State estimation of medium voltage distribution networks using smart meter measurements

CrossMark

Ali Al-Wakeel, Jianzhong Wu [*], Nick Jenkins

School of Engineering, Cardiff University, Cardiff CF24 3AA, United Kingdom

HIGHLIGHTS

- An integrated load and state estimation algorithm was developed.
- *k*-means based load estimation and IRWLS state estimation methods were applied.
- Real-time and pseudo measurements derived from smart meter measurements were used.
- Performance was investigated using 11 kV residential distribution network in England.
- Accurate estimates were obtained even with limited MV real-time measurements.

ARTICLE INFO

ABSTRACT

Distributed generation and low carbon loads are already leading to some restrictions in the operation of distribution networks and higher penetrations of e.g. PV generation, heat pumps and electric vehicles will exacerbate such problems. In order to manage the distribution network effectively in this new situation, increased real-time monitoring and control will become necessary. In the future, distribution network operators will have smart meter measurements available to them to facilitate safe and cost-effective operation of distribution networks. This paper investigates the application of smart meter measurements to extend the observability of distribution networks. An integrated load and state estimation algorithm was developed and tested using residential smart metering measurements and an 11 kV residential distribution network. Simulation results show that smart meter measurements, both real-time and pseudo measurements derived from them, can be used together with state estimation to extend the observability of a distribution network. The integrated load and state estimation algorithm was shown to produce accurate voltage magnitudes and angles at each busbar of the network. As a result, the algorithm can be used to enhance distribution network monitoring and control.

1. Introduction

The increasing use of distributed energy resources (DERs) such as distributed generators, electric vehicles, heat pumps, demand response, and energy storage brings significant uncertainties and, at high penetrations, may lead to operational difficulties in the distribution network [1]. Therefore, accurate knowledge of system states is critical for the network operator to ensure safe, prompt and cost-effective operation of the network, while making the best use of the assets [2].

At present, comprehensive real-time monitoring and control of the medium voltage (MV) and low voltage (LV) distribution networks is limited due to technical and economic constraints. In a typical distribution network, real-time measurements are provided only at the primary (33/11 kV) substations. Almost no real-time monitoring is carried out on the 11 kV or the 0.4 kV circuits [3,4]. A distribution network can therefore be described as an under-determined system. This means that the installed real-time measurements are insufficient to make the system fully observable.

In over-determined power networks with many measurements, e.g. transmission networks, state estimation [5] is used routinely to clean up the errors in a set of redundant measurements. In contrast, in under-determined distribution networks state estimation is used to find the state of the network from a limited number of measurements. A distribution network state estimator applies a minimum set of real-time measurements together with pseudo measurements to identify the operating states of the network. Pseudo measurements are obtained from load estimates using historical measurements (e.g. data collected from load surveys,

transformer kVA ratings, and customer energy meter readings) or from near real-time measurements of smart meters and measurements of Automatic Meter Reading (AMR) systems [1]. Statistical analysis, artificial intelligence, and cluster analysis methods have all been used for load estimation in power networks [6–11].

Weighted least squares (WLS) estimators [12] have been widely applied in radial distribution networks. WLS estimators exhibit a good performance only when the measurement errors and noise characteristics are known and normally distributed. The performance of WLS based state estimators is affected by outliers, gross errors in real-time and pseudo measurements, and measurement errors that do not follow normal distribution [1,13–15].

Iteratively re-weighted least squares (IRWLS) estimators have been developed to find system states more accurately and detect, identify, and eliminate the inherent errors in the measurements, network model, or system parameters [16]. IRWLS estimators, whose outputs remain insensitive to deviations in a limited number of measurements, are more robust than WLS estimators. In contrast to WLS estimation which assigns the same weight to a measurement throughout all iterations, IRWLS estimators iteratively change the measurement weight. Measurements with large residuals will have their weights reduced iteratively [17–20].

Smart metering is recognised as an important starting point in the evolution of smart grids [21]. Smart meters employ advanced metrology, control, data storage, and Information and Communication Technologies (ICT) to provide near real-time consumption information to the consumers that will help them manage their energy use, save money, and reduce greenhouse gas emissions [22]. At the same time, smart meter measurements will: enable more accurate demand forecasts, allow improved asset utilisation in distribution networks, locate outages, shorten supply restoration times and reduce the operational and maintenance costs of the networks [23,24].

Smart meters and their associated ICT infrastructure can improve the observability of distribution networks. However, their communication systems face significant technical and operational challenges [1]. The technical challenges include the lack of sufficient signal strength; the shortage of tools to detect network failure; and the indoor/outdoor placement of meters. Examples of the operational challenges include planned or unplanned maintenance of the system, software and hardware faults or malfunction of the smart meters, and customers unwilling to communicate their energy consumption data. These challenges make smart meter measurements susceptible to time delays or even temporary loss when requested by the energy suppliers or network operators [25–27]. Imprecise and lost measurements will degrade the performance or even disable a conventional state estimator.

As smart meter measurements become widely available, distribution networks will progressively evolve from being in an underdetermined to an over-determined state. Therefore, a state estimation algorithm that can work for both conditions is required. To address this requirement, an integrated load and state estimation algorithm for both under-determined and over-determined distribution networks was developed in this paper. The integrated estimation algorithm builds upon previous research on load estimation using k-means clustering [10,11] and iteratively re-weighted least squares (IRWLS) state estimation [12].

Ref. [10] presented the k-means based clustering of residential load profiles. Detailed description of both the k-means cluster analysis and the load estimation algorithm was introduced in [11]. Ref. [12] investigated the performance of the IRWLS state estimation algorithm using real-time measurements collected from an actual LV microgrid.

In this paper, the mathematical formulation of the IRWLS state estimation algorithm is presented in detail. A comprehensive case

study has been carried out to demonstrate the capabilities of the integrated estimation algorithm using a real 11 kV residential distribution network and smart meter measurements.

This paper shows that the state estimation algorithm is easy to implement and requires no prior knowledge of any variables other than primary (HV/MV) substation measurements and aggregated measurements collected from LV smart meters to reliably define the voltage magnitudes and angles at each busbar of the distribution network. Additionally, the integrated load and state estimation algorithm overcomes the requirement to have smart meter measurements in real-time by deriving pseudo measurements from past measurements of smart meters. Therefore, the integrated estimation algorithm has the capability to extend the observability and enhance monitoring, operation and control in distribution networks.

2. Framework of the state estimator

The integrated load and state estimator consists of three basic components: the k-means cluster analysis algorithm, a load estimation algorithm, and an IRWLS state estimator. The framework of the integrated estimation algorithm is illustrated in Fig. 1.

Fig. 1 shows that MV measurements (which consist of MV real-time measurements and network information) are directly input to the IRWLS state estimator. The load profiles of LV smart meters are aggregated before being applied to the integrated load and state estimation algorithm.

The time delays in delivering (and aggregating) the measurements of LV smart meters are taken into consideration through a variable (T_{Max}) which represents the maximum allowed latency of the smart metering system ICT infrastructure. If the time delay $T < T_{Max}$, the aggregated measurements of the smart meters are treated as real-time measurements and passed to the IRWLS state estimator. For any MV node, if the information from its downstream LV smart meters cannot be collected completely within T_{Max} or smart meters are not installed in all LV network premises, the load estimation algorithm will provide the estimated MV nodal load measurements to allow the state estimator to function.

An assumption was made that there are 384 residential smart meters connected to each MV/LV transformer [28]. An aggregated daily load profile was created by summing the measurements of the 384 smart meters at each half-hour time step. Eq. (1) illustrates the aggregation of smart meter measurements.

$$\mathbf{lp}_{\text{agg.daily}} = \left[\left(\sum_{i=1}^{384} \text{lp}_i(t) \right)_{t=1}, \left(\sum_{i=1}^{384} \text{lp}_i(t) \right)_{t=2}, \ldots, \left(\sum_{i=1}^{384} \text{lp}_i(t) \right)_{t=48} \right]$$

(1)

$\mathbf{lp}_{\text{agg.daily}}$ is a vector that represents the aggregated daily load profile, $\text{lp}_i(t)$ is the measurement of the ith smart meter at the tth half-hour, i is the smart meter index, i = 1, 2, 3, . . ., 384, t is the half-hour index, t = 1, 2, 3, . . ., 48.

The aggregated daily load profiles were divided into training and test sets. Details of the training and test sets are given in Section 3. The training set consists of historical (previous) load profiles, but excludes any profiles that contain negative, zero and missing half-hourly measurements. Load profiles of the training set were used as inputs to the k-means clustering algorithm to obtain a number of clusters and their corresponding centres.

The test set is made up of real-time load profiles. If the profiles of the test set

a. contain no missing measurements, these profiles are passed to the state estimator as real-time measurements;
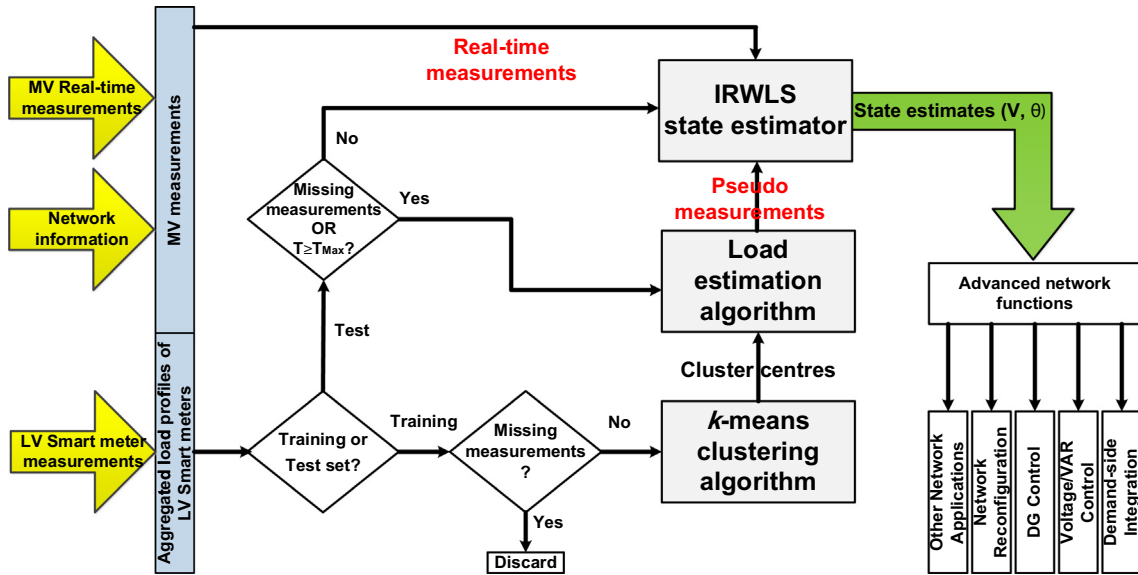
Fig. 1. Framework of the load and state estimator.

b. contain any missing measurements, the profiles are passed to the load estimation algorithm to estimate these measurements. Outputs of the load estimator (i.e. the load profiles with estimated measurements) are used as pseudo measurements by the state estimator.

Finally, an IRWLS state estimator using real-time smart meter measurements, MV real-time measurements (collected at the primary substation) and pseudo measurements on the MV side of MV/LV transformers (derived from LV smart meter measurements) was applied to define the voltage magnitude and angle at each busbar of a distribution network.

2.1. k-means cluster analysis algorithm

The k-means algorithm that was developed in [11] is used in this paper. The load profiles (at the 11 kV level) of the training set and the maximum number of clusters are the inputs of the k-means algorithm. The outputs of the k-means algorithm include the cluster centres, and load profiles assigned to their respective clusters.

The k-means clustering method iteratively groups n training load profiles (each comprised of T half-hourly measurements) into k clusters, by minimising the intra-cluster sum of squared distances between the load profiles and cluster centres. Eq. (2) shows the objective function of the k-means method.

$$\min \sum_{j=1}^{k} \sum_{\substack{i=1, \\ i \in j}}^{n} \left\| \mathbf{lp}_{\text{agg.daily},i} - \mathbf{cc}_j \right\|^2 \tag{2}$$

$\mathbf{lp}_{\text{agg.daily},i}$: is a vector representing the ith aggregated daily load profile, $i = 1, 2, 3, \ldots, n$, and $\mathbf{cc}_j$: is a vector that represents the jth cluster centre, $j = 1, 2, 3, \ldots, k$. The ith load profile is described as $\mathbf{lp}_i = [\text{lp}_i(t=1), \text{lp}_i(t=2), \text{lp}_i(t=3), \ldots, \text{lp}_i(t=T)]$. Similarly, the jth cluster centre is defined as $\mathbf{cc}_j = [\text{cc}_j(t=1), \text{cc}_j(t=2), \text{cc}_j(t=3), \ldots, \text{cc}_j(t=T)]$. At each iteration, the Average Euclidean Distance (AvED) is calculated between the load profiles and their cluster centres according to Eq. (3). As a result, each load profile is assigned to the cluster that has the nearest centre.

$$AvED(\mathbf{lp}_{\text{agg.daily},i}, \mathbf{cc}_j) = \sqrt{\frac{\sum_{t=1}^{T}(\text{lp}_i(t) - \text{cc}_j(t))^2}{T}} \tag{3}$$

The centre of a cluster is the mean values of all load profiles that are assigned to this specific cluster, calculated at each half-hourly time step. Eq. (4) defines the centre of any cluster.

$$\mathbf{cc}_{j,final} = \left[\frac{\sum_{i=1}^{\beta} \text{lp}_i(t=1)}{\beta}, \frac{\sum_{i=1}^{\beta} \text{lp}_i(t=2)}{\beta}, \frac{\sum_{i=1}^{\beta} \text{lp}_i(t=3)}{\beta}, \ldots, \frac{\sum_{i=1}^{\beta} \text{lp}_i(t=T)}{\beta} \right] \tag{4}$$

β is the number of load profiles assigned to the jth cluster.

Pycluster [29], an open source cluster analysis package, was used to develop the k-means cluster analysis algorithm in Python 2.7.

2.2. Load estimation algorithm

The load estimation algorithm applies the cluster centres (that were obtained from the clustering of the training load profiles) to estimate any missing measurements that exist in the test profiles. A Canberra distance function is used to link test profiles (with missing measurements) to the training cluster centres. Eq. (5) calculates the average value of Canberra distance between the ith test load profile and the jth training cluster centre.

$$d(\mathbf{lp}_{\text{agg.daily},i}, \mathbf{cc}_j) = \frac{\sum_{t=1}^{T} \frac{|\text{lp}_i(t) - \text{cc}_j(t)|}{|\text{lp}_i(t)| + |\text{cc}_j(t)|}}{T} \tag{5}$$

Canberra distance measures the distance between $\mathbf{lp}_i$ and $\mathbf{cc}_j$ on a rectilinear basis such that the absolute difference between any two half-hourly measurements of the load profile and cluster centre is divided by the sum of the absolute values of these two measurements. The average value is then calculated as the average Canberra distance. As a result, a set of r half-hourly measurements (one measurement to be estimated plus the r − 1 half-hourly measurements that precede it) is paired to the nearest cluster centre (that has the same duration of r half-hours). The measurements

were estimated iteratively with only one half-hourly measurement estimated in each iteration. Fig. 2 demonstrates the concept of load estimation using cluster centres.

2.3. IRWLS state estimation algorithm

Iteratively re-weighted least squares (IRWLS) estimators [19] minimise the sum of weighted squared residuals between the measured and the estimated values of the network states, as shown in Eq. (6), subject to the constraints given by the measurement equations shown in Eq. (7):

$$\min \sum_{i=1}^{m}(z_i - f_i(x))^2 W_i \tag{6}$$

$$\text{Subject to } \mathbf{z} = \mathbf{f}(\mathbf{x}) + \mathbf{e} \tag{7}$$

where $\mathbf{z}$ is the measurements vector, $\mathbf{f}$ is the vector of nonlinear measurement functions, $\mathbf{x}$ is the vector of system state variables, $\mathbf{e}$ is the vector of measurement errors, W_i is the weight of the ith measurement, i is the measurement index, and m is the number of the measurements.

Eq. (6) is formulated in the same way as the WLS estimator resulting in Eq. (8)

$$\min (\mathbf{z} - \mathbf{f}(\mathbf{x}))^T \mathbf{W}(\mathbf{z} - \mathbf{f}(\mathbf{x})) \tag{8}$$

given that $\mathbf{W}$ is the weight matrix which is a diagonal matrix such that the weights of the measurements are in the main diagonal. The weight, that is equal to the reciprocal of the variance of a measurement $(1/\sigma_i^2)$, reflects the accuracy of the measurement. Measurements are normalised with respect to their standard deviations.

$$\mathbf{s} = \frac{\mathbf{z}}{\boldsymbol{\sigma}} = \left[\frac{z_1}{\sigma_1}, \frac{z_2}{\sigma_2}, \cdots, \frac{z_m}{\sigma_m}\right] \tag{9}$$

In a general form, $\mathbf{h}(\mathbf{x}) = \mathbf{f}(\mathbf{x})/\boldsymbol{\sigma}$, i.e. for each i, $h_i(\mathbf{x}) = f_i(\mathbf{x})/\sigma_i$. Consequently, the normalised residual between the ith measurement and its calculated value is defined as:

$$R_i = |s_i - h_i(\mathbf{x})| \tag{10}$$

The diagonal elements of the weight matrix, W_i, are modified iteratively (with iteration count α) according to the relationship

$$W_i^\alpha \propto \frac{1}{R_i^\alpha} \tag{11}$$

To avoid convergence problems that might result from the division by very small residuals, the range of weights was limited between minimum and maximum weight thresholds. The minimum weight threshold was set to 0.001 while the maximum weight threshold was unity. Eqs. (12)–(14) define the criteria of measurement re-weighting.

$$R_{max}^{k} = \max_{i}(R_i^k) \tag{12}$$

If $R_i^k > 0.001 R_{max}^{k}$ then:

$$W_i^k = 0.001 R_{max}^{k}/R_i^k \tag{13}$$

else:

$$W_i^k = 1 \tag{14}$$

The IRWLS state estimator can handle real-time measurements collected from the primary/secondary substation(s), distributed generators, smart meters and pseudo measurements based on load estimates. The integration of distributed generators brings several uncertainties which will impact the operation of distribution networks. The IRWLS state estimator is capable of accommodating uncertainties arising from the deployment of DGs. Distributed generators equipped with metering devices will have their measurements input to the state estimator as (near) real-time measurements. Alternatively, any unmetered generation can be estimated using the load estimation algorithm. In this case, the estimated generation will be modelled as a negative load with relatively low weight in the state estimation algorithm. The modelling of DG is not the focus of this paper. Further details can be found in [12].

3. Smart meter measurements dataset

Load profiles using measurements from residential smart meters were used to train, test and validate the performance of the load and state estimation algorithm. The load profiles were obtained from the Irish smart metering Customer Behaviour Trials (CBT) and were accessed via the Irish Social Science Data Archive [30]. The Irish smart metering CBT is a large and statistically robust smart metering trial. This trial investigated the impact of smart meters upon the power consumption of different types of customers.

The trials extended over 18 months (between 1st July 2009 and 31st December 2010) and covered more than 4200 residential customers and 485 small-and-medium enterprises (SMEs). Smart meters installed at the customers' premises recorded the consumption data in terms of the average power consumption in half-hourly time steps. For any individual smart meter, a daily load profile consists of 48 half-hourly measurements. The first measurement was recorded at hour 00:30 and is the average power consumed between hour 00:00:00 and 00:29:59. The last measurement is the average power consumption between hours 23:30:00 and 23:59:59, and was recorded at hour 00:00.

Two weeks of measurements were used in this study. One week of measurements (between 20th July and the end of 26th July 2009) was used as a training set to train the k-means clustering algorithm. A set of measurements collected over the week from 27th July until the end of 3rd August 2009, was used to test the load and state estimation algorithm.

Daily and segmented (generated by partitioning the daily load profiles using different time windows) load profiles were used in this study. Whereas a daily load profile consists of 48 half-hourly measurements, a segmented load profile extends over a time window that is less than or equal to 24h. Fig. 3 illustrates the concept of segmented load profiles.

Fig. 3 shows that a daily load profile is segmented into 17 segments each with a time window equal to 16h. For any segmentation time window (S) in hours, provided that the segmented profiles are rolled one half-hourly step at a time, then the number of profiles is determined according to Eq. (15) [31]

$$number\ of\ segmented\ profiles = (n \times T) - 2S + 1 \tag{15}$$

where n is the number of the daily load profiles, and T is the number of half-hourly measurements per a daily load profile. The present study applies a time window of 16h to segment the daily load profiles.

4. Test distribution network

An 11 kV residential distribution network was used to investigate the performance of the integrated load and state estimation algorithm. Fig. 4 is a single line diagram of the distribution network.

The distribution network [32], that is located in England, consists of a radial feeder that has 12 busbars and 11 sections. Busbar 1 is the secondary side of the primary substation while busbars
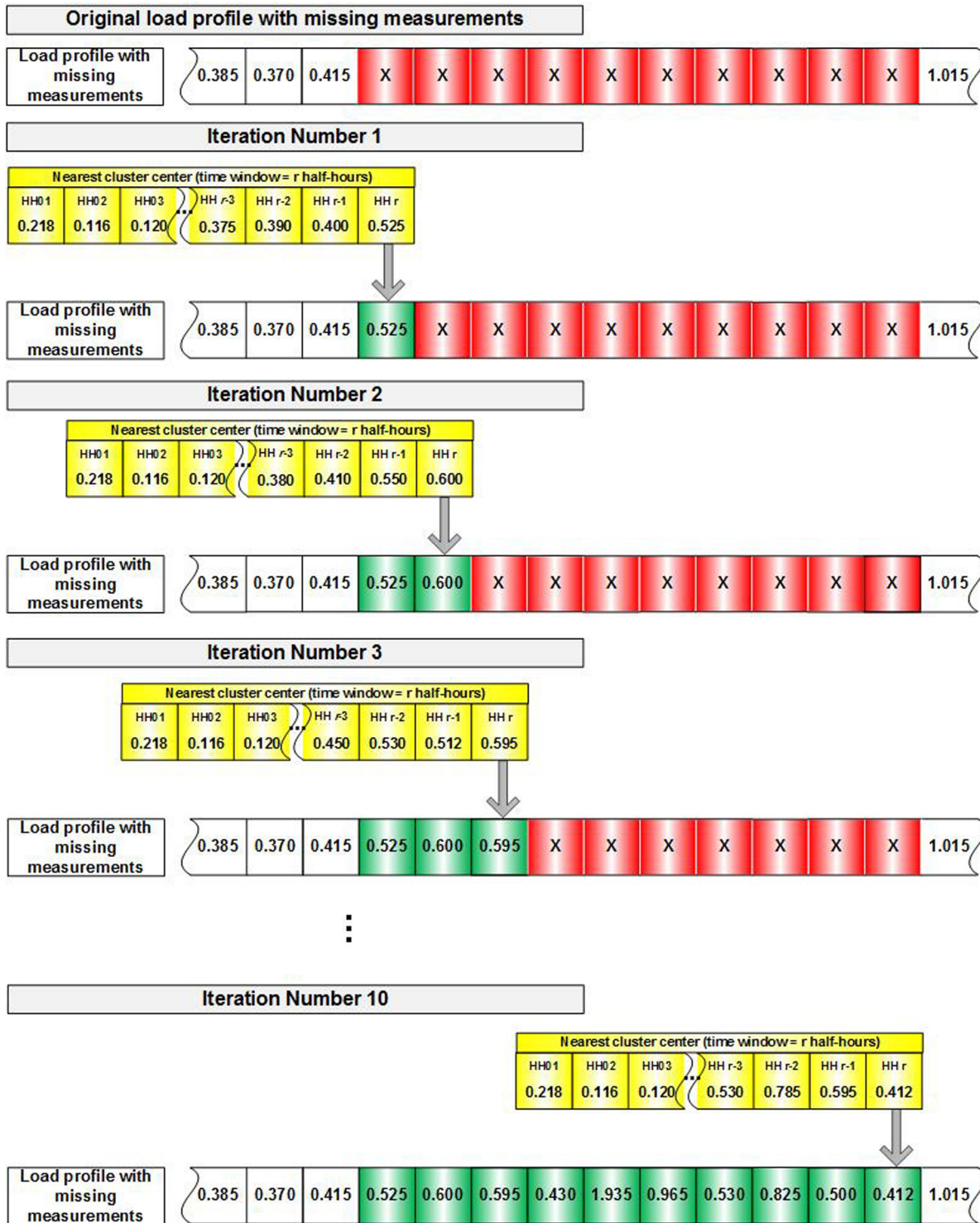
Fig. 2. Load estimation using cluster centres.

2–12 represent MV/LV transformers. The network has a peak load of 3642 kW and 1245 kVAr and an average feeder R/X ratio of 2.2.

5. Load and state estimation methodology

For both the training and test sets of load profiles, measurements from the smart meters installed at customer premises were aggregated using Eq. (1) to give the active power consumption at the LV side of the MV/LV transformer. The following steps describe

the approach adopted to obtain the (active and reactive) load profiles at the MV side of the transformer.

1. The aggregated daily load profiles at the LV level were normalised with respect to their maximum values. The normalised profiles were then scaled up to match the peak (active and reactive) power demand of each transformer.
2. A random percentage in the range between 6% and 10% [33,34] of the load demand (at the LV side of the transformer) was added to each daily load profile. This percentage reflects the

Fig. 3. Segmentation of a daily load profile into 16h segmented load profiles.

Fig. 4. Single line diagram of the 11 kV distribution network.

total (active and reactive) power loss (of service cables and the transformer). The outputs of this step are the daily load profiles at the MV side of each MV/LV transformer.

5.1. Cluster analysis of the load profiles

Using Eq. (15), the integrated load and state estimation algorithm started with dividing the daily load profiles (at the MV side of the MV/LV transformer) of the training set into a number of segmented profiles each with 16h time window. The segmented load profiles were clustered using the k-means cluster analysis algorithm (Section 2.1).

The outputs of this stage are the (active and reactive power) cluster centres at the MV level of each MV/LV transformer.

5.2. Load estimation of missing measurements

The load estimation algorithm (Section 2.2) replaced missing measurements of the test set with estimated measurements obtained from the cluster centres (Section 5.1) using Canberra distance function.

In order to simulate the loss of smart meter measurements at an MV/LV transformer, the actual values of the test load profiles (at the MV side) were replaced by zeros. For each test load profile, different durations of lost measurements (T_{loss}) from 1 to 24 consecutive hours were simulated. The load estimation algorithm

iteratively estimated the lost measurements one half-hourly measurement at a time. Using a brute-force approach, all possible combinations of (T_{loss}) hours of measurement loss were covered.

The estimated measurements at the MV (11 kV) level are applied as pseudo measurements to the IRWLS state estimator.

5.3. Integrated load and state estimation

The integrated load and state estimation algorithm was applied to estimate the voltage magnitudes and angles at each busbar[1] of the 11 kV distribution network shown in Fig. 4. The integrated estimation algorithm utilised the test profiles and made the following assumptions.

1. A set of real-time measurements is available at the MV side of the primary substation. The voltage magnitude, active and reactive powers injections at Busbar 1 and active and reactive power flows at the sending end of the first feeder segment (that connects Busbars 1 and 2) are measured in real-time.
2. The active and reactive power demand at the MV side of each MV/LV transformer are modelled as
   a. real-time measurements when the aggregated smart meter measurements are collected and applied to the IRWLS state estimator in real-time;
   b. pseudo measurements when smart meter measurements are lost. In this case, the power demand at the MV side of a MV/LV transformer is estimated using the load estimation algorithm. The estimated demand is applied as a pseudo measurement to the IRWLS state estimator.

3. The integrated estimation algorithm was applied to estimate the network operating state(s) using pseudo measurements (that were estimated due to measurement loss) in the range between 1 and 24h.

The base case operating state of the distribution network was calculated using Newton-Raphson load flow analysis. Pypower [35], a MATPOWER [36] based power system simulation package was applied to run load flow in Python 2.7.

---

[1] Except the voltage angle at Busbar 1 (slack busbar) where it is held constant at 0°.

For each busbar, *m*, of the distribution network and at each half-hourly time step, *t*, the errors between the load flow (x_{LF}) and the estimated (x_{SE}) operating states of the network were quantified in terms of the Mean Absolute Percentage Error (MAPE), the mean of the MAPE and the maximum value of the maximum APE. Eqs. (16)–(18) define the MAPE [37], mean of the MAPE and the overall maximum value of APE.

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{x_{LF}(t) - x_{SE}(t)}{x_{SE}(t)}\right| \times 100 \qquad (16)$$

$$Mean\ MAPE = \frac{1}{M}\sum_{m=1}^{M}\frac{1}{T}\sum_{t=1}^{T}\left|\frac{x_{LF}(t) - x_{SE}(t)}{x_{LF}(t)}\right| \times 100 \qquad (17)$$

$$Max.Maximum\ APE = \max_{M}\left(\max_{T}\left(\left|\frac{x_{LF}(t) - x_{SE}(t)}{x_{LF}(t)}\right|\right)\right) \times 100 \qquad (18)$$

T is the overall number of half-hourly time steps and *M* is the number of busbars of the network.

6. Results and discussion

Differences exist between the load and voltage profiles on weekdays and weekends of the test period. These differences are results of the diversified use of appliances at different times of the day which can be linked to the daily schedule of occupants within the residential premises. The integrated load and state estimation algorithm was applied to estimate the load and voltage profiles of each busbar (except the load and voltage angle of Busbar 1) of the residential network regardless of the day type (weekday or weekend).

6.1. Load estimation

The load estimation algorithm was applied to estimate the active and reactive power demand of each MV/LV transformer. Fig. 5 shows the actual and estimated load profiles of Busbar 11 for a representative weekday and weekend.

In Fig. 5, the solid black profile is the actual active power demand calculated using smart meter measurements from the domestic premises directly. For durations of lost measurements ranging between 1 and 24h, the solid red profile is the mean of all estimated profiles that were produced by the load estimation algorithm (due to the application of the brute-force approach). The dashed red[2] profiles are the mean of the minimum and maximum estimated profiles.

The performance of the load estimation algorithm using low frequency smart metering data was comprehensively analysed in [11]. Accuracy of the estimated load was reported to decrease as the duration of the estimated load increases.

6.2. Load flow analysis

For the representative weekday and weekend, Figs. 6 and 7 show the range of load flow voltage magnitudes and angles at each busbar of the residential distribution network.

The solid black profile is the mean value of the voltage magnitudes (Fig. 6) and angles (Fig. 7) while the red profiles are the minimum and maximum voltage magnitudes (and angles) calculated over the 48 half-hourly time steps during a day. For any busbar, the mean voltage magnitude and angle were obtained according to Eqs. (19) and (20)

$$Mean\_V_i = \frac{1}{T}\sum_{t=1}^{T}V_i(t) \qquad (19)$$

$$Mean\_\theta_i = \frac{1}{T}\sum_{t=1}^{T}\theta_i(t) \qquad (20)$$

The maximum and minimum voltage magnitudes and angles of any busbar are calculated according to Eqs. (21)–(24)

$$Max\_V_i = \max_{t\in T}V_i(t) \qquad (21)$$

$$Min\_V_i = \min_{t\in T}V_i(t) \qquad (22)$$

$$Max\_\theta_i = \max_{t\in T}\theta_i(t) \qquad (23)$$

$$Min\_\theta_i = \min_{t\in T}\theta_i(t) \qquad (24)$$

In Eqs. (19)–(24), *V* is the busbar voltage magnitude, *θ* is the busbar voltage angle, *i* is the busbar index, *t* is the index of the half-hourly measurement, and *T* is the overall number of half-hours within a day.

For all half-hours during a day, the minimum voltage magnitude (Fig. 6) and the most negative voltage angle (Fig. 7) were observed at Busbar 11 which is located at the end of the feeder. Busbar 11 was used in the following studies.

6.3. Integrated load and state estimation

The integrated load and state estimation algorithm was applied to estimate the operating state of the residential distribution network using the assumptions presented in Section 5.3.

The estimated voltage magnitudes and angles and the active and reactive power demands at each busbar of the network were compared with their respective values obtained from the load flow solution.

6.3.1. Accuracy of the estimated voltage magnitudes

Fig. 8 shows the load flow (to represent the actual values) and estimated voltage magnitudes of Busbar 11 for the weekday and weekend, up to 24h of pseudo measurements.

In Fig. 8, the grey profiles are the estimated voltage magnitudes of Busbar 11 while the red profile represents the voltage magnitudes obtained from the load flow analysis.

Fig. 9 illustrates a box-whisker plot [38] of the MAPE distribution of the estimated voltage magnitudes (of the test period) at Busbar 11 for 1–24h of pseudo measurements.

The red horizontal lines in Fig. 9 represent the median of the MAPE which varies between 0.028% (for one hour of pseudo measurements) and 0.05% (for a day ahead estimated load). The boxes below the red lines are the errors between the first quartile and the median of the MAPE, whereas those boxes above the red lines are the error values between the median and the third quartile of the MAPE. The whiskers are the minimum and maximum values of the MAPE. Fig. 9 shows that for 24h of pseudo measurements, the MAPE of the estimated voltage magnitudes at Busbar 11 were less than 0.08%.

6.3.2. Accuracy of the estimated voltage angles

For the weekday and weekend, the load flow and estimated voltage angles at Busbar 11, for 1–24h of pseudo measurements, are shown in Fig. 10.

Fig. 10 shows that the estimated voltage angles (grey profiles) follow the voltage angles (red profile) which were obtained from the load flow analysis.

A box-whisker representation of the MAPE of the estimated voltage angles during the test period is shown in Fig. 11.

Fig. 11 shows that as the duration of pseudo measurements increases from 1 to 24h, the median of the MAPE (of the estimated

---

[2] For interpretation of color in Figs. 5, 7–10, and 12, the reader is referred to the web version of this article.
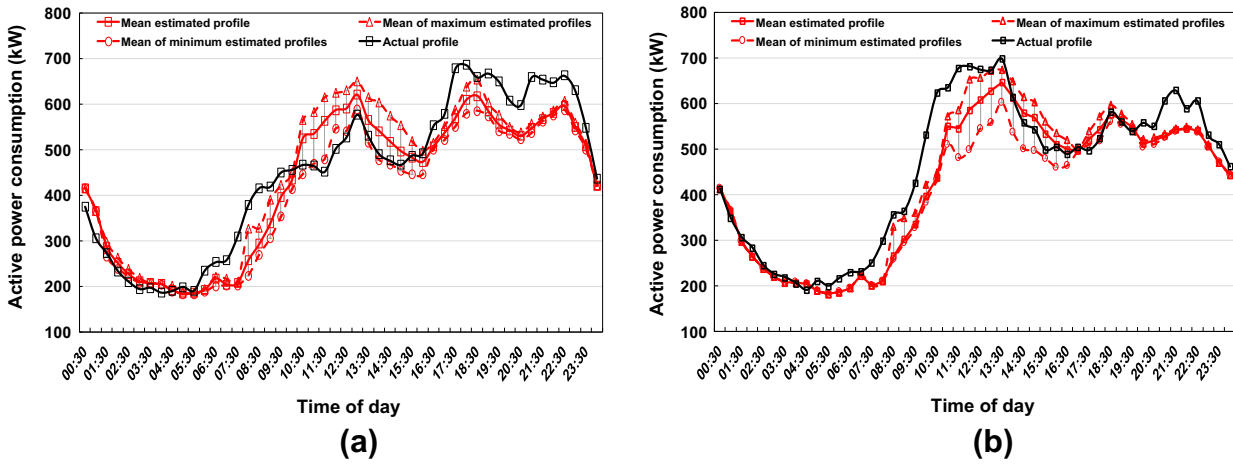
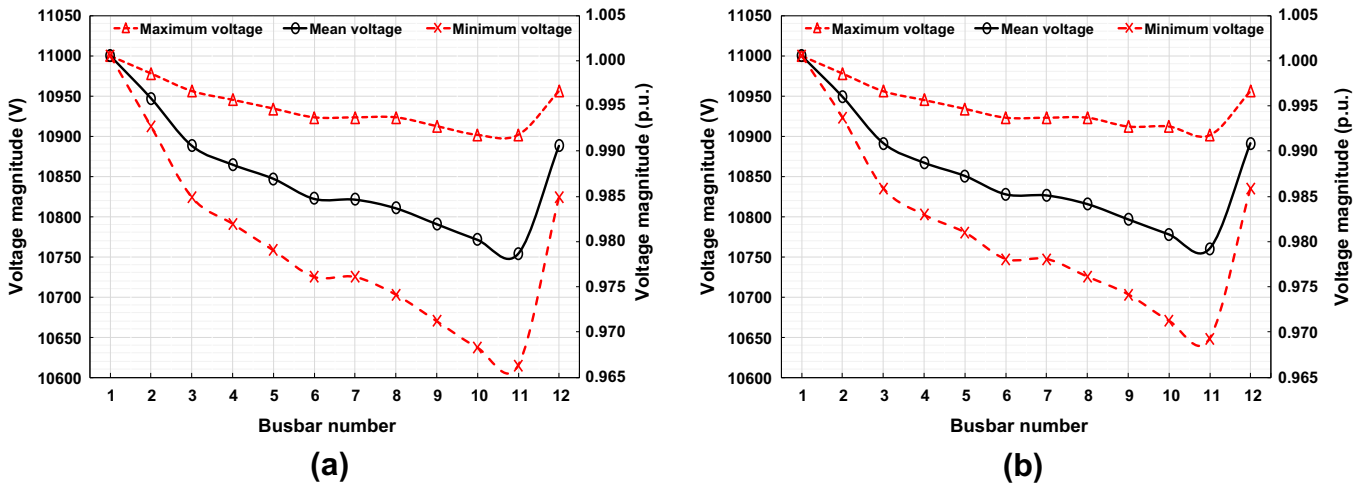**Fig. 5.** Actual and estimated active demand of Busbar 11 for (a) weekday and (b) weekend.

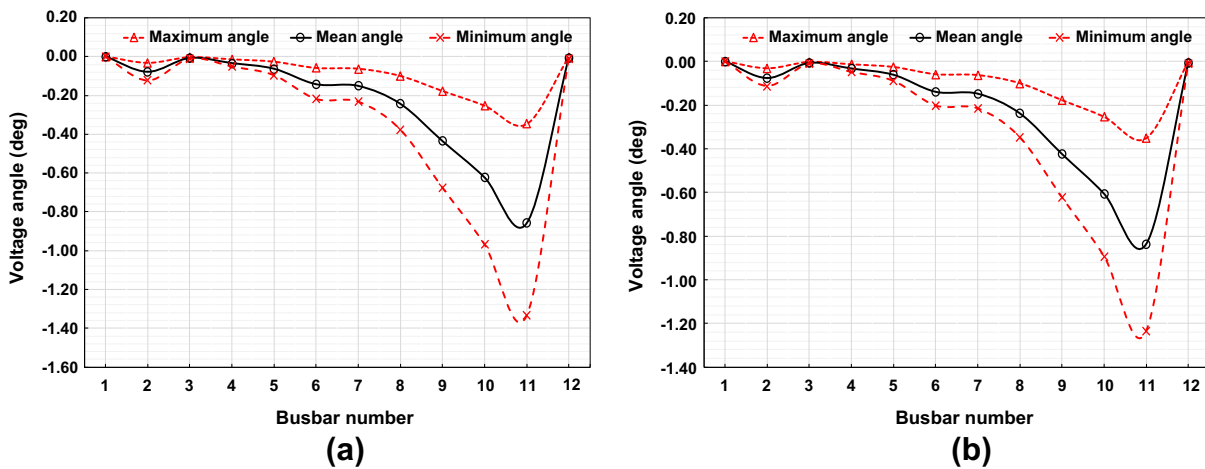**Fig. 6.** Load flow voltage magnitudes for (a) weekday and (b) weekend.

Fig. 7. Load flow voltage angles for (a) weekday and (b) weekend.

voltage angle at Busbar 11) increases from 0.4% to approximately 5.5%. Using the maximum values of the MAPE, Fig. 11 shows that the error in voltage angle estimates was around 9%.

The accuracy of load estimates affects the accuracy of the state estimation outputs. The increase in the accuracy of load estimates

results in a similar increase in the accuracy of the estimated voltage magnitudes and angles at each busbar of the network. Figs. 9 and 11 indicate that accuracies of the estimated voltage magnitudes and angles at Busbar 11 increase as a result of increasing the accuracy of the estimated load.
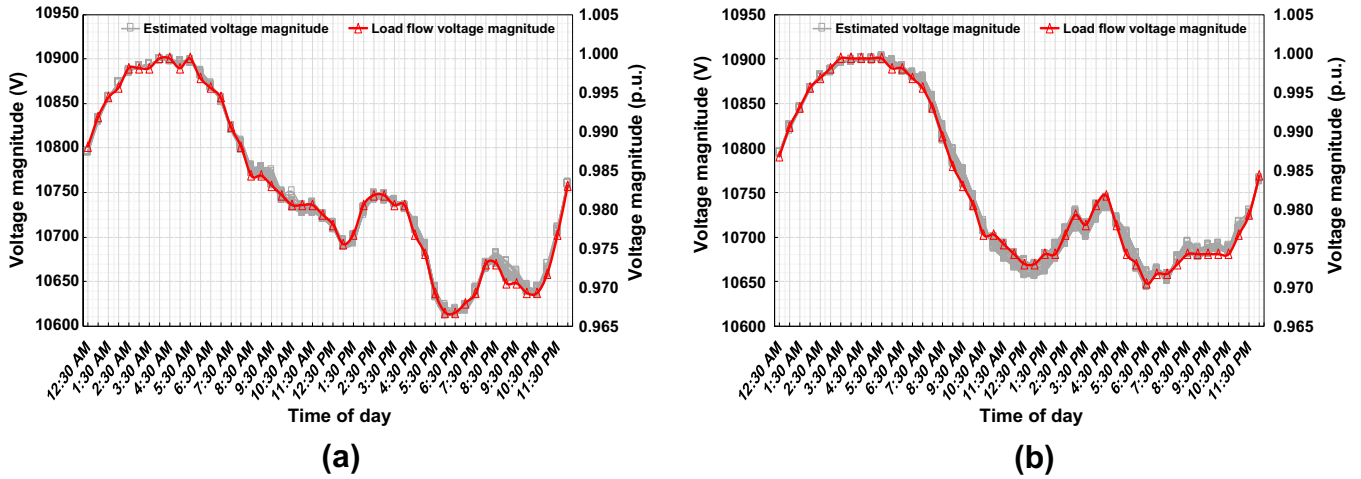
Fig. 8. Load flow and estimated voltage magnitudes for (a) weekday and (b) weekend – Busbar 11.
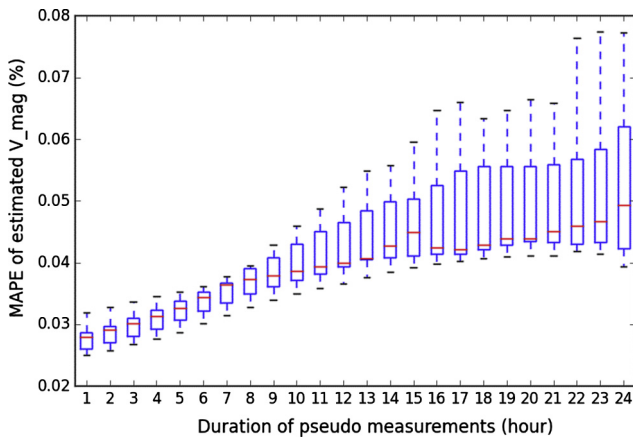
Fig. 9. MAPE distribution of Busbar 11 voltage.
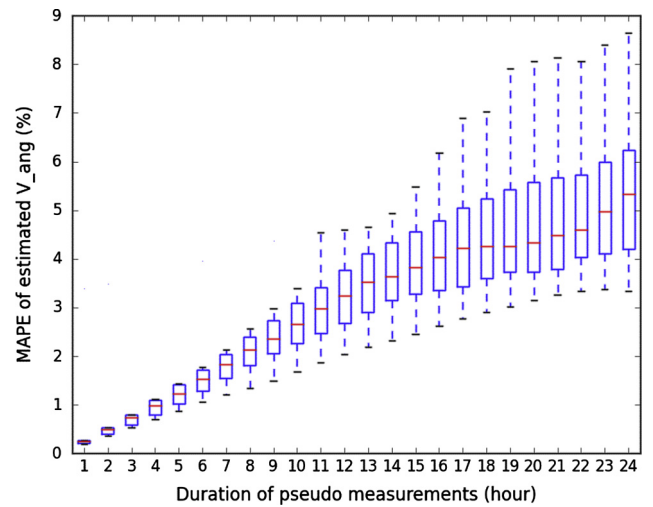
Fig. 11. MAPE distribution of Bus 11 voltage angles.

6.3.3. Accuracy of the estimated power demands

Using the estimated voltage magnitudes and angles, the state estimator calculates the active and reactive power demands at each MV/LV transformer of the residential distribution network. Fig. 12 shows the load flow and estimated active power demand of Busbar 11 on both the weekday and weekend, for 1–24h of pseudo measurements.

In Section 6.1, Fig. 5 showed the estimated active power demand of Busbar 11 resulting from the load estimation algorithm. However, Fig. 12 shows the active power demand calculated using the estimated voltage magnitudes and angles.
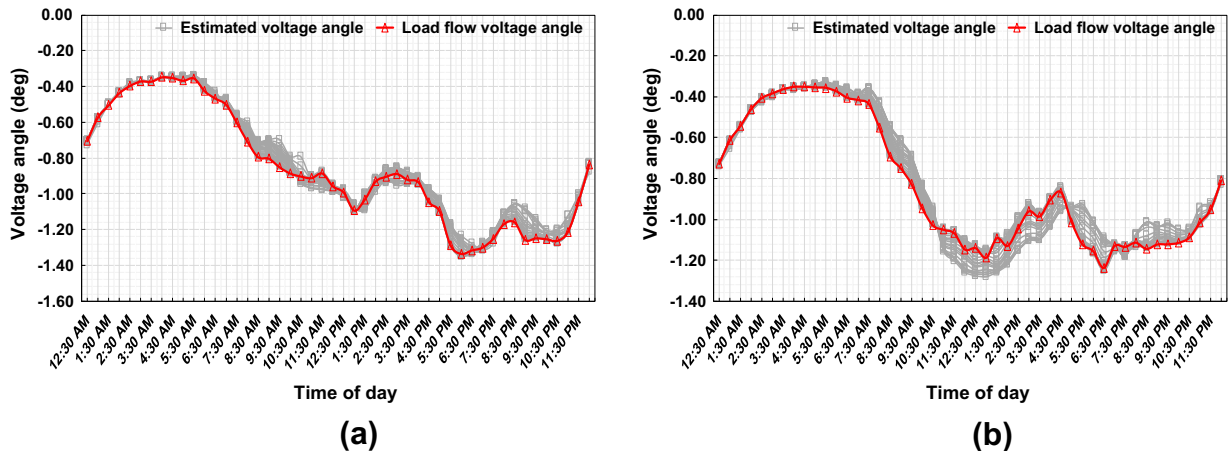
Fig. 10. Load flow and estimated voltage angles for (a) weekday and (b) weekend – Busbar 11.
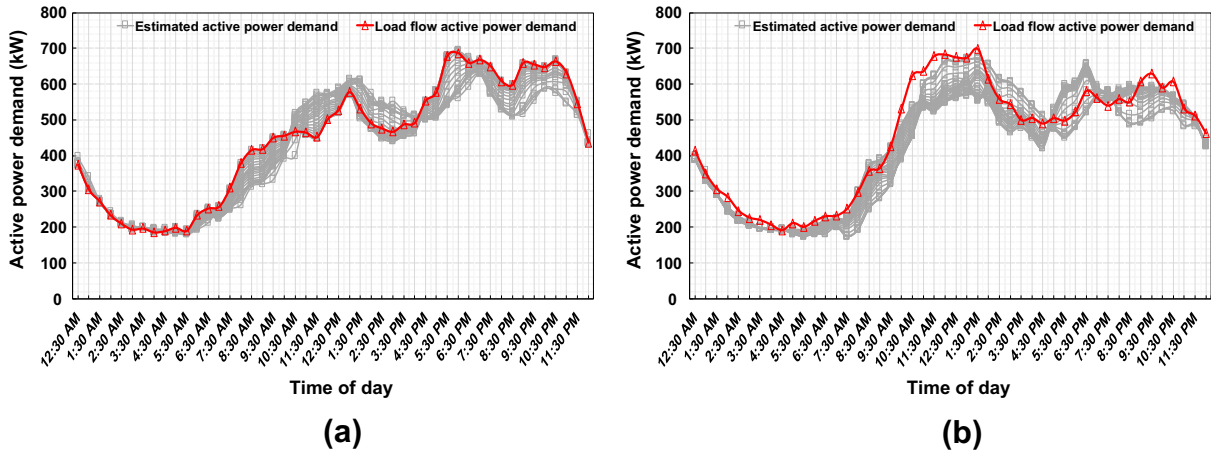
Fig. 12. Load flow vs. estimated active power demand for (a) weekday and (b) weekend – Busbar 11.
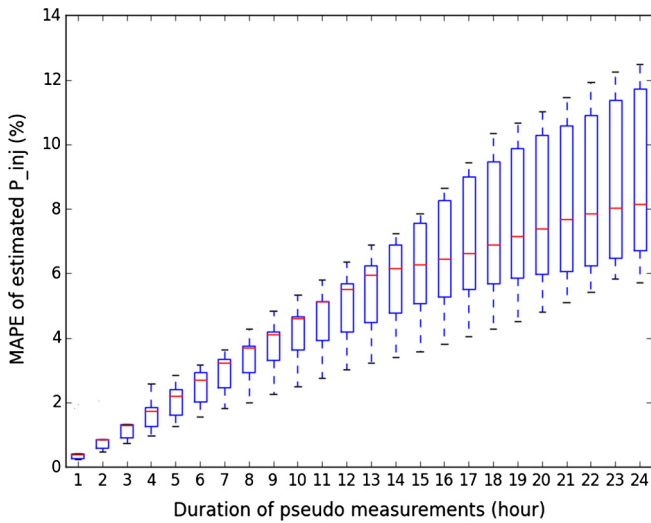
Fig. 13. MAPE distribution of active power demand at Busbar 11.

The red profile in Fig. 12 is the active power demand obtained from the load flow analysis, whereas the grey profiles represent the power demand calculated using the estimated voltage magnitudes and angles. As the duration of pseudo measurements increases from 1 to 24h, the estimated demand shown in Fig. 12 diverges from the load flow demand.

For 1–24h of pseudo measurements, Fig. 13 shows a box-whisker plot representing the MAPE distribution of the estimated active power demand at Busbar 11 during the test period.

Fig. 13 shows that up to 10h of pseudo measurements, 75% of the estimation errors were less than 5%. At the same time, a day ahead active power demand was estimated with a median MAPE that was less than 8%.

6.3.4. Sensitivity of state estimator outputs to uncertainty in LV network losses

The results shown in Figs. 8–13 were obtained assuming that the losses of the LV network and MV/LV transformers during the test period were known and were equal to their values (selected between 6% and 10% of the load demand) defined in the training period (Section 5).

In order to reflect the uncertainty in network losses at its lowest voltage level, these losses were modelled in the range of 110–200% of their actual (6–10% of the load demand) values and were simultaneously added to the test load profiles of all MV/LV transformers.

For different levels of uncertainty in the LV network losses, the integrated load and state estimation algorithm was re-run to estimate the operating state of the network using the new set of test load profiles.

The mean value of the MAPE (Eq. (17)) and the overall maximum value of absolute percentage errors (Eq. (18)) of the estimated voltage magnitudes for different levels of uncertainty in LV network losses up to 24h of pseudo measurements are shown in Fig. 14.
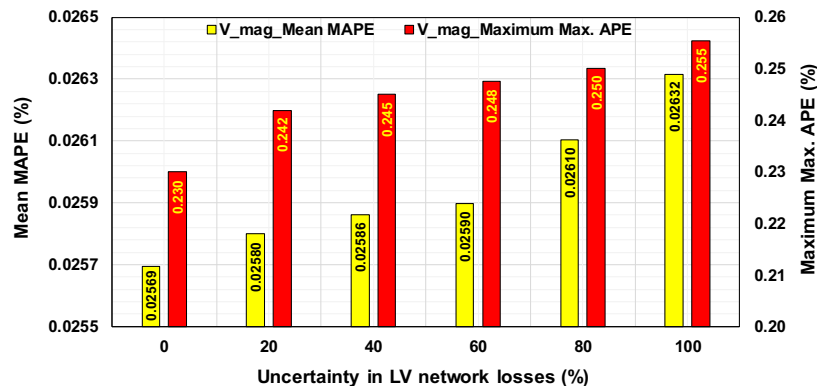

Fig. 14. Errors in estimated voltage magnitude vs. uncertainty in LV network losses.
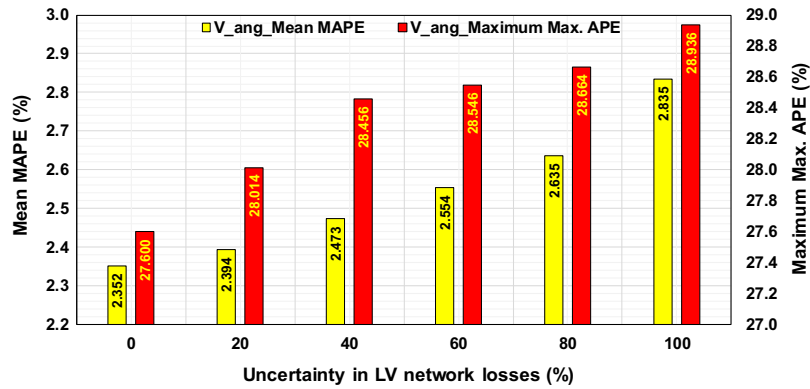
Fig. 15. Errors in estimated voltage angle vs. uncertainty in LV network losses.

Fig. 14 shows that the estimation errors increase as the uncertainty in LV network losses increases. However, the mean error of the estimated voltage magnitude remains below 0.03% while the overall maximum error attains values less than 0.26% even for 100% uncertainty in the LV network losses.

The mean MAPE and the overall maximum absolute percentage errors of the estimated voltage angles are shown in Fig. 15.

Fig. 15 shows that although the increase in the level of uncertainty in LV network losses is accompanied with an increase in the errors of estimated voltage angle, however, the mean MAPE remains below 3% and the overall maximum error less than 30%.

Despite the differences between voltage and load profiles on weekdays and weekends, the integrated load and state estimation algorithm is capable of producing reliable estimates of the load and voltage profiles at each busbar of a distribution network.

7. Conclusions

An integrated load and state estimation algorithm was developed and used to estimate the operating state, defined in terms of the busbar voltage magnitudes and angles, of a residential distribution network. The k-means based load estimation algorithm showed promising application of cluster analysis methods to extract consumption patterns from, and estimate the load of, aggregated smart meters at the 11 kV level.

The integrated estimation algorithm was tested using real-time and pseudo measurements derived from residential smart meter measurements and an 11 kV residential distribution network. Under extreme conditions of real-time measurements only at the primary substation, pseudo measurements of MV/LV transformers, and 100% uncertainty in the LV network losses, the overall maximum APE was less than 0.26% in the estimated voltage magnitudes and 29% in the estimated voltage angles. In this case, the mean values of the MAPE were less than 0.03% for the estimated voltage magnitudes and 3% for the estimated voltage angles.

Simulation results showed that the integrated load and state estimation algorithm reliably defined the operating state of a distribution network using limited MV real-time measurements and both real-time and pseudo measurements derived from smart metering data. Therefore, the algorithm can be applied to extend the observability and improve the operation and control of distribution networks.

References

[1] Wu J, He Y, Jenkins N. A robust state estimator for medium voltage distribution networks. IEEE Trans Power Syst 2013;28:1008–16. http://dx.doi.org/10.1109/TPWRS.2012.2215927.

[2] Ekanayake J, Jenkins N, Liyanage K, Wu J, Yokoyama A. Smart grid: technology and applications. John Wiley & Sons; 2012.

[3] Singh R, Pal BC, Vinter RB. Measurement placement in distribution system state estimation. IEEE Trans Power Syst 2009;24:668–75. http://dx.doi.org/10.1109/TPWRS.2009.2016457.

[4] Singh R, Pal BC, Jabr RA. Statistical representation of distribution system loads using gaussian mixture model. IEEE Trans Power Syst 2010;25:29–37. http://dx.doi.org/10.1109/TPWRS.2009.2030271.

[5] Schweppe F, Wildes J. Power system static-state estimation, Part I: exact model. IEEE Trans Power Apparatus Syst 1970;PAS-89:120–5. http://dx.doi.org/10.1109/TPAS.1970.29267.

[6] McLoughlin F, Duffy A, Conlon M. A clustering approach to domestic electricity load profile characterisation using smart metering data. Appl Energy 2015;141:190–9. http://dx.doi.org/10.1016/j.apenergy.2014.12.039.

[7] Räsänen T, Ruuskanen J, Kolehmainen M. Reducing energy consumption by using self-organizing maps to create more personalized electricity use information. Appl Energy 2008;85:830–40. http://dx.doi.org/10.1016/j.apenergy.2007.10.012.

[8] Zhou K, Fu C, Yang S. Big data driven smart energy management: from big data to big insights. Renew Sustain Energy Rev 2016;56:215–25. http://dx.doi.org/10.1016/j.rser.2015.11.050.

[9] Räsänen T, Voukantsis D, Niska H, Karatzas K, Kolehmainen M. Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data. Appl Energy 2010;87:3538–45. http://dx.doi.org/10.1016/j.apenergy.2010.05.015.

[10] Al-Wakeel A, Wu J. K-means based cluster analysis of residential smart meter measurements. Energy Procedia 2016;88:754–60.

[11] Al-Wakeel A, Wu J, Jenkins N. K-means based load estimation of domestic smart meter measurements. Appl Energy 2016. http://dx.doi.org/10.1016/j.apenergy.2016.06.046.

[12] Al-Wakeel A, Wu J, Verga M, Jenkins N. State estimation of low voltage microgrid using an iteratively reweighted least squares method. CIRED Workshop, Rome; 2014. p. 1–5.

[13] Manitsas E, Singh R, Pal B, Strbac G. Distribution system state estimation using an artificial neural network approach for pseudo measurement modeling. IEEE Trans Power Syst 2012;27:1–9.

[14] Singh R, Pal BC, Jabr RA, Vinter RB. Meter placement for distribution system state estimation: an ordinal optimization approach. IEEE Trans Power Syst 2011;26:2328–35. http://dx.doi.org/10.1109/tpwrs.2011.211877.

[15] Singh R, Pal BC, Jabr RA. Choice of estimator for distribution system state estimation. IET Gener Transm Distrib 2009;3:666–78. http://dx.doi.org/10.1049/iet-gtd.2008.048.

[16] Abur A, Exposito AG. Power system state estimation theory and implementation. New York: CRC Press; 2004.

[17] Irving MR, Sterling MJH. Robust state estimation for large scale power systems. In: Proceedings of IEE conference on power system monitoring and control, Durham, UK. p. 261–4.

[18] Clewer BC, Irving MR, Sterling MJH. Robust state estimation in power systems using sparse linear programming. IEE proceedings C – generation, transmission and distribution 1985;132:123–31.

[19] Smith MM, Powell RS, Irving MR, Sterling MJH. Robust algorithm for state estimation in electrical networks. IEE proceedings C – generation, transmission and distribution 1991;138:283–8.

[20] Irving MR, Macqueen CN. Robust algorithm for load estimation in distribution networks. IEE proceedings – generation, transmission and distribution 1998;145:499–504. http://dx.doi.org/10.1049/ip-gtd:19982176.

[21] Coelho VN, Coelho IM, Coelho BN, Reis AJR, Enayatifar R, Souza MJF, et al. A self-adaptive evolutionary fuzzy model for load forecasting problems on smart grid environment. Appl Energy 2016;169:567–84. http://dx.doi.org/10.1016/j.apenergy.2016.02.045.

[22] Chou J-S, Ngo N-T. Time series analytics using sliding window metaheuristic optimization-based machine learning system for identifying building energy consumption patterns. Appl Energy 2016;177:751–70. http://dx.doi.org/10.1016/j.apenergy.2016.05.074.

[23] Jenkins N, Long C, Wu J. An overview of the smart grid in Great Britain. Engineering 2015;1:413–21. http://dx.doi.org/10.15302/J-ENG-2015112.

[24] DCC. Building a smart metering network for Great Britain, ; 2015. p. 1–3 <https://www.smartdcc.co.uk/media/338770/15574_building_a_smart_metering_network_v3.pdf> [accessed February 28, 2016].

[25] Ofcom. Infrastructure report 2014. Ofcom's second full analysis of the UK's communications infrastructure 2014, <http://stakeholders.ofcom.org.uk/binaries/research/infrastructure/2014/infrastructure-14.pdf>.

[26] Tafazolli R. Smart metering system for the UK technologies review, ; 2013. p. 19 <https://m2m.telefonica.com/system/files_force/SM_Report_07-06-20132_9.pdf?download=1> [accessed August 27, 2015].

[27] Commission for Energy Regulation. Electricity smart metering technology trials findings report. Information paper CER11080b' 2011. p. 58 <https://www.ucd.ie/t4cms/ElectricitySmart Metering Technology Trials Findings Report.pdf> [accessed September 1, 2015].

[28] Ingram S, Probert S, Jackson K. The impact of small scale embedded generation on the operating parameters of distribution networks. PB Power, Department of Trade and Industry (DTI); 2003 <http://webarchive.nationalarchives.gov.uk/20100919181607/http://www.ensg.gov.uk/assets/22_01_2004_phase1b_report_v10b_web_site_final.pdf> [accessed March 30, 2016].

[29] de Hoon MJL, Imoto S, Nolan J, Miyano S. Open source clustering software. Bioinformatics 2004;20:1453–4. http://dx.doi.org/10.1093/bioinformatics/bth078.

[30] Irish Social Science Data Archive. Irish smart metering measurements <http://www.ucd.ie/issda/data/commissionforenergyregulationcer/> [accessed March 10, 2016].

[31] Zivot E, Wang J. Modeling financial time series with S-PLUS. Springer-Verlag; 2006. http://dx.doi.org/10.1007/978-0-387-32348-0.

[32] Electricity North West. HV network information <https://www.enwl.co.uk/secure-area/ltds-document-library/hv-network-information> [accessed June 1, 2016].

[33] OfGem. Electricity distribution losses: a consultation document 2003. p. 55 <https://www.ofgem.gov.uk/ofgem-publications/44682/1362-03distlosses.pdf> [accessed June 1, 2016].

[34] Sohn Associates Limited. Electricity distribution system losses non-technical overview; 2009. p. 41 <https://www.ofgem.gov.uk/sites/default/files/docs/2009/05/sohn-overview-of-losses-final-internet-version.pdf> [accessed February 20, 2016].

[35] Lincoln R, PSERC PSERC. Pypower 2015 <https://pypi.python.org/pypi/PYPOWER> [accessed January 22, 2016].

[36] Zimmerman RD, Murillo Sánchez CE, Thomas RJ. MATPOWER: steady-state operations, planning, and analysis tools for power systems research and education. IEEE Trans Power Syst 2011;26:12–9. http://dx.doi.org/10.1109/TPWRS.2010.2051168.

[37] Ghasemi A, Shayeghi H, Moradzadeh M, Nooshyar M. A novel hybrid algorithm for electricity price and load forecasting in smart grids with demand-side management. Appl Energy 2016;177:40–59. http://dx.doi.org/10.1016/j.apenergy.2016.05.083.

[38] Tukey JW. Exploratory data analysis. Addison-Wesley Publishing Company; 1977.