise (SNR_I = 10dB, equivalent to a 33% error in I) only doubles the error in S when compared to reasonable current noise (SNR_I = 40dB, equivalent to a 1% error in I). However, $|S|$ does exhibit sensitivity to voltage noise, a factor ignored in [18,27,28]. An equivalent increase in error in V results in a five-fold surge in error in $|S|$. This is because a given amount of noise may be a small fraction of V, but a large fraction of the much smaller quantity $V_s - V$.

The proposed method outperforms [27,28] by around 7.5 and 4 fold respectively in same-phase component of $|S|$ in 55-node network,



(a) $|S|$



(b) V_s

Fig. 4. Average Relative error of the estimated sensitivity matrix and transformer voltage versus adjustable power P_a, with different network sizes, when $SNR_I = 40$, $SNR_V = 46$, auto-correlation coefficient $\lambda = 0.8$. The legend are the same as in Figs. 3(a) and 3(b), note that the actual average power is 5 times as large.

as both methods use $1-\phi$ estimation, and the former treated V_s as a constant.

4.3. Three-phase loads

So far, all sensitivity estimation experiments have been limited to three-phase networks containing only single-phase nodes. However, even residential networks often have some three-phase loads, which are approximately balanced. A simple way to incorporate these is to treat them as three single-phase loads. This degrades the estimate of the sensitivity because it adds extra degrees of freedom to S that are not present in the data, leading to ambiguities in the solution. In further work, the problem should be reformulated to eliminate those degrees of freedom, but it is useful to understand how well the present algorithm works with a limited number of balanced three-phase loads.

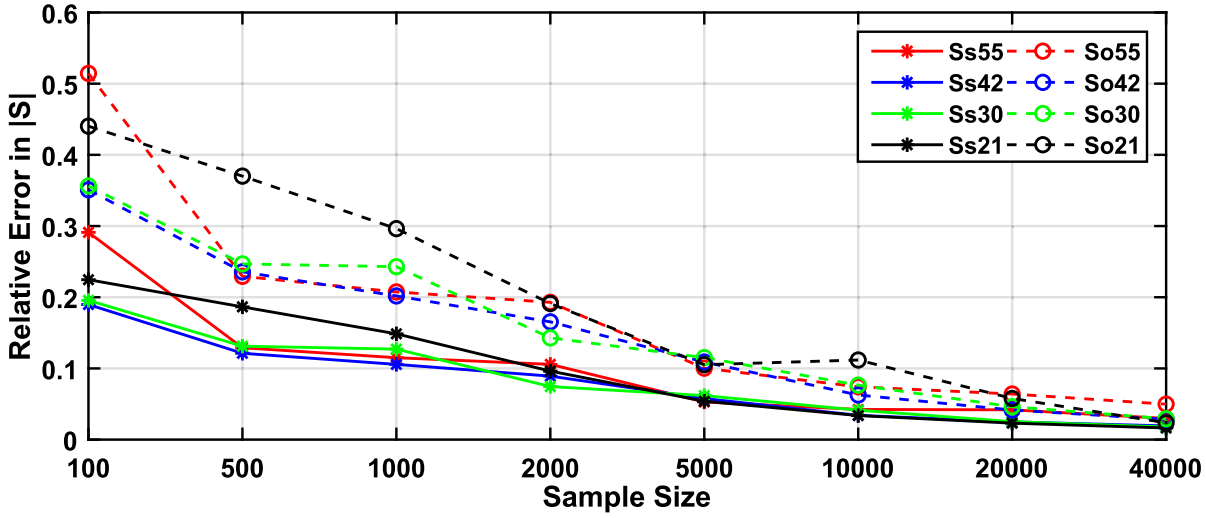
The relative error of $|S|$ was calculated for networks with 0% to 50% penetration of three-phase nodes, in steps of 12.5%. A penetration of x% consists of having $48x/100$ three-phase loads and $48(100-x)/100$ single-phase nodes. For each penetration, 50 random networks were generated, with their line parameters based on those of the EU feeder.

Fig. 6 presents the relative error of both diagonal (same-phase) and off-diagonal (other-phase) elements of estimated $|S|$ against the penetration level of three-phase nodes. As expected, the errors in $|S|$ increases as the fraction of three-phase loads increases, due to ambiguities in the additional redundant degrees of freedom in the model. This is in contrast to the degradation in most three-phase models when imbalance is introduced, because most models are designed for balanced nodes and imbalance is a form of mismatch, whereas the proposed model is for a three-phase network of single-phase loads and so the mismatch consists of having three-phase nodes.

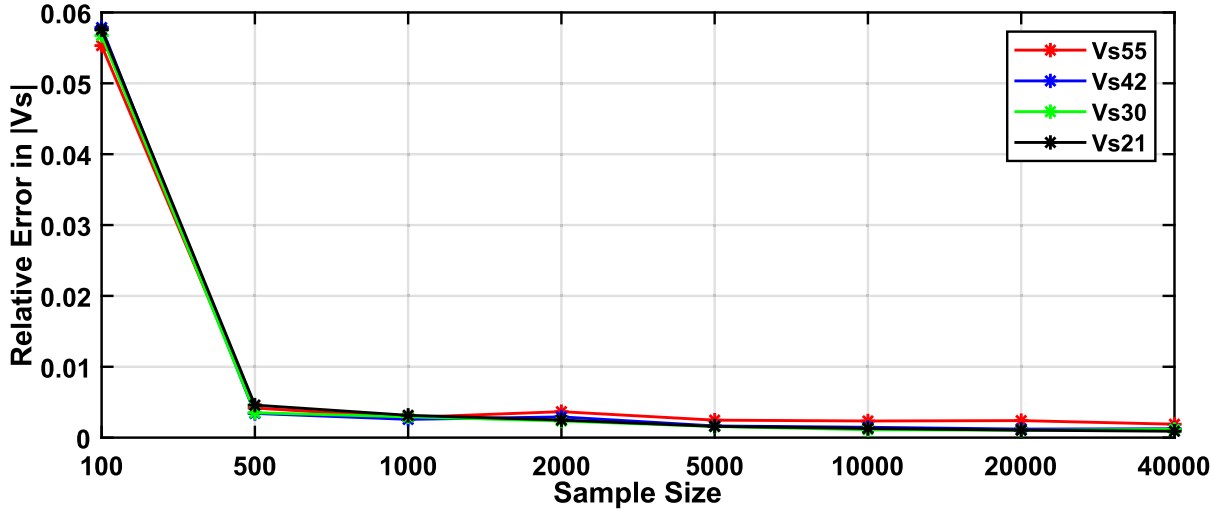
5. Evaluation of topology estimation

5.1. Single-phase topology estimation

We employ Algorithm 1 for reconstructing individual phase topologies in a 3-ϕ network, utilizing the estimated 3-ϕ voltage sensitivity coefficients. These evaluations involve the European 55-node feeder along with fifty synthetic networks of each of several sizes. This section assesses the influence of factors such as network size – 21, 30, 42,



(a) $|S|$



(b) V_s

Fig. 5. Average Relative error of the estimated sensitivity matrix and transformer voltage against sample size, with different network sizes. when $SNR_I = 40$, $SNR_V = 46$, $P_a = 400$ W auto-correlation coefficient $\lambda = 0.8$. The legend are the same as in Figs. 3(a) and 3(b).

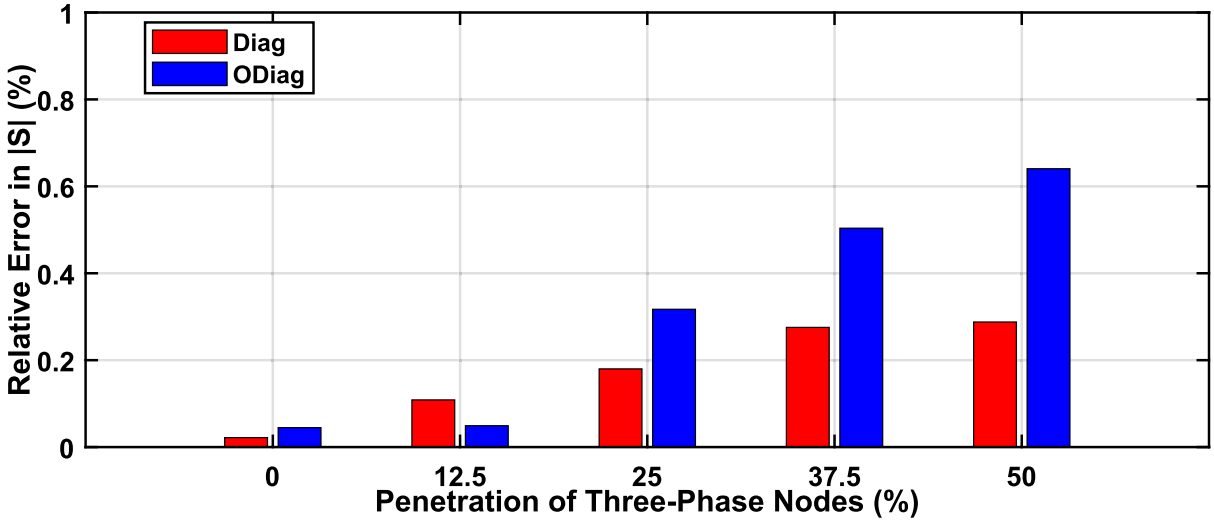


Fig. 6. Relative error of the estimated sensitivity matrix in 48-node networks with different penetration of three-phase nodes.

55-nodes –, average active load power level P_a, measurement signal-to-noise ratio *SNR*, and varying approximations on $\tilde{\theta}_v$ on the estimated network topology.

5.1.1. European 55-node feeder reconstruction

In this experiment, the synthetic load power and the LV transformer voltage are produced according to (33)–(37), with the following parameters: active power level $P_a = 400$ W, auto-correlation $\lambda = 0.8$, 40,000 samples, and applied noise for current and voltage at $SNR_I = 40$ dB and $SNR_V = 46$ dB respectively.

A comparison between the true and estimated topologies is not straightforward. For small individual instances, we show the true and estimated topology explicitly, but a numerical score is needed to evaluate the performance on larger data sets. There are various such scores, which differ in the emphasis they place on various types of mismatches: is transporting an entire subtree to a remote part of the graph a bigger or smaller difference than multiple changes that each only exchange neighbors? A natural candidate is the relative error between the true and estimated distance matrices D. However, this cannot be calculated if the topologies have different numbers of internal nodes. Instead, we show the tree graphs of one of the topologies, as well as quantifying their similarity by the DeltaCon score (DCS) introduced in [39] and used in [27] and [28]. The score is between 0 and 1, with a higher score meaning two topologies are more similar so that identical topologies have a score of 1. As a guide, topologies with a few minor structural differences would typically score 0.8–0.9, with large-scale subtree mismatch would score 0.5–0.7, and two completely unrelated topologies often score 0.25 or less. There are other quantitative measures of topological similarity, such as the normalized Total Vector Error (nTVE) used in [26] when the true topology is unknown. The Centrality Measurement (CM) used in [27] does not always accurately reflect the degree of structural discrepancy. Another alternative would be the Group Similarity proposed in [28] which seeks to weigh large-scale errors more than small-scale errors. As shown in [28], a higher Group Similarity typically corresponds to a higher DCS but is also much slower to calculate than the DCS. Consequently, we choose DCS for measuring the similarity between topologies, due to its overall accuracy and reduced computational time.

Figs. 7(a)–7(c) show the true topology (left), the topology reconstructed by the proposed method (middle), and the topology reconstructed by method in [28]. The mistakes made by both methods in each phase are emphasized by corresponding darker colors. To quantify the similarity of two graphs, we can use the Graph Edit Distance (GED) [40], which refers to the minimum number of graph operations – deleting a node/subtree, adding a node/subtree, moving a node/subtree to another place – required to transform one topology graph G to another G'. The proposed method makes one to two mistakes per phase (each consisting of moving a leaf and its parent). Moving a node from one edge to another causes one edge to split into two and another two edges to combine into one. In the topologies estimated by the proposed method (middle), the edge that has been merged is shown in magenta, and the edges resulting from the split are blue; the corresponding edges are colored the same as the true topology on the left. For example, in phase A, the parent of leaf node 11 (labeled P_{11} in Fig. 7(a)) is connected on the edge A_1A_3 (blue) in the true graph (left), but it is instead attached on the edge A_2A_3 in the estimated graph (middle). We count this as one mistake. Leaf node 16 is also attached to the wrong place, giving a GED of two mistakes. Calculating the GED is slow [40], and so we instead use the simpler DeltaCon score (DCS) when evaluating large collections of networks. As shown in the right figure of all three phases, the method in [28] not only made more mistakes than the proposed method involving more leaf nodes but also connected some subtree structures to relatively distant locations. For example, the subtree in phase A including 9,10,11 is misplaced to a distant location.

As for the DCS measure, the proposed method achieved 0.80, 0.89 and 0.82 for each phase, respectively. Comparing with 0.68, 0.65 and 0.7 scored by [28]; 0.67, 0.62, 0.67 scored by [27] (not shown in Figs. 7(a)–7(c)). To conclude, the proposed method outperformed [27, 28] by a notable margin.

5.1.2. Accuracy vs network size

This experiment evaluates the dependence of the accuracy of the estimated topology in terms DCS, as a function of the network size, and compares the effectiveness of the proposed features used for topology selection with those of [27]. Four different sizes of 3-ϕ networks – 21, 30, 42, 55 nodes – are used in this experiment, with fifty different networks of each size. The synthetic load power and the LV transformer voltage settings are identical to those in Section 5.1.1.

The results are shown in Fig. 8. The DCS (averaged over fifty networks) illustrate that the proposed method consistently achieved scores around 0.9 across all networks, although with a slight reduction in 55-node networks (refer to the fifth bar (blue) for each network size). The leftmost bar, in pale red, representing [28] and [27] performs worse by a factor of 1.1 to 1.3, and shown a decreasing trend with network size. Note that [27] and [28] shared the same topology selection features. The first and fourth bars, in pale red and red, denote the topology candidate selection based on features used in [27], which include voltage magnitude, sensitivity coefficients, voltage profile, and voltage correlation. The bars in blue show the proposed features, and the rightmost bar, in green, combines features from both the proposed method and [27]. The three sets of features achieving similar DCS, but the iterative backward-forward method applied in [27] required a 20-fold higher computation time than the proposed method. Thus, the proposed topology selection method is more efficient than that of [27].

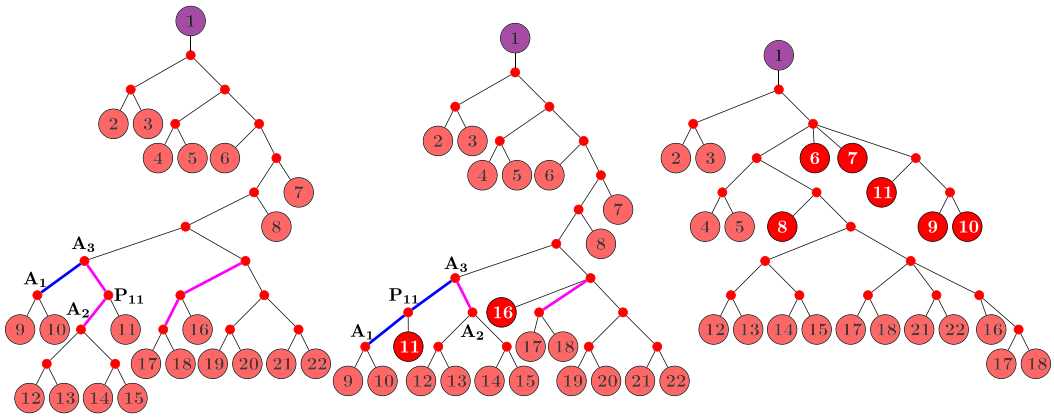
5.1.3. Accuracy vs sample size

This experiment is to evaluate the impact of number of time samples used to estimate the sensitivity on the accuracy of the estimated topology, as measured by DCS. The network and the parameter are the same as in 5.1.2, except a varying sample size from 100 to 40,000. As presented in 9, the accuracy increases steadily as the sample size increases, but plateaus below 1. Doubling the size of the network requires 10 or more times as many samples for a given accuracy.

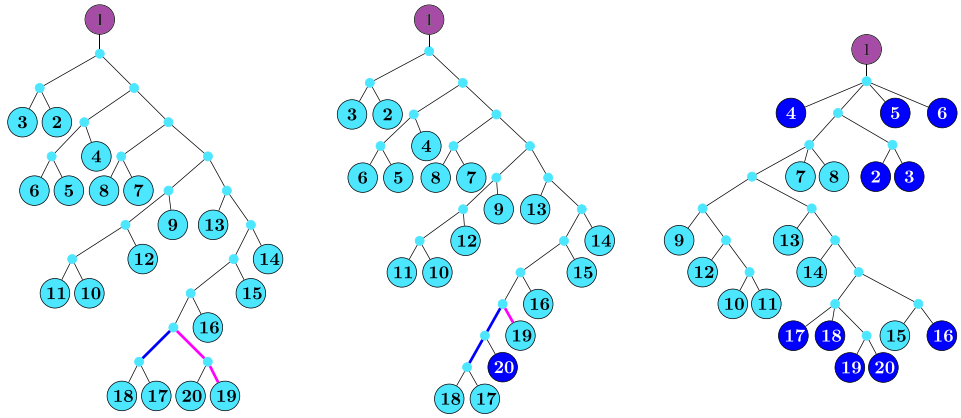
5.1.4. Accuracy vs active load power level

This experiment compares the DCS accuracy for networks with different loads. It is conducted on EU 55-node feeder as well as fifty 55-node synthetic networks. The synthetic load power and the LV transformer voltage settings are the same as in 5.1.1, other than the varying the adjustable load power level P_a.

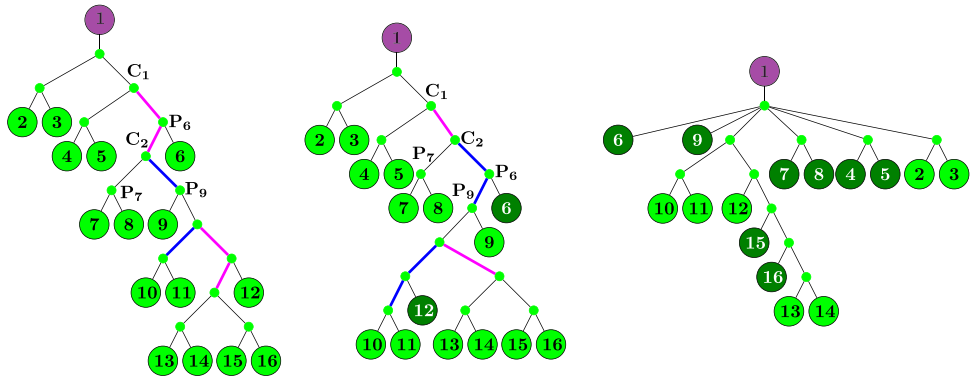
The results are depicted in Fig. 10. The proposed method in red first shows a slight increase then consistently decreases in DCS as P_a increases. This observation aligns with the pattern in 4(a), as a smaller relative error in $|S|$ leads to a more accurate topology. The topology estimated by [27] shown in green is significantly worse in DCS, degrades in performance fairly consistently as the load power increases. In contrast, the oracle (with known $\tilde{\theta}_v$) in blue is almost flat with increasing P_a after a slight initial increase. Knowing $\tilde{\theta}_v$ leads to an accurate same-phase component of $|V|$, thus the relative error on $V_s - |V|$ decreases with increasing load power. Although $|S|$ becomes more accurate as P_a increases, this marginal improvement in S has a negligible impact on topology estimation, rendering the blue line nearly flat.



(a) Phase A



(b) Phase B



(c) Phase C

Fig. 7. True and estimated topology of EU 55-node feeder. Left: True topology, Middle: Estimated topology by the proposed method, Right: Estimated topology by [28]. (a) Phase A:, (b) Phase B:, and (c) Phase C:. Mistakes in topology are highlighted.

5.1.5. Accuracy vs SNR

This experiment considers the impact of measurement noise on the accuracy of reconstruction, as measured by DCS. It also conducted on EU 55-node feeder and fifty 55-node synthetic networks, with the same setting as in 5.1.1, other than the varying current noise SNR_I, and the SNR_V, which is always 6 dB higher than SNR_I.

Fig. 11 shows that the DCS all increase with SNR. Both the proposed method and the oracle has a steeper improvement after $SNR_I = 30$ dB, the true topology starts to be reconstructed in some instances, they improve the average DCS significantly as they score 1. In the range where $SNR_I > 25$ dB, the proposed method outperforms the method presented in [27] (green), yet lags behind the benchmarking oracle (blue) by a notable gap.

5.2. Three-phase topology reconstruction

Our analysis has thus far focused on estimating three distinct 1-ϕ topologies using a 3-ϕ sensitivity matrix. However, reconstructing a unified 3-ϕ topology provides a more complete representation the interconnections between different phases. A few studies have reconstructed the 3-ϕ topologies according to [12], with most utilizing D-PMU data. The work presented in [26] stands out as the only one that used smart

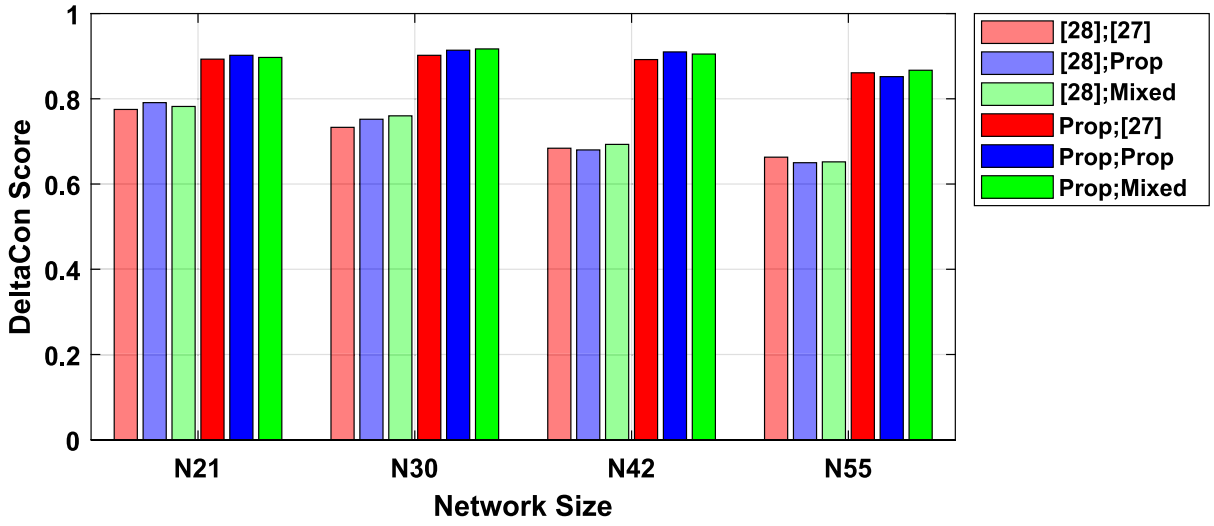


Fig. 8. Comparison of DCS of the estimated topology of [28] and the proposed method against network size. The legend shows ⟨grouping algorithm⟩;⟨selection features⟩. "Mixed" selection uses both sets of features. The proposed features improve speed without sacrificing accuracy.

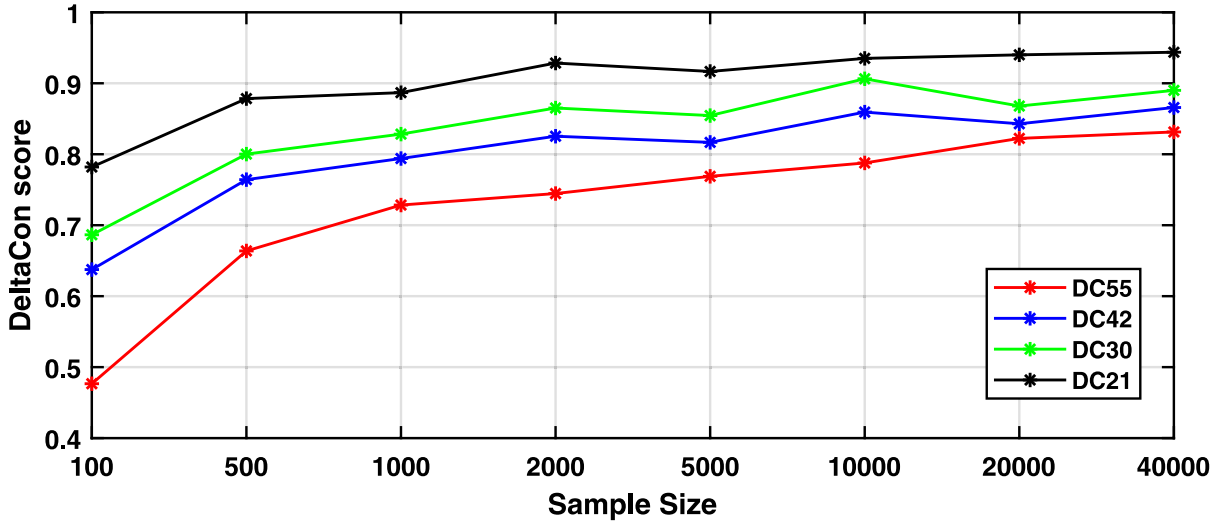


Fig. 9. Average DCS against sample size in the estimated topology of 55-node European feeder and several synthetic networks. Note the nonlinear horizontal axis.

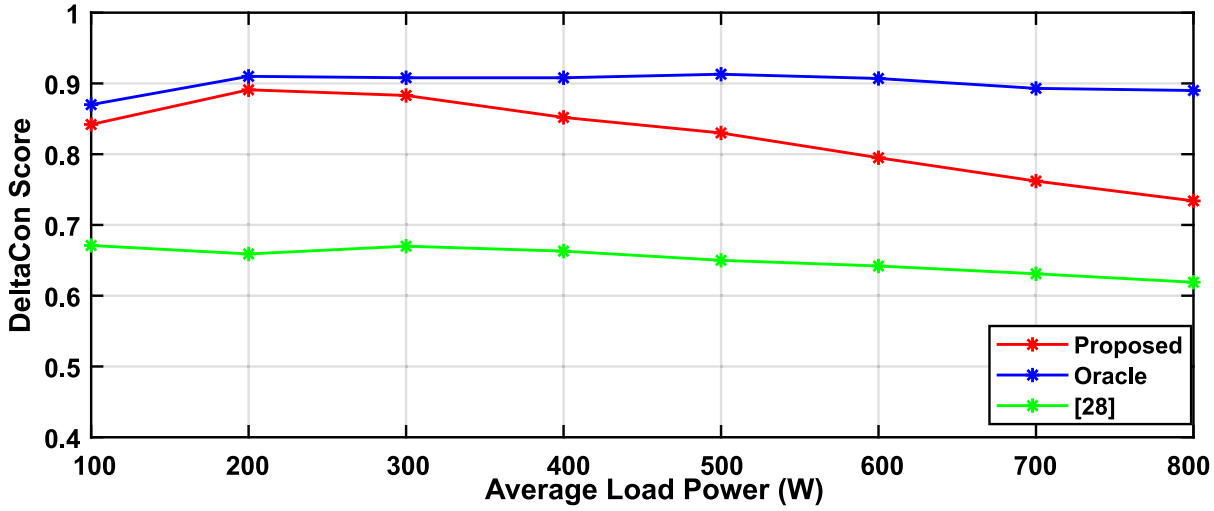


Fig. 10. Average DCS versus average load power P_a in the estimated topology of 55-node European feeder and several synthetic networks.

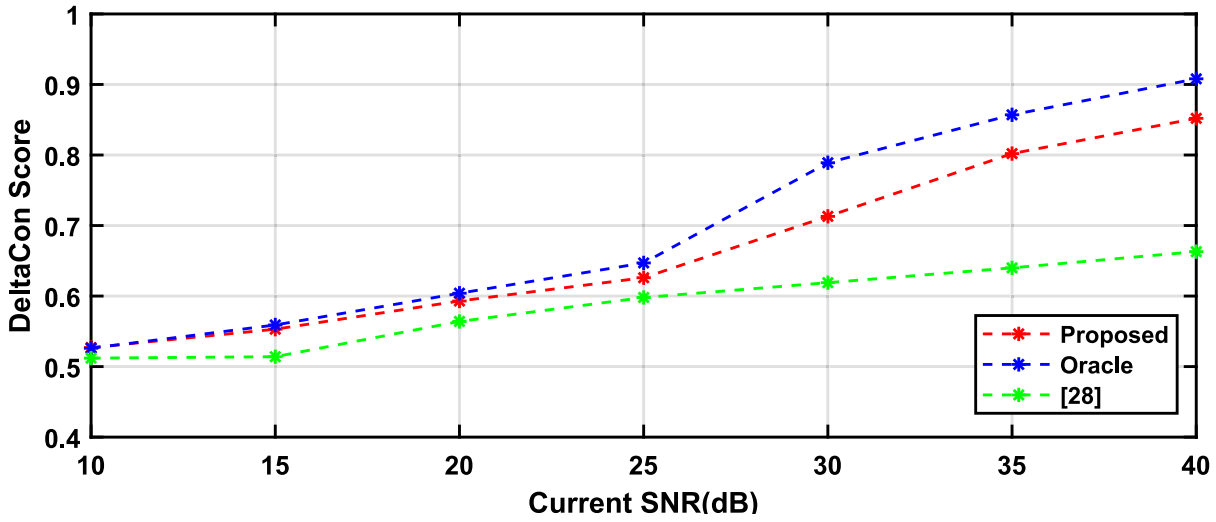


Fig. 11. Average DCS against *SNR* in the estimated topology of 55-node European feeder and several synthetic networks.

Table 3
DCS of the estimated topology of EU feeder. Estimating phases individually (1ϕ) or the unified topology (3ϕ), using the 1ϕ and full 3ϕ sensitivity estimates.

Ph.	1ϕ, S from 1ϕ	1ϕ, S from 3ϕ	3ϕ, S from 3ϕ
A	0.68	0.80	
B	0.65	0.89	0.76
C	0.7	0.82	

meter data. However, since they focused on other real networks that lacked ground truth and modeled the network as a spanning tree, drawing a comparison between the proposed method and [26] proves challenging.

To reconstruct the 3-ϕ topology, our approach uses the components of the sensitivity coefficients due exclusive to the neutral wire. However, the sensitivity coefficients estimated by our proposed method combine line and neutral in same-phase elements, while other-phase elements inherently contain only the neutral. Here, we assume that the line and neutral impedance are equal in same-phase elements, the neutral sensitivity matrix S_N can be obtained by simply halving all same-phase elements of S. As in the real network, the impedance on the line and neutral conductors should be similar to maintain balanced voltage levels and prevent excessive voltage drops, if the electrical wiring is correctly installed and properly maintained [41]. In this experiment, we reconstruct the 3-ϕ topology of 12, 21, 30, 42, and 55-node synthetic networks and the EU 55-node feeder. The inputs and parameters are the same as in Section 5.1.2. The results for synthetic networks are shown in 12; the DCS comparison of the EU 55-node feeder by different methods are depicted in Table 3.

In Table 3, a noteworthy observation is the lower performance of the 3-ϕ reconstructed network compared to its 1-ϕ counterparts. This discrepancy can be attributed to the significantly larger size of 3-ϕ networks, approximately three times the size of their 1-ϕ sub-networks. When employing the RG-backtracking algorithm on larger networks, it involves a multitude of decisions and a broader set of candidate topologies. This increases the likelihood of wrong decisions and suboptimal candidate topology selections, worsening the similarity score. As shown in Fig. 12, the similarity score is Figs. 8 and 12, the similarity score of 3-ϕ reconstructed 21-node network is similar to the 1-ϕ reconstructed 55-node network, (the 55-node network comprises roughly three 21-node networks) which demonstrates the network being 3-ϕ is not a reason of worsening performance. Table 3 further implies the proposed method is superior to [28], despite handling a network roughly three times as large.

Fig. 13 shows the computational time comparison between the proposed RG-backtracking algorithm and the algorithm in [28]. In this experiment, fifty networks of each size (250 in total) are evaluated. The sensitivity matrix is estimated by the proposed method, and converted to neutral sensitivity S_N to allow both algorithms to estimate the 3-ϕ topology. The median time taken by the proposed algorithm is two orders of magnitude less than that of the method outlined in [28]. Particularly in larger networks, the maximum processing time of the proposed algorithm is similar to or less than the median time of [28]. Notably, there are instances where [28] fails to find a topology within 24 hours (86,400 seconds, the corresponding instances are not included in the plot), a situation that has not been observed with the proposed algorithm. This substantial time reduction can be attributed primarily to a more efficient topology selection technique. While the topology selection process accounts for 82% to 98% of the total computation time in [28], it constitutes only 23% to 55% in our proposed approach. In summary, the proposed algorithm is markedly faster than the method in [28], with no loss of accuracy.

6. Conclusion and future work

This paper proposes a novel method for estimating the three-phase voltage sensitivity to current and the transformer voltage in low-voltage distribution power networks, using only smart meter measurements. This is achieved by considering the coupling effect between phases, and without the need to estimate the voltage phase shift and the quadrature component of the voltage. A series of experiments demonstrate that the proposed method outperforms state-of-the-art methods by a factor of 4 and 10 in the accuracy of same-phase sensitivity elements and transformer voltage, respectively. Moreover, the proposed method can estimate the other-phase sensitivity elements, which many Dynamic Operating Envelope researchers, such as [42] and [43], ignore. Instead, they approximate the three-phase system as three independent phases, causing some errors. Additionally, an enhanced recursive-grouping and Backtracking algorithm capable of reconstructing the full three-phase topology is presented, a feat not achieved by sensitivity-based methods in the prior art. The algorithm also features refined decision rules, more selective backtracking, and a fast topology selection technique, which improves accuracy and is two orders of magnitude faster than state-of-the-art methods in single-phase topology estimation.

This work leads to several new research directions. First, an accurate estimation of the voltage angle, possibly by iteration-optimization approach or gradient descent approach, can improve the accuracy of the voltage sensitivity and topology estimation, especially in industrial or commercial networks with larger loads. Second, the decision

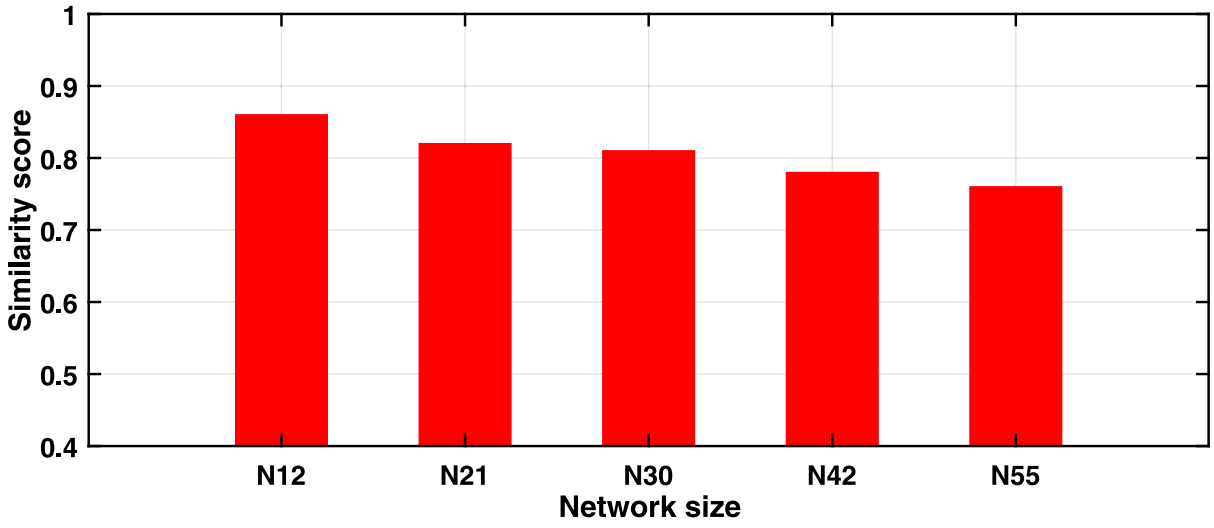


Fig. 12. Similarity score against network size in 3-ϕ topology reconstruction.

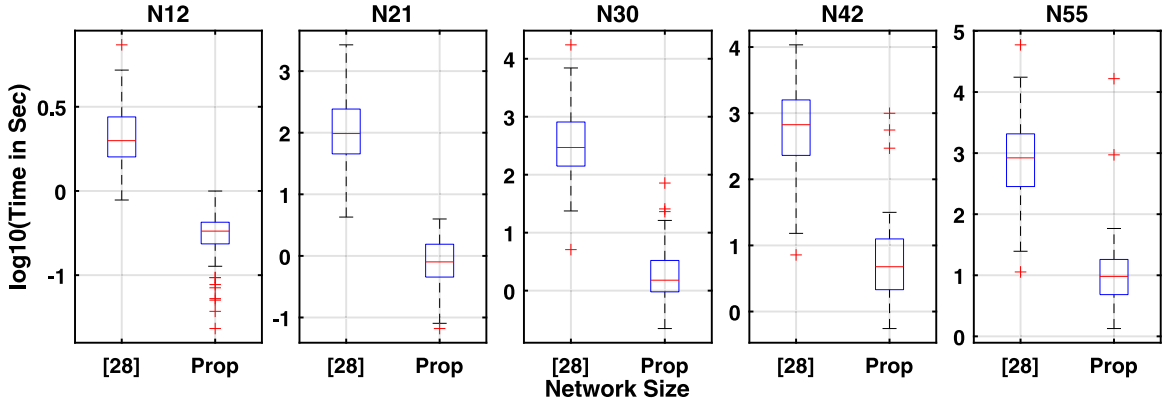


Fig. 13. Computational time comparison of the proposed topology estimation algorithm and method in [28]. In the horizontal axis, N: network size, Prop.: Proposed method, [28]: method in [28]. In the plot, red line: median time, box: 25% and 75% percentile, "+": outliers.

rules in the recursive-grouping and backtracking method were determined heuristically, and could be replaced by systematic rules based on a probabilistic model. Third, the formulation can be extended to three-phase loads by imposing suitable constraints on the sensitivity coefficients. In addition, the accurately estimated sensitivity matrix can help developing the model free Dynamic Operating Envelope and fairness problems. Finally, note that the proposed method is only applicable to radial networks, as found in Low-Voltage distribution networks, where most smart meters occur. Topology estimation of meshed network as found in Medium Voltage and High Voltage networks requires more information than is contained in phasor measurements at the customers when there may be unidentified (Steiner) nodes.

CRediT authorship contribution statement

Luxin Fang: Conceptualization, Data curation, Formal analysis, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Abu Bakr Pengwah:** Investigation, Resources, Validation. **Lachlan L.H. Andrew:** Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – review & editing. **Reza Razzaghi:** Conceptualization, Supervision, Writing – review & editing. **Mario Andrés Muñoz:** Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

This work was supported by the Australian Research Council grants DP190102134, DE230100056 and IC200100009.

References

[1] Meliopoulos AS, Polymeneas E, Tan Z, Huang R, Zhao D. Advanced distribution management system. IEEE Trans Smart Grid 2013;4(4):2109–17.
[2] Medina J, Muller N, Roytelman I. Demand response and distribution grid operations: Opportunities and challenges. IEEE Trans Smart Grid 2010;1(2):193–8.
[3] Fahim SR, Sarker SK, Muyeen S, Das SK, Kamwa I. A deep learning based intelligent approach in detection and classification of transmission line faults. Int J Electr Power Energy Syst 2021;133:107102.
[4] Terzija V, Valverde G, Cai D, Regulski P, Madani V, Fitch J, et al. Wide-area monitoring, protection, and control of future electric power networks. Proc IEEE 2010;99(1):80–93.

[5] Manditereza PT, Bansal RC. Protection of microgrids using voltage-based power differential and sensitivity analysis. Int J Electr Power Energy Syst 2020;118:105756.
[6] Liu Y, Ćetenović D, Li H, Gryazina E, Terzija V. An optimized multi-objective reactive power dispatch strategy based on improved genetic algorithm for wind power integrated systems. Int J Electr Power Energy Syst 2022;136:107764.
[7] Farajollahi M, Shahsavari A, Mohsenian-Rad H. Topology identification in distribution systems using line current sensors: An MILP approach. IEEE Trans Smart Grid 2019;11(2):1159–70.
[8] Phadke AG. Synchronized phasor measurements – a historical overview. In: IEEE/PES transmission and distribution conference and exhibition, vol. 1, IEEE; 2002, p. 476–9.
[9] Uribe-Pérez N, Hernández L, De la Vega D, Angulo I. State of the art and trends review of smart metering in electricity grids. Appl Sci 2016;6(3):68.
[10] Lovell H. Are policy failures mobile? An investigation of the advanced metering infrastructure program in the state of Victoria, Australia. Environ Plan A Econ Space 2017;49(2):314–31.
[11] Chen F, Dai J, Wang B, Sahu S, Naphade M, Lu C-T. Activity analysis based on low sample rate smart meters. In: Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining. 2011, p. 240–8.
[12] Deka D, Kekatos V, Cavraro G. Learning distribution grid topologies: A tutorial. IEEE Trans Smart Grid 2023.
[13] Moffat K, Bariya M, Von Meier A. Unsupervised impedance and topology estimation of distribution networks—Limitations and tools. IEEE Trans Smart Grid 2019;11(1):846–56.
[14] Chauhan K, Sodhi R. A robust state estimation framework for active distribution network with distribution-level PMUs. In: 2020 IEEE power & energy society general meeting. IEEE; 2020, p. 1–5.
[15] Peppanen J, Grijalva S, Reno MJ, Broderick RJ. Distribution system low-voltage circuit topology estimation using smart metering data. In: 2016 IEEE/PES transmission and distribution conference and exposition (t&d). IEEE; 2016, p. 1–5.
[16] Soumalas K, Messinis G, Hatziargyriou N. A data driven approach to distribution network topology identification. In: 2017 IEEE manchester powerTech. IEEE; 2017, p. 1–6.
[17] Liao Y, Weng Y, Liu G, Rajagopal R. Urban MV and LV distribution grid topology estimation via group lasso. IEEE Trans Power Syst 2018;34(1):12–27.
[18] Park S, Deka D, Backhaus S, Chertkov M. Learning with end-users in distribution grids: Topology and parameter estimation. IEEE Trans Control Netw Syst 2020;7(3):1428–40.
[19] Ardakanian O, Wong VW, Dobbe R, Low SH, von Meier A, Tomlin CJ, et al. On identification of distribution grids. IEEE Trans Control Netw Syst 2019;6(3):950–60.
[20] Yu J, Weng Y, Rajagopal R. PaToPa: A data-driven parameter and topology joint estimation framework in distribution grids. IEEE Trans Power Syst 2017;33(4):4335–47.
[21] Yu J, Weng Y, Rajagopal R. PaToPaEM: A data-driven parameter and topology joint estimation framework for time-varying system in distribution grids. IEEE Trans Power Syst 2018;34(3):1682–92.
[22] Bariya M, Deka D, von Meier A. Guaranteed phase & topology identification in three phase distribution grids. IEEE Trans Smart Grid 2021;12(4):3605–12.
[23] Deka D, Chertkov M, Backhaus S. Topology estimation using graphical models in multi-phase power distribution grids. IEEE Trans Power Syst 2019;35(3):1663–73.
[24] Gandluru A, Poudel S, Dubey A. Joint estimation of operational topology and outages for unbalanced power distribution systems. IEEE Trans Power Syst 2019;35(1):605–17.
[25] Liao Y, Weng Y, Liu G, Zhao Z, Tan C-W, Rajagopal R. Unbalanced multi-phase distribution grid topology estimation and bus phase identification. IET Smart Grid 2019;2(4):557–70.
[26] Li H, Weng Y, Liao Y, Keel B, Brown KE. Distribution grid impedance & topology estimation with limited or no micro-PMUs. Int J Electr Power Energy Syst 2021;129:106794.
[27] Pengwah AB, Fang L, Razzaghi R, Andrew LLH. Topology identification of radial distribution networks using smart meter data. IEEE Syst J 2021.
[28] Flynn D, Pengwah AB, Razzaghi R, Andrew LLH. An improved algorithm for topology identification of distribution networks using smart meter data and its application for fault detection. IEEE Trans Smart Grid 2023.
[29] García S, Mora-Merchán JM, Larios DF, Personal E, Parejo A, León C. Phase topology identification in low-voltage distribution networks: A Bayesian approach. Int J Electr Power Energy Syst 2023;144:108525.
[30] Shi Z, Xu Q, Liu Y, Wu C, Yang Y. Line parameter, topology and phase estimation in three-phase distribution networks with non-μPMUs. Int J Electr Power Energy Syst 2024;155:109658.
[31] Duan G, Yu Y. Power distribution system optimization by an algorithm for capacitated steiner tree problems with complex-flows and arbitrary cost functions. Int J Electr Power Energy Syst 2003;25(7):515–23.
[32] Liu Y, Zhang N, Hou Q, Botterud A, Kang C. Topology and admittance estimation: Precision limits and algorithms. 2021, arXiv preprint arXiv:2106.00532.
[33] Talkington S, Turizo D, Grijalva S, Fernandez J, Molzahn DK. Conditions for estimation of sensitivities of voltage magnitudes to complex power injections. IEEE Trans Power Syst 2023.
[34] Talkington S, Grijalva S, Reno MJ, Azzolini JA, Peppanen J. Localized structure in secondary distribution system voltage sensitivity matrices. Electr Power Syst Res 2024;226:109788.
[35] Peschon J, Piercy DS, Tinney WF, Tveit OJ. Sensitivity in power systems. IEEE Trans Power Appar Syst 1968;(8):1687–96.
[36] Choi MJ, Tan VY, Anandkumar A, Willsky AS. Learning latent tree graphical models. J Mach Learn Res 2011;12:1771–812.
[37] Deka D, Talukdar S, Chertkov M, Salapaka MV. Graphical models in meshed distribution grids: Topology estimation, change detection & limitations. IEEE Trans Smart Grid 2020;11(5):4299–310.
[38] IEEE. IEEE PES test feeder. 2023, https://site.ieee.org/pes-testfeeders/.
[39] Koutra D, Vogelstein JT, Faloutsos C. Deltacon: A principled massive-graph similarity function. In: Proceedings of the 2013 SIAM international conference on data mining. SIAM; 2013, p. 162–70.
[40] Gao X, Xiao B, Tao D, Li X. A survey of graph edit distance. Pattern Anal Appl 2010;13:113–29.
[41] IEEE. IEEE recommended practice for powering and grounding electronic equipment. 1999, p. 1–408. http://dx.doi.org/10.1109/IEEESTD.1999.90559, IEEE Std 1100-1999.
[42] Liu MZ, Ochoa LF, Wong PK, Theunissen J. Using OPF-based operating envelopes to facilitate residential DER services. IEEE Trans Smart Grid 2022;13(6):4494–504.
[43] Naughton JC. A modelling framework for virtual power plants under uncertainty [Ph.D. thesis], University of Birmingham; 2022.