A Statistical Approach to Guide Phase Swapping for Data-Scarce Low Voltage Networks

Lurui Fang, *Student Member, IEEE,* Kang Ma, Xinsong Zhang

***Abstract*—Phase swapping, which rebalances the unbalanced three-phase low voltage (LV, 415V) networks, improves network efficiency by reducing capacity waste and energy losses. A key challenge against phase swapping is that the majority of LV networks are data-scarce, i.e. there is a general lack of data in LV networks. In light of this, this paper proposes a new statistical approach to develop phase swapping guidance for data-scarce LV networks with neither time-series network measurements nor customer metering data. Firstly, given a set of data-rich LV networks (with time-series phase currents data collected at LV substations throughout a year), typical load profiles and their weights in each of the three phases are extracted by applying a non-negative matrix factorization method. Then, phase swapping guidance are developed for data-rich LV networks along with their rebalancing potentials (rebalancing potentials refer to the reduction of phase imbalance degree). Secondly, a rapid screening model is developed to efficiently identify the data-scarce LV networks with high rebalancing potentials. Phase swapping guidance are then developed for these data-scarce networks with high rebalancing potentials. Case studies reveal that the statistical approach produces effective phase swapping guidance, which reduce the phase imbalance degrees for 99% of the LV networks and the maximum reduction is 35%. Validation results show that the average reduction of the phase imbalance degree for data-scarce networks is only 14.3% less than that for data-rich networks.**

***Index Terms*—low voltage, phase imbalance, phase balancing, phase swapping, power distribution, three-phase power, statistical approach**

Nomenclature

$DPIB_{ori}$	The original phase imbalance degree
$DPIB_{Bal}$	The phase imbalance degree after rebalancing
$DPIB_{v}$	The virtual phase imbalance degree
$DPIB_{sba}$	The phase imbalance degree after rebalancing for the validation sample
e_v	The rebalancing error
$\mathbf{H}$	The matrix of weighting factors
$\mathbf{H}_{net}$	The weighting factor matrix for a data-rich network
$\mathbf{H}_{\mathrm{bm}}$	The balanced weighting factor matrix
$\mathbf{H}_{\mathrm{ds}}$	The weighting factor matrix for a data-scarce network
$\mathbf{H}_{\mathrm{bms}}$	The balanced weighting factor matrix for a data-scarce network
$h_{\emptyset}$	The weighting factors in phase $\emptyset$ ($\emptyset \epsilon \{\mathrm{a, b, c}\}$) of the data-rich network
$\mathbf{I}_{\mathbf{PI}}$	The input phase current matrix
$\mathbf{I}_{SB}$	the rebalanced time-series phase current data for a validation sample
$I_{p\emptyset,i}(t)$	The time-series phase current data for phase $\emptyset$ ($\emptyset \epsilon \{\mathrm{a, b, c}\}$) of the i_{th} data-rich LV networks
$I_{P\emptyset B}$	The rebalanced time-series phase current profiles
I_{ya}, I_{yb}, I_{yc}	The yearly average phase current data
$\mathrm{I}_{o\emptyset}(t)(\emptyset \epsilon \{\mathrm{a, b, c}\})$	The one day's phase current data
n_t	The length of the time series phase current data
n_p	The total number of phases for all data-rich LV networks
n_{net}	The number of data-rich LV networks
n_d	The number of the constituent load profiles
n_{dt}	The length of one-day' time-series phase current data
n_{dy}	The number of days throughout a year
SPSS	The statistical phase swapping matrix

I. Introduction

PHASE imbalance causes significant consequences to low voltage networks (415V, LV), e.g. extra energy losses [1], [2], additional reinforcement cost [3], risks of network nuisance tripping (because of a high zero-sequence current) [4], risks of network overloading [5], and possible damages to induction motors because of voltage imbalance [6], [7]. Phase swapping is a natural way to rebalance the three phases and resolve the above problems [4], [5], [8]. However, developing mass-scale guidance for phase swapping remains a challenge, because the majority of LV networks are data-scarce, i.e. there are no time-series phase current data throughout a year from these LV networks [9].

L. Fang, K. Ma are with the University of Bath, Bath, UK. (Correspondence author: K. Ma. E-mail: K.Ma@ bath.ac.uk).
X. Zhang is with University of Nantong, Nantong, China.

A number of references focus on developing phase swapping strategies for data-rich distribution networks. Reference [4] uses mixed integer programming to develop phase swapping strategies. References [10], [11] develop optimal phase swapping strategies using simulated annealing and immune algorithms, respectively. Reference [12] applies a fuzzy function to develop phase swapping strategies. Reference [13] applies an expert system to develop phase swapping strategies. References [14], [15] apply heuristic algorithms to develop phase swapping strategies, considering load patterns. Reference [16] develops phase swapping strategies using look-ahead optimizations, considering load uncertainties. Reference [17] summarizes different methods for developing phase swapping strategies. All the above references perform phase rebalancing based on full data, including network topology, time-series network current data, and demand data. However, these data are not normally available in most LV networks that are data-scarce.

Reference [18] uses smart meter data for phase swapping. However, the use of smart meters for phase balancing face three limitations: 1) a smart meter does not know which phase a customer is connected to, thus offering limited support for phase swapping. 2) In the UK, electricity suppliers (i.e. retailers) and distribution network operators (DNOs) are separate entities. Data protection concerns arise if suppliers are to share smart meter data with DNOs. 3) In the UK, not all customers have smart meters – the rollout of smart meters is much slower than the original plan of deploying smart meters for all customers by 2020. Reference [19] uses automated meter management (AMM) system for phase balancing. The AMM system overlaps smart meters in terms of functionality: they both provide customer side data. Therefore, the AMM system faces the same limitations as smart meters, despite that the former provides additional data compared to smart meters. Further, the deployment of the AMM system for millions of LV networks in the UK is economically infeasible.

This paper advances from existing references by extrapolating the knowledge to data-scarce LV networks with neither time-series network measurements nor customer metering data. The knowledge extrapolation was an unanswered question and it calls for a statistical approach. The extrapolation is also one of the key technical aspects of this paper. In light of this, this paper makes the following original contributions:

1) It for the first time develops phase swapping guidance for data-scarce LV networks with neither the need for any time-series network measurements nor the need for any customer-side metering data.

2) To achieve 1), this paper proposes a new statistical approach.

The statistical approach effectively overcomes the insufficient data challenge by extrapolating knowledge from a set of 800 representative data-rich LV networks with time-series phase current data to the vast population of data-scarce networks. Given the 800 data-rich LV networks, the first step is to develop a rebalancing model to rebalance the three phases of the data-rich networks by applying a non-negative matrix factorization method. The model also outputs the rebalancing potentials (the reduction of the phase imbalance degree) of these data-rich LV networks. At the 2[nd] step, a rapid screening model is developed to identify the data-scarce LV networks with high rebalancing potentials among all data-scarce networks. At the 3[rd] step, phase swapping guidance are developed for the identified data-scarce LV networks by applying the statistical rebalancing model developed in the first step. The phase swapping guidance guide the distribution network operators to reallocate loads among the three phases of data-scarce networks in order to rebalance the three phases.

The statistical approach develops phase swapping guidance for data-scarce networks, which take the majority of the LV networks in the UK, while requiring only a minimal amount of data. The approach is economically appealing in the sense that no cost in monitoring system is incurred in Scenario 1 (where only yearly average data is required) and a minimal cost in monitoring effort is incurred in Scenario 2 (where only one-day's time-series data are required). This is compared to investing in monitoring systems to collect year-round time-series data from millions of LV networks in the UK. If distribution network operators (DNOs) follow the phase swapping guidance, energy losses would be reduced and the network capacity that is wasted by phase imbalance would be released.

The rest of this paper is organized as follows: Section II presents the statistical approach. Section III performs case studies. Section IV concludes this paper.

II. Methodology

To develop phase swapping guidance for data-scarce LV networks, this paper proposes a new statistical approach. It consists of three steps. Fig. 1 shows the flowchart of the statistical approach.

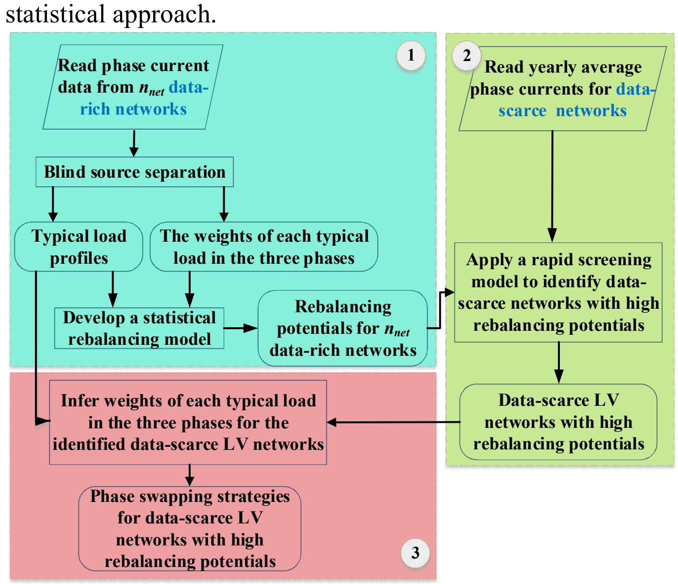


Fig. 1 Methodology of the statistical approach

We have the time-series phase current collected every 10 minutes throughout a year from the substations of 800 data-rich LV networks within Western Power Distribution (a UK DNO)'s business area. These LV networks cover approximately 10% of the population in South Wales areas with a good mixture of urban, suburban and rural areas and a good mixture of domestic and commercial loads [20]. These data come from the project "Low Voltage Network Template" and are described in details in [20].

The purpose of Stage 1 is that it derives two key variables (which are not normally available in data-scarce LV networks) from data-rich networks for the development of phase swapping guidance. These two variables are 1) time-series constituent load profiles; and 2) their weights on each of the three phases. The reason for developing a rapid screening model in Stage 2 is that it identifies the data-scarce LV networks with high rebalancing potentials. These are seriously unbalanced networks that are worth phase rebalancing. In Stage 3, the phase swapping guidance developed in Stage 1 is extrapolated from data-rich networks to data-scarce networks that have high rebalancing potentials.

A. Develop a statistical rebalancing model

For the 800 data-rich LV networks, a non-negative matrix factorization (NMF) method is adopted to extract typical load profiles and their weights in the three phases of the data-rich networks. The reason for using the NMF method is that it is a classical method to extract non-negative source signals (e.g. typical constituent load profiles in this case) and their weighting factors from mixed signals (e.g. time-series phase current data from the 800 data-rich LV networks) [21],[22]. In addition, NMF removes outliers [23]. After deriving typical constituent load profiles and their weighting factors, a statistical phase rebalancing model is developed for data-rich LV networks.

1) Extract constituent load profiles and their weights

To apply the NMF method, the time-series phase current data $\mathbf{I}_{p\emptyset,n}(t)$ from the 800 data-rich LV networks form an input phase current matrix $\mathbf{I_{PI}}$, which is given by

$$\mathbf{I_{PI}}(t) = \left[I_{p\emptyset,1}(t), I_{p\emptyset,2}(t), \cdots, I_{p\emptyset,n_{net}}(t)\right] \tag{1}$$

where $I_{p\emptyset,i}(t)$ denotes the time-series phase current data for phase $\emptyset$ ($\emptyset \epsilon \{\mathrm{a,b,c}\}$) of the i_{th} data-rich LV networks; $\mathbf{I}_{\mathrm{PI}}$ is a matrix with n_t (the length of the time series phase current data) rows and n_p (the total number of phases for all data-rich LV networks) columns; n_{net} is the number of data-rich LV networks. It is straightforward to see that $n_p = 3n_{net}$.

Given the input phase current matrix $\mathbf{I}_{\mathrm{PI}}$, the NMF method is applied to find the constituent load profiles and weighting factors in each of the phases. The relationship among $\mathbf{I}_{\mathrm{PI}}$, $\mathbf{W}$, and $\mathbf{H}$ is given by:

$$\mathbf{I}_{\mathrm{PI}} \approx \mathbf{W} \cdot \mathbf{H} \tag{2}$$

where $\mathbf{W}$, given by (3), is a matrix of n non-negative constituent load profiles; $\mathbf{H}$, given by (4), denotes the matrix of weighting factors. $\mathbf{W}$ and $\mathbf{H}$ are given by:

$$\mathbf{W} = \left[\boldsymbol{W_1}, \boldsymbol{W_2}, \cdots, \boldsymbol{W_{n_d}}\right] \tag{3}$$

$$\mathbf{H} = \left[H_1, H_2, \cdots, H_{n_d}\right]^{\mathrm{T}} \tag{4}$$

where W_i is the i_{th} constituent load profile; H_i is the weighting factors of the i_{th} load profile in each of the phases in $\mathbf{I}_{\mathrm{PI}}$; n_d is the number of the constituent load profiles. $\mathbf{W}$ is an n_t-by-n_d matrix. $\mathbf{H}$ is an n_d-by-n_p matrix.

The constituent loads are interpreted by three typical load profiles in the UK [11], [24]: low demand households, high demand households (the households with electric heating), and commercial loads. Therefore, in this paper, the number of the constituent load profiles $n_d = 3$.

To obtain $\mathbf{W}$ and $\mathbf{H}$, an optimization model is formulated, which is the key to the NMF method. The optimization model minimizes the distance between the actual phase current $\mathbf{I}_{\mathrm{PI}}$ and the reconstructed phase current ($\mathbf{W} \cdot \mathbf{H}$). This optimization model is given by [21]:

$$\min_{\mathbf{W},H_i} \sum_{i=1}^{n_p} \left\| I_{PI,i} - \mathbf{W} \cdot H_i \right\|^2 \tag{5}$$
$$s.t.\ \text{all elements of } \mathbf{W} \text{ and } H_i \geq \mathbf{0}$$

where n_p is the total number of phases for all data-rich LV networks; $I_{PI,i}$ is the i_{th} columns of the input phase current matrix $\mathbf{I}_{\mathrm{PI}}$, which was given by (1); H_i is defined in (4). The detailed procedure for deriving $\mathbf{W}$ and $\mathbf{H}$ is given in [21].

In this paper, the constituent load profiles derived above are normalized to be within a band [0, 1]. The normalization is given by

$$\mathbf{W} = \left[\frac{W_1}{W_{1,max}}, \frac{W_2}{W_{2,max}}, \cdots, \frac{W_{n_d}}{W_{n_d,max}}\right] \tag{6}$$

where $W_{i,max}$ denotes the maximum element of W_i, which is defined in (3).

Correspondingly, the matrix $\mathbf{H}$ is adjusted, as given by

$$\mathbf{H} = \left[H_1 \cdot W_{1,max}, H_2 \cdot W_{2,max}, \cdots, H_{n_d} \cdot W_{n_d,max}\right]^{\mathrm{T}} = \left[h_{\emptyset,i}\right] \tag{7}$$

where H_i and $W_{i,max}$ are defined in (4) and (6), respectively; $h_{\emptyset,i}$ is the normalized weighting factors in phase $\emptyset$ ($\emptyset \epsilon \{\mathrm{a,b,c}\}$) of the i_{th} data-rich LV network; $h_{\emptyset,i}$ is a vector of n_d rows.

2) Develop phase swapping guidance for data-rich LV networks

After deriving the constituent load profile matrix $\mathbf{W}$ and the weighting factor matrix $\mathbf{H}$, a statistical phase rebalancing model is developed for the 800 data-rich LV networks.

For a data-rich network, a weighting factor matrix $\mathbf{H}_{net}$ is given by:

$$\mathbf{H}_{net} = [h_{\mathrm{a}}, h_{\mathrm{b}}, h_{\mathrm{c}}] \tag{8}$$

where $h_\emptyset$ is the weighting factors in phase $\emptyset$ ($\emptyset \epsilon \{\mathrm{a,b,c}\}$) of the data-rich network, as defined in (7); $\mathbf{H}_{net}$ is a n_d-by-3 matrix. n_d is defined in (4).

Then, a balanced weighting factor matrix $\mathbf{H}_{\mathrm{bm}}$ is derived.

$$H_b = \frac{1}{3}\sum_{i=1}^{3} H_{net,i} \tag{9}$$

$$\mathbf{H}_{bm} = [H_b, H_b, H_b] \tag{10}$$

where $H_{net,i}$ denotes the weighting factors at the i_{th} column (phase) of $\mathbf{H}_{net}$, which is given by (8).

Secondly, a statistical phase swapping matrix (**SPSS**) is given by:

$$\mathbf{SPSS} = \left[spss_{i,j}\right] = \mathbf{H}_{bm} - \mathbf{H}_{net} \tag{11}$$

where $spss_{i,j}$ is the i_{th} row and j_{th} column of **SPSS**.

In this paper, $spss_{i,j}$ indicates that the amount (in an average power (kW)) of the i_{th} constituent load (the i_{th} column of **W**) should be moved away from the j_{th} phase ($j = 1, 2$ and 3 represent phase a, b and c, respectively) of the data-rich network. The rebalanced time-series phase current profiles for the data-rich LV network are given by:

$$I_{P\emptyset B} = I_{p\emptyset} - \mathbf{W} \cdot spss_i \tag{12}$$

where $I_{P\emptyset B}$ is a vector of n_t rows ($\emptyset \epsilon \{a, b, c\}$). $I_{p\emptyset}$ is the time-series phase current data for phase $\emptyset$, as defined in (1). Compared to (1), the subscript i of $I_{p\emptyset,i}(t)$ is dropped, because now we consider a given data-rich network. **W** is defined in (6). $spss_i$ is the i_{th} column of **SPSS**, which is given by (11).

In addition, this model is the key to calculating the rebalancing potential (i.e. the reduction of the phase imbalance degree) for the data-rich LV network. Before deriving the rebalancing potentials, the following variables are defined.

$$I_{maxP\emptyset}(t) = \max\{I_{pa}(t), I_{pb}(t), I_{pc}(t)\} \tag{13}$$

$$I_{aveP\emptyset}(t) = \left(I_{pa}(t) + I_{pb}(t) + I_{pc}(t)\right)/3 \tag{14}$$

where $I_{pa}(t)$, $I_{pb}(t)$, and $I_{pc}(t)$ are defined in (1).

$$I_{maxP\emptyset B}(t) = \max\{I_{PaB}(t), I_{PbB}(t), I_{PcB}(t)\} \tag{15}$$

$$I_{aveP\emptyset B}(t) = \left(I_{PaB}(t) + I_{PbB}(t) + I_{PcB}(t)\right)/3 \tag{16}$$

where $I_{P\emptyset B}(t)$ denotes the rebalanced time-series phase current data for phase $\emptyset$ ($\emptyset \epsilon \{a, b, c\}$) of the data-rich network.

For the data-rich network, the original phase imbalance degree ($DPIB_{ori}$) and the phase imbalance degree after rebalancing ($DPIB_{Bal}$) are given by:

$$DPIB_{ori} = \frac{1}{n_t}\sum_{t=1}^{n_t}\frac{(I_{maxP\emptyset}(t) - I_{aveP\emptyset}(t))}{I_{maxP\emptyset}(t)} \tag{17}$$

$$DPIB_{Bal} = \frac{1}{n_t}\sum_{t=1}^{n_t}\frac{(I_{maxP\emptyset B}(t) - I_{aveP\emptyset B}(t))}{I_{maxP\emptyset B}(t)} \tag{18}$$

where $I_{maxP\emptyset}(t)$ is defined in (13); $I_{aveP\emptyset}(t)$ is defined in (14); $I_{maxP\emptyset B}(t)$ is defined in (15); $I_{aveP\emptyset B}(t)$ is defined in (16). n_t is defined in (1).

Thus, the rebalancing potential RP for the data-rich LV network is given by:

$$RP = DPIB_{ori} - DPIB_{Bal} \tag{19}$$

B. Develop a rapid screening model

At this stage, a rapid screening model is developed to identify data-scarce LV networks with high rebalancing potentials. Phase balancing for these networks would lead to significant benefits in terms of reducing energy losses and saving network investment costs. Before developing the rapid screening model, a virtual phase imbalance degree ($DPIB_v$) is defined:

$$DPIB_v = \frac{\max\{I_{ya}, I_{yb}, I_{yc}\} - (I_{ya} + I_{yb} + I_{yc})/3}{\max\{I_{ya}, I_{yb}, I_{yc}\}} \tag{20}$$

where I_{ya}, I_{yb} and I_{yc} denote the average currents for phases a, b and c, respectively, throughout a year. $DPIB_v$ is a feature for both data-rich and data-scarce networks.

The reason for using $DPIB_v$ as the feature is that: 1) it is derived from yearly average phase current data, which are available for both data-rich and data-scarce LV networks [25]; 2) the alternative yearly maximum phase current data from data-scarce LV networks cannot represent the actual phase imbalance [8].

Based on the virtual phase imbalance degree and the rebalancing potentials from n_{net} data-rich LV networks, a rapid screening model is developed. It uses a quadratic function to map the virtual phase imbalance degree ($DPIB_v$) to the rebalancing potential, which is given by

$$RP = f(DPIB_v) = \theta_1 DPIB_v{}^2 + \theta_2 DPIB_v + \theta_3 \tag{21}$$

where RP is defined in (19). $DPIB_v$ is the virtual phase imbalance degree; θ_1, θ_2 and θ_3 are coefficients.

The choice of a quadratic function is justified by Fig. 4 (in the case study section), which shows an approximate quadratic relationship between the degree of phase imbalance and the rebalancing potential. In other words, the quadratic function represents an optimal tradeoff between bias and variance. The fitted quadratic function can then be used to estimate the rebalancing potentials for data-scarce LV networks.

To derive the optimal coefficients (θ_1, θ_2 and θ_3), the following optimization model is solved:

$$\min_{\theta_1,\ \theta_2,\ \theta_3} \sqrt{\frac{1}{n_{net}}\sum_{i=1}^{n_{net}}(RP_i - f(DPIB_{v,i}))^2} \tag{22}$$

where $f(DPIB_{v,i}) = \theta_1 DPIB_{v,i}{}^2 + \theta_2 DPIB_{v,i} + \theta_3$

$DPIB_{v,i}$ is the virtual phase imbalance degree for the i_{th} data-rich LV networks; RP_i denotes the derived rebalancing potentials for the i_{th} data-rich LV networks; n_{net} is the number of data-rich LV networks.

After developing the rapid screen model, rebalancing potential are estimated for data-scarce LV networks. Furthermore, a threshold variable RP_T divide rebalancing potentials into two sections: the low rebalancing potential section (LR) and high rebalancing potential section (HR). Thus, a data scarce LV network is identified as a network with high rebalancing potential, if the estimated rebalancing potential is greater than RP_T. The choice of RP_T (i.e. how "high" is a high rebalancing potential) is subjective. It requires expert's judgment and the criteria can vary from case to case. For example, RP_T can be chosen so that the "high rebalancing potential" section includes the LV networks whose rebalancing potentials are among the top 25% of all LV networks considered.

C. Develop phase swapping guidance for data-scarce LV networks with high rebalancing potentials

In this section, phase swapping guidance are developed for data-scarce LV networks with high rebalancing potentials (defined in Section – II – B). The weights of each constituent load are rebalanced in the three phases according to the developed guidance, thus approximately balancing the three phases. The key is to infer the weighting factor matrix **H** for data-scarce LV networks. When applied to the field, different distribution network operators have different types of available data: 1) one scenario is where only the yearly average phase current data (I_{ya}, I_{yb}, I_{yc}) are available; 2) the other scenario is where only one day's phase current data $\mathrm{I}_{o\emptyset}(t)$ ($\emptyset \in \{\mathrm{a,b,c}\}$) are available. The above two scenarios are considered.

1) Use the yearly average phase current data to infer the missing weighting factors

Select a data-scarce network as an example, the weighting factor matrix $\mathbf{H}_{\mathrm{ds}}$ is inferred, as given by:

$$H_{ds,i} = \left[\frac{\frac{1}{3}I_{ya}}{w_{ave.i}}, \frac{\frac{1}{3}I_{yb}}{w_{ave.i}}, \frac{\frac{1}{3}I_{yc}}{w_{ave.i}}\right] \tag{23}$$

$$\mathbf{H}_{\mathrm{ds}} = \begin{bmatrix} H_{ds,1} \\ H_{ds,2} \\ \vdots \\ H_{ds,n_d} \end{bmatrix} \tag{24}$$

where $w_{ave.i}$ is the average value of the data in the i_{th} column of **W** (given by (6)); n_d is defined in (3).

After deriving the weighting factor matrix $\mathbf{H}_{\mathrm{ds}}$, a balanced weighting factor matrix $\mathbf{H}_{\mathrm{bms}}$ is derived for the data-scarce network, which is given by:

$$H_{bs} = \frac{1}{3}\sum_{i=1}^{3} \mathbf{H}_{ds,i} \tag{25}$$

$$\mathbf{H}_{bms} = [H_{bs}, H_{bs}, H_{bs}] \tag{26}$$

where $\mathbf{H}_{ds,i}$ denotes the weighting factors at the i_{th} column of $\mathbf{H}_{\mathrm{ds}}$.

Finally, for the data-scarce LV network, a statistical phase swapping matrix ($\mathbf{SPSS_s}$) is given by:

$$\mathbf{SPSS_s} = \left[spss_{s,i,j}\right] = \mathbf{H}_{bms} - \mathbf{H}_{ds} \tag{27}$$

where $\mathbf{H}_{bms}$ and $\mathbf{H}_{ds}$ are derived in (26) and (24), respectively; where $spss_{s,i,j}$ is the i_{th} row and j_{th} column of $\mathbf{SPSS_s}$. The variable $\mathrm{spss}_{s,i,j}$ indicates that the amount of the i_{th} constituent load (the i_{th}column of **W**, as defined in (6)) should be moved away from the j_{th} phase ($j = 1, 2$ and 3 represent phases a, b and c, respectively) of the data-scarce network.

2) Use one day's phase current data to infer the missing weighting factors

Select a data-scarce network as an example, the weighting factor matrix $\mathbf{H}_{\mathrm{ds}}$, as defined in (24), is inferred through the following steps.

Firstly, for the i_{th} constituent load (the i_{th} column of **W**), the j_{th} day's load profile $W_{i,j}$ is derived, which is given by

$$W_{i,j} = \mathbf{W}((n_{dt}(j-1)+1):(n_{dt}j), i) \tag{28}$$

where $\mathbf{W}((n_{dt}(j-1)+1):(n_{dt}j), i)$ is the i_{th} column, $(n_{dt}(j-1)+1)$ to $(n_{dt}j)$ rows of the matrix **W,** which is define in (6). **W** (*x*, *y*) denotes the element in the x_{th} row and y_{th} column of matrix **W**. The "*m*:*n*" expression denotes "from *m* to *n* (inclusive)", e.g. from row $(n_{dt}(j-1)+1)$ to row $(n_{dt}j)$ inclusive of both rows. $W_{i,j}$ is a vector of n_{dt} rows; n_{dt} is the length of one-day' time-series phase current data.

Then, an average load profile that corresponds to the i_{th} constituent load is given by:

$$I_{pd,i} = \frac{1}{n_{dy}}\sum_{j=1}^{n_{dy}} W_{i,j} \tag{29}$$

$I_{pd,i}$ is a vector of n_{dt} rows; n_{dt} is defined in (28); $W_{i,j}$ is defined in (28); n_{dy} is the number of days throughout a year.

To derive the weighting factor matrix for a data-scarce network, an optimization problem is solved:

$$\begin{gathered} \min_{h_{1,\emptyset}, h_{2,\emptyset}, \cdots, h_{n_d,\emptyset}} \sqrt{\frac{1}{n_{dt}}\sum_{t=1}^{n_{dt}} (I_{o\emptyset}(t) - \mathbf{W}_{\mathrm{ds}}(t) \cdot H_{ds,\emptyset})^2} \\ s.t.\ h_{1,\emptyset}, h_{2,\emptyset}, \cdots \ and\ h_{n\_d,\emptyset} \geq 0 \\ \text{where } \mathbf{W}_{\mathrm{ds}} = [I_{pd,1}, I_{pd,2}, \cdots I_{pd,n_d}]; \\ H_{ds,\emptyset} = [h_{1,\emptyset}, h_{2,\emptyset}, \cdots, h_{n_d,\emptyset}]^T \end{gathered} \tag{30}$$

$I_{pd,i}$ is define in (29); $h_{i,\emptyset}$ denotes the weight of the i_{th} constituent load in phase $\emptyset$ ($\emptyset \in \{\mathrm{a,b,c}\}$) of the data-scarce network; $I_{o\emptyset}(t)$ is the t_{th} element of one day's phase current from the data-scarce network; $\mathbf{W}_{\mathrm{ds}}$ is a matrix with n_{dt} rows and n_d columns. $\mathbf{W}_{\mathrm{ds}}(t)$ denotes the elements in the t_{th} row of matrix $\mathbf{W}_{\mathrm{ds}}$; n_{dt} is defined in (28).

The weighting factor matrix for the data-scarce LV network $\mathbf{H}_{\mathrm{ds}}$ is given by:

$$\mathbf{H}_{\mathrm{ds}} = [H_{ds,a}, H_{ds,b}, H_{ds,c}] \tag{31}$$

where $H_{ds,\emptyset}$ ($\emptyset \in \{\mathrm{a,b,c}\}$) is defined in (30).

After deriving $\mathbf{H}_{\mathrm{ds}}$, the statistical phase swapping matrix ($\mathbf{SPSS_s}$) is given by (25) – (27). $\mathbf{SPSS_s}$ presents the phase swapping matrix. $\mathbf{SPSS_s}$ has the same meaning as that explained immediately after Equation (27).

D. Method for validation

Before validation, the rebalancing potential *RP* for $n_{net}(n_{net} = 800)$ data rich LV networks are derived in stage 1 (explained in Section II – A) as the accurate RP.

Then, to validate the developed phase swapping guidance for data-scarce LV networks, *k*-fold cross-validation ($k = 10$ in this paper) is used. Firstly, the 800 data-rich LV networks are randomly partitioned into k equal sized groups. Then, the LV networks from one of the k groups are held out as the validation samples (treat them as if there were data-scarce LV networks), the LV networks from the remaining $k-1$ groups are used as the training samples (data-rich LV networks). The training samples are used to develop the statistical rebalancing model and rapid screening model as explained in Section – II – A and B. Then, the phase swapping guidance are developed for the validation samples.

For the first scenario (explained in Section II – C – 1)), the rebalanced time-series phase current data $\mathbf{I}_{SB}$ for a validation sample is given by:

$$\mathbf{I}_{SB} = [I_{pa}, I_{pb}, I_{pc}] - \mathbf{W} \cdot \mathbf{SPSS_s} \tag{32}$$

where **W** is defined in (6); $\mathbf{SPSS_s}$ is defined in (27); $I_{p\emptyset}$ is defined in (1) ($\emptyset \in \{\mathrm{a,b,c}\}$); Compared to (1), the subscript *i* of $I_{p\emptyset,i}(t)$ is dropped, because now we consider a given network.

The corresponding phase imbalance degree after rebalancing is given by:

$$DPIB_{sba} = \frac{1}{n_t}\sum_{t=1}^{n_t}\frac{(I_{maxSB}(t) - I_{aveSB}(t))}{I_{maxSB}(t)} \quad (33)$$

where $I_{maxSB}(t)$ is the maximum element in the t_{th} row of $\mathbf{I}_{SB}$, which is defined in (32); $I_{aveSB}(t)$ is the average value of the elements in the t_{th} row of $\mathbf{I}_{SB}$, which is defined in (32); n_t is defined in (1).

The validation rebalancing potential for the validation sample is given by:

$$RP_v = DPIB_{ori} - DPIB_{sba} \quad (34)$$

where $DPIB_{ori}$ denotes the original phase imbalance degree for the validation sample (given by (17)); $DPIB_{sba}$ denotes the phase imbalance degree after rebalancing for the validation sample (given by (33)).

Through the previous validation, the rebalancing error for the i_{th} validation sample e_v is given by

$$e_v = \frac{RP - RP_v}{RP} \quad (35)$$

where RP (given by (19)) is the accurate rebalancing potential for the validation sample; RP_v is the rebalancing potential (given by the proposed statistical rebalancing method) for the validation sample. This error indicates the validity of the phase swapping performance for the data-scarce LV networks.

The previous process repeats until each of the k groups has been held out as the validation samples. After k iterations, the rebalancing potentials (i.e. the reduction of phase imbalance degree) are derived for all data-scarce LV networks. The rebalancing errors are given by (35).

For the second scenario (explained in Section II – C – 2)), the validation process in each iteration of the k-folds validation is presented in Fig. 2.

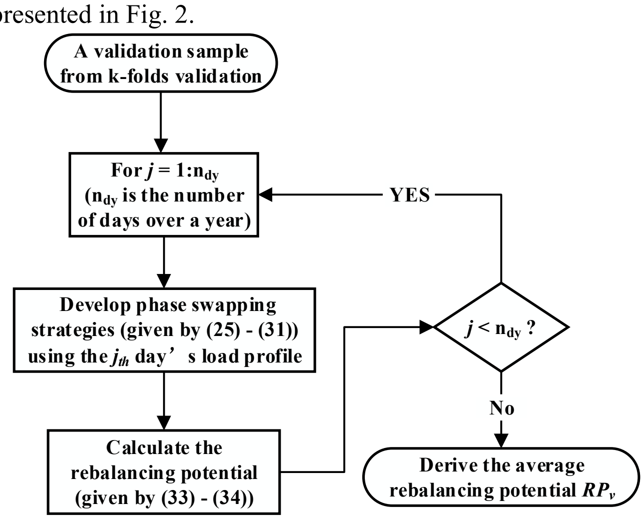


Fig. 2 Flowchart of the validation process for Scenario 2)

After deriving RP_v for the validation sample, the rebalancing error is given by (35). The above process repeats until each of the k groups has been held out as the validation samples.

III. Case Studies

This section presents the numerical results. The results from the statistical rebalancing modelling and rapid screening modelling for data-rich networks are given in Sections IV – A and B, respectively. Section IV – C presents the phase swapping results for data-scarce networks. A discussion is presented in Section IV – D.

A. Results from the statistical rebalancing model

In the first step, three constituent load profiles are extracted and normalized: low demand households, high demand households and commercial loads. Fig. 3 presents the three constituent load profiles throughout a week.

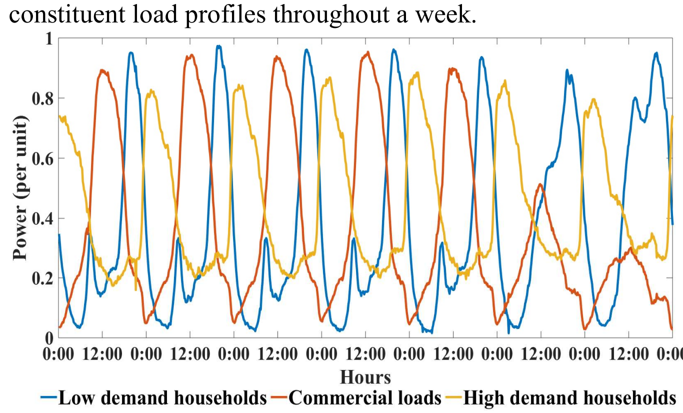


Fig. 3 The constituent load profiles throughout a week (Monday to Sunday)

Commercial loads and low demand households share the same characteristic: the profile is different between workdays and weekends. For commercial loads, the weekend peak load is approximately 1/3 of the workday peak. For low demand households, the first peak of the weekend's load is approximately twice of the first peak of the workday's load.

Select a data-rich network as an example. If the average power of these typical customers (e.g. low demand household, high demand household and commercial load) are 0.2 kW, 0.4 kW and 0.8 kW, respectively, the phase swapping guidance are presented as follows:

TABLE I
A STATISTICAL PHASE SWAPPING GUIDANCE

Unit: kW, number of loads

	Phase a	Phase b (kW)	Phase c (kW)
Low demand households	–5.01, –25	4.43, 22	0.57, 3
High demand households	–2.61, –7	3.10, 8	–0.49, 1
Commercial loads	–2.90, –4	0.62, 1	2.28, 3

In TABLE I, a negative number indicates the amount of the constituent load that should be moved away to other phases; a positive number indicates the amount of the constituent load that should be taken in from other phases. For example, for phase a, 25 low demand households (which sums up to an average power of 5.01 kW), 7 high demand households (which sums up to an average power of 2.61 kW), and 4 commercial loads (which sums up an average power of 2.90 kW) should be moved away to other phases. If phase swapping strictly follows the guidance in Table I, the amount of each constituent load in the three phases is rebalanced, thus balancing the three phases. For example, after phase swapping, the average power of LDH is 15 kW in phases a, b and c, respectively, but it was 10.8 kW, 20.28 kW and 16.42 kW before phase swapping. The accurate rebalancing potential (RP) is 0.115.

If the derived phase swapping guidance is followed, it significantly reduces the degree of phase imbalance for the 800 data-rich LV networks. It reduces the average and maximum phase imbalance degree by 34% and 40%, respectively. In addition, 387 (48.3%) data-rich LV networks have rebalancing potentials greater than 0.05. Five (0.67%) out of the 800 data-rich LV networks have negative rebalancing potentials, indicating that phase swapping actually increases the degree of phase imbalance. The reason for this is that these few LV networks have no consistent direction of imbalance among the three phases. In other words, phase swapping is not applicable to these five networks.

B. Results from the rapid screening model

Fig. 4 shows a quadratic mapping from the virtual degree of phase imbalance to the rebalancing potential, based on the data from the 800 data-rich LV networks. Such a mapping is the rapid screening model.

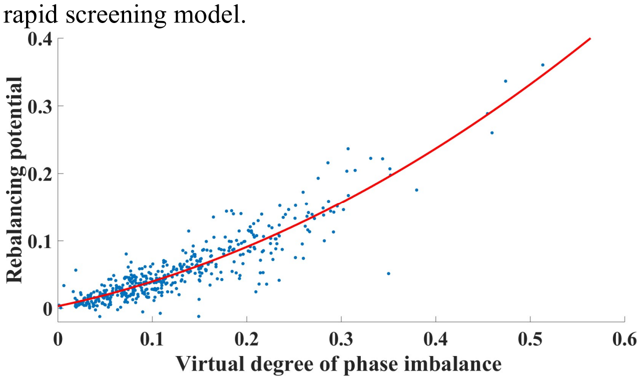


Fig. 4 The rapid screening model

The majority of the networks have $DPIB_v$ values that fall in the range of [0, 0.3] and rebalancing potentials that fall in the range of [0, 0.2]. In this paper, the threshold of the rebalancing potential is $RP_T = 0.05$. 48.3% of the LV networks have rebalancing potentials greater than the threshold – they have high rebalancing potentials.

Based on the data set of the 800 networks, the mapping incurs a root-mean squared error (RMSE) and a mean absolute percentage error (MAPE) of 0.0214 and 16.3%, respectively.

C. Phase swapping guidance for data-scarce LV networks with high rebalancing potentials

To rebalance the data-scarce LV networks with high rebalancing potentials (explained in Section III – B), two scenarios are considered: 1) the scenario with yearly average phase current data; and 2) the scenario with one day's phase current data. Phase swapping guidance are developed for the two scenarios. The results are validated by 10-fold cross-validation.

1) Phase swapping results for data-scarce networks with yearly average phase current

Select the data-rich network (the example in Section – III – A) as a validation sample (treat this data-rich network as if it were data-scarce by ignoring its time-series data), this network has yearly average phase currents $[I_{ya}, I_{yb}, I_{yc}] = [90.43\text{A}, 168.48\text{A}, 144.63\text{A}]$. The weighting factors matrix is calculated as follows:

$$\mathbf{H}_{\text{ds}} = \begin{bmatrix} 81.74 & 159.27 & 108.36 \\ 81.74 & 159.27 & 108.36 \\ 81.74 & 159.27 & 108.36 \end{bmatrix}$$

If the average power of these typical customers (e.g. low demand household, high demand household and commercial load) are 0.2 kW, 0.4 kW and 0.8 kW, respectively, the phase swapping guidance are presented as follows:

TABLE II
A STATISTICAL PHASE SWAPPING GUIDANCE
Unit: kW, number of loads

	Phase a	Phase b	Phase c
Low demand households	–3.52, –17	2.85, 14	0.67, 3
High demand households	–3.37, –9	2.72, 7	0.65, 2
Commercial loads	–4.26, –5	3.43, 4	0.81, 1

The meaning of negative numbers and positive numbers are explained after Table I. For example, for phase b, 14 low demand households (which sums up to an average power of 2.85 kW), 7 high demand households (which sums up to an average power of 2.72 kW), and 4 commercial loads (which sums up an average power of 3.43 kW) should be taken in from other phases. The rebalancing potential for this network is 0.0959. The rebalancing error e_v (defined in (35)) is 16%. It indicates that, for this validation sample, the rebalancing potential RP_v (given by the proposed method) is 16% lower than the accurate RP.

Through validation, the average rebalancing error is 19.33% in Scenario 1). If the phase swapping implementation strictly follows the developed guidance, the practical benefits (including network reinforcement cost reduction [3] and energy loss reduction [26]) from phase swapping are shown as follows:

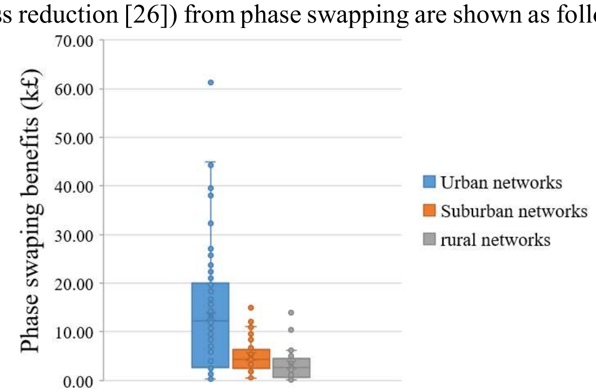


Fig. 5 Practical benefits form phase swapping

In Fig. 5, the average and maximum phase swapping benefits are: 1) £12,232 and £61,304, respectively, for urban LV networks; 2) £4,940 and £20,372, respectively, for suburban LV networks; and 3) £3,280 and £13,919, respectively, for rural LV networks.

2) Use one day's phase current data to infer the missing weighting factor factors

In this scenario, two average load profiles (given by (25)) are considered: 1) average workday profile; 2) average weekend profile. The average one day's load profiles are shown as follows:

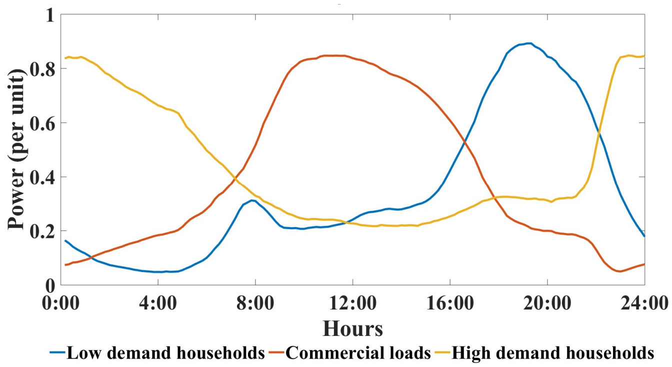


Fig. 6 Average workday profiles of constituent loads

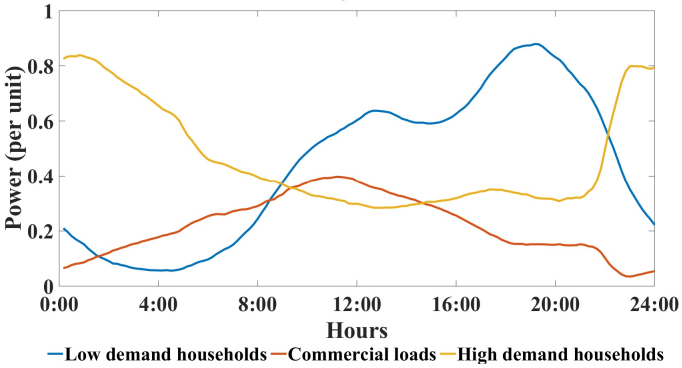


Fig. 7 Average weekend profiles of constituent loads

Select the data-rich network (the example in Section III – A) as a validation sample (treat this data-rich network as if it were data-scarce), this network has one workday's phase current data. If the average power of these typical customers (e.g. low demand household, high demand household and commercial load) are 0.2 kW, 0.4 kW and 0.8 kW, respectively, the phase swapping guidance are presented as follows:

TABLE III
A STATISTICAL PHASE SWAPPING GUIDANCE

Unit: kW, number of loads

	Phase a (kW)	Phase b (kW)	Phase c (kW)
Low demand households	–6.90, –35	6.14, 31	0.76, 4
High demand households	–3.34, -8	4.04, 10	–0.70, –2
Commercial loads	–2.12, –3	–0.35. –1	2.47, 4

The meaning of negative numbers and positive numbers are explained after Table I. For example, for phase a, 35 low demand households (which sums up to an average power of 6.90 kW), 8 high demand households (which sums up to an average power of 3.34 kW), and 3 commercial loads (which sums up an average power of 2.12 kW) should be moved away to other phases. The rebalancing potential for this validation sample is 0.1037. The rebalancing error e_s (defined in (35)) is approximately 10.1%. It indicates that, for this validation sample, the rebalancing potential RP_v (given by the proposed method) is 10.1% lower than the accurate RP.

Through validation, the average rebalancing errors are: 1) 14.3%, using the workday's phase current data; 2) 31.9%, using the weekend's phase current data for the data-scarce networks. In this case, using the workday's phase current data as the feature for the data-scarce network results in a greater reduction of the phase imbalance degree, compared with using the weekend's phase current data. If the phase swapping implementation strictly follows the developed guidance, the practical benefits (including network reinforcement cost reduction [3] and energy loss reduction [26]) from phase swapping are shown as follows:

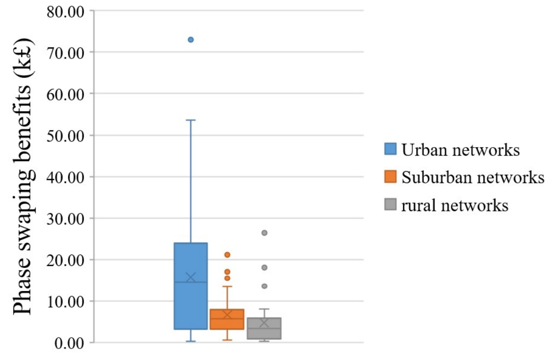


Fig. 8 Practical benefits form phase swapping

In Fig.8, the average and maximum phase swapping benefits are: 1) £15,884 and £72,981, respectively, for urban LV networks; 2) £6,674 and £21,125, respectively, for suburban LV networks; and 3) £4,710 and £26,457, respectively, for rural LV networks.

The average rebalancing errors of 19.33% and 14.3% are acceptable because our statistical approach requires very few data from data-scarce networks: in Scenario 1), only yearly average phase currents are required; in Scenario 2), only one-day's time-series phase currents are required. There is a tradeoff between the data requirement from data-scarce networks and the accuracy (measured by the rebalancing error) of the phase swapping guidance derived by the statistical approach. The more data required from these networks, the lower rebalancing error we will get, but more costs will be needed to obtain the additional data. In addition, the practical benefits presented in Fig. 5 and Fig. 8 demonstrate the effectiveness of the statistical approach.

D. *Implementation*

The developed approach is not a full phase balancing strategy but rather provides an important guidance for phase balancing for data-scarce LV networks at a minimal cost of monitoring. The guidance does not specify which connection point on the feeder is to be phase swapped and how, as this depends on the specific LV feeder topology and customers' phase connectivity which vary from case to case. Rather, the approach provides the following key guidance for phase swapping:

1) whether any given data-scarce network suffers from a serious phase imbalance or not;

2) whether any given data-scarce network has a phase imbalance direction or not. A phase imbalance direction refers to the existence of a particular phase that is consistently heavier (or lighter) than the other phases. Phase swapping is only applicable where there is a phase imbalance direction.

3) given any data-scarce network, move what load profiles from which phase to which phase in order to achieve near-balanced three phases (the results are presented in Tables II and III in the paper).

In other words, the above 1) and 2) inform whether any given LV network is worthy of phase swapping or not. If yes, to develop a network-specific phase swapping strategy, the following steps should be taken:

1) The DNOs should firstly obtain the network topology.

2) The type of each customer should be identified. The customers' phase connectivity should also be obtained.

3) Determine the points of phase swapping so that the phase swapping strategy closely follows the 3rd guidance as mentioned above.

E. Discussions

The developed statistical approach addresses a problem that no existing method can address: developing phase swapping guidance for data-scarce LV networks with neither network monitoring nor any metering from the customer side. The phase swapping guidance derived through the statistical approach serves as a benchmark. Future research can compare the effectiveness of their guidance with the one in this paper.

The statistical approach is designed to be generic. To apply this method to other countries, it requires the following steps: 1) Collect yearly time-series phase current data from *N* number of LV networks (*N* should be at least 800 hundred). These LV networks should be representative enough. 2) Set the number of constituent loads, so that the constituent loads are interpretable in that country, e.g. the low demand households in the UK. Then, the proposed method can be applied.

It should be noted that there are millions of LV networks in the UK alone. The fact that the statistical approach only requires the training data from 800 representative networks is not demanding, compared to requiring full data from each of the millions of networks in the UK. Furthermore, distribution network operators (DNOs) can reasonably monitor the full data for 800 networks at a moderate cost.

Phase balancing is most needed at LV (11kV/415V) substations. This is because a substation is a critical load node seen by higher-level networks. Phase balancing at the substation would prevent phase imbalance and its consequences from propagating to higher-level networks. Depending on individual circumstances, phase balancing may be extended beyond substations onto critical nodes on LV feeders. However, phase balancing at every node of the LV network is neither necessary nor feasible. Therefore, phase swapping is not required at all connection points but are only required at critical nodes, e.g. the substation. This significantly relieves the burden of phase balancing on a mass scale. Further, not all LV substations need phase balancing, only those with serious phase imbalance need balancing, thus further relieving the burden of phase balancing.

This paper uses the average load profiles to approximate customers' loads. This approximation is justified in the following way: 1) the phase swapping guidance derived by the statistical approach which uses the average load profiles turns out to have satisfactory accuracy (the accuracy values are presented in Section III). This implicitly justifies the use of the average load profiles. 2) The derived load profiles are typical load profiles for different types of customers (e.g. low demand households, high demand households and commercial loads) because the non-negative matrix factorization has clustering property.

Training data of less durations are also used to develop the phase swapping guidance. The accuracies of the guidance are presented in Table IV:

TABLE IV
EFFECTIVENESS COMPARISON OF DEVELOPED PHASE SWAPPING GUIDANCE

	Year-round data	Half year data	One month data	One week data
Rebalancing error	14.3%	16.9%	22.11%	27.4%

The above results prove that our current practice of using year-round data yields the least rebalancing error. A greater rebalancing error indicates a lower reduction of phase imbalance (hence less effectiveness of the phase swapping guidance) for the data-scarce networks.

The statistical approach does not divide the training data from the 800 LV networks into urban, suburban, and rural groups, because: 1) If the division were made, the training data in each group would be insufficient, thus compromising the accuracy of the developed phase swapping guidance; and 2) the case studies yield phase swapping guidance of satisfactory accuracies for data-scarce LV networks. This in turn justifies the practice of training the model on the 800 LV networks as a whole rather than dividing these networks into three groups.

Although the average phase current values are not yet collected from all LV networks, they can be obtained via the following means at minimal costs:1) The average phase current values can be derived from energy meter data if the energy consumption is recorded per phase. 2) The average phase current values can be obtained from the protection systems, which monitor the network operation status over time. 3) A recent project, OpenLV, sponsored by Western Power Distribution and undertaken by EA Technology, monitors a range of LV (11kV/415V) substations and the collected data include the average phase current values [27]. This paper advocates the collection of the average phase current values for the purpose of developing phase swapping guidance.

The developed approach yields effective phase swapping guidance with satisfactory accuracy for the sample networks we have. These networks have a low penetration of PVs and EVs, representing the status quo in Western Power Distribution's business areas. To account for increasing single-phase PVs and EVs, the approach can be adapted by updating the average load profiles to account for single-phase PVs and EVs. This also requires the monitoring of representative PV/ EV-rich, data-rich substations. Then the developed approach can learn the knowledge and extrapolate it to PV- or EV-rich, data-scarce substations. This is part of the future work.

IV. CONCLUSIONS

This paper addresses an unresolved problem for distribution network operators (DNOs): develop phase swapping guidance for data-scarce low voltage (415V, LV) networks with neither time-series network measurements nor customer metering data. To achieve this, this paper develops a new statistical phase swapping approach, extrapolating knowledge from 800 representative data-rich networks to data-scarce networks. This approach produces phase swapping guidance that guides the DNOs to reallocate typical loads (e.g. low demand households, high demand households and commercial loads) among the three phases, thus rebalancing the three phases of data-scarce networks.

Case studies are performed to validate the statistical phase swapping approach, which achieves effective reductions of the phase imbalance degrees for data-scarce networks: the reduction of phase imbalance degree is only 14.3% lower than that for data-rich networks. If DNOs follow the phase swapping guidance produced by the statistical approach, energy losses would be reduced and the network capacity wasted by phase imbalance would be released, while only a minimal amount of data is required.

References

[1] J. D. Watson, N. R. Watson, and I. Lestas, "Optimized dispatch of energy storage systems in unbalanced distribution networks," *IEEE Transactions on Sustainable Energy,* vol. PP, no. 99, pp. 1-1, 2017.

[2] L. F. Ochoa, R. M. Ciric, A. Padilha-Feltrin, and G. P. Harrison, "Evaluation of distribution system losses due to load unbalance," in 15th Power Systems Computation Conference PSCC 2005, 2005, pp. 1-4.

[3] K. Ma, R. Li, and F. Li, "Quantification of Additional Asset Reinforcement Cost From 3-Phase Imbalance " *IEEE Transactions on Power Systems,* vol. 31, no. 4, pp. 2885 - 2891 July, 2016.

[4] Z. Jinxiang, C. Mo-Yuen, and Z. Fan, "Phase balancing using mixed-integer programming [distribution feeders]," *IEEE Transactions on Power Systems,* vol. 13, no. 4, pp. 1487-1492, 1998.

[5] C.-C. Kuo, and Y.-T. Chao, "Energy management based on AM/FM/GIS for phase balancing application on distribution systems," *Energy Conversion and Management,* vol. 51, no. 3, pp. 485-492, 2010.

[6] D. Singh, R. K. Misra, and S. Mishra, "Distribution system feeder re-phasing considering voltage-dependency of loads," *International Journal of Electrical Power & Energy Systems,* vol. 76, pp. 107-119, 2016.

[7] G. Grigoraş, and M. Gavrilaş, "Phase swapping of lateral branches from low-voltage distribution networks for load balancing," in 2016 International Conference and Exposition on Electrical and Power Engineering (EPE), 2016, pp. 715-718.

[8] "HV and LV Phase Imbalance Assessment," https://www.spenergynetworks.co.uk/userfiles/file/HVandLVPhaseImbalanceAssessment16.pdf.

[9] 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, 2015.

[10] J. Zhu, G. Bilbro, and C. Mo-Yuen, "Phase balancing using simulated annealing," *IEEE Transactions on Power Systems,* vol. 14, no. 4, pp. 1508-1513, 1999.

[11] M. Huang, C. Chen, C. Lin, M. Kang, H. Chuang, and C. Huang, "Three-phase balancing of distribution feeders using immune algorithm," *IET Generation, Transmission & Distribution,* vol. 2, no. 3, pp. 383-392, 2008.

[12] R. A. Hooshmand, and S. Soltani, "Fuzzy Optimal Phase Balancing of Radial and Meshed Distribution Networks Using BF-PSO Algorithm," *IEEE Transactions on Power Systems,* vol. 27, no. 1, pp. 47-57, 2012.

[13] C. Lin, C. Chen, H. Chuang, M. Huang, and C. Huang, "An Expert System for Three-Phase Balancing of Distribution Feeders," *IEEE Transactions on Power Systems,* vol. 23, no. 3, pp. 1488-1496, 2008.

[14] L. Chia-Hung, C. Chao-Shun, C. Hui-Jen, and H. Cheng-Yu, "Heuristic rule-based phase balancing of distribution systems by considering customer load patterns," *IEEE Transactions on Power Systems,* vol. 20, no. 2, pp. 709-716, 2005.

[15] G. Mahendran, "Multi-objective Unbalanced Distribution Network Reconfiguration through Hybrid Heuristic Algorithm," *Journal of Electrical Engineering & Technology,* vol. 8, 2013.

[16] S. G. Xinbo Geng, Le Xie "Robust Look-ahead Three-phase Balancing of Uncertain Distribution Loads," https://arxiv.org/pdf/1810.00425.pdf.

[17] K. Wang, S. Skiena, and T. G. Robertazzi, "Phase balancing algorithms," *Electric Power Systems Research,* vol. 96, pp. 218-224, 2013.

[18] I. Mendia, S. Gil-López, J. Del Ser, A. G. Bordagaray, J. G. Prado, and M. Vélez, "Optimal Phase Swapping in Low Voltage Distribution Networks Based on Smart Meter Data and Optimization Heuristics." pp. 283-293.

[19] G. ANTOINE, L. D. ALVARO, and G. ROUPIOZ, "LARGE SCALE PHASE BALANCING OF LV NETWORKS USING THE AMM INFRASTRUCTURE " in 21st International Conference on Electricity Distribution Frankfurt, 2011.

[20] "LV network templates for a low-carbon future," https://www.westernpower.co.uk/docs/Innovation/Closed-projects/Network-Templates/LVNT-Appendix-A-Knowledge-Management.aspx.

[21] D. D. Lee, and H. S. Seung, "Algorithms for Non-negative Matrix Factorization," in Advances in Neural Information Processing Systems 13 (NIPS 2000), 2000.

[22] O. Berné, C. Joblin, Y. Deville, J. D. Smith, M. Rapacioli, J. P. Bernard, J. Thomas, W. Reach, and A. Abergel, "Analysis of the emission of very small dust particles from Spitzer spectro-imagery data using blind signal separation methods," *Astronomy & Astrophysics,* vol. 469, pp. 575-586, 2007.

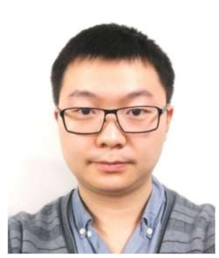
[23] C. Ding, X. He, and H. D. Simon, "On the equivalence of nonnegative matrix factorization and spectral clustering " in SIAM International Conference on Data Mining, 2005.

[24] J.-P. Zimmermann, M. Evans, J. Griggs, N. King, L. Harding, P. Roberts, and C. Evan. "Household Electricity Survey A study of domestic electrical product usage "; https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/208097/10043_R66141HouseholdElectricitySurveyFinalReportissue4.pdf.

[25] S. Electric. "Sepam™ Series 20 Protective Relays User's Manual," https://www.schneider-electric.com/resources/sites/SCHNEIDER_ELECTRIC/content/live/FAQS/221000/FA221290/en_US/63230-216-208C1_Sepam_Series_20_User_Manual.pdf.

[26] W. H. Kersting, *Distribution System Modeling and Analysis, 4th Edition*, pp. 39-77: CRC Press, taylor& Francis Group, 2017.

[27] R. Ash, T. Butler, R. Potter, and D. Hollingworth. "OPEN LV," https://openlv.net/wp-content/uploads/2017/10/OpenLV-Measurement-Points-V1.0.pdf.



Lurui Fang received his B.Eng. degree in Electrical Power Engineering at Chongqing University of Science and Technology in 2013. He received the MSc degree in Electrical Power Engineering at University of Southampton in 2014. He is currently a PHD student at University of Bath. His research focuses on the consequences analysis of three phase imbalance for low voltage (0.4kV) networks.

Kang Ma is working as a lecturer at University of Bath. His research focuses on the operation and planning of low voltage distribution networks. He worked as an R&D engineer at China Electric Power Research Institute (Beijing) from 2011 to 2014. He received his PhD degree in Electrical Engineering from the University of Manchester (U.K.) in 2011 and his B.Eng. degree from Tsinghua University (China) in 2007.

Xinsong Zhang received the B.Eng. degree in electrical engineering from the Xi'an University of Technology, Xi'an, China, in 2002, the M.Sc. degree in electrical engineering from Xi'an Jiaotong University, Xi'an, in 2005, and the Ph.D. degree from Hohai University, Nanjing, China, in 2013. He joined the Faculty of Nantong University, China, in 2006. From 2018 to 2019, he was an Academic Visitor with the Department of Electronic and Electrical Engineering, University of Bath, Bath, U.K. He is currently an Associate Professor with the School of Electrical Engineering, Nantong University. His research interests include power systems operation and planning, wind power integration, and energy storage systems.