Automating the Verification of the Low Voltage Network Cables and Topologies

Maizura Mokhtar, *Member, IEEE*, Valentin Robu, David Flynn, Ciaran Higgins, Jim Whyte, Caroline Loughran, Fiona Fulton

***Abstract*—Low Voltage (LV) networks are increasingly required to cope with challenges they were not designed for, requiring for more active network management (ANM). Crucially, ANM solutions require the availability of accurate network information. In practice, available data on LV networks can be incomplete, a problem often overlooked in prior ANM research. For example, in the UK and many developed countries, the lifetime of distribution networks assets spans several decades, with some of the available asset data gathered and maintained over many years. This can often lead to incomplete cable data being available to network operators. To overcome this, we propose a novel machine learning technique to autonomously approximate the missing cable information in LV networks. Our proposed algorithm uses a tree-based search methodology, which approximates the missing cable's cross section area (XSA) data based on rules engineers used when designing the LV networks. We validate our approach using a large database of real LV networks, where some of the cables' XSA are treated as unknown and used as ground truth to evaluate the accuracy of the predictions. Moreover, we propose a mechanism that scores the confidence level of the prediction, information which is then presented to the human network planners.**

***Index Terms*—LV networks, Cables, Asset management, Network trees (graphs), Machine learning**

NOMENCLATURE

u	Cable with unknown data
a_m	Cable with known data selected to approximate u
P_u	Asset path with u
P_{a_m}	Asset path with a_m
S_{a_m}	Score which measures similarity between asset paths P_u and P_{a_m}.
M	The total number of cables with known data selected to approximate u, a_m, $m \in M$.
a_x	Cable with known data with the maximum score among S_{a_m}, $m \in M$.
S_{a_x}	The score for a_x.
R	The total number of occurrence that a_m is selected to approximate u.
a_{m_r}	The r^{th} occurrence that a_m is selected to approximate u. $r \in R$.

I. INTRODUCTION

The reactive and passive approach to LV network management may no longer be suitable with the predicted change to the users energy needs, specifically, the changes to the domestic electricity demand [1]. The changes ranging from the increase in penetration of embedded renewables; the charging of electric-powered vehicles (EVs) at homes [2] and the use of electric heating. Historically, distribution network operators' or DNOs' option for LV networks reinforcements were to add more capacity from new assets or to manually reconfigure the networks when warnings are reported. The former comes at a high cost, and the later if not resolved in time can lead to faults resulting in loss of electricity connection. Active network management (ANM) is often identified as an option to help resolve this issue [3], [4], [5], [6]. Enabling ANM for the LV Networks, however, is historically unfeasible. ANM may now be possible because of the broader visibility of the LV network through the use of smart meters and the broader infrastructure initiatives, i.e. for communication and cyber-security. There is also an increasing use of Geographical Information Systems (GIS), which collate and store the LV network topologies and assets data within a single platform, which also ease in enabling ANM for LV networks.

In more detail, one of the key tools required for LV networks ANM is a Power System State Estimation (PSSE) tool. Given a network topology and energy demand information, PSSE estimates and simulates the most likely state of the networks [7], [8], and for LV networks ANM, informing how best to manage the energy import from Distributive Energy Resources (DER) [9], [10], [11], [12], [13], and/or the scheduling of Electric Vehicles (EVs) charging [14], [15], [16]. PSSE can also be used to evaluate alternative LV network configurations when the need for network reconfiguration is predicted. LV network reconfiguration is beneficial, for example, if the predicted increase in energy demand resulted from the potential increase of EVs charging at home in the evening after their use have resulted in significant voltage constraints violations. If the number of EVs requiring charging is high and demand side management (DSM) is unable to effectively schedule their charging, network reconfiguration may be required.

There is an abundance PSSE tools described in literature, and PSSE has also become an industry accepted tool. However, PSSE was developed and is commonly used for simulating the Medium Voltage (MV) and above. At present, the effectiveness of PSSE decreases significantly when used for simulating the LV networks [17], [18], [19], [20], [21]. As these studies indicate, PSSE is currently ineffective for the LV networks, not only because of the lack of observability in the networks, but also due to the nature of the network themselves: dynamic radial networks with asymmetry between the phase loading; high variability in the type of cables used; and higher cables

M. Mokhtar, V. Robu and D. Flynn is with School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, EH14 4AS, Scotland. e-mail: {m.mokhtar,v.robu,d.flynn}@hw.ac.uk

C. Higgins is with Derryherk Ltd., UK.

J. Whyte is with Notsoanalytic Ltd., UK.

C. Loughran and F. Fulton is with Smart Meter Systems, SP Energy Networks, Blantyre, UK

Manuscript received Month Day, 2019.

impedance ratio (R/X) [22]. Furthermore, one of the most common issues identified, which can significantly impact the LV networks state estimations, is the possible incomplete cables data within the networks [8], [23].

A. Related work: Validation of LV network cable data

In recent years, DNOs have begun to address the lack of observability by using data from the mass smart meter roll-out. Yet, the incompleteness of the LV cables information may be harder to resolve, as obtaining the complete LV assets information can be a challenging and expensive task, if it was handled manually.

The difficulties are due both to the high volume and significant age of typical network assets, which can result in missing data. Significant errors can also result from undocumented LV assets changes, because of their historic maintenance and repair cycles [20], [21], [24]. The reporting of the assets traditionally relied on the effectiveness of human operators performing repairs to report on the most recent asset information. Furthermore, in many real networks, the majority of the LV assets are underground, posing a significant challenge in verifying the available assets information [24], [25].

There are many studies described in literature aim to perform PSSE using smart meter data [18], [26], [27], [28]. This includes the need to infer LV network topologies [29]. Authors of [30] and [31] use voltage clustering to identify customers' phase on the network and their source of supply. Graph-theoretic interpretation [32], and the combination of linear coupled power flow (LC-PF) model using smart meter data and recursive grouping algorithm [24], [33] are some of the methods proposed to identify the network topology. These studies assume that near-full to full coverage of smart meters are provided. This is not often the case, resulting in the limited uptake of these techniques by DNOs in practice. Crucially, these studies also assume that impedance of all cables are available. In practice, however, missing cable data is an issue that needs to be resolved *a-priori*.

The aim of our work is to approximate the missing cable data within the LV networks, in order to enable effective analysis of the networks. Incomplete cable data can result in erroneous network capacity estimation and can affect the outcome of many LV network analysis, from PSSE, ANM, including DER management and DSM [9], [10], [13], [15], [16], to network sensitivity analysis [34] and network topology assessments [35].

Furthermore, accurate reporting of LV networks' cable data are not only required for the network's PSSE, but it is also vital for the health management of the networks. In the UK, there have been a number of government-funded initiative to improve the accuracy of assets and cables reporting for the LV networks. Majority of this work concentrates on the health assessments of the cables. Example initiatives are: (i) the work conducted by Electricity North West that concentrates on the development of hardware solutions aimed to identify and capture the condition of cables within the LV network [36]. (ii) IQA has conducted manual inspection of SP Energy Network's cables [37] to identify at-risks networks. The limitations of these example methodologies are the potential high man hours required to inspect and identify the cables on the network [38]; more so when one wish to identify the missing cable data. Data analytic can reduce this requirement. This paper proposed a novel machine learning technique that can address the issue related to missing cable data, specifically the missing cable cross section area (XSA). The novel technique together with smart meter data can in-turn reduce the need for the indicated high man hours, whereby erroneous correlation between the LV network impedance, indicated by the cable XSA, and its voltage distribution can be indicative of errors.

We propose a tree-based search methodology that first represents the LV network as a tree describing its cable make-up and connectivity, and the branch of the tree with the unknown cable are compared against other branches on the same tree or other trees, so that those that are similar to the one with the unknown cable is the candidate suitable to approximate the unknown cable. A scoring scheme is also proposed to indicate its suitability and how the scores are used to rank the options if multiple options are provided. This paper also discusses and evaluates the search criteria to approximate the unknown cable. The proposed method improves and revises our preliminary method initially presented in [39], whereby a new scoring scheme with the scores between 0 to 1 is proposed in this paper to provide better indication of choice if multiple options for approximations are found. This paper also aims to validate the proposed method via ground truth evaluation, whereby known cable values in the LV networks were defined as unknown and are to be approximated. This analysis was not performed in the initial publication.

This method is to be implemented as part of the business-as-usual continual improvement program for LV monitoring and asset data management practice for one of UK's DNO. The method is part of a bigger goal of providing data analytical checks that can react to ongoing data management process failure and adapt to historical issues which are harder to fix manually. Feedback from the DNO has indicated the significant time-saving advantage of the proposed method, essential for the transitioning of UK's DNO to Distribution System Operator that provides active management of the LV networks.

The paper is divided into 6 sections; Section II describes how the LV network is modelled as a tree and how the trees generated are used to approximate the unknown cables information. Section III presents the experiment used to evaluate the effectiveness of the proposed methodology. Section IV presents the results from the experiment. Section V indicates the potential benefits of the proposed method. Section VI concludes the paper.

II. Methodology

A. Representing an LV network as a network tree graph

For this paper, we defined an LV network as a collection of LV circuits that share the same source or transformer. LV circuit is defined as a collection of cables that connect the endpoints to a common electricity source or transformer. Figure 1 shows an example LV network consisting of 3 LV circuits.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2019.2941722, IEEE Transactions on Smart Grid

Fig. 1: The LV network and its asset path tree.

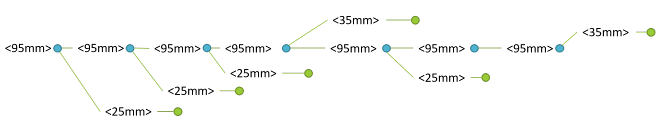


Fig. 2: The conventional display of the network tree for the boxed area of the LV network in Fig. 1.

A good representation of the LV network is required when approximating the unknown or missing cables information. A tree diagram, consisting of V nodes and E edges is chosen to provide this representation. An edge $e_{i,j} \in E$ indicates either the connection point that links one cable (segment) i to another j or the connection point that links a cable i to an endpoint j. An endpoint can either be a consumer of electricity on the network or an open point connection. A node $v_i \in V$ represents the cable (segment) between two connection points or edges. Each node v_i is associated with the properties z_i that indicates the electrical impedance of the cable, its construction type and its cable length, or any other attributes that describe the properties of the cable.

Conventionally, an LV network tree displays the nodes and edges for all connection points between cables. The complexity of the network tree, therefore, can be indicative of the complexity of the LV network that it represents. For example, a long mains 95mm cable that connects to two 35mm and four 25mm service cables shown in the boxed area in Fig. 1 is displayed as the branch of the tree as shown in Fig. 2.

B. Asset path and the asset paths tree

A network tree consisting of all the connection point within the network may have information unnecessary to approximate the missing cable information. For example, it is unnecessary to represent the long 95mm mains cable in Fig. 1 as multiple nodes on a tree as shown in Fig. 2 [40]. Because of this, the term asset path is introduced to compress the complex network tree down to its most important information [39].

An asset path is defined as the collection of cables that connects the energy source to the customer(s). Fig. 1a shows an example asset path tree that is a compressed network tree representation of the LV network shown in Fig. 1. Like the actual LV network, for which it acts as a "digital twin", the tree shows how the cables located in the middle of the circuits (or the mains cables) are connected to one another. However, the connections to the cables at the end of the network (or the service cables) will differ between the asset path tree, the conventional network tree and the actual network.

In the example shown in the boxed area in Fig. 1, the endpoints connected at the end of the two 35mm and four 25mm service cables, because of their similar cross-section area, these endpoints are categorised into two; one for the 35mm cables and the other for the 25mm cables. These two types of cables categories are each represented as a node on the asset path tree, despite their actual connection locations and lengths. These nodes are connected at the end of the mains 95mm cable's node, which these two cables types are all connected to. Because of this, the branch on the network tree shown in Fig. 2 is simplified so that the long 95mm cable can be represented as a single node connected to the two nodes representing the 25mm and 35mm service cables. The filled blue circles in Fig. 1a indicate the connection points between the cables' nodes. The endpoints are represented as nodes that are connected at the end of its representative cable's node. These are the green filled diamonds shown in Fig. 1a.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2019.2941722, IEEE Transactions on Smart Grid

If one traces the tree from the root node of a circuit (identified in Fig. 1a as <Fuse>) to an endpoint (the filled green diamond), the list of cables that creates the connection between the source (a closed fuse in a substation) to the endpoint is therefore the asset path.

C. Asset path to approximate the unknown cable

Network designers often follows a set of guidelines when designing the LV network. The combinations of cables are often selected to meet the aggregated energy demand requirements of customers and the customers distance from the energy source or substation. Based on this understanding, two assumptions are made when devising the methodology to approximate the missing cable information [39].

First, if the unknown portion of an LV circuit (identified as <*UNKNOWN*> in Fig. 1a) shares the same connection point (the filled blue circle) with another cable, the unknown cable can be approximated with the other cable if it is of similar length and loading. This is based on the design rule which states that *a specific cable that is of specific length is selected for a specific type of distributed customer loading*.

If the unknown cable does not share its connection point with another cable (see unknown cable shown in Fig. 1a), or the available cable options have large dissimilarities in their lengths and loading, the asset path with the unknown cable is then compared against other asset paths on the same circuit or on other circuits that have similar cable configurations (i.e. similar combinations of lengths and loading). This assumption is justified by the design rule which states that *a specific combination of cables and their lengths (asset path) is often selected for a specific type of distributed customer loading*.

D. Scoring the selected asset paths

Equation (1) is proposed to score the similarities, S_{a_m}, between two asset paths:

1) the asset path P_u that has the unknown cable data u.
2) the other asset path P_{a_m} selected to approximate u.

$$S_{a_m} = \left(\frac{\sum_{n=1}^{U_n - 1} \frac{min(D_{a_n}, D_{u_n})}{max(D_{a_n}, D_{u_n})} + \frac{min(E_{a_{a_m}}, E_{u_u})}{max(E_{a_{a_m}}, E_{u_u})}}{U_n} \right) \times \left(\frac{\sum_{n=1}^{U_n - 1} \frac{min(C_{a_n}, C_{u_n})}{max(C_{a_n}, C_{u_n})} + \frac{min(L_{a_{a_m}}, L_{u_u})}{max(L_{a_{a_m}}, L_{u_u})}}{U_n} \right) \quad (1)$$

U_n is the number of cable types in asset paths P_u, up to and including the unknown cable u. Therefore, the asset paths $\{P_{a_1}, P_{a_2}, ..., P_{a_m}, ..., P_{a_M}\}$, M is the number of options selected to approximate u, should have $U_n - 1$ cable types from the source that are the same as to P_u. In the example shown in Fig. 1a, there are $M = 2$ asset paths that meets this criteria to approximate the indicated unknown cable type <UNKNOWN>.

If u is the cable located at the end of the circuit, $E_{a_{a_m}}$ is the mean length for cable type a_m in P_{a_m} that are connected to the load(s). E_{u_u} is the mean length for u in P_u. If u is located in the middle of the circuit, $E_{a_{a_m}}$, instead, is the mean length for where the start of a_m (mains cable) connects to other cable at the customer end. E_{u_u} is similar to $E_{a_{a_m}}$, but for u in P_u. D_{a_n} is the maximum length for the (mains) cable n in P_{a_m}. D_{u_n} is the maximum length for the same cable n in P_u.

C_{a_n} is the number of cables connected to the cable n, towards the consumer end and in P_{a_m}. C_{u_n} is similar to C_{a_n}, but for the same cable n in P_u. If u is located at the end of the circuit, $L_{a_{a_m}}$ is the number of loads connected to a_m in P_{a_m} and L_{u_u} is similar to $L_{a_{a_m}}$, but for u in P_u. If u is located in the middle of the circuit, $L_{a_{a_m}}$ is the number of cables connected to a_m and L_{u_u} is similar to $L_{a_{a_m}}$ but for u in P_u. If u and a_m are nodes on the same branch, i.e. share the same connection point (the filled blue circle in Fig. 1a), $D_{a_n} = D_{u_n}$ and $C_{a_n} = C_{u_n}$, therefore:

$$\frac{min(D_{a_n}, D_{u_n})}{max(D_{a_n}, D_{u_n})} = \frac{min(C_{a_n}, C_{u_n})}{max(C_{a_n}, C_{u_n})} = 1 \quad (2)$$

If there are multiple choices available to approximate u in P_u, the score S_{a_m} will guide the selection, with $m \in M$, where M is the number of asset path(s) P_{a_m} selected for comparison. The cable a_x in P_{a_x} is selected to approximate u in P_u because a_x has the highest score S_{a_x} (3) compared against others ($x \in M$).

$$S_{a_x} = max(\{S_{a_1}, S_{a_2}, ..., S_{a_m}, ..., S_{a_M}\}) \quad (3)$$

The cable a_x is selected to approximate u in P_u with the assumption that P_u is most similar to P_{a_x}.The closer the score S_{a_x} to 1, the higher the similarities between P_{a_x} and P_u, and a_x can best approximate u. This is inline with the two assumptions derived from the LV network design rules:

1) *a specific cable that is of specific length is selected for a specific type of distributed customer loading*.
2) *a specific combination of cables and their lengths (asset path) is often selected for a specific type of distributed customer loading*.

E. Scoring multiple similar asset paths

A specific asset path P_{a_m} can be found multiple times, both/either in the same circuit as P_u and/or in a different circuit altogether; $\{P_{a_{m_1}}, P_{a_{m_2}}, ..., P_{a_{m_r}}, ..., P_{a_{m_R}}\}$, $r \in R$ and R is the number of occurrence that the asset path P_{a_m} is found with cable type a_m that can approximate u. The score $S_{a_{m_r}}$, (1) with r omitted from the equation, may differ at each match occurrence r of $P_{a_{m_r}}$, especially if the $P_{a_{m_r}}$ is found on another circuit. The requirements of $P_{a_{m_r}}$ may differ from one circuit to another, resulting in the differences in scores $S_{a_{m_r}}$ in (4).

If $R > 1$, the mean scores for S_{a_m} calculated at each match occurrence r (4) is used to find the best approximated cable type a_x that has the highest score S_{a_x} in (3).

$$S_{a_m} = mean([S_{a_{m_1}}, S_{a_{m_2}}, ..., S_{a_{m_r}}, ..., S_{a_{m_R}}]) \quad (4)$$

If a particular asset path P_{a_m} are found multiple times ($R > 1$) and have similar length and loading to each other, they will have their scores $S_{a_{m_r}}$, $r \in R$, closer to 1. This will result in the mean of the scores S_{a_m} also closer to 1 (4), and will therefore be selected to approximate u. This brings us to our

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2019.2941722, IEEE Transactions on Smart Grid

third assumption which indicates that *the more often a particular combination of cables (asset path) are found, the higher is the likelihood that the combination of cables is effective for a particular distributed customer loading requirement*.

This scoring scheme differs from the initial scheme presented in [39], to limit the score S_{a_m} between 0 to 1. The new scheme provides a clearer indication of similarity in comparison to the initial scoring scheme presented in [39], which resulted in the scores to be > 1.

F. Defining the search space

The number of LV networks available to perform the comparison analysis can vary. The number can be as big as from one particular region or country, or as small as a single settlement or a specific postcode area. If a large search space is provided to approximate u in P_u, this can be computationally expensive. Additional complication can also arise when dealing with large search space. An example complication will be discussed in Sec. IV-C.

The search to approximate u can be classified into 3 main types (i) the local search, (ii) the clustered search and finally, if needs be, (iii) global search.

1) Local search

As shown in Fig. 1a, a transformer can have multiple fuses, each is providing electricity to its own circuit. An LV network connected to a specific transformer can therefore be defined as a combination of circuits that share the same transformer. Local search compares the asset path P_u with u with all other asset paths in its own circuit and in the circuits that shares the same transformer as P_u. Ideally, the local search is performed first to approximate u. This is based on our fourth design principle which states that *the circuits that share the same source are of balanced distributed customer loading, and therefore will have the common combination of cables used at specific lengths* (asset paths), which the unknown cable u can be approximated from. For the network in Fig. 1a, the circuit with u can be compared against two other circuits that share its transformer.

2) Clustered and global search

Local search is beneficial for circuits that share their transformer with others, as it has its 'nearest neighbours' to compare against. However, a large proportion of the circuits do not fall within this category. Typically, majority of the transformers only have one fuse and provides electricity to just one circuit. For these circuits, there is a higher likelihood that u may not be approximated using local information only.

If the local search is unable to find a suitable approximation, i.e., no similar asset path is found or that the scores $\{S_{a_1}, S_{a_2}, ..., S_{a_m}, ..., S_{a_M}\}$, M is the number of cable XSA options to approximate the unknown cable XSA u have low scores, the search space is to be widened to include more circuits, ideally, geographically closer to the circuit with u. This, for example, in the same postcode area or in the same county.

If the defined area is small, complication resulted from large search space may not be an issue, and all circuits provided can be included in the analysis (or global search). For example, if a radius of 0.7km in a suburban area in Scotland is selected for analysis, only 26 circuits are to be analysed, therefore all circuits can be included in the search space (global search). If the radius is to increase to 1km with 225 circuits, 2km with 1,800 circuits, or 5km radius with 22,576 circuits, the computation time can be significantly higher, if all circuits in the radius are to be included in the search space. Table I indicates the approximated run times for analysing and approximating missing cables' XSA in 1,800 circuits and 22,576 circuits. The descriptions of full and partial asset paths are described in Sec. II-G.

TABLE I: Run time analysis for large number of LV circuits

No. of circuits	Full asset path & local search	Partial asset path & local search	Full asset path & clustered search	Partial asset path & clustered search	Full asset path & global search	Partial asset path & global search
1,800 circuits	16.5 mins	25.4 mins	37.25 mins	2.40 hrs	2.55 hrs	6.47 hrs
22,576 circuits	84 mins	146.7 mins	6.42 hrs	31.96 hrs	6.4 days	> 7 days

Furthermore, the larger the radius, the higher the likelihood of dissimilarities between the circuits. For example, 5km radius from Edinburgh or Glasgow city centre can encompass both rural and urban areas. The dissimilarities may not add significant benefits to the results of approximations despite the additional computation time. To minimise this complexity, the search space is best curtailed to only those that are of similar characteristics to the circuit with u.

The properties selected for the clustering are: (i) the circuit length; (ii) the number of load; and (iii) the number of cable types on the longest asset path. The asset path comparison is then performed for the circuits that are in the same cluster as the circuit with u. Because of the similarity between the circuits that are clustered together, the selected asset path P_{a_x} will be similar to P_u, and the score S_{a_x} will be closer to 1. This indicates for a high confidence that u is that of a_x.

In our analysis, the k-medoid algorithm [41] is used for clustering, specifically the Matlab $kmedoids$ function [42], which utilise a variant of the Lloyd's iteration as described in [43] to find the medoids. Through this method, 20 clusters are found.

There is the likelihood that circuits that share a source (transformer and local search) do not belong in the same cluster. When this occur, all the circuits from the same transformer will be assigned to the cluster for which majority of these circuits do belong to. This is to ensure that these circuits are evaluated using the same set of information.

G. Full versus Partial Asset Paths

The analysis is performed in two parts:

1) First, to approximate u located at the end of the circuits. These cables are often the missing service cable data.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2019.2941722, IEEE Transactions on Smart Grid

2) Second, to approximate u located in the middle of the circuit, typically the missing mains cables data.

These two parts are further split into two methods:

1) Full asset path, where the comparison requires all the cables in P_u and P_{a_m} to match except for u in P_u and a_m in P_{a_m}. A strict match is therefore required before a score can be calculated.
2) Partial asset path, which uses only the cable previously connected to u (at the source end) and another cable prior to that cable (as shown in Fig. 3). Hence, a less strict match is required before the score is calculated.

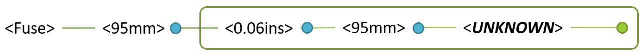


Fig. 3: Full asset path vs. partial asset path (boxed area).

The reason for these two options can be seen in the results figures presented in Sec. IV. A much higher percentages of successful approximation of u are shown when the partial asset path was used, instead of the full asset path. The key reason for this is that the strict match of all but u is needed when the full asset path method is used. This especially when there are > 2 cable types in the asset path which connects u from the source, which as shown in Fig. 3 (with 3 cable types prior to u). In another example, if some of the service cables, especially the cables at the end of the long LV circuit path in Fig. 1 are instead with unknown XSA, the number of cable types per asset path prior to the service cables will be > 2.

III. Experimental Setup

Approximately 4,000 circuits were provided for the analysis. These circuits are randomly selected from Central Belt region of Scotland (UK), and are a combination of domestic and commercial circuits located in urban, suburban and rural area. The dataset lists the nodes V and edges E data for the LV network trees $G(V, E)$ as shown in Fig. 2. A node v, $v \in V$ is the cable segment in the circuit. The unknown cable data in the dataset is the cable's cross section area (XSA), and is to be approximated. This missing data is important and is required to calculate the LV circuit's ratings and impedance. Incorrect calculation can lead to incorrect understanding of circuit capacity and risk.

To evaluate the success of the algorithm, cable segments with known XSA were selected and defined as unknown u. These set of cable segments act as ground truth, to evaluate if a_x is that of its original cable XSA. Half of the 4,000 provided circuits were selected at random for their cable segments with known XSA to be defined as unknown. For each of the selected circuit, the cable segments were also selected at random.

These circuits also have existing unknown cables XSA within them. The length and location of the cables with existing unknown XSA varies from one circuit to another. For majority of the circuits, the length of the existing unknown cables' XSA are $<50\%$ of the total length of the circuit. Because of this, the number of known cable segments selected for ground truth evaluation per circuit will also vary. This ranges between 2 to 15 cable segments per circuit. The selected cable segments defined as unknown may also be next to each other.

The experiment is performed in two stages: (i) local search and (ii) clustered search. Global search was not performed because of the high number of circuits provided for analysis, resulting in significantly long computation times. Furthermore, as indicated in Sec. II-F, because of the varied nature of the circuits provided (both urban and rural circuits), no significant benefits can be achieved with global search.

For the experiment, half of the provided circuits are selected at random for their cable segments with known XSA to be defined as unknown u. The experiment is repeated 5 times, to provide us with different locations of unknown cables XSA in the circuits at each iteration. The repetition enables for $>$10,000 circuit evaluations. Each circuit evaluation calculates the percentage of successful approximation of the ground truth cable XSA. A successful approximation is when a_x is the cable actual XSA before it was defined as unknown and has the highest match score S_{a_x} (3).

The random selection of the circuits is performed once for the initial search criteria of local search with full asset path. To have a consistent benchmark, this same sets of circuits are then re-used when performing the local search with the partial asset path. They are also re-used for the clustered search, for both full and partial asset path.

IV. Experimental Evaluation

The effectiveness of the algorithm will not only depend on the size of the search space, but also the quality of the data. If a circuit contains higher proportion of existing unknown cables within the circuit, the likelihood for the unknown cable to be approximated will decrease, as there is insufficient information available to approximate the unknown cable.

The results corroborate this statement, with the tables and figures in this section show the median percentages of successful approximation of the ground truth (y-axis) reduces as the proportions of unknown cables in the circuit, both existing and defined (x-axis), were to increase, both when full and partial asset path were used and when using local or clustered search. The percentage of unknown cables is calculated based on the number of cable segments with existing and defined unknown XSA against the total number of cable segments in the circuit. A cable segment is the cable node v in the asset path tree in Fig. 1a.

A. Local search

As indicated in Sec. III, the likelihood of the unknown cables approximated will also decrease with the increase in the number of cables in the asset path. This is as shown in Table II, with the lowest median percentages of successful ground truth approximation when u are located at the end of the asset path and when full asset path is used. When the full asset path method is used, a strict match for all but u in the two asset paths P_u and P_{a_m} are required and in the same order. The unknown cable u at the end of the asset path are typically the service cables on a long LV circuit connected via multiple mains cable types from the source. u located in the middle are the mains cable with missing XSA. The results show that

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2019.2941722, IEEE Transactions on Smart Grid

TABLE II: Median percentages of successful approximations (confidence interval), with local search.

	$u <$ 10%	$u <$ 25%	$u <$ 50%	$u <$ 75%	$u <$ 100%
Full asset path & u in the middle	77.78% (2.38%)	71.43% (1.45%)	71.43% (1.20%)	69.23% (1.22%)	66.67% (1.35%)
Partial asset path & u in the middle	80.0% (2.47%)	75.0% (1.44%)	71.43% (1.16%)	71.43% (1.27%)	66.67% (1.35%)
Full asset path & u at the end	62.5% (5.52%)	60.0% (2.63%)	50.0% (2.04%)	50.0% (2.10%)	50.0% (2.67%)
Partial asset path & u at the end	66.67% (5.14%)	62.5% (2.75%)	57.14% (2.06%)	50.0% (2.19%)	50.0% (2.67%)

TABLE III: Median percentages of successful approximations (confidence interval), with clustered search.

	$u <$ 10%	$u <$ 25%	$u <$ 50%	$u <$ 75%	$u <$ 100%
Full asset path & u in the middle	82.58% (6.38%)	76.92% (1.51%)	75.0% (1.13%)	75.0% (1.07%)	75.0% (1.29%)
Partial asset path & u in the middle	80.0% (2.61%)	80.0% (1.32%)	75.0% (1.11%)	75.0% (1.12%)	75.0% (1.29%)
Full asset path & u at the end	66.67% (4.59%)	66.7% (2.34%)	62.5% (2.13%)	60.0% (2.0%)	60.0% (2.23%)
Partial asset path & u at the end	75.0% (3.02%)	75.0% (2.33%)	71.43% (1.97%)	66.67% (1.80%)	66.67% (2.09%)

the median percentages increases when partial asset path was used and when u is at the end. This is because the strict match of the full asset path method are no longer required. Only the two cable types prior to u are needed to match to enable the approximation.

Because of this also, the median percentages of successful approximation are similar when u is in the middle, when full and partial asset paths are used. This is because the number of cable types prior to the u mains cable are lower compared to when u are at the end. There are also with less variability of how the mains cables are typically connected as they are often chosen to meet a specific requirements. As a result, this decreases the number of options available for approximations, increasing the likelihood of successful approximation.

The results show that the percentages of successful ground truth approximation are lower for when u is at the end of the asset path, both when full and partial asset path are used. This, in comparison to those in the middle. This is because of the high variability of how service cable are to be connected, as how they are connected depends on how far the cables (customers) are from the source. This for example as shown in Fig. 1. Furthermore, there is a ratio of 1.74:1 existing unknown cables located at the end of the asset paths, in comparison to those located in the middle.

To increase the percentages of successful ground truth approximation, a larger search space is required. Results shown in Fig. 4 show that for the circuits which shares its transformers with >1 other circuits, its median percentages of successful approximation have increased. This, especially for the unknown cables u located in the middle of the asset paths and when either full asset path or partial asset path were used, with 50% success for both when the circuits do not share their transformer, to 81.82% of success for full asset path and 83.33% for partial asset path when the circuits share their transformer with ≥ 8 other circuits. When u are at the end and when partial asset path are used, the percentage of success went from 50% when the circuits do not share their transformers to 72% of success when the circuits share with ≥ 8 others. This indicates the benefit of extending the search criteria to include more circuits for comparison, increasing the likelihood of successful match.

B. Increasing the search space using clustered information

The search space can be increased depending on the data availability and preference. LV circuits provide electricity for different uses, depending on the type and the number of loads (users) that they are connected to. Clustering the circuits ensures that those circuits that are of similar properties are only selected for comparison. As indicated in Sec. II-F2, all the circuits from the same transformer will be assigned to the cluster for which majority of these circuits do belong to. This is to ensure that these circuits are evaluated using the same set of information. Clustering can also identify abnormal circuits' topology (outliers), which can also be indicative of 'at-risk' connections or resulted from data error. An example of this is when a cluster only consists of abnormally long (length) circuits when compared against others.

Figure 5 shows that the median percentages of successful ground truth approximation increases when there is an increase in the search space, with Table III shows the percentages of successes when using clustered search. More so for the selected cables located at the end of the circuit, and when partial asset paths were used.

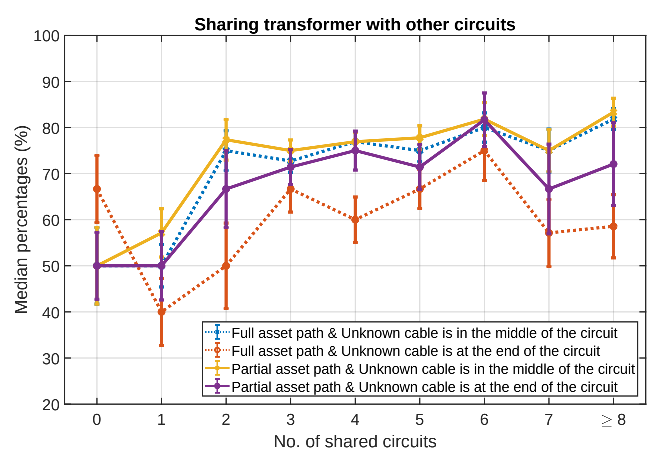


Fig. 4: Median percentages of successful ground truth approximation for circuits that share their transformers.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2019.2941722, IEEE Transactions on Smart Grid

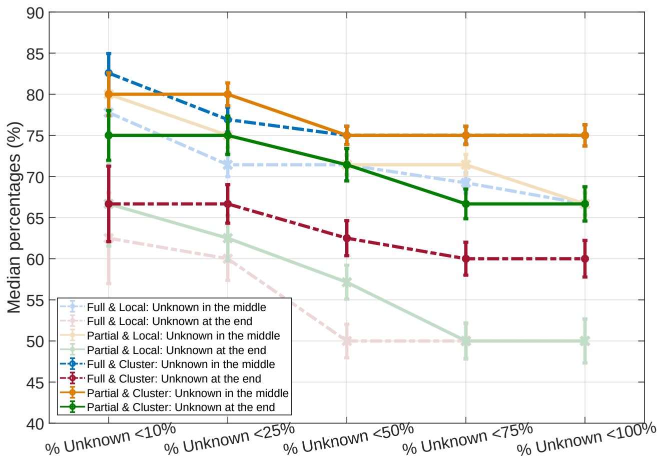


Fig. 5: The median percentages of successful approximation for varying percentage of unknown cable u in the circuit.

C. Cable dissimilarities

As the search space increases, there will be more options available to approximate the unknown cable u. The results indicated in Table III indicated the median percentages of successful ground truth approximation when a_x with the maximum score S_{a_x} is that of the actual cable XSA (ground truth). Table IV shows the median percentages of success if the ground truth cable XSA is one of the cables options $\{a_1, a_2, ..., a_m, ..., a_M\}$ that were selected as valid M number of options to approximate the cable XSA, when the valid XSA was defined as unknown u.

When comparing the percentage of success when a_x with the maximum score S_{a_x} is that of the actual cable XSA against when the ground truth cable XSA is one of the valid M options indicated, the results show that there are increases to the median percentages. This is as summarised in Fig. 6.

This indicates that for some LV circuits, the asset path with the ground truth cable XSA is an available option from the M selected asset paths $\{P_{a_1}, P_{a_2}, ..., P_{a_m}, ..., P_{a_M}\}$, however, there is another asset path in the circuit or other circuits that may have similar asset path attributes compared to the one with the ground truth value. Further insight into the dissimilarity of choice when the selected cable a_x with the maximum score S_{a_x} is not the ground truth can provide indication of risk of the existing cables' combination or asset path. This is one of our future work.

V. Benefits of the proposed method

As indicated in Sec. I, Active Network Management (ANM) may be the most cost-effective solution to address the potential risk of the increase in electricity demand to the LV networks. ANM requires the use of PSSE to simulate the state of the network based on new input and control options. Effective use of PSSE in turn requires high visibility of the LV networks and the correct reporting of the network assets, specifically the cables, and topology. The mass roll-out of smart meters and the use of Geographical Information System or GIS that collates the LV network topology information can provide this opportunity. However, missing cable data can affect the effectiveness of the ANM controls and analysis. The algorithm proposed in this paper will approximate any missing cable XSA required to enable effective simulation of the LV networks, specifically the calculation of the network topology to identify the risks and capacities.

TABLE IV: Median percentages of successful approximations (confidence interval), when the ground truth is one of the M valid options.

	$u <$ 10%	$u <$ 25%	$u <$ 50%	$u <$ 75%	$u <$ 100%
Full asset path & u in the middle	85.71% (2.91%)	81.82% (1.27%)	83.33% (1.01%)	83.33% (1.18%)	83.33% (1.11%)
Partial asset path & u in the middle	90.0% (2.04%)	88.89% (1.07%)	88.89% (0.86%)	88.89% (0.79%)	90.0% (0.74%)
Full asset path & u at the end	71.43% (3.97%)	66.7% (2.33%)	66.7% (2.36%)	66.7% (2.16%)	66.7% (2.00%)
Partial asset path & u at the end	85.71% (3.40%)	80.91% (1.97%)	80.0% (1.97%)	80.0% (1.80%)	80.0% (1.67%)

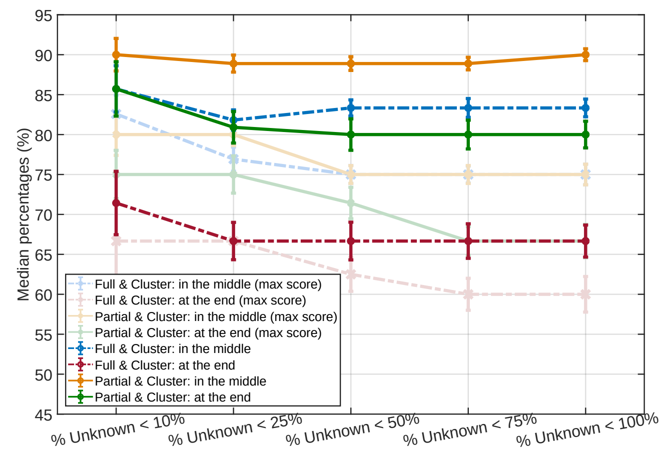


Fig. 6: The median percentages of successful approximation when ground truth is a_x with S_{a_x} only (max score) or that it is also one of the M options available (a_m, $m \in M$).

A network engineer can validate the approximated missing cable information, by comparing the outputs of the voltage distribution from PSSE with that recorded by the smart meters. Validated LV networks assets and topologies are necessary to ensure the most effective ANM solutions can be proposed.

VI. Conclusion and Future work

The demand for electricity is predicted to increase and the predicted increase may be higher than what the existing LV networks can accommodate for. Reinforcements to the LV networks are therefore necessary to ensure risks are appropriately managed. The most cost-effective reinforcement is to enable the Active Network Management (ANM) of the LV networks; previously designed with the passive 'fit-and-forget' approach to network management.

The mass roll out of smart meters and the growing use of GIS are the enabler to this capability. However, there is the challenge of ensuring that the correct LV network information are reported in the GIS. This is because, missing asset information, historical or resulting from on-site operators failing to follow rigorous data management processes, are some of the most common issues identified. This requires for a cost-effective data repair solution. This paper proposes the algorithm that autonomously approximate the missing cable data, specifically the missing cross section area cable data. The proposed algorithm uses a tree-based search methodology, which approximates the missing cable data based on rules engineers used when designing the LV network topology. Known cables data are redefined as unknown, and act as the ground truth. The algorithm is able to successfully approximate the ground truth data, and also providing a mechanism that scores the confidence level for the choices made to approximate the missing cable data.

In the future, we aim to include operators' expert knowledge to the scoring of the cable choices, for example, the date of installation and repair cycles, and the cable preferences used. Our future work will also evaluate the use of smart meter data, specifically the voltage data, to validate the cable choice, as well as to identify errors in the data. How voltage is distributed across the circuit depends on the total line impedance. Errors can be identified if these two data are uncorrelated.

Other clustering methods, for example k-means++ and hierarchical clustering are also to be explored in the future. Alternatively, clustering can be performed based on the geographical area, if the specific coordinate of the transformers are provided, by analysing the LV circuits within a small radius, for example, 0.5km radius from a chosen transformer, and to perform local search first, followed by global search if local search is unable to approximate all the unknown cable data. If the initial search with the small radius was unable to approximate the unknown cables, the radius is then increased gradually until a maximum search criterion is met. This, for example with the increments of 0.5km radius from a reference transformer until the maximum radius of 2km is met.

Acknowledgment

The authors would like to thank the funders Innovate UK (B16N12241) and OFGEM (NIA SP Energy Networks - SPEN0016).

References

[1] L. F. Ochoa and P. Mancarella, "Low-carbon lv networks: Challenges for planning and operation," in *2012 IEEE Power and Energy Society General Meeting*, July 2012, pp. 1–2.

[2] "Global ev outlook 2017: Two million and counting," 2017. [Online]. Available: https://www.iea.org/publications/freepublications/publication/GlobalEVOutlook2017.pdf

[3] Z. Bao, W. Qiu, L. Wu, F. Zhai, W. Xu, B. Li, and Z. Li, "Optimal multi-timescale demand side scheduling considering dynamic scenarios of electricity demand," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 2428–2439, May 2019.

[4] C. Le Floch, S. Bansal, C. J. Tomlin, S. J. Moura, and M. N. Zeilinger, "Plug-and-play model predictive control for load shaping and voltage control in smart grids," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 2334–2344, May 2019.

[5] Y. Liu, J. Li, and L. Wu, "Coordinated optimal network reconfiguration and voltage regulator/der control for unbalanced distribution systems," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 2912–2922, May 2019.

[6] P. Li, H. Ji, C. Wang, J. Zhao, G. Song, F. Ding, and J. Wu, "Optimal operation of soft open points in active distribution networks under three-phase unbalanced conditions," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 380–391, Jan 2019.

[7] A. Abur and A. G. Expsito, *Power System State Estimation: Theory and Implementation*. CRC Press, 2004.

[8] M. Pau, E. Patti, L. Barbierato, A. Estebsari, E. Pons, F. Ponci, and A. Monti, "Low voltage system state estimation based on smart metering infrastructure," in *2016 IEEE International Workshop on Applied Measurements for Power Systems (AMPS)*, Sep. 2016, pp. 1–6.

[9] S. Weckx and J. Driesen, "Optimal local reactive power control by pv inverters," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 4, pp. 1624–1633, Oct 2016.

[10] A. Navarro-Espinosa and L. F. Ochoa, "Probabilistic impact assessment of low carbon technologies in lv distribution systems," *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 2192–2203, May 2016.

[11] M. Andoni, V. Robu, W.-G. Frh, and D. Flynn, "Game-theoretic modeling of curtailment rules and network investments with distributed generation," *Applied Energy (Elsevier)*, vol. 201, pp. 174 – 187, 2017.

[12] V. Robu, G. Chalkiadakis, R. Kota, A. Rogers, and N. R. Jennings, "Rewarding cooperative virtual power plant formation using scoring rules," *Energy (Elsevier)*, vol. 117, pp. 19 – 28, 2016.

[13] G. Chicco, "Overview and performance assessment of the clustering methods for electrical load pattern grouping," *Energy*, vol. 42, no. 1, pp. 68 – 80, 2012, 8th World Energy System Conference, WESC 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360544211008565

[14] V. Robu, E. H. Gerding, S. Stein, D. C. Parkes, A. Rogers, and N. R. Jennings, "An online mechanism for multi-unit demand and its application to plug-in hybrid electric vehicle charging," *J. Artif. Intell. Res.*, vol. 48, pp. 175–230, 2013.

[15] Y. Song, Y. Zheng, and D. J. Hill, "Optimal scheduling for ev charging stations in distribution networks: A convexified model," *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 1574–1575, March 2017.

[16] A. H. Zaidi, K. Sunderland, and M. Conlon, "Role of reactive power (statcom) in the planning of distribution network with higher ev charging level," *IET Generation, Transmission Distribution*, vol. 13, no. 7, pp. 951–959, 2019.

[17] G. A. Taylor, M. R. Irving, N. Nusrat, R. Liao, and S. Panchadcharam, "Smart distribution network operation: Emerging techniques and standards," in *2011 IEEE Power and Energy Society General Meeting*, July 2011, pp. 1–6.

[18] A. Abdel-Majeed and M. Braun, "Low voltage system state estimation using smart meters," in *2012 47th International Universities Power Engineering Conference (UPEC)*, Sept 2012, pp. 1–6.

[19] S. Lefebvre, J. Prvost, and L. Lenoir, "Distribution state estimation: A necessary requirement for the smart grid," in *2014 IEEE PES General Meeting — Conference Exposition*, July 2014, pp. 1–5.

[20] J. Yu, Y. Weng, and R. Rajagopal, "Patopa: A data-driven parameter and topology joint estimation framework in distribution grids," *IEEE Transactions on Power Systems*, vol. 33, no. 4, pp. 4335–4347, July 2018.

[21] K. Dehghanpour, Z. Wang, J. Wang, Y. Yuan, and F. Bu, "A survey on state estimation techniques and challenges in smart distribution systems," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 2312–2322, March 2019.

[22] U. Kuhar, M. Panto, G. Kosec, and A. vigelj, "The impact of model and measurement uncertainties on a state estimation in three-phase distribution networks," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 3301–3310, May 2019.

[23] K. Christakou, M. Paolone, and A. Abur, "Voltage control in active distribution networks under uncertainty in the system model: A robust optimization approach," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 5631–5642, Nov 2018.

[24] S. Park, D. Deka, and M. Chcrtkov, "Exact topology and parameter estimation in distribution grids with minimal observability," in *2018 Power Systems Computation Conference (PSCC)*, June 2018, pp. 1–6.

[25] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of smart meter data analytics: Applications, methodologies, and challenges," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 3125–3148, May 2019.

[26] J. Huang, V. Gupta, and Y. F. Huang, "Electric grid state estimators for distribution systems with microgrids," in *2012 46th Annual Conference on Information Sciences and Systems (CISS)*, March 2012, pp. 1–6.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2019.2941722, IEEE Transactions on Smart Grid

[27] A. Primadianto and C. N. Lu, "A review on distribution system state estimation," *IEEE Transactions on Power Systems*, vol. 32, no. 5, pp. 3875–3883, Sept 2017.

[28] L. Wang, D. H. Liang, A. F. Crossland, P. C. Taylor, D. Jones, and N. S. Wade, "Coordination of multiple energy storage units in a low-voltage distribution network," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 2906–2918, Nov 2015.

[29] Y. Liu, N. Zhang, Y. Wang, J. Yang, and C. Kang, "Data-driven power flow linearization: A regression approach," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 2569–2580, May 2019.

[30] V. Arya, T. S. Jayram, S. Pal, and S. Kalyanaraman, "Inferring connectivity model from meter measurements in distribution networks," in *Proceedings of the Fourth International Conference on Future Energy Systems*, ser. e-Energy '13. New York, NY, USA: ACM, 2013, pp. 173–182. [Online]. Available: http://doi.acm.org/10.1145/2487166.2487186

[31] R. Mitra, V. Arya, B. Sullivan, R. Mueller, H. Storey, and G. Labut, "Using analytics to minimize errors in the connectivity model of a power distribution network," in *Proceedings of the 2015 ACM Sixth International Conference on Future Energy Systems*, ser. e-Energy '15. New York, NY, USA: ACM, 2015, pp. 179–188. [Online]. Available: http://doi.acm.org/10.1145/2768510.2768533

[32] S. J. Pappu, N. Bhatt, R. Pasumarthy, and A. Rajeswaran, "Identifying topology of low voltage (lv) distribution networks based on smart meter data," *IEEE Transactions on Smart Grid*, vol. PP, no. 99, pp. 1–1, 2017.

[33] D. Deka, S. Backhaus, and M. Chertkov, "Structure learning in power distribution networks," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 3, pp. 1061–1074, Sep. 2018.

[34] Z. Zhang, L. F. Ochoa, and G. Valverde, "A novel voltage sensitivity approach for the decentralized control of dg plants," *IEEE Transactions on Power Systems*, vol. 33, no. 2, pp. 1566–1576, March 2018.

[35] Y. Li and P. J. Wolfs, "Taxonomic description for western australian distribution medium-voltage and low-voltage feeders," *IET Generation, Transmission Distribution*, vol. 8, no. 1, pp. 104–113, Jan 2014.

[36] "Enwl009 - cable health assessment for low voltage cables," 2018. [Online]. Available: https://www.enwl.co.uk/innovation/smaller-projects/network-innovation-allowance/enwl009---cable-health-assessment-for-low-voltage-cables/

[37] "Innovation funding incentive annual report," 2015. [Online]. Available: https://www.scottishpower.com/userfiles/document_library/IFIAnnualReport2014-2015.pdf

[38] F. Provoost, F. Smits, and B. Jansen, "Fully automated calculations in both mv and lv networks," *CIRED - Open Access Proceedings Journal*, vol. 2017, no. 1, pp. 2304–2307, 2017.

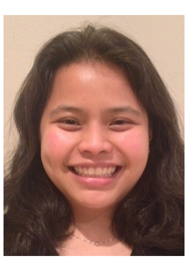
[39] M. Mokhtar, V. Robu, D. Flynn, C. Higgins, J. Whyte, and F. Fulton, "Automated verification of lv network topologies," in *2018 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, Oct 2018, pp. 1–6.

[40] D. Ablakovi, I. Dafi, R. A. Jabr, and B. C. Pal, "Experience in distribution state estimation preparation and operation in complex radial distribution networks," in *2014 IEEE PES General Meeting — Conference Exposition*, July 2014, pp. 1–5.

[41] L. Kaufman and P. J. Rousseeuw, *Finding groups in data: an introduction to cluster analysis*, 2nd ed., ser. Wiley Series in Probability and Statistics. Hoboken, NJ, USA: Wiley-Interscience, 2005.

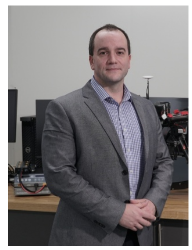
[42] "k-medoids clustering." [Online]. Available: https://uk.mathworks.com/help/stats/kmedoids.html

[43] H.-S. Park and C.-H. Jun, "A simple and fast algorithm for k-medoids clustering," *Expert Systems with Applications*, vol. 36, no. 2, Part 2, pp. 3336 – 3341, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S095741740800081X



Maizura Mokhtar Maizura Mokhtar is the data scientist for the Smart Metering team in SP Energy Networks. She is responsible for transforming the smart meter data into meaningful information, including identifying foreseeable risks to the LV networks and developing new techniques to enable their active management. She has a BEng (Hons) from Nottingham Trent University, an MSc (Eng) from the University of Sheffield, and a PhD from the University of York.

Valentin Robu Valentin Robu is an Associate Professor at Heriot-Watt University, UK, where he is the Co-Director of the Smart Systems research group. He is also a Research Affiliate in the Center for Collective Intelligence at Massachusetts Institute of Technology (MIT), US. He has published over 80 papers in top-ranked journals, conferences and edited volumes, in both the areas of Artificial Intelligence and Electrical Engineering. He is a Co-Investigator in several large-scale energy and AI-related projects, such as CESI (the UK National Centre for Energy Systems Integration), ORCA Hub (UK Offshore Robotics for Certification of Assets Hub), CEDRI (Community-Scale Energy Demand Reduction in India) and Reflex (Responsive Energy Flexibility demonstration project in Orkney Islands). He is Principal Investigator and academic lead of the Knowledge Transfer Partnership (KTP) project NCEWS (Network Constraints Early Warning Systems) with SP Energy Networks.



David Flynn Professor David Flynn is a director and founder of the Smart Systems Group (SSG) at Heriot-Watt University. The SSG are co-founders of the UKs EPSRC CESI, the worlds largest offshore robotics hub for the certification of assets (EPSRC, ORCA), as well as the UKs largest Whole System Energy Demonstrator project (Innovate UK, ReFELX). Prof. Flynn has over 130 publications and several international patents. Prof. Flynn received the BEng (Hons) in electrical and electronic engineering, the MSc (Hons) in microsystems, and the PhD in microscale magnetic components from Heriot-Watt University, Edinburgh, UK. He is an Institute of Engineering and Technology (IET) Scholar as a recipient of the IET Leslie H Paddle Prize.

Ciaran Higgins Ciaran Higgins is Director of Derryherk Ltd, who provided the software platform to convert SP Energy Networks' network data into a node-and-edge model. Ciaran has 20 years engineering experience, originally as an electronic engineer before moving over to the energy sector where his interest ranges from the modelling and deployment of low carbon technologies, particularly in urban environments, through to network analysis and the impact of controllable loads on the network to deliver network, carbon and customer benefit. He has an MEng in Electronics and Electrical Engineering from the University of Glasgow, an MSc in Renewable Energy and Distributed Generation from Heriot-Watt University and an MSc in Urban Design from the University of Strathclyde.

Jim Whyte Jim Whyte has recently move on from a 30 years career in a leading UK DNO, SP Energy Networks, part of the multinational energy group, Iberdola. Starting as an general operational engineer, his career then specialized in connection and design, before the last 10 years spent in Low Carbon transition led innovations. After leaving SP Energy Networks, he set up his own consultancy, Notsoanalytic. His speciality is in DNO data improvement and data analytical transformation requirements to facilitate future network management of the coming tsunami of Low Carbon Technologies.

Caroline Loughran Caroline has been working in the Utilities industry for 3 years and is currently a Smart Metering Consultant within the SP Energy Networks Smart Systems team. Having previous experience and expertise in the development and testing of SaaS (Software as a Service) tools, and managing the large volumes of data contained within these, Caroline is now responsible for analysing and reporting on the data received from Smart Meters. She also manages the toolset being developed to help understand the network using the influx of data from Smart Meters and LCT devices. Caroline has a BSc (Hons) in Mathematics from the University of Glasgow.

Fiona Fulton Fiona Fulton has over 25 years of experience in the development of telecoms, control and data systems, latterly focusing on smart energy systems. She has worked in a range of roles from technical design through to management of corporate-wide innovation strategies. She currently leads SP Energy Networks Smart Systems programme, focusing on smart meter enablement, as well as other data related innovation. She holds a MEng in Electronics, an MSc in Sustainable Energy Systems and is a Chartered Engineer.