Article:

Poursharif, G., Brint, A.T. orcid.org/0000-0002-8863-407X, Black, M. et al. (1 more author) (2017) Using smart meters to estimate low voltage losses. IET Generation, Transmission and Distribution. ISSN 1751-8687

https://doi.org/10.1049/iet-gtd.2017.1300

Using Smart Meters to Estimate Low Voltage Losses

Goudarz Poursharif[1*], Andrew Brint[1], Mary Black[2], Mark Marshall[2]

[1] Management School, University of Sheffield, 1 Conduit Road, Sheffield, S10 1FL, UK
[2] Northern Powergrid Ltd., 98 Aketon Road, Castleford, West Yorkshire, WF10 5DS, UK
[*]g.poursharif@sheffield.ac.uk

Abstract: Losses on low voltage networks are often substantial. For example, in the UK they have been estimated as being 4% of the energy supplied by low voltage networks. However, the breakdown of the losses to individual conductors and their split over time are poorly understood as generally only the peak demands and average loads over several months have been recorded. The introduction of domestic smart meters has the potential to change this. How domestic smart meter readings can be used to estimate the actual losses is analysed. In particular, the accuracy of using 30 minute readings compared with 1 minute readings, and how this accuracy could be improved, were investigated. This was achieved by assigning the data recorded by 100 smart meters with a time resolution of 1 minute to three test networks. Smart meter data from three sources were used in the investigation. It was found that 30 minute resolution data underestimated the losses by between 9% and 24%. By fitting an appropriate model to the data, it was possible to reduce the inaccuracy by approximately 50%. Having a smart meter time resolution of 10 minutes rather than 30 gave little improvement to the accuracy.

1. Introduction

Losses on low voltage networks can comprise a significant portion of the total network losses. For example, [1] estimated the (technical) losses on the UK's low voltage networks as 4% of the electric energy supplied to low voltage customers, with the majority of this being due to conductor 'copper' losses. Besides the monetary and environmental costs of generating this lost energy, the extra currents can contribute to voltage and capacity problems. A detailed knowledge of when and where these losses are occurring would allow better assessments of network efficiencies, assist with planning improved networks, help with operational decision making, improve identification of non-technical losses (i.e. unmetered and inaccurately metered loads [2]) and aid in the setting of network configurations [3]. Additionally, accurately evaluating the reduction in low voltage losses that embedded generation, such as photovoltaics, can provide is important for assessing the benefits of this generation [4]. However, the widespread lack of monitoring on low voltage networks makes it difficult to be precise about the location, timing and even the size of these losses, with consequent problems for decision making. The roll out of smart meters to domestic customers in the UK which is due for completion in 2025, will provide the distribution network operators (DNOs) with an order of magnitude more information about the loads on their low voltage circuits. This paper analyses how this information can be used to improve the estimates of the low voltage conductor losses. In particular, it investigates:

- How much lower the estimated losses using 30 minute intervals are compared with using 1 minute intervals.
- How to adjust the 30 minute loss calculation to get closer to the 1 minute value.
- The benefit for loss calculations of the smart meters reporting 10 minute averages rather than 30 minute averages.

It is found that fitting a suitable model to the 30, 60 and 120 minute readings reduces by approximately 50% the error from calculating the losses using only the 30 minute readings.

The next section reviews approaches to estimating low voltage losses. Section 3 describes the smart meter data sets and the test networks that were used in the analysis. Section 4 looks at how the calculated losses depend on the time interval size used by the smart meters, before Section 5 analyses how accurately the losses based on 30 minute smart meter data, can be used to estimate the losses that would have been calculated if the smart meter time resolution had been 1 minute. The paper ends with a brief discussion of the implications of the work.

2. Background on low voltage losses

Although low voltage conductor losses are a major component of electricity network technical losses, accurately estimating them has been difficult. A simple approach to estimating the losses on a low voltage network is to measure the difference between the energy supplied by the distribution transformer and the metered energy used by the customers. However, as [3] points out, until the advent of smart meters, the energy usage from each customer had been measured over long periods, e.g. 6 months or a year, and these periods were extremely unlikely to be the same for all customers on a low voltage circuit, e.g. the starting day of the period will vary for different customers. Even if the metering periods for all customers did coincide, the breakdown of the losses to individual conductors and time periods would not be known. Hence the fair allocation of the costs resulting from these poorly known low voltage losses, is difficult. This is a particular issue when evaluating the benefits from embedded generation reducing current sizes around the network and hence the size of the losses [5], [6], [7], [8], or assessing the consequences of new load types such as electric vehicles [9].

A common approach to estimating low voltage losses on a circuit has been to use a loss factor [10], [11]. This factor is used to multiply the peak load losses to give the average losses. Its attraction is that maximum demand is often measured (or estimated) for low voltage circuits, and so it provides a straightforward way to estimate the losses. However, it provides only a rough estimate of the losses as the relationships between the peak demand and the peak losses, and between the peak losses and the average (or total) losses, are very dependent on the circuit's characteristics and the shape of the load curves at different points on the circuit. [10] argues that using the average demand rather than the maximum demand is better as it reflects a period of time rather than one time instant.

An alternative approach estimates a low voltage circuit's losses by matching the circuit with a set of benchmark circuits. Various features can be used for the matching, for example, [3] use the main feeder length, the length of branches, the number of branches, customer information and conductor sizes. The approach relies on the benchmark circuits having been modelled in detail, and so their calculated loss values are regarded as being accurate. [12] notes that a weakness of this approach is that no two circuits are exactly the same, and so the matching may not be valid.

A general weakness of these approaches is that they just provide a single figure for the losses rather than providing a geographical and temporal breakdown of the losses. [13] notes that breaking down the losses is becoming more important due to decentralised generation and the move towards smart grids. Existing standard load profiles were combined with smart meter data to look at the consequences of simplifications such as using mean or peak loads. [13] found that existing loss estimation approaches had particular problems in low density rural areas. These branches were also sensitive to the time resolution of the data used to calculate losses with losses calculated using one second mean values being up to 20% higher than those calculated using 15 minute values.

Having customer smart meter readings available for a low voltage circuit will allow the "copper" losses to be estimated using a load flow analysis [6], [10]. Not only will this avoid the coarse approximations involved in using loss factors and allow a temporal breakdown of the losses, but the consequences of phase imbalance [9] and embedded generation can be accounted for. However, although their measurement time periods are much shorter than the months or years of the meters that they are replacing, the typical measurement time periods of 15, 30 and 60 minutes [14] mean that the losses calculated using smart meter data underestimate the true losses [13], [15]. This relatively poor time resolution when estimating low voltage losses, is a particular problem when assessing the impact of recent and future developments, such as high levels of photovoltaic generation [16], [17]. So as to investigate the effect of the time period length on the calculated losses, [15] considered the losses from a single appliance switching on and off at random. For short time periods, i.e. in terms of seconds rather than minutes, the underestimation was modelled (and validated) as being a linear function of the time period. As the time period increases, the assumption of there being at most a single switching event (either off to on or on to off) in any time interval breaks down and the relationship stops being linear. Comparing the summed demands from between 1 and 22 dwellings indicated that the relationship between losses calculated using a one minute resolution and larger time resolutions became closer to a linear one as the number of dwellings increased (Figure 7 in [15]).

3. The data

The main question that this paper addresses is how well the losses calculated at a time resolution of 30 minutes (i.e. the time resolution of the UK's smart meters) can be adjusted to estimate the losses that would have been calculated if the time period resolution had been one minute. As the accuracy of the estimate depends on the loads and the network topology being analysed, 3 smart meter data sets and 3 very different test networks were investigated.

3.1 Smart meter data sets

Three smart meter data sets with a time resolution of 1 minute were used in the investigation:

- The Customer-Led Network Revolution (CLNR) data [18]. This project was carried out in the UK from 2011 to 2014. 53 sample dates between December 2012 and May 2014 were chosen for which there were at least 100 customers with demand readings for the whole day. Figure 1 shows the first two smart meter profiles in the data set for two consecutive Wednesdays in March 2014.
- The UMass Trace Repository data [19]. The data used were the readings from 114 single-family apartments in the USA for the first 350 days of 2016. Figure 2 shows the first two smart meter profiles in the data set for two consecutive Wednesdays in March 2016.
- The UK Data Archive data [20]. This data set comprises data from a small number of smart meters covering a very diverse range of customer types. Three Wednesdays and one Saturday in 2008 were chosen for the analysis. Only 22 complete profiles were available for these days, and so profiles from these customers for adjacent Wednesdays and Saturdays were used to produce a set of 100 customers for each of the four days

The three data sets are available for free public download – details are given in [19], [20] and [21].

The smart meter loads were modelled as being independent of the customer's voltage, i.e. the customer current was the smart meter load divided by the voltage at this point in the network (see Section 5). Coarser time intervals for the smart meters were modelled by averaging the 1 minute readings over the coarser time interval for each customer. For example, for a 15 minute interval a customer's readings for minutes 00:00, 00.01, …, 00:14 were averaged to give the value for the first 15 minute period, and then 00:15 to 00:29 were averaged for the next period.

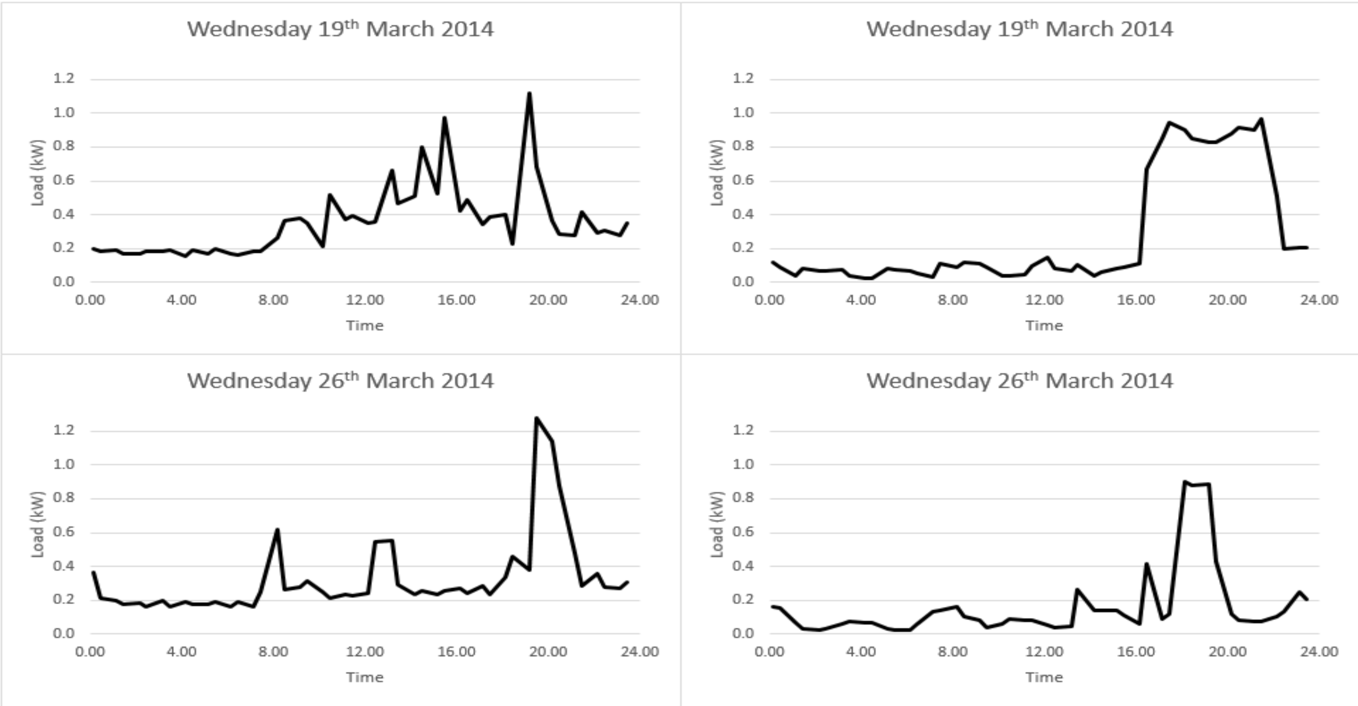

Fig. 1. The first two smart meter profiles in the CLNR data set [18] for two Wednesdays in the 2nd half of March 2014.

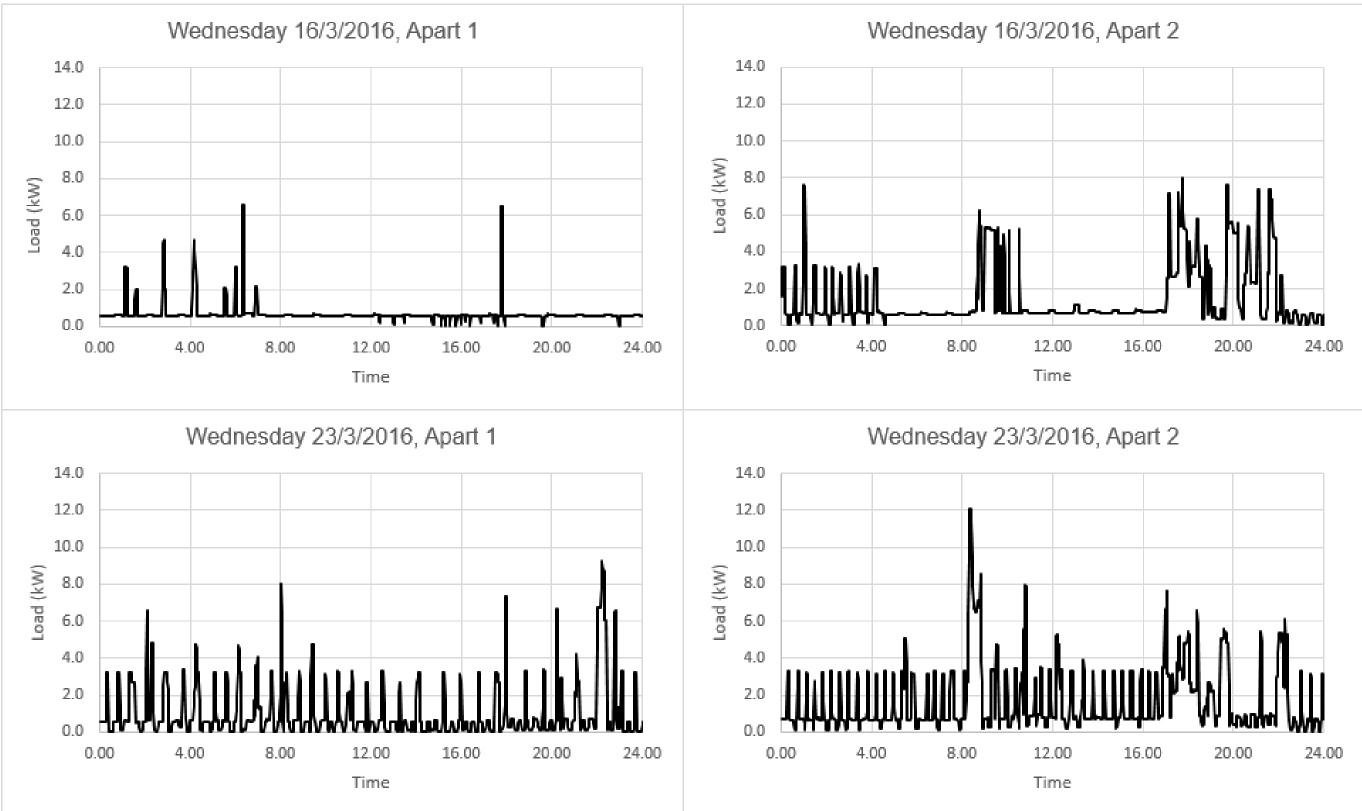

Fig. 2. The first two smart meter profiles in the UMass data set [19] for two Wednesdays in the 2nd half of March 2016.

3.2 Test networks

As losses depend on the size of the phase and neutral currents around the network, the ratio of the losses calculated using 1 minute intervals to the losses calculated using 30 minute intervals will vary with the network topology. Therefore, three test networks with very different topologies were used for the analysis (see Figure 3).

a) A single tee with the customers split in the ratio 30:30:40 on the branches.

b) A linear network.

c) A high branching network with just one customer on each branch.

Networks (b) and (c) are the two extremes with network (a) in the middle. For each network, the customers were evenly spaced out. The length of the branch sections was chosen so that the highest voltage drop for the three phase single tee network with the 100 CLNR loads was just under 6%.

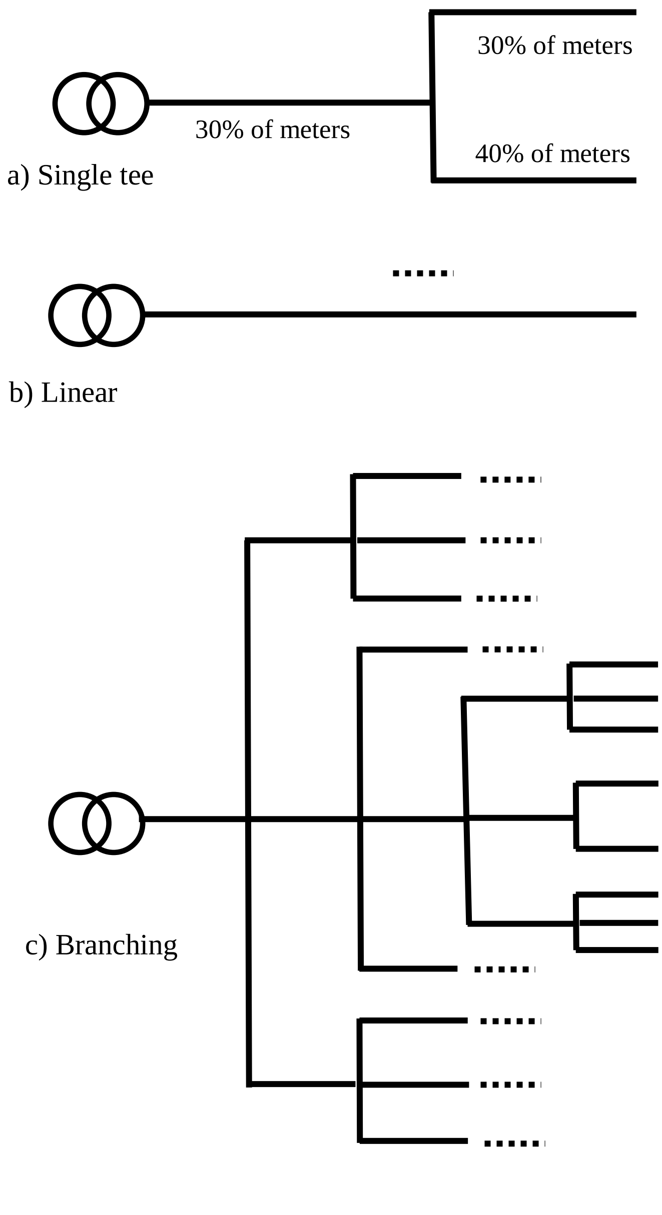

Fig. 3. The test networks.

4. The effect of current variability on loss calculations

Averaging a current over a time interval makes it constant over the interval and so reduces the spikiness of the current. As losses are proportional to the square of the current, this means that the losses calculated using the average current are lower than the actual losses [10], [15]. Therefore, the variability of the total current from a group of customers over a time period directly affects the accuracy of the estimated losses.

4.1. Very high spikiness: One narrow spike

If there is a single narrow column of current of time duration (width) D and with a current value of I, and the current is zero outside of this column, then for a time interval of width $w \geq D$ that contains all of the current spike (and assuming a resistance of 1):

- The actual loss is I^2D
- The average current is $\frac{ID}{w}$, and so the loss calculated using the average current over the time interval is $\frac{I^2D^2}{w}$

Hence, when $w \geq D$, the calculated loss from using the average current decreases at a rate of $\frac{1}{w}$.

4.2. Very low spikiness: constant current

The other extreme case is when the current has the constant value I over the whole period. Then the calculated loss over any time period is simply I^2 times the length of the period. So over a fixed time interval, the calculated loss is independent of the size and number of periods the fixed interval is divided into.

4.3. Smooth with a linear trend

Besides short term variations in the current (spikes), for the longer time periods, e.g. 30 minutes, there may be a distinct increase or decrease in the current over the period. We will consider a whole time period of width 1 and a linearly increasing current:

$$l_t = c + m \times t$$

where l_t is the current at time $t \in [0, 1]$.

If the time period is split up into n equal intervals, then the calculated loss is (assuming a resistance of 1) (1)

$$\sum_{j=1}^{n}\left(\int_{x=\frac{j-1}{n}}^{\frac{j}{n}}\left\{c+m\left(\frac{j-\frac{1}{2}}{n}\right)\right\}^2 dx\right)$$
$$= \frac{1}{n}\sum_{j=1}^{n}\left(c^2+2cm\left\{\frac{j-\frac{1}{2}}{n}\right\}+\frac{m^2\left\{j-\frac{1}{2}\right\}^2}{n^2}\right)$$
$$= c^2 + cm + \frac{m^2}{3} - \frac{m^2}{12n^2} \quad (1)$$

Consequently, as the width, $w = \frac{1}{n}$, of the time intervals that the time period is divided up into, increases, the loss calculated using the average interval current decreases in line with w^2.

4.4. Loss dependency on the time resolution in practice

Sections 4.1, 4.2 and 4.3 have shown that depending on the circumstances, the calculated losses as the time interval resolution increases may stay the same (constant current), decrease as the reciprocal of the interval width (single narrow spike) or decrease with the square of the width (linear trend). In practice the relationship will be a mixture of these and other effects. Therefore, an empirical approach is taken to analysing the effect of the time interval width on the calculated losses.

Modelling all the loads as being located at the same point, Figure 4 shows the losses calculated from adding different numbers of the CLNR smart meters together for Wednesday the 13th of February 2013, and then varying the time interval resolution. So, for example, the hollow square symbol at a meter time interval of 90 minutes indicates that for the group of 3 meters, the ratio of the losses calculated using the average current over 90 minutes to the losses calculated using the average current over one minute, is 0.64. Hence all the symbol “curves” start off with a y value of one when the smart meter time interval (the x value) is one minute. (The losses were calculated by assuming that all the meters were grouped together at one node, were all on the same phase and the average current over the time interval was used in the loss calculation. Calculating the losses for meters distributed across phases and in different locations will be considered in Section 5.) The 1 meter curve has a very steep, near linear decrease over the first 4 minutes that is not present in the other curves. Apart from this, although the shapes of the curves are very similar, the rate of change (the gradient) becomes much less as the number of meters increases. This has consequences when estimating the losses for networks as the number of customers using a section of conductor will vary, decreasing as you move away from the substation.

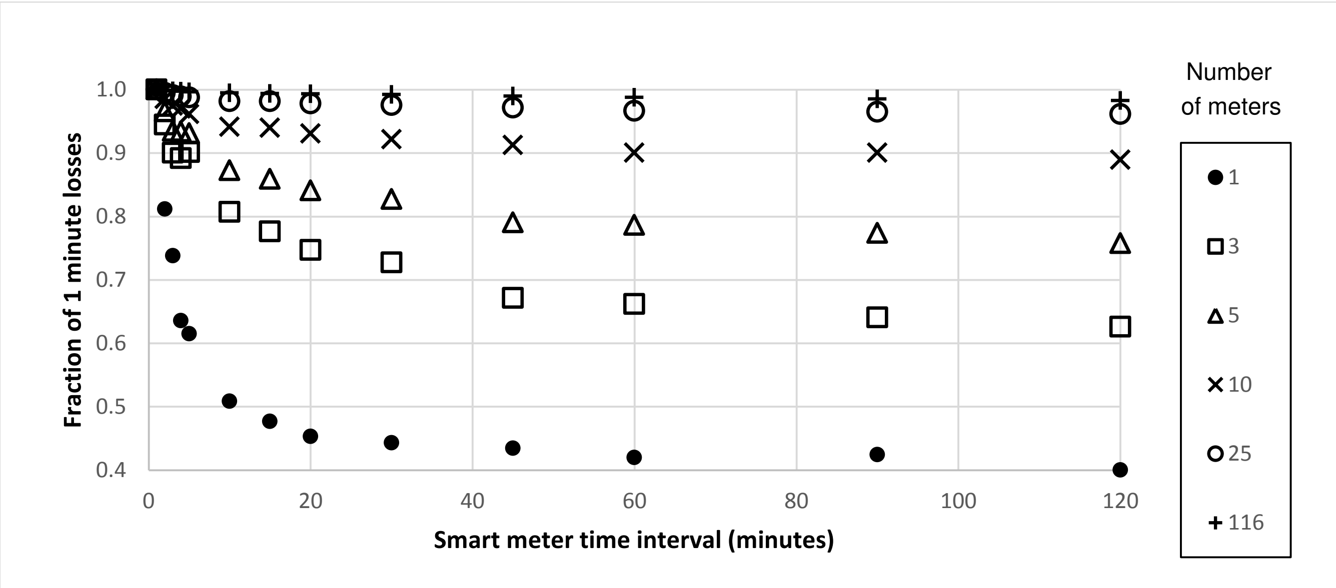

Fig. 4. How the losses from combining different numbers of meters alter as the meter interval lengthens. 13th February 2013. Each symbol type indicates a different number of meters. As the y-value is the calculated loss using time periods of x minutes divided by the loss from using time intervals of 1 minute, all the curves start with a fraction of 1.0 for a time interval of 1 minute.

5. Estimation of network losses

The smart meters were equally spaced over the networks shown in Figure 3. The substation voltage was set at 250 volts and the voltage at each smart meter was determined by iteratively calculating the voltages and the currents throughout the network. The losses in each time period were then calculated by summing the phase and neutral losses in each branch. Finally, the losses for the day were calculated by summing the losses for the day’s time periods.

5.1. Losses model

The model chosen to fit to the time interval width, t, and losses data, L, was (2)

$$L = \beta \div t^{\alpha} \qquad \text{where } \alpha \in (0, 1) \tag{2}$$

α determines the shape of the curve while β is the loss estimate for time resolution t=1.

Given the losses L_1 and L_2 at times t_1 and t_2, then (3)

$$\alpha = \log\left(\frac{L_1}{L_2}\right) + \log\left(\frac{t_2}{t_1}\right) \tag{3}$$

Hence, α can be determined followed by β.

The form of the model in equation 2 was selected as

- The few values that t can take (e.g. 30, 60 and 120 minutes) means that the number of parameters (i.e. α and β) needs to be low.
- The curves in Figure 4 all have this approximate shape, e.g. using α=0.003 gives a good approximation to the 116 meter curve while α=0.24 does similarly well for the 1 meter curve.
- The extreme cases for a single demand spike are α=1 for a very narrow spike and α=0 for a very broad spike (see Sections 4.1 and 4.2).
- Equation 3 means that it is straightforward to implement, and so it could be a suitable choice in practice.

5.2. Seasonal and day effect

The CLNR data set of 53 dates was used to investigate whether different values of α should be used for different days of the week and seasons of the year. For each day, the losses were calculated for different time interval resolutions for the single tee test network shown in Figure 3, and the model in equation 2 was fitted to the data. For the day of the week, performing a single factor Analysis of Variance test gave a P-value of 0.45, and so this factor is not considered any further. However, for the comparison of the αs for winter and spring, the Analysis of Variance P-value was less than 0.001. The alpha values calculated for each day for this case are shown in Figure 5.

Seven approaches for estimating the 1 minute losses using the losses from either longer time periods or from different days, were analysed for the 53 dates. Approaches (i) to (v) involve fitting the model $L = \beta \div t^{\alpha}$ to the data, while approaches (vi) and (vii) simply scale the 30 minute losses. In approaches (i) to (v), the value of α to use for a day is determined before the value of β is calculated. α is chosen to minimise the difference between the predicted and actual losses for the 30, 60 and 120 minute intervals using equation 3.

i. **All 1 minute** For all the days excluding the one being estimated, fit α for each day separately using the time interval sizes of 1, 5, 10, 15, 30, 60 and 120. The α used for the estimation day is the average of the 52 αs for the other days.

ii. **Season 1 minute** This is the same as the “All 1 minute” approach except that the averaging for α is only carried out over the days in the same season as the estimation days, e.g. winter days if the day being predicted is a winter day.

iii. **Current day 30 minute** This approach fitted the model of equation 2 only using the losses for the 30, 60 and 120 minute intervals on the prediction day.

iv. **All 30 minute** For each of the 53 days, α was chosen in the same way as in part (iii). The average of these 53 αs was then used for the estimation day and the value of β that gave the best fit at times 30, 60 and 120 for this α was then determined.

v. **Season 30 minute** This approach is the same as the “All 30 minute” approach except that instead of using all 53 days, only those days in the same season as the estimation day are used.

vi. **Scaling the 30 minute losses** For all the days excluding the one being estimated, the ratio of the 1 minute losses to the 30 minute losses is calculated. The 1 minute losses for the estimation day are estimated by multiplying the day’s 30 minute losses by the average of these ratios.

vii. **Seasonal scaling of the 30 minute losses** This is the same as the “Scaling the 30 minute losses” approach except that only the ratios for the days in the same season as the estimation day are used.

The results for the approaches are shown in Figure 6 and Table 1 where, for each day, the absolute percentage error (APE) from the one minute losses for the day was calculated for each of the methods. The APE was calculated as (4)

$$\text{Absolute Percentage Error (APE)} = 100 \times \text{absolute value} \left(\frac{\text{actual loss} - \text{predicted loss}}{\text{actual loss}} \right) \tag{4}$$

Hence, the means and standard errors in Table 1 are over 53 observations. Over the 53 days, the 30 minute loss values summed over each day underestimated that day’s 1 minute losses by, on average, 24%. Comparing this with the 30 minute values in Figure 4, where 3 meters gave a difference of 27% and 5 meters gave a difference of 17%, suggests that having small numbers of customers on individual phases near the tips of the network, has a large impact on the ratio of the 1 minute losses to the 30 minute losses. Performing a single factor Analysis of Variance test on the 7 approaches, gave a P-value less than 10^{-6}, i.e. differences in the performances between some of the approaches are highly significant.

Table 1. The absolute percentage errors (APEs) for the *approaches’ predictions of the 1 minute time resolution* losses for the 53 days, i.e. each mean is based on 53 observations.

	Approaches	Mean	Standard error
i.	All 1 minute	6.2	0.8
ii.	Season 1 minute	5.9	0.7
iii.	Current day 30 minute	15.3	1.7
iv.	All 30 minute	14.1	0.9
v.	Season 30 minute	13.9	0.9
vi.	Scaling of 30 minute	6.5	0.8
vii.	Seasonal scaling of 30 minute	6.1	0.7

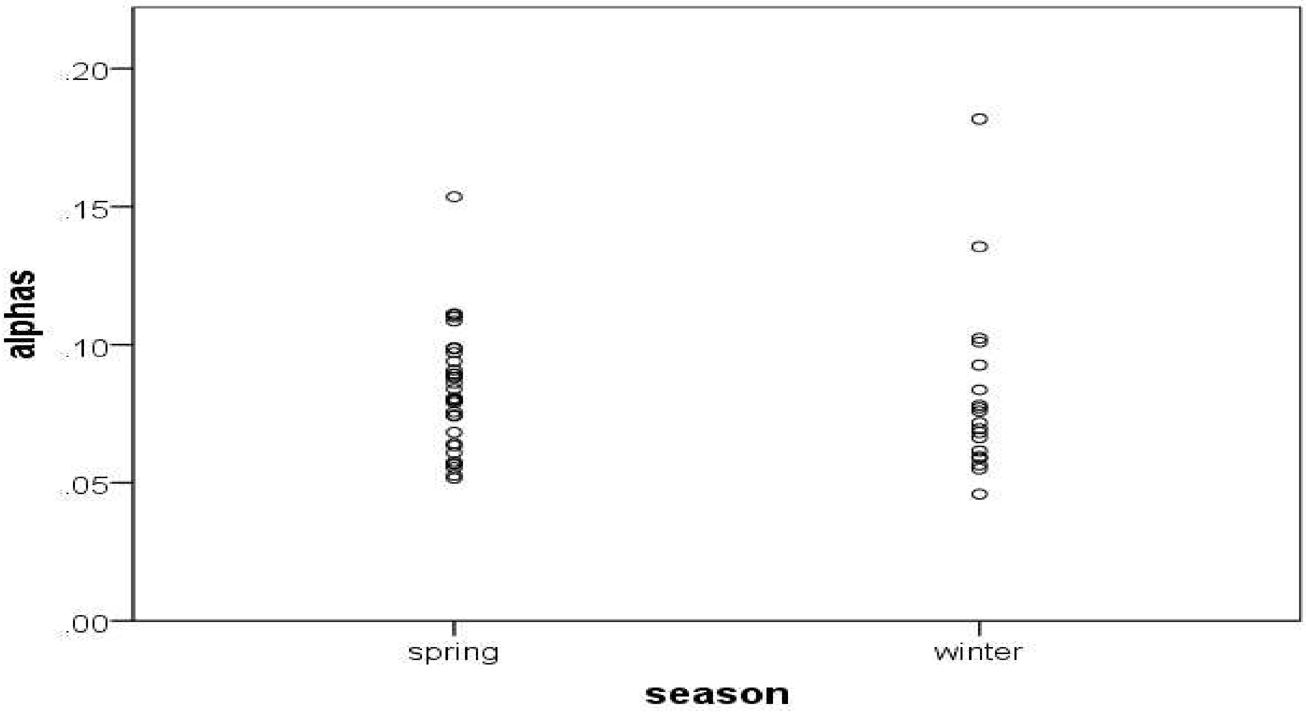

Fig. *5*. The optimal αs for estimating the 1 minute resolution losses using L = $\beta \div t^{\alpha}$ for each of the 53 days in Section 5.2 grouped by seasons of the year.

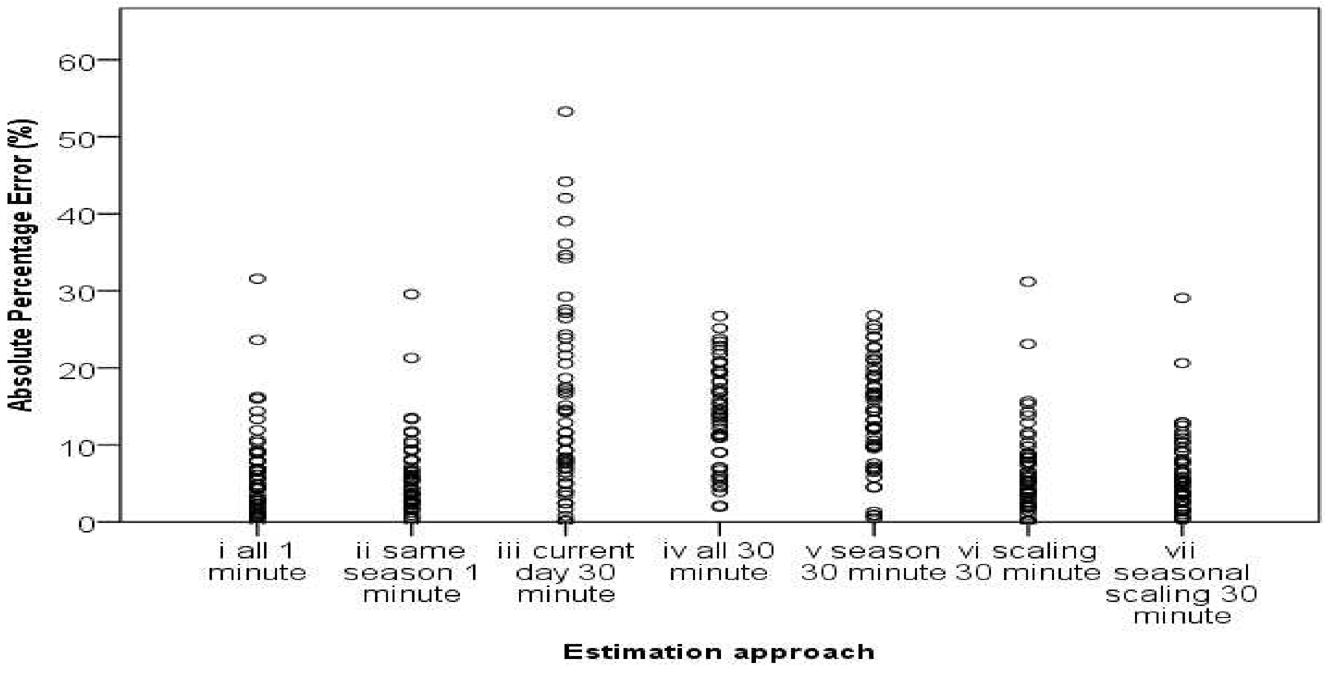

Fig. *6*. Absolute Percentage Errors (APEs) from the 1 minute resolution losses for each of the 53 days given by the different approaches.

As the 1 minute losses usually will not be available in practice for any of the days, approaches (i), (ii), (vi) and (vii) provide more of a performance target rather than a generally applicable approach. Performing the Analysis of Variance test on the 3 approaches that do not use knowledge of 1 minute losses (i.e. (iii), (iv) and (v)) gave a two sided P-value of 0.52. Consequently, the differences between approaches (iii), (iv) and (v) are not significant. In practice, (iv) or (v) would seem more appropriate as the value of α could be calculated once and then used for future days while approach (iii) calculates α afresh for each day.

Paired two sample tests for differences between the means between approaches (i) and (vi) and between approaches (ii) and (vii) (i.e. comparing the curve fitting approach with the scaling approach when 1 minute data for similar circuits is known) gave two sided P-values of 0.01 and 0.11 respectively, and so while the former is significant the latter is not. Hence, if 1 minute loss data is available for other days, then there is possibly some indication that the model of equation 2 is better than simply scaling the 30 minute loss value.

5.3. Assessing approach (iv)

The lack of a significant benefit from modelling at the seasonal level in Section 5.2, means that in practice the simpler approach of making no distinction between the seasons is likely to be the preferred approach. The robustness of the performance of approach (iv) along with the inaccuracy of simply using the 30 minute loss value to estimate the 1 minute losses, were analysed using the 114 smart meters from the UMass data set and the three test networks of Figure 3. Besides randomly allocating the meters to the phases for these networks, the situation where the linear network was single phase was also analysed. Finally, how the performance depends on the network size is investigated by reducing the number of smart meters to 85 (75%) and 29 (25%). Table 2 gives the mean over the 350 days of the Absolute Percentage Error (APE) for approach (iv) along with the corresponding MAPEs from using the 30 minute values to estimate the 1 minute losses.

In Table 2, approach (iv) gives just under a 50% improvement over using the 30 minute interval values when estimating the 1 minute losses. Table 2 also shows the loss estimates are significantly worse when there are few meters on a branch.

Table 2. The Mean Absolute Percentage Errors (MAPE) of the estimates from approach (iv). The MAPEs from the 30 minute losses are given in the brackets.

	114 meters	85 meters	29 meters
Single tee, 3	4.3 (8.6)	5.4 (10.9)	9.7 (20.5)
Branching, 3	7.9 (16.2)	9.0 (19.0)	12.0 (28.5)
Linear, 3 phase	4.2 (8.7)	5.2 (10.7)	9.9 (21.3)
Linear 1 phase	1.0 (2.1)	1.3 (2.7)	3.4 (6.7)

The UK Data Archive smart meter data set [20] was also used to assess the benefit of approach (iv). Three Wednesdays and one Saturday in 2008 were chosen for the analysis. Only 22 complete profiles were available for these days, and so profiles from these customers for adjacent Wednesdays and Saturdays were used to produce a set of 100 customers for each of the four days to populate the single tee network in Figure 3. How the calculated losses for each day vary with different smart meter time intervals is shown in Figure 7. The values are given as a fraction of the losses calculated using 1 minute time intervals.

Approach (iv) gave a MAPE of 4.8% compared with a MAPE of 19.0% from using the 30 minute intervals to estimate the 1 minute losses.

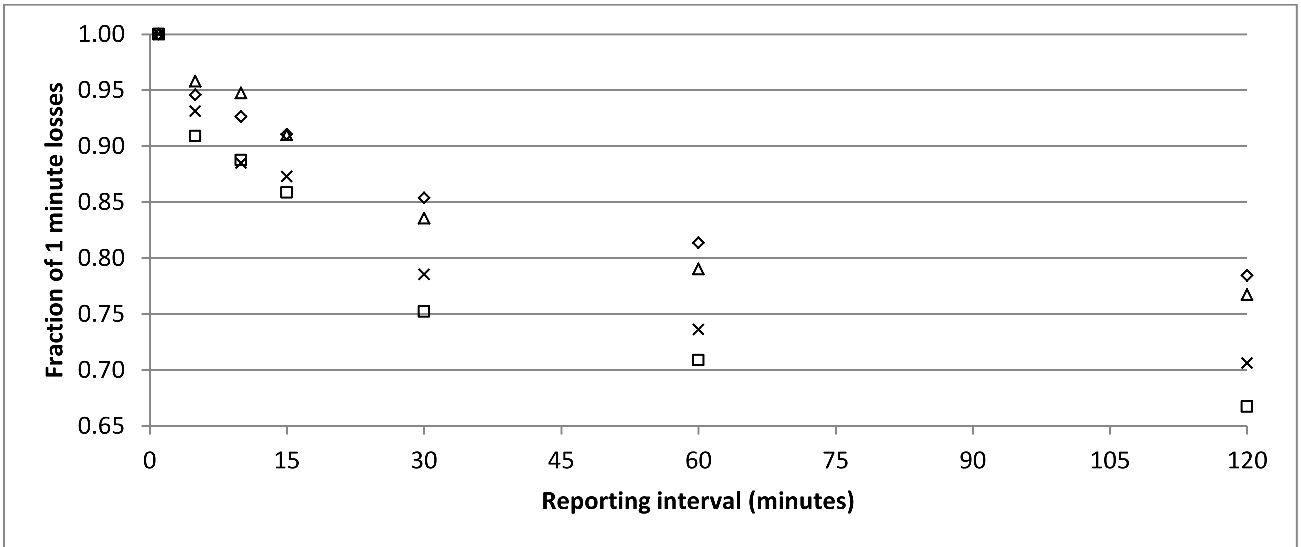

Fig. 7. How the calculated losses at different time resolutions compare with those calculated at a resolution of 1 minute for 4 different days using the UK data archive data set of smart meter customer curves [20]. The symbols △, ◇, □ and × denote the different days.

5.4. 10 minute versus 30 minute time intervals

Most smart meters allow the interval the load is averaged over to be configured in software, but shorter time intervals increase the communication overhead. Therefore, the benefit of having a smart meter time interval resolution of 10 minutes as opposed to the UK's 30 minutes was investigated by applying approach (iv) in the cases of 10, 30, 60 & 120 minutes and 30, 60 & 120 minutes. Table 3 gives the results for the CLNR and UMass data sets. Although having 10 minute data generally leads to better estimates, the improvement is limited and the MAPE is only being reduced by 10% or less.

Table 3. The MAPEs from using approach (iv) with 10 minute and 30 minute intervals.

	Using 10	Using 30 minutes
CLNR Single tee, 3	13	15
UMass Single tee, 3	3.9	4.3
UMass Branching, 3	7.9	8.8
UMass Linear, 3	4.0	4.2
UMass Linear 1	1.0	1.0

5. Discussion

The time resolution for the three data sets used in this research was 1 minute. As the calculated losses will be higher if a finer time resolution is used, the estimates of the 1 minute losses underestimate the actual losses. If the energy supplied by the substation is not available for improving the 1 minute estimates, then the results in [15] suggest fitting a straight line to the two lowest time resolutions available, i.e. the estimates for 1 and 2 minute resolutions, to estimate the actual losses. However, the very limited availability of data with time intervals below 1 minute meant that it was not possible to analyse the accuracy of the approximation.

The analysis has assumed that all customers have smart meters. In practice, there may be some customers with older "legacy" meters. By the completion of the smart meter roll out in the UK in 2025, the number of these meters is likely to be small. Approach (iv) can be used to estimate the losses for networks with a small number of non-smart meters by assigning appropriate smart meter profiles to these customers. A sensitivity analysis can then be carried out by carrying out the analysis with different profiles.

The investigation assumed that the values recorded by working smart meters were 100% accurate. Although concerns have been raised about smart meter accuracy, the analysis reported in [22] found that 99.9% of meters had an accuracy of between ±0.5%. Hence, smart meter inaccuracy is unlikely to have a significant effect on the investigation's findings.

The voltage at each customer and the peak currents in each cable section are other important low voltage network performance measures that data from smart meters could help to estimate. However, these depend on the maximum and minimum values during the smart meter period rather the case of losses where the dependence is on the sum over the period. Therefore, their estimation has not been considered in this paper.

6. Conclusions and implications

When a high percentage of customers have smart meters, it offers distribution network operators a low cost way to estimate low voltage network losses. However, how the losses calculated using the average currents over a time interval, vary as the time interval size varies depends on factors such as the spikiness of the demands, the relative sizes of the different spikes, and the presence or absence of trends in the demands. Hence, just using the 30 minute average currents that stem from smart meter readings, will underestimate the actual losses. The investigation found the following:

- <u>30 minute estimates of losses</u> – The absolute percentage error (APE) of the daily estimate from using 30 minute smart meter data to estimate the 1 minute losses, depends on the number of smart meters on each branch, with a lower APE when this number is higher. For the single tee network, the three smart meter data sets gave MAPE values of 9%, 14% and 24%. While these provide an indication of the order of magnitude of the underestimation of the losses when 30 minute intervals are used, there will be considerable variation between different network topologies, customer types and days.
- <u>Improving on the 30 minute value</u> – Fitting the model of equation 2 to the 30, 60 and 120 minute readings, reduced the error from using the 30 minute values to estimate the 1 minute losses by around 50%. For example, the reductions for the single tee network with the CLNR, UMass and UK data archive smart meter data sets were respectively 41%, 50% and 75%.
- <u>Benefit of 10 minute smart meter intervals</u> – Table 3 shows that using 10 minute smart meter intervals improves the losses estimate, but the improvement in the accuracy is relatively low compared with the overall inaccuracy. For the single tee network, the improvements for the CLNR and UMass data sets were respectively 13% and 9%.

Combined with the straightforward nature of approach (iv), these results mean that approach (iv) is a practical way to improve the loss estimates calculated using smart meter data.

7. Acknowledgments

We would like to thank Northern Powergrid Ltd. for sponsoring and supporting this research project. We would also like to thank the CLNR project team, the UK Data Archive and the UMass Trace Repository for providing the Smart Meter datasets necessary to carry out this research. The Customer-Led Network Revolution (CLNR) was carried out by British Gas, Northern Powergrid, Durham Energy Institute, the Newcastle University, EA Technology, and Low Carbon Network (LCN) fund from 2011-2014. This paper has been significantly improved by the comments of three anonymous referees.

8. References

[1] Sohn Associates "Electricity Distribution Systems Losses: Non-Technical Overview" Ofgem 2009. Available from: https://www.ofgem.gov.uk/ofgem-publications/43519/sohn-overview-losses-final-internet-version.pdf

[2] Yip, S.-C., Wong, K., Hew, W.-P., Gan, M.-T., Phan, R.C.-W. & Tan, S.-W.: "Detection of energy theft and defective smart meters in smart grids using linear regression", Electrical Power and Energy Systems, vol. 91, 2017, pp. 230-240, DOI: 10.1016/j.ijepes.2017.04.005

[3] Dashtaki, A. K. and Haghifam, M. R.: "A new loss estimation method in limited data electric distribution network", IEEE Transactions on Power Delivery, vol. 28, no. 4, 2013, pp. 2194–2200, DOI: 10.1109/TPWRD.2013.2273103

[4] Thomson, M. and Infield, D. G.: "Impact of widespread photovoltaics generation of distributed systems", IET Renewable Power Generation, vol. 1, no. 1, 2007, pp. 33-40, DOI: 10.1049/iet-rpg:20060009

[5] Costa, P. M. and Matos, M. A.: "Loss allocation in distribution networks with embedded generation", IEEE Transactions on Power Systems, vol. 19, no. 1, 2004, pp. 384–389, DOI: 10.1109/TPWRS.2003.820698

[6] Carpaneto, E., Chicco, G. and Akilimali, J.: "Branch current decomposition method for loss allocation in radial distribution systems with distribution generation", IEEE Transactions on Power Systems, vol. 21, no. 3, 2006, pp. 1170–1179, DOI: 10.1109/TPWRS.2006.876684

[7] Jagtap, K. M. and Khatod, D. K.: "Loss allocation in radial distribution networks with different load models and distributed generation", IET Generation, Transmission, and Distribution, vol. 9, no. 12, 2015, pp. 1275-1291, DOI: 10.1049/iet-gtd.2014.0884

[8] Jagtap, K. M. and Khatod, D. K.: "Loss allocation in distribution network with distributed generations", IET Generation, Transmission, and Distribution, vol. 9, no. 13, 2015, pp. 1628-1641, DOI: 10.1049/iet-gtd.2014.0923

[9] Frame, D. F., Ault, G.W. and Huang, S.: "The uncertainties of probabilistic LV network analysis", IEEE Power and Energy Society General Meeting, July 2012, DOI: 10.1049/cp.2012.0760

[10] Queiroz, L. M. O., Roselli, M.A., Cacellucci, C. and Lyra, C.: "Energy Losses Estimation in Power Distribution Systems", IEEE Transactions on Power Systems, vol. 27, no. 4, 2012, pp. 1879–1887, DOI: 10.1109/TPWRS.2012.2188107

[11] Oliveira, M. E. and Padilha-Feltrin, A.: "A top-down approach for distribution loss evaluation", IEEE Transactions on Power Delivery, vol. 24, no. 4, 2009, pp. 2117–2124, DOI: 10.1109/TPWRD.2009.2014266

[12] Dortolina, C. A. and Nadira, R.: "The loss that is unknown is no loss at all: A top-down/bottom-up approach for estimating distribution losses", IEEE Transactions on Power Systems, vol. 20, no. 2, 2005, pp. 1119–1125, DOI: 10.1109/TPWRS.2005.846104

[13] Brandauer, W., Köck, K., Wakolbinger, C. and Fickert, L.: "Loss calculation and optimisation in low voltage networks", CIRED 22nd International Conference on Electricity Distribution, Stockholm, Sweden, June 2013, paper 1186, DOI: 10.1049/cp.2013.1108

[14] McKenna, E., Richardson, I. and Thomson, M.: "Smart meter data: Balancing consumer privacy concerns with legitimate applications", Energy Policy, vol. 41, 2012, pages 807-814, DOI: 10.1016/j.enpol.2011.11.049

[15] Urquhart, A. J. and Thomson, M.: "Impacts of demand data time resolution on estimates of distribution system energy losses", IEEE Transactions on Power Systems, vol. 30, no. 3, 2015, pp. 1483–1491, DOI: 10.1109/TPWRS.2014.2349157

[16] Navarro, A., Ochoa, L.F., and Mancarella, P.: "Learning from residential load data: Impacts on LV network planning and operation", IEEE/PES Transmission and Distribution: Latin America Conference and Exposition, 2012, pp. 1-8, DOI: 10.1109/TDC-LA.2012.6319101

[17] Navarro, A., Ochoa, L.F., and Reynolds, D.: "Monte Carlo-based assessment of PV impacts on real UK low voltage networks", IEEE Power and Energy Society General Meeting, 2013, pp. 1-5, DOI: 10.1109/PESMG.2013.6672620

[18] Bulkeley, B., Matthews, P., Whitaker, G., Bell, S., Wardle, R., Lyon, S., & Powells, G.: "Domestic Smart Meter Customers on Time of Use Tariffs" Customer-Led Network Revolution CLNR-L243, 2015, http://www.networkrevolution.co.uk/resources/project-library/

[19] UMass Trace Repository: "Smart data set for sustainability", 2017, University of Massachusetts Amherst http://traces.cs.umass.edu/index.php/Smart/Smart

[20] Richardson, I. & Thomson, M.: "One-Minute Resolution Domestic Electricity Use Data, 2008-2009" [computer file]. Colchester, Essex: UK Data Archive [distributor], October 2010. SN: 6583, http://dx.doi.org/10.5255/UKDA-SN-6583-1

[21] CLNR Project Data: "Domestic smart meters on time of use tariffs", 2017, http://www.networkrevolution.co.uk/resources/project-data/

[22] EEI-AEIC-UTC: "Smart meters and smart meter systems: a metering industry perspective", EEI- AEIC-UTC white paper, March 2011, www.eei.org/issuesandpolicy/grid-enhancements/Documents/smartmeters.pdf