Utility-Scale Estimation of Additional Reinforcement Cost from 3-Phase Imbalance Considering Thermal Constraints

Kang Ma, Ran Li, Furong Li, *Senior Member, IEEE*

***Abstract*—Widespread three-phase imbalance causes inefficient uses of low voltage (LV) network assets, leading to additional reinforcement costs (ARCs). Previous work that assumed balanced three phases underestimated the reinforcement cost throughout the whole utility by more than 50%. Previous work that quantified the ARCs was limited to individual network components, relying on full sensory data. This paper proposes a novel methodology that will scale the ARC estimation at a utility level, when the data concerning the imbalance of circuits or transformers are scarce. A novel statistical method is developed to estimate the volume of assets that need to be invested by identifying the relationship between the triangular distribution of circuit imbalance and that of circuit utilization. When there are more data available in future, accurate probability distributions can be constructed to reflect the network condition across the whole system. In light of this, two novel generalized ARC estimation formulas are developed that account for generic probability distributions. The developed methodology is applied to a real utility system in the UK, showing that: 1) three-phase imbalance leads to ARCs that are even greater than the reinforcement costs in the balanced case; 2) a 1% increase in the demand growth rate, the maximum degree of imbalance (DIB) and the maximum nominal utilization rate leads to over 10%, approximately 1% and 2% increases in the ARCs, respectively; and 3) the ARC is not sensitive to the minimum DIB values and the minimum nominal utilization rates.**

***Index Terms*—power distribution, power system economics, power transformer, low voltage, network investment, three-phase imbalance**

I. Introduction

WIDESPREAD three-phase imbalance causes inefficient uses of low voltage (LV) network assets [1, 2]. When facing the three-phase imbalance issue, a common approach for the UK distribution network operators (DNOs) is to perform network reinforcements when the asset capacity is reached, i.e. a passive approach. This leads to a network reinforcement cost being higher than if three phases were balanced – the difference is the additional reinforcement cost (ARC) as a result of three-phase imbalance. Three-phase imbalance causes inefficient uses of LV feeders (referred to as feeders) and LV transformers (referred to as transformers) in different ways: for a feeder, the phase with the greatest power among the three phases uses up the per-phase capacity when the other two phases still have spare capacities, thus triggering network reinforcements earlier than if the three phases were balanced [1]; for a transformer, three-phase imbalance induces neutral line power, which reduces the available capacity of the transformer, causing network reinforcements to be earlier than the balanced case [1]. Such inefficient uses of assets cause the ARCs. Quantifying ARCs on a utility scale would help DNOs to appraise whether it is sensible to take the passive network investments approach or a proactive phase balancing approach [3-5]. However, the challenges towards the ARC estimation at a utility scale are: 1) the extensiveness of LV networks with varying characteristics across different regions [6]; and 2) the limited monitoring data available [7]. This paper focuses on the estimation of the ARCs driven by thermal constraints on a utility scale, based on the sensory data of hundreds of LV substations.

A number of papers and reports investigated the LV network investment costs, but very few of them focuses on a utility scale. Reference [6] estimated utility-scale network investment costs for the first time, but with serious limitations: although reference [6] used triangular distributions to model loading levels, it did not model the degree of three-phase imbalance; nor did it identify the relationship between the loading level and the degree of imbalance – it assumed that LV networks have balanced three phases. Distinct from [6], this paper addresses the new problem of a utility-scale ARC estimation under three-phase imbalance by using triangular distributions to model the degree of imbalance and by developing a novel numerical method to establish the relationship between the degree of imbalance and the loading level. We will later demonstrate that Reference [6] seriously underestimated the reinforcement costs by wrongfully assuming balanced three phases.

Other papers that focus on LV network investments [8-10], LV network expansion planning [11, 12], and smart network planning strategies [13-15] all assumed balanced three phases; and they are not on a utility scale. References [2, 16] identified the impact of three-phase imbalance on network reinforcements for individual feeders qualitatively.

References [1, 17] which estimated the ARCs for individual network assets have the following limitations:

1) The ARC in [1, 17] was defined differently from that in this paper. References [1, 17] defined the ARC as the present value of a future investment cost, where the time span is from now to when the network margin is used up under long-term demand growth, i.e. over the lifetime of the asset. This paper, however, defines the ARC within a fixed interval in the future, e.g. the next 5 years. Apart from the interest in understanding assets' lifetime ARC, DNOs are also interested in the ARC defined in this paper, i.e. the ARC

This work was sponsored the Engineering and Physical Science Research Council, UK, under Grant RC-EE-1057.

Kang Ma, Ran Li, and Furong Li are with University of Bath. (e-mail: K.Ma@bath.ac.uk).

within a fixed interval in the future, because this will help DNOs make rolling business plans for the next *N* years.

2) References [1, 17] were limited to calculating their version of the ARCs for individual LV network components with full sensory data that are available only for a minority of network components. The methods in [1, 17] do not work if the full data are not available. In contrast, the new method in this paper requires minimum data from the DNO for the calculation of the ARC on a whole system level. Also, the methods in [1, 17] experience scalability issue when the problem involves the whole system with millions of networks, because those methods calculate the ARC on an asset by asset basis. For the whole system with millions of assets, if the methods in [1, 17] are applied, the calculation has to be repeated for millions of times, causing considerable computation burden. In contrast, the method in this paper is not on an asset by asset basis. Later we will develop two formulas, one calculating the ARC for all feeders in the whole system and the other calculating the ARC for all transformers. In summary, this paper resolves the challenges of both data deficiency and scalability by proposing i) triangular distributions to model the degree of three-phase imbalance (which is also distinct from [6]); and ii) a novel numerical method (in Section IV) that are fundamentally different from the methodologies in [1, 17].

There is a lack of a method to estimate the ARCs from three-phase imbalance at a utility level with limited sensory data. To scale up the ARC calculation onto a utility level with scarce information concerning the circuit/transformer imbalance, we for the first time model the full distributions of circuit/transformer imbalance across the whole system with limited information; we develop a novel statistical method to estimate the ARC by identifying the relationship between the triangular distribution of circuit imbalance and that of circuit utilization; and we propose two novel generalized ARC formulas in the integral form that account for generic probability distributions.

The major advantage of our approach is that it effectively addresses the scalability and data deficiency barriers that a utility-scale ARC estimation faces.

The rest of this paper is organized as follows: Section II presents the overall procedure for estimating the utility-scale ARCs; Section III models the degree of imbalance (DIB) at a utility level by three steps; Section IV calculates the ARCs for both feeders and transformers; Section V presents a case study; and Section VI concludes the paper.

II. Overall Procedure for Utility-Scale ARC Estimation

LV network reinforcements are driven by thermal and voltage constraints due to either demand growth or generation growth [6]. This paper considers thermal constraints as the cost driver.

This work is based on three-phase LV networks popular in the UK [18]. The methodology, however, is not limited to the UK networks, but is applicable to European-style three-phase LV networks. The methodology has the following characteristics: 1) it models the degree of imbalance (DIB) at a utility level by triangular distributions to overcome the data deficiency barrier; 2) it identifies the relation between nominal utilization rates and the DIB thresholds, and incorporates the thresholds into the triangular distribution of the DIB; and 3) based on the triangular distributions of the nominal utilization rates and the DIB values, the method calculates the proportion of assets of which: i) the thermal capacities are reached under three-phase imbalance; and ii) the thermal capacities would not be reached if the three phases were balanced.

The input data are the three-phase voltages and currents data over a year monitored at the secondary side of 642 LV (11kV/400V) substations (referred to as the substations) with a sampling interval of 15 minutes. Based on the input data, the overall procedure of the statistical approach is introduced as follows: 1) Derive the triangular distributions of the DIB values for the feeders and transformers of three substation groups, i.e. the rural, suburban, urban groups. 2) Identify the proportion of assets that contribute to the ARC, based on the triangular distribution of the asset nominal utilization rates [6] and that of the DIB values. 3) Calculate the utility-scale ARCs for both the feeders and transformers of the three substation groups, based on the proportion of the ARC-contributing assets.

Step 1) is further divided into three sub-steps, as introduced in Section III. Steps 2 – 3) are introduced in Section IV.

It should be noted that the methodology proposed in this paper applies to demand-dominated LV networks with a low penetration of renewable generation.

III. Utility-Scale Modelling of Degree of Imbalance

The major challenge for deriving the distribution of DIB is the lack of data at a utility level with millions of customers. This section addresses the challenge by three steps: 1) calculate two characteristic DIB values (one for the feeders and the other for the transformer) for each substation; 2) group the characteristic DIB values into three substation groups by applying k-means clustering to the peak demands of the sample substations; and 3) model the DIB values of each substation group as triangular distributions.

A. Determine Characteristic DIB Values for Substations

The two mathematical definitions of the degree of imbalance (DIB), one for feeders and the other for transformers, are both defined under an instantaneous peak demand in [1].

In this paper, the peak demand P_{peak} is defined as a range for each substation. The peak demand range is given by

$$P_{peak} \in [K_{peak} P_{\max}, P_{\max}] \tag{1}$$

where K_{peak} ($K_{peak} < 1$) and $P_{\max}$ denote the lower threshold of the peak demand range and the maximum instantaneous power of the year, respectively.

A time point when the power falls within this range is regarded as a peak time point. Each substation therefore has a series of DIB values, each corresponding to a peak time point. A DIB for the feeders of a substation is expressed as the function of a peak time point *t*, as opposed to the original time-independent version in [1].

$$d_{\mathrm{IB\_f}}(t) = \frac{\max\{P_{\mathrm{a}}(t), P_{\mathrm{b}}(t), P_{\mathrm{c}}(t)\} - \frac{P_{\mathrm{total}}(t)}{3}}{P_{\mathrm{total}}(t)} \tag{2}$$

where $P_{\emptyset}(t)$ ($\emptyset = \mathrm{a,b,c}$) denotes the power on phase $\emptyset$ at time *t*; $P_{\text{total}}(t)$ denotes the total power of the three phases at time *t*, i.e. $P_{\text{total}}(t) = \sum_{\emptyset=\text{a,b,c}} P_\emptyset(t)$.

A DIB for the transformer of a substation is also given as the function of a peak time point *t,* as opposed to the original time-independent version in [1].

$$d_{\text{IB\_T}}(t) = \frac{P_{\text{n}}(t)}{P_{\text{total}}(t)} = \frac{\sqrt{P_a(t)^2+P_b(t)^2+P_c(t)^2-P_a(t)P_b(t)-P_a(t)P_c(t)-P_b(t)P_c(t)}}{P_a(t)+P_b(t)+P_c(t)} \quad (3)$$

where $P_{\text{n}}(t)$ denotes the neutral line power at time *t*. Other variables have the same definitions as in (2).

It should be noted that the definition of DIB in [1] and the definition of voltage imbalance in [19] both involve the deviation from average values, but the former is based on power whereas the latter is based on voltage. Furthermore, unlike in [20-22], the definition of DIB does not involve symmetrical components.

Two series of DIB values can be derived for each substation from the raw data, the first for the feeders and the second for the transformer: $\text{D}_{\text{IB\_f}} = \{d_{\text{IB\_f}}(t_1), d_{\text{IB\_f}}(t_2), \dots, d_{\text{IB\_f}}(t_{\text{n}})\}$; $\text{D}_{\text{IB\_T}} = \{d_{\text{IB\_T}}(t_1), d_{\text{IB\_T}}(t_2), \dots, d_{\text{IB\_T}}(t_{\text{n}})\}$, where t_i denotes a peak time point.

For each substation, a characteristic DIB value ($\overline{d_{\text{IB\_f}}}$) is calculated as the mean of $d_{\text{IB\_f}}(t_i)$ where $i = 1 \dots n$. Similarly, for each substation, a characteristic DIB ($\overline{d_{\text{IB\_T}}}$) is calculated as the mean of $d_{\text{IB\_T}}(t_i)$ where $i = 1 \dots n$.

Each substation *i* therefore has two characteristic DIB values, i.e. $\overline{d_{\text{IB\_f}i}}$ and $\overline{d_{\text{IB\_T}i}}$, the former for the feeders and the latter for the transformer. They are temporal means.

B. DIB Values Grouping

LV networks serve a diverse range of customers. The percentages of different types of customers and the population/demand density vary considerably across regions, corresponding to highly diverse LV feeder utilization rates.

It is a common practice in the UK to classify LV networks into three categories: rural, suburban, and urban substations [23]. The sample substations are clustered into three groups by applying k-means clustering to the peak demands of these substations, based on the assumption that the three substation groups have the following ascending order of peak demands: rural, suburban, and urban substation clusters. The ascending order of peak demands for rural, suburban, and urban substation groups is justified by: 1) the increasing population density for rural, suburban, and urban areas; and 2) the consistency with the ascending order of the UK's prevalent capacities for rural, suburban, and urban transformers [24].

The k-means clustering of the substations is to find

$$\text{argmin} \sum_{j=1}^{3} \sum_{x \in S_j} (x - P_{peakj})^2 \quad (4)$$

where S_j and P_{peakj} denote substation group *j* and the peak demand of substation group *j*, respectively.

Each substation *i* exclusively belongs to one of the three substation groups S_j, and the two characteristic DIB values ($\overline{d_{\text{IB\_f}i}}$ and $\overline{d_{\text{IB\_T}i}}$) of substation *i* are mapped to substation group S_j: $\overline{d_{\text{IB\_f}i}}, \overline{d_{\text{IB\_T}i}} \rightarrow S_j$.

C. Triangular Distribution of DIB

Triangular distribution is a 'lack-of-information' distribution, which allows a full probability distribution to be established with only three parameters, i.e. the maximum, the minimum, and the mode [25].

The major challenge in modelling the utility-scale DIB distribution is data deficiency – the cost of obtaining the data for all LV networks within a utility's business zone is prohibitively high. From the sample data, the useful information that can be derived are the maximum DIB values, the minimum ones, and the mode for both the feeders and transformers of each of the three substation groups. Furthermore, the capability of the triangular distribution to represent skewness in both the left and the right directions makes it a suitable choice for this application [25] (symmetric distributions such as normal distributions are not suitable because they are unable to model the skewness nature). Because of the above reasons, triangular distribution is suitable for the modelling of the DIB values on a utility scale.

For feeders, the maximum and the minimum of the DIB values (denoted as $d_{\text{IB\_fmax}j}$ and $d_{\text{IB\_fmin}j}$, respectively) are obtained for each substation group S_j.

$$d_{\text{IB\_fmax}j} = \max\{\overline{d_{\text{IB\_f}i}}\} \text{ where } i \in S_j \quad (5)$$

$$d_{\text{IB\_fmin}j} = \min\{\overline{d_{\text{IB\_f}i}}\} \text{ where } i \in S_j \quad (6)$$

where $\overline{d_{\text{IB\_f}i}}$ and S_j denote the DIB for the feeders of substation *i* and substation group *j*, respectively.

The mode of the DIB for the feeders of substation group S_j, $d_{\text{IB\_fmod}j}$, is found by the following steps: 1) divide the set of $\overline{d_{\text{IB\_f}i}}$ (where $i \in S_j$) into 10 sections with equal length; 2) find the modal section *m*, i.e. the section which the most number of $\overline{d_{\text{IB\_f}i}}$ falls into. In other words, section *m* is the one with the maximum frequency. And 3) calculate the mode $d_{\text{IB\_fmod}j}$ by the formula introduced in [26].

$$d_{\text{IB\_fmod}j} = l + \frac{f_m - f_1}{(f_m - f_1) + (f_m - f_2)} \times h \quad (7)$$

where l and f_m denote the lower boundary and the frequency of section *m*, respectively; f_1 and f_2 denote the frequency of the sections preceding section *m* and following section *m*, respectively; and h denotes the section length.

Similarly, the maximum and the minimum of the DIB values (denoted as $d_{\text{IB\_Tmax}j}$ and $d_{\text{IB\_Tmin}j}$, respectively) are found for the transformers of each substation group S_j.

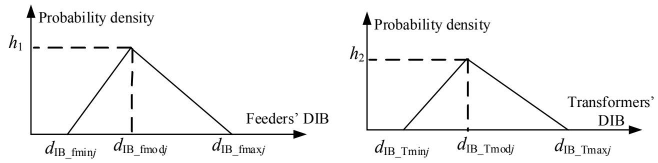
$$d_{\text{IB\_Tmax}j} = \max\{\overline{d_{\text{IB\_T}i}}\} \text{ where } i \in S_j \quad (8)$$

$$d_{\text{IB\_Tmin}j} = \min\{\overline{d_{\text{IB\_T}i}}\} \text{ where } i \in S_j \quad (9)$$

where $\overline{d_{\text{IB\_T}i}}$ denotes the DIB for the transformer of substation *i*.

The mode of the DIB for transformers, $d_{\text{IB\_Tmod}j}$, is calculated in a similar way to $d_{\text{IB\_fmod}j}$ in (7).

From the above procedure, two triangular distribution functions, one for the feeders' DIB values and the other for the transformers' DIB values, are obtained for each substation group S_j , as illustrated in Fig. 1. These distributions are essentially the spatial distributions of temporal means of the DIB – the Central Limit Theorem and normal distributions are not applicable in this case.



(a) Triangular distribution of the feeders' DIB values of substation group S_j
(b) Triangular distribution of the transformers' DIB values of substation group S_j

Fig. 1 Triangular distribution of the DIB values

IV. UTILITY-SCALE ARC CALCULATION

Before calculating the utility-scale ARC, it is first necessary to define the ARC in the context of this paper: an ARC is defined as the investment cost of an asset (a feeder or a transformer) that meets the following criteria when three-phase peak demand occurs: 1) its thermal capacity is reached under three-phase imbalance; and 2) its thermal capacity would not be reached if the three phases were balanced. The definition has a practical implication: by the time point concerned (i.e. the time point when the ARC is calculated), the DNO would not have to pay an investment cost to reinforce such an asset if the three phases were balanced; but the DNO actually has to pay an investment cost (i.e. the ARC) for this asset under three-phase imbalance. The assets that meet the above two criteria are the ARC-contributing assets.

A utility-scale ARC is the summation of the individual ARCs over the utility. The key to a utility-scale ARC calculation is to find the percentage of assets that meet the above two criteria. This section presents a novel methodology to calculate the utility-scale ARC based on the triangular distribution of the DIB values and that of the nominal utilization rates. The methodology is elaborated for both feeders and transformers.

It should be noted that thermal capacity is originally expressed as apparent power (kVA). However, because phasor measurements are absent in low voltage networks and that smart meters monitor real power only, there is a lack of reactive power data in low voltage networks. This paper therefore transforms thermal capacities that are expressed by apparent power (kVA) into those that are expressed by real power (kW), assuming an average load power factor of 0.95. All calculations are based on real power.

A. Utility-Scale ARC Calculation for Feeders

This section presents the method to calculate the ARC for the feeders of any given substation group *j*.

The nominal utilization rate of a feeder (denoted as *U*) is defined as the total power of the three phases over the capacity of the feeder under peak demand [1]. It is essentially the peak loading level when the three phases are assumed to be balanced. According to [6], *U* is modelled as a triangular distribution, as shown in Fig. 2.

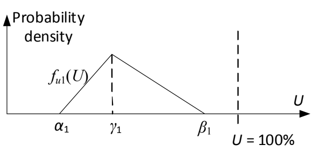


Fig. 2 The triangular distribution of feeders' nominal utilization rates

The probability density function of U is $f_{u1}(U)$. α_1, β_1, and γ_1 denote the minimum U, the maximum U, and the mode of the distribution, respectively. The vertical dashed line in Fig. 2 corresponds to $U = 100\%$. After *N* years' demand growth, the distribution of *U* shifts towards the right, as shown in Fig. 3.

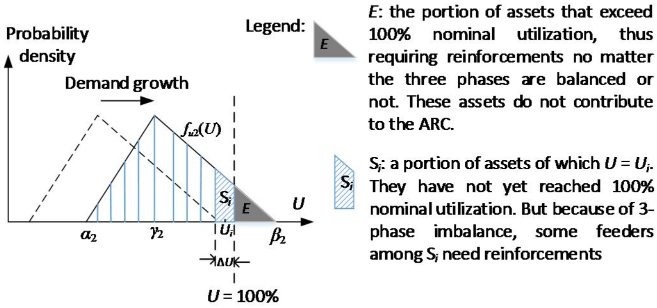


Fig. 3 The triangular distribution of feeders' nominal utilization rates after demand growth

Suppose that the annual demand growth rate is *r*. The maximum, the minimum, and the mode of the nominal utilization rate are denoted as α_2, β_2, and γ_2, respectively. By the end of the *N*th year, α_2, β_2, and γ_2 are given by

$$\begin{bmatrix}\alpha_2\\ \beta_2\\ \gamma_2\end{bmatrix} = \begin{bmatrix}\alpha_1\\ \beta_1\\ \gamma_1\end{bmatrix}(1+r)^N \tag{10}$$

where r denotes the annual demand growth rate.

In this paper, we choose *N* = 5 for the base case because it is common for DNOs in the UK to make 5-year business plans in advance [6, 27]. However, it should be noted that the methodology is generic in terms of the choice of *N*.

The new probability density function of the nominal utilization rate is $y = f_{u2}(U)$, defined in the range $[\alpha_2, \beta_2]$.

$$f_{u2}(U) = \begin{cases} \frac{2}{(\beta_2-\alpha_2)(\gamma_2-\alpha_2)}\cdot U - \frac{2\alpha_2}{(\beta_2-\alpha_2)(\gamma_2-\alpha_2)} & \text{when } \alpha_2 \le U \le \gamma_2 \\ \frac{2}{(\beta_2-\alpha_2)(\gamma_2-\beta_2)}\cdot U - \frac{2\beta_2}{(\beta_2-\alpha_2)(\gamma_2-\beta_2)} & \text{when } \gamma_2 < U \le \beta_2 \end{cases} \tag{11}$$

In this case, there is $\beta_2 > 100\%$. According to Fig. 3, the area of *E* represents the proportion of assets of which the nominal utilization rate $U \geq 100\%$. They need to be reinforced no matter their phases are balanced or not. In other words, the investment costs of these assets are not triggered by three phase imbalance, hence these assets do not contribute to the ARC.

Divide the range of $[\alpha_2, 100\%)$ into *M* sections, where *M* is a very large number (*M* > 1000). For each section i with a length of ΔU, the nominal utilization rate is taken as the middle point of this section, denoted as U_i. The area S_i enclosed by section i and $f_{u2}(U)$, as depicted in Fig. 4, is given by

$$S_i = f_{u2}(U_i)\Delta U \tag{12}$$

S_i represents a proportion of feeders of which $U < 100\%$. However, among S_i, some feeders of which the 'heaviest' phase (the phase with the maximum power) has already reached the per-phase capacity, would trigger reinforcements – they contribute to the ARC which arises from three-phase imbalance. As a proportion among S_i, these feeders meet the two criteria: 1) their thermal capacities are reached under three-phase imbalance; and 2) their thermal capacities would not be reached if the three phases were balanced.

Section *i* corresponds to a DIB threshold $D_{IB\_thre\_i}$, which is a function of U_i; when a feeder from S_i has a DIB that exceeds $D_{IB\_thre\_i}$, the feeder requires reinforcements, thus contributing to the ARC; when a feeder from S_i has a DIB that is lower than $D_{IB\_thre\_i}$, the feeder does not need reinforcements yet, thus not contributing to the ARC.

$D_{IB\_thre\_i}$ is obtained by the following two steps: firstly, the time horizon of a feeder in the imbalanced case (the number of years it takes for the peak demand of the 'heaviest' phase to reach the per-phase capacity of the feeder) was referenced from our previous paper [1]:

$$n_{\mathrm{IB}} = -\frac{\log U_i + \log(3D_{IB} + 1)}{\log(1 + r)} \tag{13}$$

where U_i, D_{IB} and r denote the nominal utilization rate, the DIB, and the annual demand growth rate, respectively.

Then, $D_{IB\_thre\_i}$ is obtained by solving the equation $n_{\mathrm{IB}} = 0$ (a zero time horizon means that the peak demand of the 'heaviest' phase reaches the capacity of that phase), subject to the natural boundary of $D_{IB\_thre\_i}$, i.e. $0 \leq D_{IB\_thre\_i} \leq 2/3$. Therefore,

$$D_{IB\_thre\_i} = g(U_i) = \min\{\frac{1}{3U_i} - \frac{1}{3}, \frac{2}{3}\} \tag{14}$$

Equation (14) shows that $D_{IB\_thre\_i}$ is a function of U_i.

Incorporate $D_{IB\_thre\_i}$ into the triangular distribution of the feeders' DIB for substation group j, as depicted in Fig. 4.

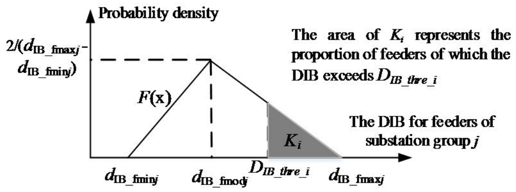


Fig. 4 The portion of feeders of which the DIB exceeds the threshold

The area of K_i represents, among S_i, the proportion of feeders of which the DIB exceeds $D_{IB\_thre\_i}$. K_i is given by

$$K_i = \int_{D_{IB\_thre\_i}}^{d_{IB\_fmaxj}} F_{DIB}(x)\mathrm{d}x = \int_{g(U_i)}^{d_{IB\_fmaxj}} F_{DIB}(x)\mathrm{d}x \tag{15}$$

where $F_{DIB}(x)$ is the probability density function of the triangular distribution of the feeders' DIB values for substation group j. $F_{DIB}(x)$ is derived in Section III. $d_{\mathrm{IB\_fmax}j}$, quantified in (5), denotes the feeders' maximum DIB value for substation group j. Equation (15) shows that K_i is a function of U_i.

The area of K_i (the grey area) in Fig. 4 represents the proportion of feeders that contribute to the ARC because they meet the two criteria detailed at the beginning of this section: Criterion 1) is met because these feeders are imbalanced enough that they have at least one phase of which the power reaches the thermal capacity of that phase; Criterion 2) is met because, if the three phases were perfectly balanced, each phase would have a loading level of U_i ($U_i < 100\%$) that is less than 100%.

Considering section i, the feeders that contribute to the ARC take the following percentage out of all feeders of substation group j.

$$P_i = S_i K_i = f_{u2}(U_i)[\int_{g(U_i)}^{d_{IB\_fmaxj}} F_{DIB}(x)\mathrm{d}x]\Delta U \tag{16}$$

P_i is a function of U_i.

The total proportion of feeders that contribute to the ARC is the summation of P_i for all sections.

$$P_{fj} = \sum_{i=1}^{M} P_i \tag{17}$$

where subscript f denotes feeders; subscript j denotes substation group j.

Among all feeders of substation group j, P_{fj} is the proportion of feeders of which: 1) the thermal capacities are reached under three-phase imbalance; and 2) the thermal capacities would not be reached if the three phases were balanced.

Therefore, the ARC for all the feeders of substation group j is given by

$$ARC_{fj} = N_{fj} C_{fj} P_{fj} \tag{18}$$

where subscript f denotes feeders; N_{fj} denotes the number of feeders belonging to substation group j; C_{fj} denotes the average investment cost of an individual feeder of substation j.

The above process calculates the ARC by discretizing functions $f_{u2}(U)$, $g(U)$, and $F_{DIB}(x)$, which are defined in (11), (14), and (15), respectively. This is the practical method adopted in the case study.

As a generalization from the above method, the ARC is presented as a generic formula in the integral form:

$$ARC_{fj} = N_{fj} C_{fj} \int_{\alpha_2}^{100\%} f_{u2}(U)[\int_{g(U)}^{d_{IB\_fmaxj}} F_{DIB}(x)dx]dU \tag{19}$$

where $f_{u2}(U)$ is the probability density function of the nominal utilization rate; $g(U)$ is defined in (14); and $F(x)$ is the probability density function of the feeders' DIB values.

Equation (19) is generic in the sense that it does not specify that $f_{u2}(U)$ or $F_{DIB}(x)$ has to represent any particular probability distributions. In other words, $f_{u2}(U)$ and $F_{DIB}(x)$ represent generic probability distributions. In this work, they are realized as triangular distributions: $f_{u2}(U)$ and $F_{DIB}(x)$ are defined in (11) and (15), respectively. The use of triangular distributions is justified in Section III – C.

B. Utility-Scale ARC Calculation for Transformers

The process to calculate the transformers' ARC is similar to that for feeders. The ARC of the transformers for substation group j is given as a generic formula:

$$ARC_{Tj} = N_{Tj} C_{Tj} \int_{\alpha_{T2}}^{100\%} f_{uT2}(U)[\int_{g_T(U)}^{d_{\mathrm{IB\_Tmax}j}} F_T(x)dx]dU \tag{20}$$

where subscript T denotes transformers. The nominal utilization rate of a transformer is denoted as U. N_{Tj} denotes the number of transformers belonging to substation group j; C_{Tj} denotes the average investment cost of an individual transformer of substation group j; α_{T2} denotes the minimum utilization rate of the transformers after N years' demand growth; $d_{\mathrm{IB\_Tmax}j}$, quantified in (8), denotes the transformers' maximum DIB value for substation group j; $f_{uT2}(U)$ is the probability density function of the probability distribution (e.g. triangular distribution) of the nominal utilization rates for the transformers, after N years' demand growth; $F_T(x)$ is the probability density function of the probability distribution (e.g. triangular distribution) of the transformers' DIB values, derived in Section III; and $g_T(U)$ is the transformer's threshold DIB as a function of the transformer's nominal utilization rate.

The following process derives $g_T(U)$, which is incorporated into (20):

The original formula for calculating the time horizon of a transformer under imbalance (the number of years it takes for the three-phase power plus the neutral line power to reach the transformer capacity) was referenced from our previous paper [1]:

$$n_{\text{T\_IB}} = -\frac{\log U + \log(D_{\text{IB\_T}} + 1)}{\log(1 + r)} \tag{21}$$

where U, $D_{\text{IB\_T}}$ and r denote the nominal utilization rate, the DIB of the transformer, and the annual demand growth rate, respectively.

The DIB threshold is obtained by solving the equation $n_{\text{T\_IB}} = 0$, subject to the natural boundary of the DIB threshold, i.e. [0, 1.0]. Therefore,

$$g_T(U) = \min\{\frac{1}{U} - 1, 1\} \tag{22}$$

Similar to the feeders, the ARC of the transformers of substation group j can be calculated by discretizing (20), i.e. discretizing functions $f_{uT2}(U)$, $g_T(U)$, and $F_T(x)$.

C. Utility-Scale ARC Calculation under Uncertainties from Growing Electric Vehicles and Heat Pumps

Emerging heat pumps (HPs) and electric vehicles (EVs) are a key contributor to future demand growth in the UK. The uncertainty in the speed of their rollouts adds to the uncertainty of the annual demand growth rate r defined in (10). To reflect this uncertainty, the demand growth rate r is modelled as a random variable using confidence interval estimate: r is within a range at a confidence level α.

$$r \in [r_{\text{lower}}, r_{\text{upper}}] \quad \text{at confidence level } \alpha \tag{23}$$

Where r_{lower} and r_{upper} denote the lower and upper bounds of the estimate, respectively. It should be noted that $r_{\text{lower}} > 0$.

According to Fig. 3, such an uncertain r translates to an uncertain 'shift' of the triangular distribution to the right. The ARC first increases then decreases with the increase of r. This is because, when the triangular distribution shifts towards the right in Fig. 3, the area of E (the proportion of assets that exceed 100% nominal utilization) which does not contribute to the ARC increases with r, leading to an eventual decrease of the proportion of assets that may contribute to the ARC. It should be emphasized that the increase of r always lead to an increase in the total reinforcement cost of which the ARC is a component, but not necessarily lead to an increase in the ARC. In theory, the maximum ARC for feeders of substation group j occurs when $r = r_{fm}$ which satisfies:

$$\left.\frac{dARC_{fj}(r)}{dr}\right|_{r=r_{fm}} = 0 \tag{24}$$

Where $ARC_{fj}(r)$ is the ARC expressed as a function of r for feeders of substation group j.

Therefore, the ARCs for feeders fall in the range of:

$$ARC_{fj} \in [\min\{ARC_{fj}(r_{\text{lower}}), ARC_{fj}(r_{\text{upper}})\}, ARC_{fj}(r_{fm})] \quad \text{At confidence level } \alpha \tag{25}$$

Similarly, the maximum ARC for transformers of substation group j occurs when $r = r_{Tm}$ which satisfies:

$$\left.\frac{dARC_{Tj}(r)}{dr}\right|_{r=r_{Tm}} = 0 \tag{26}$$

Where $ARC_{Tj}(r)$ is the ARC as a function of r for transformers in substation group j.

Therefore, the ARCs for transformers fall in the range of:

$$ARC_{Tj} \in [\min\{ARC_{Tj}(r_{\text{lower}}), ARC_{Tj}(r_{\text{upper}})\}, ARC_{Tj}(r_{Tm})] \quad \text{At confidence level } \alpha \tag{27}$$

It is clear that the resultant ARCs are also confidence interval estimates.

Apart from the uncertainty in the demand growth rate r, the connections of EVs and HPs to the network will also bring uncertainties to phase imbalance. They may aggravate or reduce the degree of three-phase imbalance for individual networks: if connected to a phase with the least load among the three phases, they may reduce phase imbalance; otherwise, they may increase three-phase imbalance, depending on individual network circumstances. The uncertainties for individual networks are aggregated up to a utility scale, where the impact of EVs and HPs on phase imbalance is uncertain.

To reflect the uncertainties, the mode value of the degree of imbalance for feeders of substation group j is assumed to change by a percentage β during the study period, where β is modelled as a random variable using confidence interval estimate.

$$\beta \in [\beta_{\text{lower}}, \beta_{\text{upper}}] \quad \text{at confidence level } \alpha \tag{28}$$

Where β_{lower} and β_{upper} denote the lower and upper bounds of β, respectively. It should be noted that β can be positive or negative, representing either an increase or a decrease of phase imbalance at a utility scale.

$$d_{\text{IB\_fmod\_end}j} = d_{\text{IB\_fmod\_start}j}(1 + \beta) \tag{29}$$

Where $d_{\text{IB\_fmod\_end}j}$ denotes the feeder's mode DIB value at the end of the study period for substation group j. $d_{\text{IB\_fmod\_start}j}$ denotes the feeder's mode DIB value at the beginning of the study period for substation group j.

According to Fig. 4, the uncertainties in DIB values translate to an uncertain change of the triangular distribution skewness either to the left or to the right.

Given the study period of N years and the starting DIB mode, the ARC is expressed as a function of β by substituting (29) into (19). The ARC monotonically increases with β. The maximum ARC occurs when $\beta = \beta_{\text{upper}}$. Therefore, the ARCs for feeders fall in the range of:

$$ARC_{fj} \in [ARC_{fj}(\beta_{\text{lower}}), ARC_{fj}(\beta_{upper})] \quad \text{At confidence level } \alpha \tag{30}$$

Similar to feeders, the mode value of the degree of imbalance for transformers in substation j is assumed to change by an uncertain percentage β during the study period. The equations are the same as (28) – (29) except that the transformer's mode DIB replaces the feeder's mode DIB. The ARCs for transformers fall in the range of:

$$ARC_{Tj} \in [ARC_{Tj}(\beta_{lower}), ARC_{Tj}(\beta_{upper})] \quad \text{At confidence level } \alpha \tag{31}$$

V. APPLICATION OF THE PROPOSED METHOD

The methodology introduced in previous sections is applied to a real utility case, based on two sources of data: the network data from [6]; and the three-phase voltages and currents data from the substations. The power factor is set as 0.95. The annual growth rate of the peak demand is 2.1% [6].

A. Determine Characteristic DIB Values for Each Substation

The two characteristic DIB values, one for the feeders and the other for the transformers, are calculated for each substation.

The characteristic DIB values are plotted as histograms in Fig. 5.

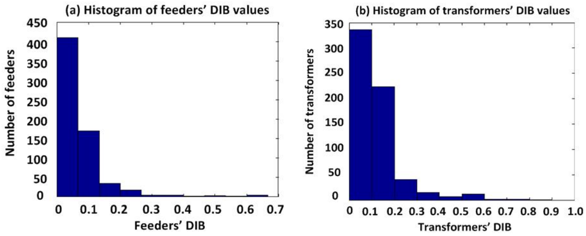


Fig. 5 The histograms of the DIB values for both the feeders and the transformers

Fig. 5 shows that the majority (>90%) of the substations have feeders' DIB values of less than 0.2 and transformers' DIB values of less than 0.3.

B. Substation Clustering

The substations are clustered into three groups by applying k-means clustering to their peak demands. The centroids (as the peak demands) of these clusters are 103.15kW, 234.55kW, and 437.80kW, corresponding to rural, suburban, and urban substation groups, respectively.

C. Triangular Distribution of DIB

Before estimating the ARCs, it is necessary to first compute the three key parameters for each triangular distribution function of the DIB values. Table I presents the maximum values, the minimum values, and the mode for the triangular distributions of the DIB values for the three substation groups.

TABLE I
THE PARAMETERS FOR THE TRIANGULAR DISTRIBUTIONS OF THE DIB VALUES

Feeders' DIB	Feeders' max DIB	Feeders' min DIB	Feeders' mode DIB
Rural	0.6528	0.0032	0.0510
Suburban	0.2820	0.0019	0.0407
Urban	0.1600	0.0039	0.0338
Transformers' DIB	**Transformers' max DIB**	**Transformers' min DIB**	**Transformers' mode DIB**
Rural	0.8544	0.0013	0.1255
Suburban	0.5407	0.0037	0.0832
Urban	0.4436	0.0069	0.0321

Table I shows that the feeders' and transformers' minimum DIB values for each substation group are close to zero – in each group, there are substations of which the three-phase power is almost balanced. In all three substation groups, the modal values are closer to the minimum values than to the maximum. This is consistent with the histograms in Fig. 5: the majority of the substations have a slight-to-moderate degree of three-phase imbalance.

Table I also shows that the maximum DIB values of the rural group are higher than those of the suburban and urban groups. This is because a minority (<10%) of rural substations exhibit a severe degree of phase imbalance, thus significantly affecting the maximum DIB values of the rural group as well as the shape of the triangular distributions of the rural DIB values.

D. Utility-Scale ARC Estimation for the Base Case

All input data, including circuits' costs per km and each transformer's cost for urban, suburban and rural areas are referenced from [6]. The period concerned is 5 years (from 2011 to 2015). During this period, the annual demand growth rate is 2.1%, as presented above. According to the definition of the ARC in Section IV, the utility-scale ARC in this context represents, by the end of 2015, the investment costs of the utility's assets (feeders and transformers) of which the thermal capacities are reached under three-phase imbalance but would not be reached if the three phases were balanced. Table II presents the ARCs for each substation group and each type of assets. The ARC results are also depicted in Fig. 6. They demonstrate that the ARC of the urban group, dominated by that of the urban feeders, considerably outweighs the ARCs of the suburban and rural groups. This is due to the high average investment cost per urban feeder, which is 2.73 times of that per suburban feeder and 2.05 times of that per rural feeder. The ARC of the rural group is the lowest among the ARCs of the three groups, because the number of rural transformers and feeders are less than 40% and 20% of that of suburban transformers and feeders, respectively.

TABLE II
UTILITY-SCALE ARCS FOR THE RURAL, SUBURBAN, AND URBAN ASSETS

ARCs	Feeders (£k)	Trans (£k)	Subtotal (£k)
Rural	22,421.00	5,165.00	27,586.00
Suburban	59,562.00	32,226.00	91,788.00
Urban	132,105.00	40,612.00	172,717.00
Subtotal (£k)	214,088.00	78,003.00	292,091.00

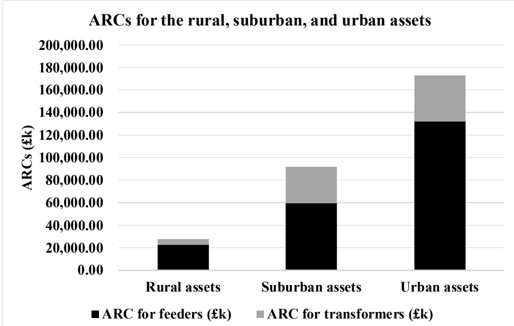


Fig. 6 The ARCs for the rural, suburban, and urban assets

Table III presents the percentages of feeders/transformers that contribute to the ARC, out of the total number of feeders/transformers of the corresponding group.

TABLE III
THE PERCENTAGES OF THE ARC-CONTRIBUTING ASSETS OUT OF THE TOTAL NUMBER OF ASSETS

	ARC-contributing feeders (%)	ARC-contributing transformers (%)
Rural	27.80%	10.86%
Suburban	16.04%	9.28%
Urban	16.02%	12.53%

The ARC results presented in Table II are significantly higher than the reinforcement costs (RCs) reported in [6], which assumed balanced three phases. This phenomenon is justified by the following reasons:

1) The three-phase imbalance is non-trivial for the rural, suburban, and urban groups. This is reflected by the percentage of the ARC-contributing assets out of the total number of assets for each substation group in Table III. Reference [6], however, ignores three-phase imbalance

completely, thus leading to a serious underestimation of the reinforcement costs.

2) According to [6], the reinforcement cost is zero at the beginning of the study period because no asset reaches its thermal capacity yet, and the reinforcement cost is solely contributed by the assets that reach the thermal capacities during the study period. However, according to this research, even at the beginning of the study period, i.e. when year $N = 0$, the ARCs are still positive for the three substation groups. This is because, when year $N = 0$, there are assets of which the thermal capacity is reached under three-phase imbalance. Without any reinforcements in between, by the end of the year $N = 5$, the ARCs are contributed by i) those ARC-contributing assets that pre-exist when year $N = 0$; and ii) the new ARC-contributing assets that emerge during the study period.

It should be noted that the study in Section V – D is performed on the base case with a deterministic demand growth rate and deterministic DIB values. Later in Section V – F, we will conduct the ARC analyses based on uncertain demand growth rates and DIB values.

E. Sensitivity Analysis

One of the key parameters for estimating the ARCs is the annual demand growth rate. We calculate the ARCs under different annual demand growth rates to investigate their impact on the ARCs. Five annual demand growth rates are considered for the 5-year study period from the beginning of 2011 to the end of 2015: 1.2%, 1.5%, 1.8%, 2.1%, and 2.3%. The ARCs under these annual demand growth rates are depicted in Fig. 7.

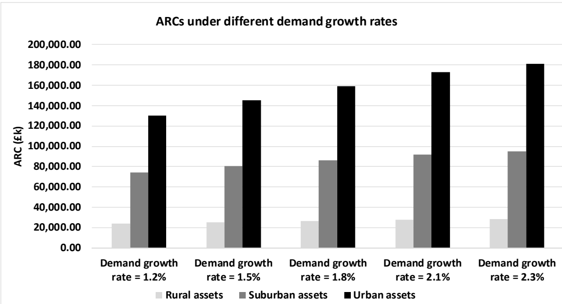


Fig. 7 The ARCs under different annual demand growth rates

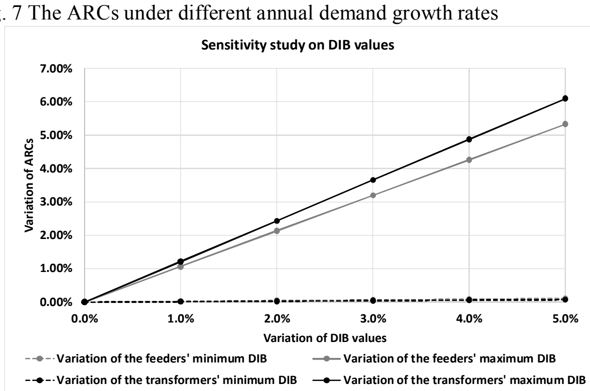


Fig. 8 Sensitivity study on DIB values

Within the given range of annual demand growth rates, the ARCs for rural, suburban and urban assets increase by 18.3%, 27.8%, and 39.5%, respectively, when the annual demand growth rate increases from 1.2% to 2.3%. The results demonstrate that the ARC is highly sensitive to the annual demand growth rate for the 5-year period.

The maximum and minimum DIB values for both the feeders and the transformers have an impact on the ARCs. The impact of the above parameters on the ARC is demonstrated on urban assets. Fig. 8 shows the impact of these parameters on the ARC. The results show that the minimum DIB values have a negligible impact on the ARC: an increase of the minimum DIB values by 5.0% causes the ARC to increase by less than 0.20%. In contrast, the increase of the maximum DIB values by 5.0% causes the ARC to increase almost linearly by 5.0% – 6.0%.

The maximum and minimum utilization rates also have an impact on the ARC. Fig. 9 shows the impact of these parameters on the ARC for urban assets, including feeders and transformers. The results show that the variations of the minimum utilization rate have a little impact on the ARC: an increase of the minimum utilization rate by 5.0% results in a 0.95% increase of the ARC. The variations of the maximum utilization rate, however, have a much greater impact on the ARC: an increase of the maximum utilization rate by 5.0% causes the ARC to increase by approximately 10.8%, but the speed of such an increase is decreasing.

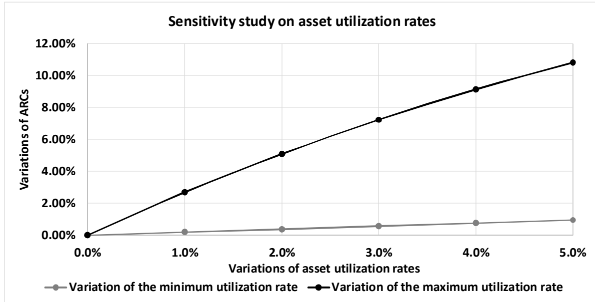


Fig. 9 Sensitivity study on asset utilization rates

The impact of the average individual asset investment cost on the ARC is also investigated. It is clear from (19) and (20) that the ARCs are linear functions of the average individual asset (feeder and transformer) investment costs C_{fj} and C_{Tj} for substation group j. The ARC results under different average investment costs C_{fj} per feeder per km is plotted in Fig. 10, where all other data are the same as in the base case.

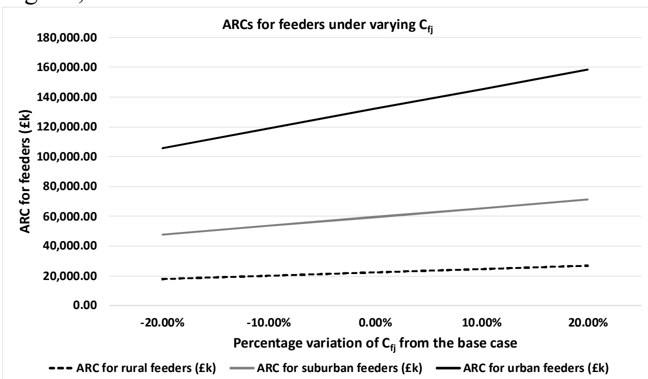


Fig. 10 The ARCs for feeders under different average investment costs per feeder per km

The results in Fig. 10 demonstrate the linear relationship between the ARC and C_{fj}.

The ARC results under different average investment costs C_{Tj} for an individual transformer is plotted in Fig. 11, where all other data are the same as in the base case.

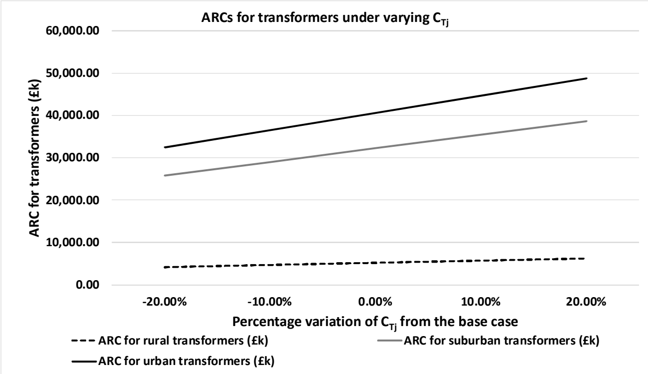


Fig. 11 The ARCs for transformers under different average investment costs per transformer

The results in Fig. 11 demonstrate the linear relationship between the ARC and C_{Tj}.

The impact of the study period *N* on the total ARC (the summation of the ARCs for rural, suburban, and urban groups) is investigated. *N* is chosen to be 5, 8, 10, and 15 years. The ARC results are plotted in Fig. 12, where all other data are the same as in the base case.

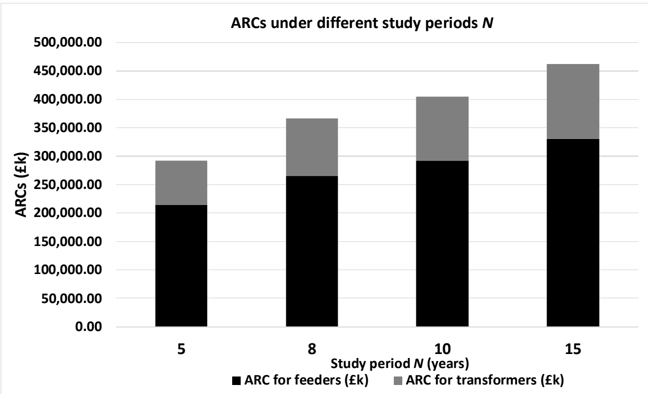


Fig. 12 The ARCs under different study periods *N*

The ARC results in Fig. 12 show that the total ARC increases with the study period *N*. This is justified because when *N* increases from 5 years to 15 years, the nominal utilization rates increase due to a greater amount of load growth, corresponding to a 'right-shift' of the triangular distribution in Fig. 3. Given that other conditions remain unchanged, the increase of the nominal utilization rates leads to a greater proportion of ARC-contributing assets. On the other hand, the increase of the area *E* in Fig. 3 is not yet enough to offset the increase of the proportion of ARC-contributing assets. Therefore, the ARC shows an increasing trend when *N* increases from 5 years to 15 years.

F. Utility-Scale ARC Estimation under Uncertain Demand Growth Rate and Degree of Imbalance

Emerging electric vehicles and heat pumps bring uncertainties to both the demand growth rate and the degree of imbalance. In this section, the ARCs are calculated under the uncertainties from EVs and HPs.

The ARCs are calculated under uncertain demand growth rate. The demand growth rate is assumed to be within the range $r \in [1.2\%, 5.0\%]$ at confidence level $\alpha = 95\%$. The ARC results are presented in Table IV at the same confidence level α.

TABLE IV
THE ARCS UNDER UNCERTAIN DEMAND GROWTH RATE

	ARC for feeders (£k)	ARC for transformers (£k)
Rural	[19,949 28,498]	[4,051 8,147]
Suburban	[49,656 82,033]	[24,875 49,164]
Urban	[100,089 199,694]	[29,897 63,956]
Total (for rural, suburban, and urban groups)	[169,694 310,225]	[58,823 121,267]

The total ARC for both feeders and transformers in all three substation groups is within the range of [£228,517k £431,492k] at confidence level α.

The ARCs are also calculated under uncertain degrees of imbalance. The percentage of change in mode DIB from the start to the end of the study period is β, as defined in (28). β is assumed to be within the range $\beta \in [-10.0\%, 10.0\%]$ at a confidence level $\alpha = 90\%$. The ARC results are presented in Table V at the same confidence level α.

TABLE V
THE ARCS UNDER UNCERTAIN DEGREE OF IMBALANCE

	ARC for feeders (£k)	ARC for transformers (£k)
Rural	[22,242 22,600]	[5,085 5,245]
Suburban	[58,727 60,405]	[31,743 32,713]
Urban	[127,475 134,768]	[40,322 40,904]
Total (for rural, suburban, and urban groups)	[208,444 217,773]	[77,150 78,862]

The total ARC for both feeders and transformers in all three substation groups is within the range of [£285,594k £296,635k] at confidence level α.

G. Discussions

Three-phase power imbalance is a major concern in a three-phase four-wire power system [3]. Furthermore, it is the power that gradually approaches the thermal limits of network assets in the long run [6, 28, 29], thus affecting reinforcement costs. This paper is therefore concerned about power and its imbalance over the three phases.

This paper is a major advancement from [1] and [6]. Reasons are detailed in Section I. The proposed methodology is ideally suited for estimating the ARCs throughout a utility's LV networks with limited data. Moreover, it should be noted that the methodology in this paper does not increase input data or extract substitute data from elsewhere. Rather, the methodology overcomes the barrier of insufficient input data and provides informative estimation of the ARC.

Up till now, triangular distributions are suitable because: the existing level of data is insufficient to support complex alternative probability distributions, e.g. Beta and Gamma. Furthermore, normal (or Gaussian) distribution is not suitable because: i) it is unable to model the skewness of the DIB distribution; and ii) the Central Limit Theorem does not apply to the spatial distribution of the temporal means of DIB. Skew Gaussian distribution again requires more data to specify its parameters which is beyond the existing level of data can cope. In future, with more sensory data becoming available, the data will serve the purpose of refining the parameters of the triangular distributions; and they would enable the derivation of more complex distribution functions for DIB and the utilization rates, thus improving the accuracy of the ARCs.

The research assumes that, when the thermal capacity of an asset is reached under three-phase imbalance, the DNO would perform network reinforcements – this is a common practice in the UK [6], but not the least-cost solution. There are alternative solutions that could potentially save costs, e.g. proactive phase balancing. The ARC quantified in this paper serves as a benchmark cost for future comparisons between the conventional reinforcement solution and alternative smart solutions.

Demand growth may increase, decrease, or have no impact on the degree of three-phase imbalance both for individual networks and at a utility scale. For an individual network, if an extra demand is connected to the 'heaviest' phase (the phase already having the greatest load among the three phases), then it will increase three-phase imbalance; if an extra demand is connected to the 'lightest' phase (the phase having the least power among the three phases), then it will likely reduce three-phase imbalance; there is also the possibility that the extra demand neither increases nor reduces three-phase imbalance. How future demand growth will affect the degree of three-phase imbalance on a utility scale is a cutting-edge research question. To the best of our knowledge, there is no answer to this question at the moment. In this paper, we propose two scenarios for this question: one is the base case scenario, where we have assessed the ARC based on the current state of imbalance projected into the future if DNOs take no actions to rebalance three-phase supply; another scenario, as presented in Section IV – C and V – F, is that there is uncertainty associated with the DIB at the end of the study period. We have modelled the DIB using confidence interval estimate, which is a classic mathematical tool for modelling uncertainty.

This paper assesses the ARCs under the uncertainties from growing EVs and HPs, where the demand growth rate and the DIB values are modelled as random variables using confidence interval estimates.

Considering demand growth in demand-dominated networks represents the base case (also the status quo in the UK [18]). This paper investigates three-phase imbalance for demand-dominated networks. This forms the basis to build upon for the future when we consider a high penetration of renewable generation over the planning period, taking into account differing degrees of uncertainties between renewables and classical demands.

Unlike DIB values which are calculated in this paper, it should be noted that the assets' nominal utilization rates are input data directly referenced from [6]. The calculations are based on the power factor of 0.95 (lagging) in [6]. Therefore, by referencing the nominal utilization rates from [6], our paper inherits the assumption of a 0.95 (lagging) power factor from by that paper.

VI. Conclusions

This paper presents a novel scalable methodology to estimate the additional reinforcement cost (ARC) from three-phase imbalance on a utility scale, modelling the distribution of circuit imbalance on a utility scale where the level of information needed to accurately specify complex distributions is not available. This paper develops a novel numerical method to estimate the volume of assets that contribute to the ARC and the ARC itself by identifying the relationship between the triangular distribution of circuit imbalance and that of circuit utilization. When the information is made more readily available in future, accurate probability distributions can be constructed to reflect the network condition across the whole system. In light of this, the paper comes up with two novel generalized ARC formulas in the integral form that account for generic probability distributions. The observations are:

1) The ARCs are at least comparable to the reinforcement costs of the balanced case.

2) The sensitivity of the ARC to the following parameters is listed in a decreasing order: demand growth rate > the maximum degree of imbalance (DIB) values and the maximum nominal utilization rates > the minimum DIB values and the minimum nominal utilization rates.

3) The proposed method effectively overcomes the scalability and data deficiency barriers. It is applicable to a utility's LV networks with millions of customers and limited sensory data.

The method provides the ARC as a benchmark cost under conventional passive network management. The ARC can be used to compare the costs of future alternative solutions.

References

[1] K. Ma, R. Li, and F. Li, "Quantification of Additional Asset Reinforcement Cost From 3-Phase Imbalance," *Power Systems, IEEE Transactions on,* vol. PP, pp. 1-7, 2015.

[2] J. Zhu, M. Y. Chow, and F. Zhang, "Phase balancing using mixed-integer programming [distribution feeders]," *Power Systems, IEEE Transactions on,* vol. 13, pp. 1487-1492, 1998.

[3] S. Yan, S.-C. Tan, C.-K. Lee, B. Chaudhuri, and S. Y. R. Hui, "Electric Springs for Reducing Power Imbalance in Three-Phase Power Systems," *IEEE Transactions on Power Electronics,* vol. 30, pp. 3601-3609, 2015.

[4] M. W. Siti, D. V. Nicolae, A. A. Jimoh, and A. Ukil, "Reconfiguration and Load Balancing in the LV and MV Distribution Networks for Optimal Performance," *Power Delivery, IEEE Transactions on,* vol. 22, pp. 2534-2540, 2007.

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

[6] Y. Zhang, F. Li, Z. Hu, and G. Shaddick, "Quantification of low voltage network reinforcement costs: A statistical approach," *Power Systems, IEEE Transactions on,* vol. 28, pp. 810-818, 2013.

[7] R. Li, C. Gu, F. Li, G. Shaddick, and M. Dale, "Development of Low Voltage Network Templates Part I: Substation Clustering and Classification," *Power Systems, IEEE Transactions on,* vol. 30, pp. 3036-3044, 2015.

[8] A. M. Cossi, R. Romero, and J. Mantovani, "Planning and Projects of Secondary Electric Power Distribution Systems," *Power Systems, IEEE Transactions on,* vol. 24, pp. 1599-1608, 2009.

[9] A. Navarro and H. Rudnick, "Large-Scale Distribution Planning Part I: Simultaneous Network and Transformer Optimization," *Power Systems, IEEE Transactions on,* vol. 24, pp. 744-751, 2009.

[10] T. Sugita, D. Iioka, Y. Yokomizu, T. Matsumura, N. Hatakeyama, T. Kuriyama, and T. Ootaki, "Low-voltage distribution network planning taking account of power loss cost in genetic algorithms," in *Future Power Systems, 2005 International Conference on*, 2005, pp. 5-10.

[11] I. Ziari, G. Ledwich, A. Ghosh, and G. Platt, "Optimal distribution network reinforcement considering load growth, line loss, and reliability," *Power Systems, IEEE Transactions on,* vol. 28, pp. 587-597, 2013.

[12] T. Asakura, T. Genji, T. Yura, N. Hayashi, and Y. Fukuyama, "Long-term distribution network expansion planning by network reconfiguration and generation of construction plans," *Power Systems, IEEE Transactions on,* vol. 18, pp. 1196-1204, 2003.

[13] B. Zeng, J. Zhang, X. Yang, J. Wang, J. Dong, and Y. Zhang, "Integrated Planning for Transition to Low-Carbon Distribution System With Renewable Energy Generation and Demand Response," *Power Systems, IEEE Transactions on,* vol. 29, pp. 1153-1165, 2014.

[14] S. Favuzza, G. Graditi, M. G. Ippolito, and E. R. Sanseverino, "Optimal Electrical Distribution Systems Reinforcement Planning Using Gas Micro Turbines by Dynamic Ant Colony Search Algorithm," *Power Systems, IEEE Transactions on,* vol. 22, pp. 580-587, 2007.
[15] Z. Hu and F. Li, "Cost-Benefit Analyses of Active Distribution Network Management, Part II: Investment Reduction Analysis," *Smart Grid, IEEE Transactions on,* vol. 3, pp. 1075-1081, 2012.
[16] J. Zhu, G. Bilbro, and M.-Y. Chow, "Phase balancing using simulated annealing," *Power Systems, IEEE Transactions on,* vol. 14, pp. 1508-1513, 1999.
[17] K. Ma, F. Li, and R. Aggarwal, "Quantification of Additional Reinforcement Cost Driven by Voltage Constraint Under Three-Phase Imbalance," *Power Systems, IEEE Transactions on,* vol. PP, pp. 1-9, 2016.
[18] 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," *Power Systems, IEEE Transactions on,* vol. PP, pp. 1-13, 2015.
[19] A. v. Jouanne and B. Banerjee, "Assessment of voltage unbalance," *IEEE Transactions on Power Delivery,* vol. 16, pp. 782-790, 2001.
[20] N. C. Woolley and J. V. Milanovic, "Statistical Estimation of the Source and Level of Voltage Unbalance in Distribution Networks," *IEEE Transactions on Power Delivery,* vol. 27, pp. 1450-1460, 2012.
[21] M. Sun, S. Demirtas, and Z. Sahinoglu, "Joint Voltage and Phase Unbalance Detector for Three Phase Power Systems," *IEEE Signal Processing Letters,* vol. 20, pp. 11-14, 2013.
[22] T. Routtenberg, Y. Xie, R. M. Willett, and L. Tong, "PMU-Based Detection of Imbalance in Three-Phase Power Systems," *IEEE Transactions on Power Systems,* vol. 30, pp. 1966-1976, 2015.
[23] G. Strbac and J. McDonald. (2005). *United Kingdom Generic Distribution System*. Available: http://www.sedg.ac.uk/Workshop%2018%20Oct%202005/Poster%20-%20UKGDS%20and%20Typical%20Networks.pdf
[24] (2013, 03 Feb 2016). *Distribution Long Term Development Statement*. Available: http://www.spenergynetworks.co.uk/userfiles/file/ScottishPower_Manweb_Long_Term_Development_Statement_2013.pdf
[25] M. Evans, N. Hastings, and B. Peacock, *"Triangular Distribution" in Statistical Distributions*, Third ed.: New York: Wiley, 2000.
[26] S. Basu and A. DasGupta, "The Mean, Median, and Mode of Unimodal Distributions: A Characterization," *Theory of Probability & Its Applications,* vol. 41, pp. 210 -- 223, 1997.
[27] *Network Investment Plan 2010 -- 2015*. Available: http://www.enwl.co.uk/docs/about-us/network-investment-plan.pdf
[28] F. Li and D. L. Tolley, "Long-Run Incremental Cost Pricing Based on Unused Capacity," *Power Systems, IEEE Transactions on,* vol. 22, pp. 1683-1689, 2007.
[29] (2013). *RIIO-ED1 HV and LV network investment analysis*. Available: http://www.spenergynetworks.co.uk/userfiles/file/201303_A2_4_TNEI_HV_LV_network_investment_analysis.pdf

APPENDIX A: MATHEMATICAL PROOF

In the appendix, we present the mathematical proof as to why the feeders within area *E* in Fig. 3 does not meet Criterion 2) of the ARC definition. Criterion 2) is: when the three-phase peak demand occurs, the feeder's thermal capacity would not be reached if its three phases were balanced.

When the three-phase peak demand occurs, denote the power for each phase as P_A, P_B, and P_C.

For any feeder within area *E* in Fig. 3, its nominal utilization rate is greater than 100%. This gives

$$\frac{P_A + P_B + P_C}{C_{3\emptyset}} > 100\% \tag{32}$$

If the three phases were fully balanced, then

$$P_A = P_B = P_C \tag{33}$$

Substituting (35) into (34), there is

$$\frac{3P_A}{C_{3\emptyset}} > 100\% \tag{34}$$

This gives

$$P_A > \frac{C_{3\emptyset}}{3} \tag{35}$$

Given that the capacity for each phase $C_{\text{single}\emptyset} = \frac{C_{3\emptyset}}{3}$, there is

$$P_A > C_{\text{single}\emptyset} \tag{36}$$

Phase A power exceeds the capacity for this phase. Therefore, Criterion 2) is not met.

Kang Ma is now working as a lecturer at University of Bath. His research focuses on three-phase unbalanced low voltage networks. He worked as an R&D engineer at China Electric Power Research Institute (Beijing) from 2011 to 2014, during which time he developed the first version of the reliability assessment module for a distribution network planning platform. This platform has been widely applied to over 20 provincial grid companies in China. He received his PhD degree in Electrical Engineering from the University of Manchester (U.K.) and his B.Eng. degree from Tsinghua University (China).

Ran Li received his B.Eng. degree in electrical power engineering from University of Bath, U.K, and North China Electric Power University, Beijing, China, in 2011. He received the Ph.D. degree from University of Bath, in 2014 and became a lecturer in Bath from 2015. His major interest is in the area of big data in power system, deep learning and power economics.

Furong Li (SM'09) was born in Shannxi province, China. She received the B.Eng. degree in electrical engineering from Hohai University, Nanjing, China, in 1990 and the Ph.D. degree from Liverpool John Moores University, Liverpool, U.K., in 1997. She is a Professor and the director of Center for Sustainable Power Distribution, University of Bath, U.K. Her major research interest is in the area of power system economics and markets.