Dirichlet Sampled Capacity and Loss Estimation for LV Distribution Networks with Partial Observability

Rory Telford, Bruce Stephen, *Senior Member, IEEE*, Jethro Browell, *Member, IEEE*, and Stephen Haben

***Abstract*—With low voltage (LV) distribution networks increasingly being re-purposed beyond their original design specifications to accommodate low carbon technologies, the ability to accurately calculate their actual spare capacity is critical. Traditionally, within the Great Britain (GB) power system, there has been limited monitoring of LV distribution networks, making this difficult. This paper proposes a method for estimating spare capacity of unmonitored LV networks using demand data from customer Smart Meters. In particular, the proposed method infers existing LV network capacity, as well as losses, across scenarios where only a limited number of customers have Smart Meters installed. Typical daily load profiles across customers with Smart Meters are learned using a Dirichlet sampled Gaussian mixture model (GMM). Learned profiles are then applied to all unmetered customers to estimate network parameters. Method accuracy is assessed by comparing estimations with simulated, fully observed, LV network models. The method is also compared to benchmark models for establishing unobserved demand profiles. Overall, results in the paper show that the proposed method outperforms benchmark models in terms of accurately assessing substation headroom, particularly in scenarios where only 10-50% of customers have Smart Meters installed.**

***Index Terms*—Maximum co-incident demand, radial feeders, distribution network losses**

I. Introduction

DISTRIBUTION networks at the low voltage (LV) level have traditionally featured minimal observability and, consequently, the true nature of loads in the *'last mile'* of the system is unclear. Conventionally, distribution network operators (DNOs) across the power system of Great Britain (GB) have estimated LV demand using a variety of methods, including: an After Diversity Maximum Demand (ADMD) procedure [1]; aggregating demand using a generic load profile model [2] and the use of other statistical means defined in the ACE49 standard [3]; or, the deployment of low-cost, low-accuracy maximum demand indicators.

However, the re-purposing of neighbourhood level distribution feeders to accommodate embedded generation and electrified heat and transport technologies has left many DNOs unclear as to how close this infrastructure is to its design limits [4]. Historical lack of monitoring of loads, uncertainty in generation penetrations and extreme heterogeneity of networks at this level of the power system complicate the issue further. In particular, challenges exist in determining how much remaining capacity is available under periods of peak demand — a problem borne out of the lack of understanding of loads at LV residential level, as well as how system losses may contribute to reducing total spare network capacity.

Across the GB power system, demand from residential/domestic customers accounted for approximately a third of the total energy consumption in 2019 [5], yet relatively little is known about these consumers [6]. The absence of LV network monitoring, coupled with quarterly customer billing, does not permit the accumulation of meaningful historical power-usage data. However, Advanced Metering Infrastructure (AMI), or *Smart Meters* [7], currently being rolled out to LV customers across GB provide utility companies with power-usage data at 30 minute intervals. These domestic power-usage readings could then be used by DNOs for improved LV network management and utilization.

It is, however, unlikely that there will be, at least in the near to mid-term future, complete coverage of Smart Meters across all LV customers, while the integration of extensive LV network monitoring would incur significant capital expenditure for a DNO. Accordingly, this paper proposes a method that uses only available Smart Meter consumer demand data to estimate remaining LV network capacity. The proposed method is based on a data driven strategy and uses a Dirichlet sampled [8, 9] Gaussian mixture model (GMM) [10] to model end-use demand from all LV customers who do not have a Smart Meter installed. The choice of using a Dirichlet sampled GMM is motivated by the assumption that multiple load sub-profiles exist within an LV network, but with a noisy composition. The Dirichlet distribution captures uncertainty over compositional variables; these are characterized by every coordinate in their sample space summing to a constant. Accordingly, the GMM mixture weights could be further sampled to determine the probability of an unobserved customer having a particular sub-load profile. This is one of the main contributions of this paper. Furthermore, the method estimates network losses, and their resultant impact on capacity, using only basic network information.

The method is assessed in terms of its capability to accurately estimate remaining capacity, even in scenarios where there is limited penetration of Smart Meters. Evaluation metrics are calculated after applying the proposed and benchmark methods to models of real LV networks. The only data required by DNOs to implement the proposed methodology are a limited sample of Smart Meters, and network Geospatial Information System (GIS) data, which is presently produced and stored by DNOs for asset management purposes. This is in contrast to previous research [11–13] attempting to characterize LV network parameters, which has generally relied on either extensive network monitoring or detailed information

R. Telford is with the Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, Scotland, G1 1RD e-mail: rory.telford@strath.ac.uk

Manuscript received XXXX

relating to network architectures and customer connectivity.

The paper is organized as follows: Section II proposes the data driven method for estimating network capacity with minimal Smart Meter observability. Section III reviews existing research and current practices for assessing remaining capacity and losses in LV networks. Section IV details the operational data sets and network models that comprise the case study against which the proposed method is evaluated and tested. Results of these tests are given in Section V along with commentary on the practicalities of implementing the proposed method as well as its value to DNOs. The paper is concluded in Section VI.

II. Estimation of LV Network Capacity and Losses

This section describes the proposed method for estimating remaining LV network capacity through modeling both peak end-use demand and associated technical losses within the system during times of peak demand. Traditionally there has been limited metering of LV networks due to limited operational risk and relatively slow-paced and predictable demand growth. The proposed method is outlined here in the context of LV networks within the GB system, although it may be applied to any unmonitored network of interest where only a percentage of connected customers have Smart Meters installed. A high level overview of the method is provided in Fig. 1. The method assumes a certain percentage of LV customers have Smart Meters installed, from which clusters of daily load profiles can be learned using a GMM. Unobserved customer demand is then estimated using these clusters and the categorical probabilities that an unobserved customer will be within each cluster, which are inferred through sampling of a Dirichlet distribution. Modeled end-use demand is then combined with estimated network characteristics, including feeder lengths and impedances, to infer network losses. Subsection II.A outlines general concepts of analytically determining LV network headroom on unmonitored LV networks, while the data-driven method for modeling unobserved customer end-use demand is described in Subsection II.B — this is the main contribution of this paper. Additionally, the method for approximating losses based on modeled customer end-use demand and feeder line length is described in Subsection II.C.

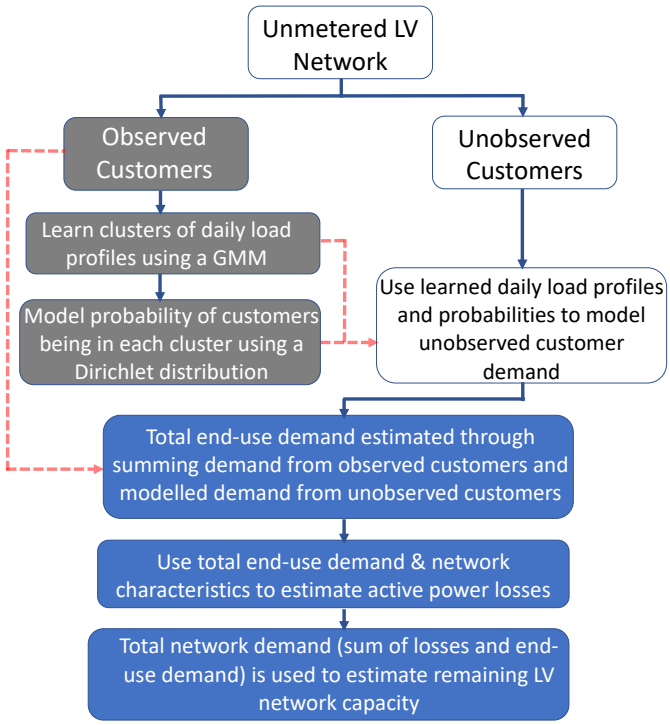

Fig. 1. Overview of the proposed method for estimating remaining LV network capacity and losses.

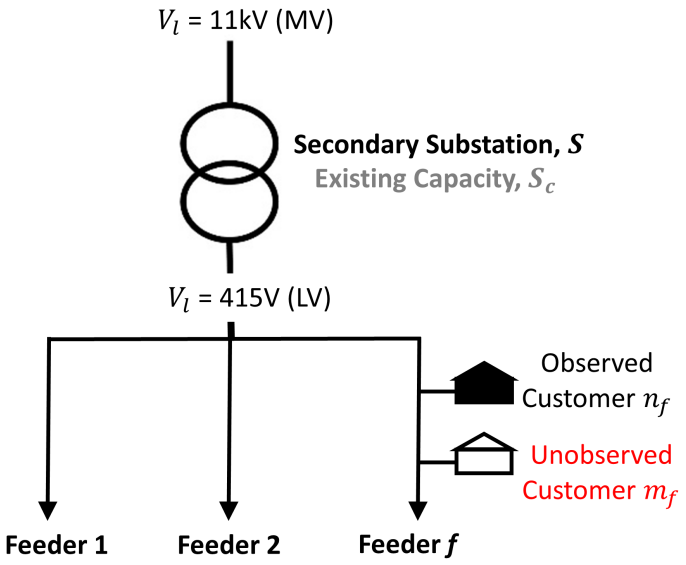

Fig. 2. General architecture of LV networks across the GB power system.

A. LV Networks in GB Power System

The general radial architecture of LV networks across GB is summarised in Fig. 2. From Fig. 2, the secondary substation S transforms line voltage V_l from 11kV medium voltage (MV) to 415V (LV). On the LV side of the secondary substation, radial three-phase feeders $f \in \{1, \ldots, F\}$, distribute power to residential and/or commercial/light industrial customers, where F is the total number of feeders. Note that, throughout this paper, LV network capacity refers to the total capacity of the secondary substation, S_c.

Assuming partial uptake of Smart Meters by LV customers, feeder f will have N_f observed customers (customers with Smart Meters) and M_f unobserved customers (customers without Smart Meters). The total active end-use demand, D_{fh}, on f at time h, is equivalent to

$$D_{fh} = \sum_{n=1}^{N_f} p^o_{nfh} + \sum_{m=1}^{M_f} p^u_{mfh}, \tag{1}$$

where p^o is observed demand, p^u is unobserved demand, p^o_{nfh} and p^u_{mfh} are demands for the n^{th} and m^{th} customers on feeder f respectively and $h \in \{1, \ldots, H\}$ where H is the total number of observed time-steps. The active load losses, L_{fh} of three-phase feeder f at time h can be estimated by

$$L_{fh} = 3\left(\sum_{k=1}^{T_p}\left(I_{ph} - \sum_{t=0}^{k-1} I_{ph}^{(t)}\right)^2 R_{pk}\right), \quad (2)$$

where I_{ph} is the single-phase current at the head of feeder phase p. In the absence of information regarding customer connections on feeder f, demand would be assumed to be evenly distributed across phases and I_{ph} can be estimated by

$$I_{ph} = (\frac{D_{fh}/3}{V_l/\sqrt{3}}). \quad (3)$$

Line resistance R_{pk} between customers k and $(k-1)$ is determined through

$$R_{pk} = Z_{pk} - Z_{p(k-1)}, \quad (4)$$

$$Z_{pk} = b_{fpk} R_u, \quad (5)$$

where b_{fpk} is the line length from the head of the feeder to customer k and R_u is the unit resistance of the feeder. T_p is the total number of customers connected to phase p and $k \in \{1, \ldots, T_p\}$. k is an ordered index, with $k = 1$ being the first customer connected along the radial phase length and $k = T_p$ being the final customer connection - this concept is illustrated in Fig. 3. $I_{ph}^{(k)}$, the current drawn from customer k, is determined through

$$I_{ph}^{(k)} = \frac{P_{ph}^{(k)}}{(V_l/\sqrt{3})}, \quad (6)$$

where $P_{ph}^{(k)}$ is the active power demand from customer k and $I_{ph}^{(0)} = 0$. If customer k is observed, $P_{ph}^{(k)}$ would be applied directly from Smart Meter data; however, if customer k is unobserved, $P_{ph}^{(k)}$ would be modeled according to the methodology summarised in Fig. 1 and described in detail in Section II.B. Note that Smart Meter data describes energy consumption in kWh at a particular time resolution, typically half-hourly [14]; half-hourly energy consumption data would be multiplied by a factor of two to derive average active power $P_{ph}^{(k)}$ across the observed time period.

Equation (2) accounts for the reduction of radial feeder current, using Kirchoff's law [15], in the calculation of active line losses - this concept, where current reduction along the feeder length is modeled as proportional to the upstream consumer current demand, is illustrated in Fig. 3. In the event that the location of each customer along each phase is unknown, the distance between each customer, and hence the resistance between them, would be assumed to be uniformly distributed along the LV feeder length and f_k would be determined through

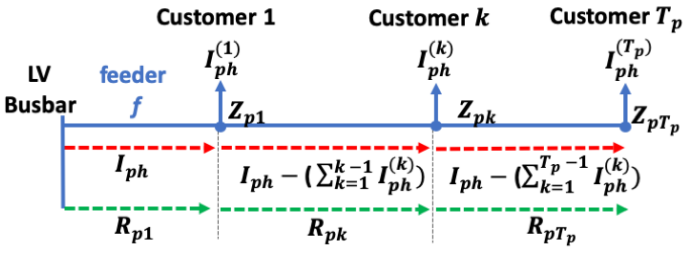

Fig. 3. Modeling the reduction of phase current along a radial LV feeder length.

$$b_{fpk} = \frac{f_{\text{length}}}{T_p} k, \quad (7)$$

where f_{length} is the total length of the feeder. Similarly, the location of customers with and without Smart Meters installed would be assumed to be randomly distributed along the feeder. Furthermore, network data relating to the types of cable used within a particular feeder, such as [16], are used to determine R_u.

Thus, the remaining network capacity C_{est} available at substation S with existing total capacity S_c can be estimated through

$$C_{\text{est}} = S_c - \max_h(\sum_{f=1}^{F}(D_{fh} + L_{fh})). \quad (8)$$

Central to the estimation of C_{est} in partially observed networks is modeling the demand of all unobserved customers, which will now be described.

B. Modeling Demand of Unobserved Customers

The data-driven method proposed in this paper for modeling unobserved feeder demand is summarised in Fig. 4. This methodology uses a Gaussian Mixture Model (GMM) to infer a set of P typical daily load profiles from the observed customers connected to a feeder f. The GMM mixture weights, π_τ, may be used to define the probability that an unobserved customer will have daily load profile $\mathbf{p} \in P$. However, this method proposes using the mixture weights as a hyper parameter of a Dirichlet distribution. The parametrized Dirichlet is then itself sampled to yield a categorical probability distribution, $\hat{\pi_\tau}$ for unobserved customers. The sampled distribution and learned daily profiles P are then used to synthesise a general daily load profile for each $m_f \in \{1, \ldots, M_f\}$. Observed load profiles are partitioned into weekday and weekend sets, W_1 and W_2 respectively, to train two separate GMMs for a feeder [17]. Aspects of using a GMM to learn daily load profiles and sampling from a Dirichlet distribution to compute the categorical distributions, are further described as follows.

1) Gaussian Mixture Model Training: The GMM is implemented to cluster daily load profiles that possess similar characteristics [18]. Within a GMM, each base distribution in the mixture is a multivariate Gaussian with mean $\boldsymbol{\mu_\tau}$ and covariance matrix Σ_τ and has the form

$$\phi(\mathbf{x_f}; \theta) = \sum_{\tau=1}^{T} \pi_\tau \mathcal{N}(\mathbf{x_f}; \boldsymbol{\mu_\tau}, \Sigma_\tau), \quad (9)$$

where T is the total number of cluster centroids and $\tau \in \{1, \ldots, T\}$; $\mathbf{x_f}$ are observed training variables, and the resulting likelihood is a weighted sum of each mixture component represented by mixing weights π_τ that satisfies $0 \leq \pi_\tau \leq 1$

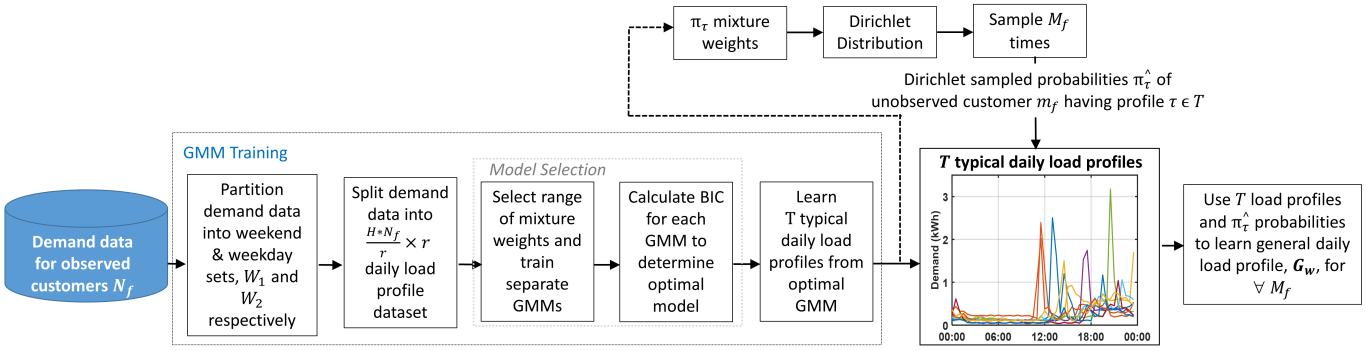

Fig. 4. Modeling end-use demand of unobserved customers connected to a LV feeder.

and $\sum_{\tau=1}^{T} \pi_\tau = 1$. The methodology proposed in this paper uses a GMM to cluster daily profiles across customers where demand is observed at r time steps throughout the day. To determine different clusters across all observed profiles, the observed data set is primarily processed into a training set of dimension $\frac{H \cdot N_f}{r} \times r$, where $\frac{H \cdot N_f}{r}$ is the number of days of training observations. Each cluster within the data is represented by a multivariate Gaussian of dimension r — for example, if demand is observed at 30 minute time resolution, each cluster would be an $r = 48$ dimension multivariate Gaussian.

The parameters of the GMM (mixture weights, means and covariances) are trained using the observations and the Expectation-Maximization (EM) algorithm [19]. The τ^{th} daily load profile is defined to be mean $\boldsymbol{\mu_\tau}$. A principle issue with GMMs concerns the fact that the number of mixture components (or clusters) on which to train the model has to be defined *a priori*. To overcome this, it is proposed that GMMs with a range of mixtures are defined and trained. The optimal model would then be selected through determining the model that minimises the Bayesian Information Criterion (BIC) [20] — BIC is used to select the model that optimises the number of parameters in terms of both model fit and complexity.

2) Sampling from the Dirichlet Distribution: The proposed method treats π_τ as a prior probability that is used to parametrize a Dirichlet distribution [8]. Dirichlet parameter, α, is represented by π_τ, which is a vector of length T describing the proportion of observed training cases attributed to mixture τ. The Dirichlet density function, given by

$$Dir(\upsilon; \alpha) = \frac{1}{B(\alpha)} \prod_{\tau=1}^{T} \upsilon_\tau^{\alpha_\tau - 1}, \tag{10}$$

where B is the multivariate Beta function [19], is then sampled M_f times to establish a compositional distribution, $\hat{\pi_\tau}$, for unobserved customers

$$\hat{\pi_\tau} \sim Dir(\upsilon; \alpha, w), \tag{11}$$

where $w \in \{W_1, W_2\}$. This process captures variance across the GMM mixture weights and considers error in the original estimate. A general daily load profile, $\mathbf{G_w}$, for each unobserved customer can then be determined from

$$\mathbf{G_w} = \sum_{\tau=1}^{T} (\hat{\pi_\tau} * \boldsymbol{\mu_\tau}). \tag{12}$$

This allows complete end-use demand, $\mathbf{D_f}$, across all unobserved and observed customers on feeder f to be modeled.

C. Modeling Feeder Losses

As outlined in Equations (2) to (6), load losses across partially observed feeders may be estimated by combining modeled end-use demand with feeder topology and electrical circuit data. This subsection summarises a method for inferring feeder line length based on feeder area (or footprint) — estimation of line length is then used to calculate feeder load losses. This method assumes that DNOs will have basic information relating to spatial characteristics of the LV networks that they operate, as well as the types of cable used within each feeder, which is available within GIS databases, and the associated cable parameters.

As an example, Fig. 5 illustrates spatial parameters of an LV feeder extracted from GIS information of an LV feeder within a distribution network in the North West of England, which is part of the GB power system. Specifically, Fig. 5 highlights the feeders' maximum and minimum lateral ($X_{\max}, X_{\min}$) and longitudinal ($Y_{\max}, Y_{\min}$) distances. The basic method proposed for inferring total length of the feeder, $\hat{f_{\text{length}}}$, using these four spatial points is defined as

$$\hat{f_{\text{length}}} = \max\left\{(X_{\max} - X_{\min}), (Y_{\max} - Y_{\min})\right\}. \tag{13}$$

Other methods, including the square root of the sum of squares, were heuristically evaluated. However, after comparison, inferring line length as being equivalent to either the maximum lateral or longitudinal distance, as outlined in Equation (7), provided an accurate estimation across a diversity of feeders. Inferred feeder length, $\hat{f_{\text{length}}}$, is then applied to Equation (5) to estimate total line resistance to each customer connection, and thus model feeder losses under certain loading conditions.

III. Existing Models of Unobserved Load & Network Losses

The previous section proposed a method for estimating unobserved end-use customer demand and network losses, and discussed how these can be used for estimating LV network capacity. This section summarises previous research and methods that have attempted to characterize these aspects. Benchmark models for estimating unobserved loads are also introduced. The effectiveness of these models in estimating LV network capacity is compared to the proposed methodology in Sections 4 and 5.

A. Generalisation of Load Profiles

A number of prior works have recognised the need for accommodating heterogeneity in end use load profiles irrespective of the level of the network at which they occur [21, 22]. The use of models that cluster load data to recover these sub-profiles has become commonplace: Chicco *et al.* [21] utilized k-Means clustering on particular features of load data in a similar manner to [23]. Giasemidis *et al.* [24] proposed a method that assigned unmonitored customers load profiles from a sample of smart meter profiles based on similar energy usage and socio-demographics. Stephen *et al.* [18] utilized the raw daily load profile as a high dimensional variable, with 48 half hourly load measurements comprising each dimension. This high dimensional space was then partitioned using a GMM, which has the added advantage over k-Means in that it does not just recover the sub-profiles inherent in the data as a set of mean vectors, but also captures their corresponding variances and the proportion of times in the sample in which they occur.

In terms of combining Dirichlet processes with clustering models, as proposed in this paper, Power *et al.* [9, 25] developed a model for synthesising solar generation and demand data by sampling feature clusters, established using a k-means model, with a Dirichlet process. In particular, [9] noted that

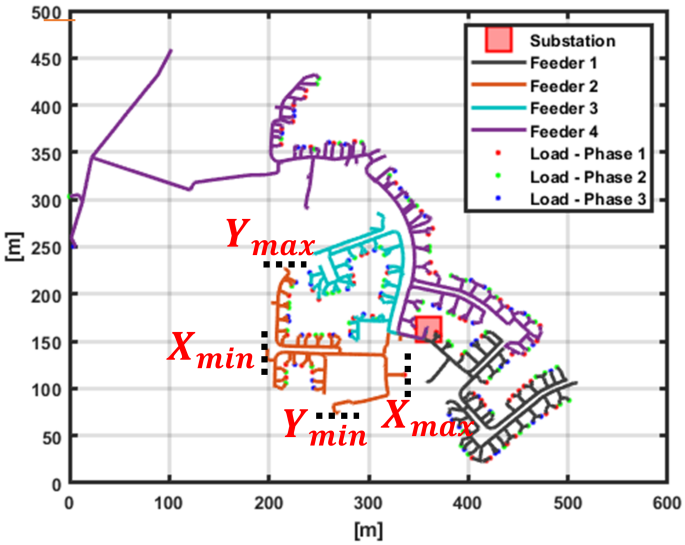

Fig. 5. Plot of LV network in GB using GIS data.

while the proportions of any observable customers having certain characteristics in any data set can be known, this may not be a reliable estimate of the true proportions across all observed and unobserved customers.

1) Relation of LV Load to Weather & Socio-demographics: the effect of weather on electrical load is not fully understood, with recent papers examining the impact [26] via forecasting methodologies showing that the relationship may not be as strong (or causal) at the local residential level as it is at national and larger regional levels. Further to this, other papers have shown that energy demand is not strongly correlated to socio-demographics [27] and other non-energy characteristics, e.g. house size [28]. While correlations exist between certain characteristics and demand behavior, previous research has concluded that they tend to be weak and do not describe intra-day behaviors [29, 30].

B. Estimation of Network Headroom and Feeder Losses

Previous methods proposed for estimating LV network operational parameters tend to rely on the availability of extensive network topology and connectivity information [12, 13] as well as detailed customer metering data [31]. For example, Chen *et al.* [32] trained an artificial neural network (ANN) to model feeder losses. However, this multi-stage methodology primarily uses fully observed customer demand to run power flow models — customer demand data and simulated losses are then used to train an ANN. Urquhart *et al.* [11, 33] circumvent the requirement for detailed network topology information, although their method does involve additional network metering at both upstream and downstream locations. Verderhol [12] generalised LV losses through detailed modeling and simulation of representative LV networks and [13] developed a loss model based on clustering feeders with similar line length and connected customers.

Various research has assessed the impact that the integration of renewable energy will have on the electrical distribution system [34, 35]. In terms of estimating remaining capacity at LV substations, Li *et al.* [36, 37] developed representative LV network templates for an area within the GB system to provide an indicative view of power flows at any given substation without the need for monitoring. Lee [38] used monitoring at select locations to learn key features of value to a DNO — learning was extended to predict features, including substation headroom, at unmonitored substations.

C. Benchmark Models

This section summarises two benchmark models for estimating LV customer end-use demand that are used to compare against the method proposed in this paper. These two models include.

1) Localised Load Profile Averaging: In the presence of available Smart Meters, it makes sense to utilize their data to inform local load characteristics with the assumption that premises in a particular area will be of similar construction and occupation behavior. This model averages half-hour advances across all observed customers, and days, to create a generic daily load profile.

2) Typical Load Profiles: System operators often use a typical daily load profile to describe a particular end use customer class [2]. At LV level, a typical seasonal profile is provided — a different profile is attributed to customers that are assumed to have electrified heating installed. While these general profiles will not necessarily resemble an exact customer profile, it will accommodate the peaks expected in their behavior. Using a typical profile it is assumed that load diversity can be accounted for.

IV. Evaluation of Capacity & Loss Estimation

This section describes the process of testing the proposed method's ability to estimate LV network capacity and losses. This process centered around the comparison of estimations with deterministic spare capacity and loss results obtained from simulated, fully observed, LV network models. The main stages included:

- **Simulation of 25 fully observed networks**: this represents the ideal, although impractical, case where all LV customers are monitored. The simulations established the ground truth of the remaining available capacity and associated losses for each network.
- **Reduction of customer observations**: the number of observed customers for each network was reduced iteratively to 80%, 50%, 20% and 10% respectively. A business as usual (BAU) scenario, where 0% of customers are monitored, was also developed. The demand from unobserved customers at each level of reduction was modeled using the method proposed in this paper, as well as the two benchmark methods described in Section III.C, with the exception of the BAU scenario where demand was modeled using only the Typical Profile model. Estimated remaining capacity was calculated through a combination of the observed and the modeled unobserved load.
- **Error analysis**: various metrics were calculated by comparing the modeled results with the simulated results to establish the accuracy of the different methods for estimating capacity and loss parameters in partially observed LV networks.

These aspects are described in further detail in the following sub-sections.

A. Distribution Network Models & Simulation

In order to test the capacity and loss estimation techniques, models of secondary substation LV networks from an actual distribution network in the surrounding area of North West England were developed using network and GIS data [39]. 25 different networks were modeled and simulated using openDSS software [40]. The generic architecture of each network is summarised in Fig. 2, with the number of three-phase feeders across the 25 substations ranging between 2 and 9. In total, 110 LV feeders across the 25 networks were modeled - these feeders have previously been characterized in [41]. Network model data and parameters, including existing substation capacities, can be accessed from [42]. Fig. 5 illustrates a 4-feeder LV network modeled in openDSS, with individual single-phase loads (customers) connected at various points along each 3-phase feeder. There is a variety of customer numbers connected to each feeder, ranging between 15 and 220. Power flows for each network were simulated at half-hour advances under a scenario where customer end-demand is fully observed. Customer demands were populated across each network model using Smart Meter data.

1) Smart Meter Dataset: A Smart Meter dataset comprising half-hourly demand measurements across 283 separate residential customers [43] was used to populate each network model. This data was recorded across 84 consecutive days between January and March 2010 in the Republic of Ireland, and thus comprised 4032 half-hour advances for each customer. Individual customers across each feeder were randomly assigned a unique demand profile from within the dataset.

2) Network Simulation: Simulation of each fully observed network model enabled the following features to be determined:

i) half-hourly simulated power flows across the 84 day period, $\mathbf{f}_{\text{sim}}$
ii) feeder peak total demand, including losses, f_{peak}
iii) actual feeder line length, f_{length}
iv) and, total feeder line losses across the 84 day period, f_{loss}

These provide the ground truth for the studies on feeder demand and loss analysis that follow.

B. Customer Observation Scenarios

Scenarios of specific levels of observed customers on each LV feeder were used to assess accuracy of capacity and loss estimation. Specifically, four separate scenarios relating to customers that have Smart Meters installed were developed, with demand from the remaining, unobserved customers, modeled using the method described in Section 2: to fully test accuracy across a range of cases, scenarios of 10%, 20%, 50% and 80% of customer observability were used across each of the 110 LV feeders. Note that, to ensure that there is sufficient observations to train the GMM, a minimum of two customers have to be observed. Hence, in LV feeders that have between 15 and 19 customers, the 10% observability scenario was actually slightly increased. To implement this, observed customers within each scenario were assigned a subset of the unique smart meter profiles assigned to each feeder customer during simulation of the fully observed scenario, as described in the previous subsection.

1) Business as Usual Scenario: The BAU scenario assumed that no Smart Meters were installed on a feeder. This amounts to zero observability, meaning that the method proposed in this paper, and the average benchmark model, did not apply - all premises in the BAU case therefore had load represented by the Typical Load Profile model only. This scenario provides the base case against which any penetration of Smart Meters can be justified.

2) GMM Training & Model Selection: As discussed in Section II, GMMs with different numbers of mixture components were trained for each feeder (and under each observation scenario), and the BIC was used to select the optimal model [17, 18]. Each GMM was trained using 30 days of half-hourly

demand data for each observed customer and mixtures ranging from 2 to 70. The optimal model for each case was determined by the number of mixture components that minimised the BIC.

To compare estimation accuracy with that of other methods, the benchmark models, described in Section III, were also used to model load from remaining unobserved customers across each scenario.

C. Metrics of Accuracy and Effectiveness

Three criteria were used to assess estimation performance. These were as follows: a comparison of estimated peak feeder demand with simulated peak feeder demand for each of the 110 LV feeders; a comparison of actual feeder line length with inferred feeder line length for each feeder; comparison of estimated existing network (substation) capacity with simulated existing capacity.

In the case of peak feeder demands, percentage error, $\mathcal{E}_{\text{peak}}$, defined as

$$\mathcal{E}_{\text{peak}} = \frac{\max_h(D_{fh}) - f_{\text{peak}}}{f_{\text{peak}}} \times 100\%, \tag{14}$$

was calculated for each feeder and under each customer observation scenario, where $h \in \{1, \ldots, H\}$, $H = 4032$ and D_{fh} is defined in Equation (1). In instances where $\mathcal{E}_{\text{peak}}$ is negative, peak feeder demand has been underestimated, while a positive $\mathcal{E}_{\text{peak}}$ indicates an overestimated peak feeder demand.

Similarly, the accuracy of inferred feeder line length, which is an important metric for determining copper losses, is defined as a percentage of actual feeder length

$$\mathcal{E}_{\text{length}} = \frac{\hat{f_{\text{length}}} - f_{\text{length}}}{f_{\text{length}}} \times 100\%. \tag{15}$$

Accuracy of estimated feeder losses throughout the 84 day period is compared with simulated energy losses, using

$$\mathcal{E}_{\text{loss}} = \frac{(\sum_{h=1}^{H} L_{fh}) - f_{\text{loss}}}{f_{\text{loss}}} \times 100\%, \tag{16}$$

where L_{fh} is defined in Equation (2).

Remaining network capacity is estimated through inference of maximum co-incident demand across all feeders connected to a substation. A percentage error, $\mathcal{E}_{\text{capacity}}$, that compared estimated capacity, C_{est} with simulated spare capacity, C_{sim} was used to assess accuracy, and is defined as

$$\mathcal{E}_{\text{capacity}} = \frac{C_{\text{est}} - C_{\text{sim}}}{C_{\text{sim}}} \times 100\%, \tag{17}$$

where C_{est} was outlined in Equation (8) and C_{sim} is

$$C_{\text{sim}} = S_c - \max(\sum_{f=1}^{F} \mathbf{f}_{\text{sim}}). \tag{18}$$

This metric was calculated across all 25 modeled networks, where F is the total number of individual feeders connected to each network: a positive value of $\mathcal{E}_{\text{capacity}}$ indicates an overestimation of remaining network capacity while a negative value indicates an underestimate.

V. Results & Analysis

The three metrics described in Section IV.C were used to assess accuracy of the proposed method, as well as compare performance with benchmark models, which were described in Section III.C. Results of this analysis are presented in this section. Methods were assessed and validated across a total of 25 modeled real LV networks. The operational benefit of improving estimation of existing network capacity is also discussed.

A. Estimation of Feeder Peak Total Demand

Fig. 6 compares the error distribution of estimated total peak feeder demand across different customer observation scenarios for the proposed Dirichlet sampled GMM as well as the average and typical profile benchmark models. From Fig. 6, the median negative errors of the proposed method indicate that peak feeder demands tend to be underestimated across all observation scenarios. As the percentage of observed customers on feeders increase, the variance of errors reduces around zero. Under the 10 and 20% scenarios of observed customers, results in Fig. 6 highlight that the median percentage error of estimates of the proposed method is superior to the benchmark models. For example, the mean percentage error of the Dirichlet sampled GMM under 10% customer observation is -6% in comparison to -15% for the average model and -19% for the typical profile model. As observability reached 80% of customers, the magnitude of the mean errors across all three models converged. However, the mean error of the average and typical profile models was positive, indicating an overestimated peak demand. Minimal outlying errors are produced by the Dirichlet sampled GMM in comparison to both benchmark models, emphasising the potential reduced level of risk involved with basing operational decisions on the outputs of the proposed method.

Under BAU observability, headroom estimates with a Typical Profile Model produced a median error of -25%. Intuitively,

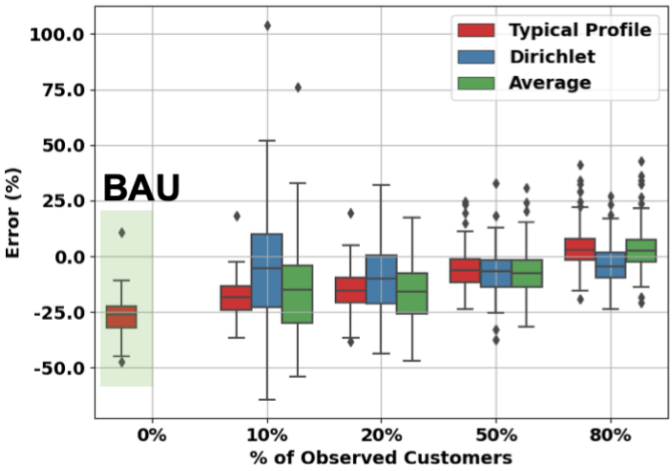

Fig. 6. Error Distribution of estimated peak feeder demands using a: Dirichlet sampled GMM; Average Model; and, a typical profile model. The typical profile model was also used to model customer demand across a BAU scenario where 0% of customers are observed.

feeder peak estimate errors improve as the number of observed customers installed Smart Meters increase, highlighting the benefit of greater observability, even at limited penetrations, to model peak feeder demands.

B. Estimation of Feeder Length & Line Losses

LV feeder cable assets are numerous, making explicit knowledge of their exact lengths impractical to obtain. Section II.C elaborated upon an approach to estimating cable lengths from spatial GIS data that network owners will hold as a matter of operational course. Fig. 7 (a) illustrates the distribution of feeder length estimation errors, A_L, across the 110 LV feeders. Overall, the basic method outlined in Section II.C, estimated 42% of the LV feeders to be within +/-25% of the actual value defined in the network models. The majority of errors (45%) were positive, meaning that feeder line length, and hence line impedance, is generally overestimated.

Figs. 7 (b) and (c) compare distributions of estimated feeder loss errors for the three models throughout the 84 day analysis period under a scenario of 10% feeder customers observability. Error distributions shapes are generally consistent between estimation models and also, pertinently, with feeder length estimate errors, indicating that loss errors are driven through misrepresentation of feeder line lengths as opposed to inaccuracies in estimating consumer end-use demand. Fig. 7 (d) shows the distribution of loss estimation errors calculated using 100% end-use demand observability and the inferred feeder length - while outlying errors are slightly reduced, the general shape of the distribution highlights that loss errors are driven by feeder length estimation errors. Statistically, the GMM/Dirichlet model had a mean estimate error of -10.8% with a standard deviation of 65%, in comparison to -1.3% and 62% respectively for the typical profile model.

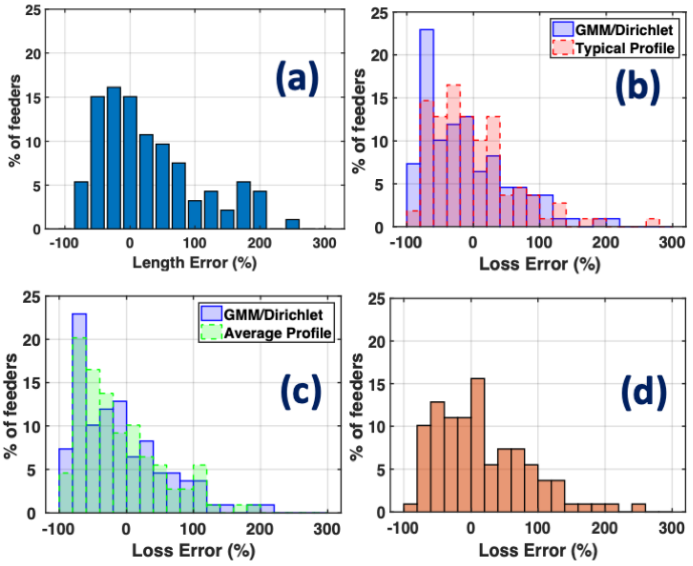

Fig. 7. (a) Distribution of inferred feeder length errors across 110 separate LV feeders (b) comparison of feeder loss estimate error distributions between GMM/Dirichlet model and average model (c) GMM/Dirichlet model and typical profile model and (d) loss errors when calculating with fully observed end-use demand and inferred feeder length. Loss error distributions are similar for each model, suggesting error is mainly driven by feeder length estimate errors as opposed to estimations of end-use demand.

C. Estimation of Remaining LV Network Capacity

Results of estimated LV network capacity using the Dirichlet sampled GMM across the 25 modeled networks are provided in Fig. 8. This figure highlights both the percentage error estimation under each customer observation scenario as well as how often the proposed method out performs the benchmark models — in cells shaded black, either the average profile or typical profile benchmark model have provided a more accurate estimation. Across the 100 considered cases (4 observation scenarios for each of the 25 networks), the proposed method was superior 56% of the time. In scenarios where 10–20% of network customers are observed, the method also outperforms the two benchmark models 60% of the time. In particular, in instances where only 20% of customers have Smart Meters installed, the method has improved accuracy in 16 out of 25 (64%) of cases. The magnitude of error under the minimal 10% observation scenario does vary, with errors as low as -0.4% (Network 9) to outliers of 97.3% (Network

TABLE I
% OF CASES WHERE PROPOSED METHOD OUTPERFORMS THE BENCHMARK MODELS.

Benchmark	10% Observation	20% Observation	50% Observation	80% Observation
Missing filled with Average Model	80	92	84	44
Missing filled with Typical Profile	60	64	68	36
Both	56	64	68	36

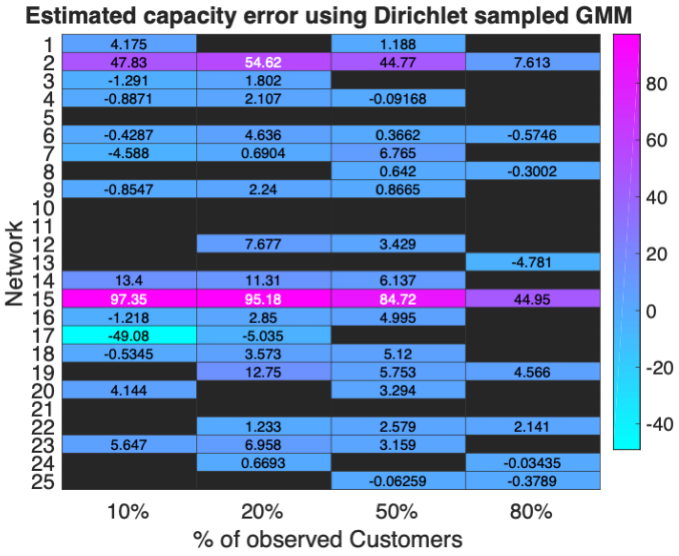

Fig. 8. Heat map of estimated existing network capacity error using a Dirichlet sampled GMM. Black shaded boxes are scenarios where the proposed methodology has been outperformed by either the Average or Typical Profile benchmark models — this mainly occurs as the percentage of observed customers increases.

15) — in practical terms, the utilisation of such models for capacity estimation should factor in uncertainty, and there may be instances where ensembles of different models may be used to reduce estimation error.

Table 1 summarises these results, as well as outlining comparisons of the proposed method with the individual benchmark models. Individual comparisons show that the proposed method is significantly superior at estimating network capacity than the Average benchmark model. In comparison to the Typical Profile model, the proposed method is more accurate in at least 60% of the test cases across each of the 10-50% customer observability scenarios.

Results in Table 1 and Fig. 8 emphasise that application of the proposed method improves accuracy of LV network capacity estimations, particularly in cases where only limited customers have Smart Meters installed.

D. Operational Benefit

Distribution networks are changing beyond their original design specification. Low carbon technologies are redefining the magnitude and diversity of peak loads resulting in capacity being reduced to critical level and thermal limits approached. Owing to the scale of distribution network infrastructure, monitoring feeders to anticipate this is prohibitively costly leaving asset owners unclear as to where reinforcement may be required in the near future. Smart Meters only have the potential to inform part of the headroom picture with penetrations being less than 100 percent and in remote rural areas, where electrification of heat is already challenging some infrastructure, penetrations are minimal. Going forward, network owners will ultimately favour approaches that do not commit them to the additional expense of monitoring to inform network investment need and will require analytics to leverage the value of existing monitoring equipment. Utilising what data is available to inform remaining headroom is therefore attractive to network owners to undertake broad brush assessments of asset reinforcement programmes. The contributed method is agnostic to network topology and load shape making it ideally suited for the heterogeneity encountered on distribution feeders. Smart Meter data and GIS are both routinely held digital assets meaning that a pipeline into a cloud based implementation of the proposed method could automatically assess all of a network owners LV distribution feeders for the need for reinforcement on a regular basis, possibly ahead of the end of each financial year, ensuring upgrades are appropriately budgeted for and assets are operated within their limits.

The method has been outlined within this paper using residential Smart Meter data. Operationally, LV network management should also consider other types of customer classes. Going forward, Smart Meter data from a variety of customer types would be included and combined with network operators' prior knowledge on the customer composition of each network to further improve accuracy of estimating network capacity and associated losses.

Furthermore, the relative features of the LV network should be accounted for. The studies outlined in the paper have focused on residential areas in which a feeder would typically service a large number of smaller loads. However, within urban areas, there is a greater likelihood that residential tower blocks and office complexes will be powered by a dedicated LV feeder (or, even, substation). Additional testing of the proposed methodology that accounts for such areas should therefore be undertaken as relevant data becomes available.

The method has been presented here in the context of utilizing half-hourly load measurements. Although Smart Meters do record demand at significantly higher resolution, there are various factors, including regulatory restrictions on DNO access and issues with data collection and storage that meant higher resolution load data was not considered. Higher resolution data, while useful for capturing power quality events and transients, does not sufficiently enhance capacity and loss estimations across all LV feeders within a DNOs remit.

The paper has focused on estimation of active power losses, with the assumption that consumed (or apparent) power is active power; however, capturing of consumers' reactive import data, a standard of modern Smart Meter technology, will improve DNOs knowledge on the distribution of active and reactive losses in particular areas of the network. This aspect will be a focus of future work.

VI. Conclusion

This paper has contributed an analytics driven means of harnessing incomplete Smart Meter data and using it to infer the nature of missing demand measurements at LV level. In particular, the paper outlined a method, based on a Dirichlet sampled GMM, of characterising typical end-use customer demand profiles inherent within observed customers. The proposed method, unlike existing approaches in literature, does not require LV network monitoring to assess system losses and only uses basic spatial information relating to network architecture. Models of unobserved customer demand enables estimation of existing capacity and losses across unmetered LV networks.

The paper assessed accuracy of the proposed method by comparing estimated capacity and losses with those obtained deterministically through simulation of 25 separate models of real LV networks within the north west of England – customer demand within these models was populated using Smart Meter data [43]. The method was also assessed against two benchmark models for modeling unobserved LV customer demand. Overall, the proposed method was shown to estimate LV capacity with more accuracy than both benchmark models, including in 60% of the cases where 20% or less customers have Smart Meters installed. The results also showed that, when 50% or less customers have Smart Meters installed, the proposed method increased accuracy in 64% of the test cases in comparison to the Typical Profile model, which is the industry standard in GB. The introduction of low carbon technologies has increased uncertainty with regards to operation and management of LV networks, and this paper has contributed a method that can improve forward visibility of potential issues through use of existing customer Smart Meters. The fact that the method is superior to other methods when only limited customers have Smart Meters installed, as is

the case today, highlights its potential use to network operators in the short-term.

ACKNOWLEDGMENTS

This work is supported by the EPSRC via the Analytical Middleware for Informed Distribution Networks (AMIDiNe) project (EP/S030131/1). Jethro Browell is supported by an EPSRC Innovation Fellowship (EP/R023484/1). **Data Statement:** Underlying LV network and Smart Meter data can be accessed via [42] and [43], respectively.

REFERENCES

[1] C. Barteczko-Hibbert, "After diversity maximum demand (admd) report," *Report for the 'Customer-Led Network Revolution' project: Durham University*, 2015.

[2] Elexon, "Profiling," https://www.elexon.co.uk/operations-settlement/profiling/.

[3] Energy Networks Association, "Report on statistical method for calculating demands and voltage regulations on lv radial distributions systems," 1981.

[4] "Flexible Networks for a Low Carbon Future - Work Package 1: Detailed Network Monitoring ," Scottish Power Energy Networks, Tech. Rep., 2015, https://www.spenergynetworks.co.uk/userfiles/file/Detailed_Network_Monitoring_Methodology_and_Learning_Report.pdf.

[5] "Energy Trends: UK electricity," UK Government: Department for Business, Energy Industrial Strategy, Tech. Rep., 2012, https://www.gov.uk/government/statistics/electricity-section-5-energy-trends.

[6] P.-H. Li, I. Keppo, M. Xenitidou, and M. Kamargianni, "Investigating uk consumers'heterogeneous engagement in demand-side response," *Energy Efficiency*, 2020.

[7] Department for Business Energy and Industrial Strategy, "Smart Metering Implementation Programme - progress report for 2018," https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/767128/smart-meter-progress-report-2018.pdf.

[8] T. Minka, "Estimating a dirichlet distribution," 2000.

[9] T. Power and G. Verbic, "A nonparametric bayesian model for forecasting residential solar generation," in *2017 Australasian Universities Power Engineering Conference (AUPEC)*, Nov 2017, pp. 1–6.

[10] D. P. G. McLachlan, S.Ng, "On clustering by mixture models," *Exploratory Data Analysis in Empirical Research. Studies in Classification, Data Analysis, and Knowledge Organization*, pp. 141–148, 2003.

[11] A. Urquhart, M. Thomson, and C. Harrap, "Accurate determination of distribution network losses," *CIRED - Open Access Proceedings Journal*, vol. 2017, no. 1, pp. 2060–2064, 2017.

[12] M. I. Verdelho, R. Prata, P. Carvalho, and J. Machado, "Impact of pv distributed generation on edp distribuição lv grid losses," *CIRED - Open Access Proceedings Journal*, vol. 2017, no. 1, pp. 2342–2345, 2017.

[13] A. K. Dashtaki and M. R. Haghifam, "A new loss estimation method in limited data electric distribution networks," *IEEE Transactions on Power Delivery*, vol. 28, no. 4, pp. 2194–2200, Oct 2013.

[14] "Smart Metering Equipment Technical Specifications 2 Version 3.1," UK Government: Department for Business, Energy Industrial Strategy, Tech. Rep., 2018.

[15] C. C. B. Oliveira, N. Kagan, A. Meffe, S. Jonathan, S. Caparroz, and J. L. Cavaretti, "A new method for the computation of technical losses in electrical power distribution systems," in *16th International Conference and Exhibition on Electricity Distribution, 2001. Part 1: Contributions. CIRED. (IEE Conf. Publ No. 482)*, vol. 5, 2001, pp. 5 pp. vol.5–.

[16] Electricty North West Ltd., "LV Capacitor Placement Study, Table C1," 2013, https://www.enwl.co.uk/globalassets/innovation/lv-busbars/lv-busbars-closedown/lv-busbars-appendix-18.pdf.

[17] B. Stephen, X. Tang, P. R. Harvey, S. Galloway, and K. I. Jennett, "Incorporating practice theory in sub-profile models for short term aggregated residential load forecasting," *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1591–1598, July 2017.

[18] B. Stephen, A. J. Mutanen, S. Galloway, G. Burt, and P. Järventausta, "Enhanced load profiling for residential network customers," *IEEE Transactions on Power Delivery*, vol. 29, no. 1, pp. 88–96, Feb 2014.

[19] K. P. Murphy, *Machine Learning*. The MIT Press, 2012.

[20] G.Mclachlan and D.Peel, *Finite Mixture Models*. Wiley, 2000.

[21] G. Chicco, R. Napoli, F. Piglione, P. Postolache, M. Scutariu, and C. Toader, "Load pattern-based classification of electricity customers," *IEEE Transactions on Power Systems*, vol. 19, no. 2, pp. 1232–1239, May 2004.

[22] J. Kwac, J. Flora, and R. Rajagopal, "Household energy consumption segmentation using hourly data," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 420–430, 2014.

[23] A. Mutanen, M. Ruska, S. Repo, and P. Jarventausta, "Customer classification and load profiling method for distribution systems," *IEEE Transactions on Power Delivery*, vol. 26, no. 3, pp. 1755–1763, July 2011.

[24] G. Giasemidis, S. Haben, T. Lee, C. Singleton, and P. Grindrod, "A genetic algorithm approach for modelling low voltage network demands," *Applied Energy*, vol. 203, pp. 463 – 473, 2017.

[25] T. Power, G. Verbič, and A. C. Chapman, "A nonparametric bayesian methodology for synthesizing residential solar generation and demand data," *IEEE Transactions on Smart Grid*, vol. 11, no. 3, pp. 2511–2519, 2020.

[26] S. Haben, G. Giasemidis, F. Ziel, and S. Arora, "Short term load forecasting and the effect of temperature at the low voltage level," *International Journal of Forecasting*, vol. 35, no. 4, p. 1469–1484, Oct 2019.

[27] S. Haben, M. Rowe, D. V. Greetham, P. Grindrod, W. Holderbaum, B. Potter, and C. Singleton, "Mathematical solutions for electricity networks in a low carbon future," in *22nd International Conference and Exhibition on Electricity Distribution (CIRED 2013)*, 2013, pp. 1–4.

[28] J. Morley and M. Hazas, "The significance of difference: Understanding variation in household energy consumption," in *ECEEE Summer School*, 2011, pp. 2037–2046.

[29] F. McLoughlin, A. Duffy, and M. Conlon, "Characterising domestic electricity consumption patterns by dwelling and occupant socio-economic variables: An irish case study," *Energy and Buildings*, vol. 48, pp. 240 – 248, 2012.

[30] "Household electricity survey: A study of domestic electrical product usage," Intertek, Tech. Rep., 2012.

[31] G. Poursharif, A. Brint, M. Black, and M. Marshall, "Using smart meters to estimate low-voltage losses," *IET Generation, Transmission Distribution*, vol. 12, no. 5, pp. 1206–1212, 2018.

[32] C. S. Chen, C. H. Lin, M. Y. Huang, H. D. Chen, M. S. Kang, and C. W. Huang, "Development of distribution feeder loss models by artificial neural networks," in *IEEE Power Engineering Society General Meeting, 2005*, June 2005, pp. 164–170 Vol. 1.

[33] A. J. Urquhart and M. Thomson, "Impacts of demand data time resolution on estimates of distribution system energy losses," *IEEE Transactions on Power Systems*, vol. 30, no. 3, pp. 1483–1491, May 2015.

[34] H. Al-Saadi, R. Zivanovic, and S. F. Al-Sarawi, "Probabilistic hosting capacity for active distribution networks," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 5, pp. 2519–2532, Oct 2017.

[35] T. Aziz and N. Ketjoy, "PV penetration limits in low voltage networks and voltage variations," *IEEE Access*, vol. 5, pp. 16 784–16 792, 2017.

[36] R. Li, C. Gu, F. Li, G. Shaddick, and M. Dale, "Development of low voltage network templates—part i: Substation clustering and classification," *IEEE Transactions on Power Systems*, vol. 30, no. 6, pp. 3036–3044, Nov 2015.

[37] ——, "Development of low voltage network templates—part ii: Peak load estimation by clusterwise regression," *IEEE Transactions on Power Systems*, vol. 30, no. 6, pp. 3045–3052, Nov 2015.

[38] T. E. Lee, "Predicting key features of a substation without monitoring," *Mathematics-in-Industry Case Studies*, vol. 8, no. 1, p. 3, Aug 2017. [Online]. Available: https://doi.org/10.1186/s40929-017-0013-z

[39] A. Navarro, L. Ochoa, R. Shaw, and D. Randles, "Reconstruction of low voltage networks: From gis data to power flow models," in *23rd International Conference on Electricity Distribution CIRED 2015*, 6 2015, pp. 1–5.

[40] Electric Power Research Institute, "What is opendss?" https://www.epri.com/#/pages/sa/opendss?lang=en.

[41] V. Rigoni, L. F. Ochoa, G. Chicco, A. Navarro-Espinosa, and T. Gozel, "Representative residential lv feeders: A case study for the north west of england," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 348–360, 2016.

[42] Electricity North West, "Low Voltage Network Solutions," https://www.enwl.co.uk/zero-carbon/innovation/smaller-projects/low-carbon-networks-fund/low-voltage-network-solutions/ Accessed on 05.2020.

[43] Commission for Energy Regulation (CER)., "CER Smart Metering Project - Electricity Customer Behaviour Trial, 2009-2010 [dataset]," 2012, http://www.ucd.ie/issda/data/commissionforenergyregulationcer/.

Rory Telford received the BEng, MSc and PhD degrees in electronic and electrical engineering in 2008 and 2011, and 2017 respectively, from the University of Strathclyde, Glasgow, U.K. He is currently a Research Associate within the Institute for Energy and Environment, University of Strathclyde. His research interests include application of AI techniques, data-driven fault diagnostics, and power system modeling and analysis.

Bruce Stephen (M'09–SM'14) received the BSc degree in aeronautical engineering from the University of Glasgow, UK, in 1997, the MSc degree in computer science from the University of Strathclyde, UK, in 1998 and the PhD degree in information retrieval from the University of Strathclyde in 2005. He currently holds the post of Research Fellow within the Institute for Energy and Environment at the University of Strathclyde, UK. His research interests include power system condition monitoring, renewable integration, characterizing low voltage network behavior and characterization of demand in electrical distribution networks. Dr Stephen is a Chartered Engineer and a Fellow of the Higher Education Academy.

Jethro Browell (S'15–M'16) received the M.Phys. degree in mathematics and theoretical physics from the University of St Andrews, U.K., in 2011, and the Ph.D. degree in wind energy systems from The University of Strathclyde, U.K., in 2015. He is a Lecturer in the University of Strathclyde's Institute for Energy and Environment where his research interests span all aspects of energy forecasting and associated decision-making in power system operation and energy markets. Dr. Browell is an EPSRC Innovation Fellow and a member of the IEEE Power Energy Society.

Stephen Haben Stephen Haben is a visiting researcher fellow at the Mathematical Institute in the University of Oxford where his research is focused on load forecasting for low voltage energy systems, battery storage control and energy analytics. He received his PhD in 2011 in the conditioning and preconditioning of variational data assimilation for numerical weather prediction systems