The University of Manchester

Title: **Deliverable 3.7 "Characterisation of LV Networks"**

Synopsis: This document describes the methodology adopted to characterise 141 LV networks (i.e., 232 feeders) based on their network parameters as well as monitoring data. This characterisation process (or taxonomy) is used to identify those LV networks or feeders that are most representative.

Document ID: UoM-ENWL_LVNS_Deliverable3.7v08

Date: 14th July 2014

Prepared For:
Rita Shaw
Future Networks Engineer
Electricity North West Limited, UK

John Simpson
LCN Tier 1 Project Manager
Electricity North West Limited, UK

Dan Randles
Technology Development Manager
Electricity North West Limited, UK

Prepared By:
Valentin Rigoni
The University of Manchester
Sackville Street, Manchester M13 9PL, UK

Revised By:
Dr Luis(Nando) Ochoa
The University of Manchester
Sackville Street, Manchester M13 9PL, UK

Contacts:
Dr Luis(Nando) Ochoa
+44 (0)161 306 4819
luis.ochoa@manchester.ac.uk

Valentin Rigoni
+44 (0)161 306 4794
valentinrigoni@yahoo.com.ar

Executive Summary

This report corresponds to the “Characterisation of LV Networks” part of the Low Carbon Network Fund Tier 1 project “LV Network Solutions” run by Electricity North West Limited (ENWL).

The aim of the LV Network Solutions project is to provide ENWL with greater understanding of the characteristics, behaviour, and future needs of their low voltage networks. This will be based on the analysis of data gathered by appropriate monitoring schemes to be deployed on hundreds of LV feeders and substations, and the assessment of the corresponding computer-based network models in current and future scenarios.

In particular, the procedure to identify statistically representative LV feeders in the North West of England is presented. This work provides a unique set of representative feeders thoroughly validated which can be used as test (representative) cases for analysing the impact of Low Carbon Technologies (LCT) and LV Network solutions.

A set of 383 feeders with network data and the correspondent monitored data was gathered from ENWL. After a filtering process (i.e. noise and outliers) this initial number of feeders was reduced to 232 obtaining the definitive data base for a clustering process. A macro partition of the 232 feeders is presented dividing them in terms of the presence of Distributed Generation (DG). Two groups, a first one of 156 feeders with-out DG penetration and a second one of 76 feeders with DG penetration, were clustered separately and results analysed. A final set of 11 clusters (families) and their representative feeders was obtained.

The opportunities in terms of understanding ENWL’s LV Networks and analysing the impacts of different technologies are really promising. The whole population of LV feeders can be divided in a small set of representative feeders that can relate their characteristics and behaviours to all the feeders belonging to their same family. This reduces considerably the complexity of the assessing the impacts of LCT on all LV networks.

Table of Contents

1 Introduction

As part of the transition towards a low carbon economy, Electricity North West Limited (ENWL), the Distribution Network Operator of the North West of England, is involved in different projects funded by the Low Carbon Network Fund. The University of Manchester is part of the Tier 1 project “LV Network Solutions”.

The objective of this project is to provide ENWL with greater understanding of the characteristics, behaviour, and future needs of their LV networks. This will be based on the analysis of data gathered by appropriate monitoring schemes to be deployed on hundreds of LV feeders and substations, and the assessment of the corresponding computer-based network models in current and future scenarios.

The following report contains a taxonomy approach that characterises a set of 232 real feeders provided by ENWL to The University of Manchester. This taxonomy also considers the busbar monitoring data for the corresponding substations.

The report contains:

- The information and characteristics of a large set of LV feeders supplied by ENWL and the corresponding data treatment.
- The application of clustering techniques aimed at grouping the population of feeders defined by a series of features or attributes according to their similarity.
- A rigorous validation process that determines the best clustering algorithm and the optimal number of clusters.
- A compilation of clusters comprised by the different feeders coherently related; and,
- A set of representative feeders across the North West of England.

This report describes the creation process of a set of representative feeders for the north west of England in base of topological and monitored data by applying clustering techniques. All data has been refined in order to increase the quality of results and each step of the process has been mathematically validated so results could have a consistent backup.

A set of initial 232 LV feeders was partitioned in terms of the presence of DG. Two groups, a first one of 156 feeders with-out DG penetration and a second one of 76 feeders with DG penetration, were clustered separately and results analysed. A final set of 11 clusters (families) and their representative feeders were obtained were only 3 of them resulted from the consideration of PV panels’ penetration. This small but statistically representative set of feeders can be used for further LCT impact assessments or LV network solutions implementation expecting the results to be meaningful for the LV network. The behaviour of certain technologies can be match to specific feeders’ types.

2 Network and Monitoring Data

2.1 Network Data

Similarly to Deliverable 1.2 “Tool for Translating Network Data from ENWL to OpenDSS” the network data originally provided by ENWL as Geographic Information System (GIS) files was processed with the software ArcGIS.

An automatic process was implemented in Python to extract the network information. In particular, the “LLF” attribute (related to the CDCM tariff mapping code, Table 1 was extracted from the MPAN files. This was done to increase the granularity of customer types as opposed to the previously adopted “Profile Class” attribute.

Table 1. ENWL customers’ profiles

CDCM TARIFF MAPPING CODE	DESCRIPTION	PROFILE CLASS
511	DOMESTIC UNRESTRICTED	PC 1
531	DOMESTIC TWO RATE	PC 2
581	DOM OFF PEAK (RELATED MPAN)	PC 2
591	SML NONDOM O/P (REL MPAN)	PC 4
631	SMALL NON DOM UNRESTRICTED	PC 3
661	SMALL NON DOM TWO RATE	PC 4
721	NHH UMS	PC 1 – 8
751	LV MEDIUM NON DOMESTIC	PC 5 – 8
752	LV SUB MEDIUM NON DOMESTIC	PC 5 – 8
753	HV MEDIUM NON DOMESTIC	PC 5 – 8
961	LV GENERATION NHH	PC 8

The automated process was applied to 141 LV networks with underground feeders. The topology and corresponding network data was obtained for 628 feeders. For the clustering process presented in this report, only the following attributes were considered:

- Number of customers;
- Number of customers per ENWL profile class;
- Total main and service cable lengths [m]; and,
- Number, capacity [kW] and type of declared DG units.

From the above 141 LV networks only 127 of them could be modelled in OpenDSS (power flow simulator), i.e., impedance of cables were fully available. Given the importance of impedances to describe network characteristics, only the subset of feeders with OpenDSS models was considered (550 out of 628).

In order to eliminate feeders that might create ‘noise’ for the clustering technique, a subsequent filtering process was applied by eliminating those with less than 6 customers (too small) or a main total cable length shorter than 5 m (might not even be a feeder). 27 feeders were excluded.

The final number of LV feeders suitable for the clustering analysis is 523.

2.2 Monitoring Data

Monitoring data collected at the busbar of the selected feeders include parameters such as:

- RMS voltages per phase;
- Active and reactive power per phase;
- Current magnitudes per phase and neutral;
- Total Harmonic Distortion (THD); and,
- Ambient and transformer temperature.

Based on data availability, the period adopted is winter considering weekdays of January, February and December 2013. January and February 2014 could not be used due to lack of data. This data was only limited to 388 feeders of those selected in the previous section.

In addition, many of the weeks and days within the selected months lacked continuity of data, i.e., blocks of minutes or hours without data. A fifth of the 388 feeders was affected by this issue. To ensure the largest sample possible, it was necessary to solve the above issue. For this, similar days (within the week, month or season) were used to 'merge' the available data and provide full daily continuity.

It is important to highlight that the monitoring data had different sampling rates (e.g., 1 minute, 5 minutes and 10 minutes) for different feeders. Consequently, for consistency, 1 minute and 5 minute data was averaged to produce 10 minute profiles.

From the 523 LV feeders selected in the previous section, 383 were chosen based on monitoring data availability (considering the merging process). These feeders have each 10 minute resolution voltages, active and reactive power, currents, THD and temperatures for a 2013 winter weekday.

3 Feeders validation and data cleansing

For the proper functioning of the clustering algorithm any noise or uncommon cases have to be eliminated [13].

3.1 Feeders Validation

Similarly to Deliverable 3.2 “Production of Validated Networks”, a feeders’ validation was proposed in order to eliminate any sort of incorrect network data. The consumed energy coming from the monitors was compared with an estimation based on each feeder’s customers’ number and type. The energy consumed by the customers was assessed using their corresponding ENWL’s Elexon-based profiles. Generation associated to declared Photo Voltaic (PV) panels was as well estimated in base on corresponding solar irradiance values from the Photovoltaic Geographical Information System data base [17].

Two consumption periods were compared and the maximum error identified: 1) the energy consumed during the 24hs of day, and 2) the consumed energy between 5pm and 8pm.

The elimination process of noise and outliers must be very careful as we can take the risk of eliminating real data that just presents particular characteristics. The definition of boundaries had to be done by taking into account that the error between the monitored data and the ENWL Elexon-based profile may present varied ranges.

In order to define a boundary, a set of 32 substations where network data had been validated by ENWL’s Load Allocation Tool described in Deliverable 3.4 “Review of ENWL’s load allocation tool”. Results in terms of error are presented in Table 2.

Table 2. 32 substations’ error

Substation	Err E(Elexon)
1	17.86%
2	23.18%
3	10.55%
4	11.18%
5	8.23%
6	5.71%
7	10.78%
8	13.64%
9	8.26%
10	7.48%
11	8.75%
12	24.48%
13	30.28%
14	55.25%
15	13.25%
16	10.04%

Substation	Err E(Elexon)
17	70.27%
18	8.83%
19	8.65%
20	6.42%
21	49.01%
22	45.23%
23	39.01%
24	54.93%
25	69.74%
26	62.52%
27	56.83%
28	56.57%
29	43.27%
30	56.94%
31	10.34%
32	67.40%

	Error
max	70.27%
min	6.00%
average	30.15%
median	20.52%
stdev	22.73%

In order to identify if the error from the substations followed a normal distribution probability, the error for the 32 substation was standardized with the z-score formula:

$$z = \frac{x - \mu}{\sigma}$$

Where:

- μ is the mean of the population.
- σ is the standard deviation of the population.

Once standardized, data was tested with “The one-sample Kolmogorov-Smirnov test” proposed in [18]. It is a nonparametric test used to decide if a sample comes from a population with a specific distribution.

The test supposes that the substations’ error follows a normal distribution probability $F_0(x)$ which is, for each value of x the proportion of substations having an error equal or less to x. The cumulative step-function of the 32 substations’ error is expected to be closed to this hypothetical distribution. It is based on drawing the hypothetical Cumulative Distribution Function (CDF) on a graph and two curves at a distance $d_\alpha(N)$, with N equal to the number of samples, above and below the first one. If the cumulative distribution function for the N samples passes outside of the band at any point the test rejects (at the α level of significance) the hypothesis that the true distribution function is $F_0(x)$.

Figure 1 shows the similarity between the empirical CDF of the centred and scaled error vector and the CDF of the standard normal distribution.

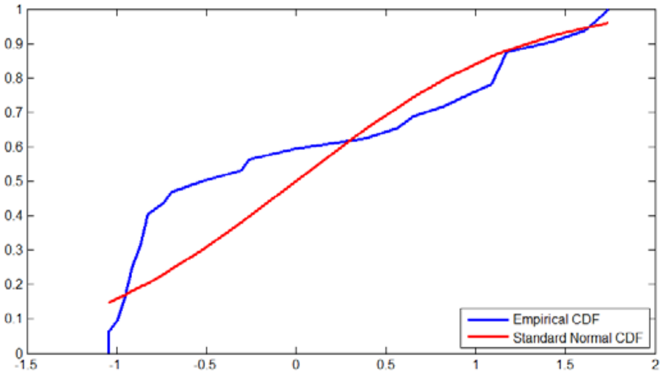


Figure 1. Empirical CDF vs Standard Normal CDF

The test was run with MATLAB with a 5% significance level and resulted positive which means that the initial hypothesis of a normal distribution probability was accepted.

Figure 2 shows the Probability Distribution Function of the 32 substations’ error under the assumption of a normal PDF.

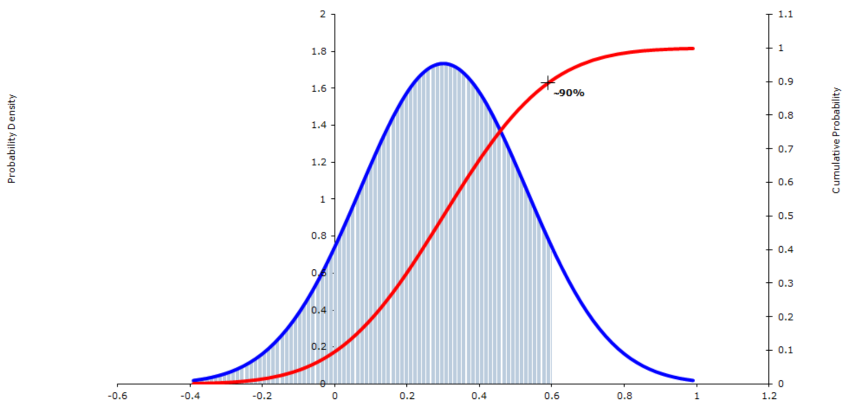


Figure 1. PDF for the 32 substations’ error

According to the PDF, the error’s limit was set to 60%. This value corresponds to ~90% of the CDF. The decision of limiting the curve to 90% was taken considering the exclusion of possible outliers.

Consequently:

- If the error is smaller than 60% then the model is valid

$$E_{3\emptyset\,(all\ day)} \& \, E_{3\emptyset\,(5-8pm)} \leq \; 60\% \; \rightarrow Feeder\ is\ valid$$

From the 383 LV feeders selected in the previous section, 246 were validated.

3.2 Data Cleansing

Due to some issues with different feeder features a data cleansing was applied. The following features were analysed.

- Type of Customer. Feeders presenting non-identifiable customers were eliminated. 8 feeders were eliminated.
- Voltages. Figure 3 shows the box plot of the average voltage level for the 250 feeders. Substations “216102” (3 feeders) and “338843” (1 feeder) were excluded as the busbar voltages were considerably low (205.6 V and 210.7 V).

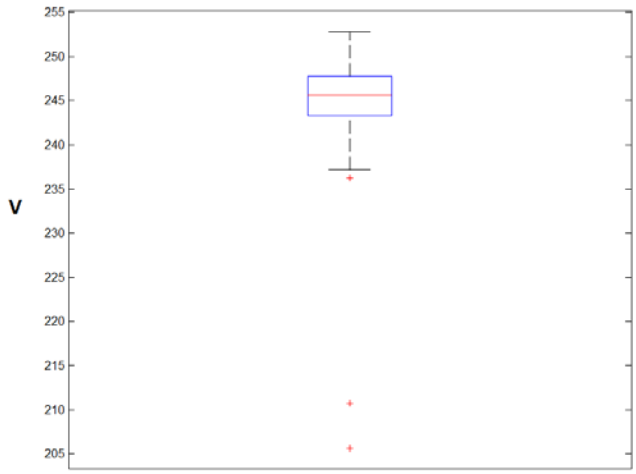


Figure 2. Voltage level Boxplot

- Type of DG. Feeders 445349908 and 445349909 from substation 177963 were removed as they had micro Combined Heat and Power (µCHP) installations.

As a result of the data cleansing, 14 feeders were removed leading to a final set of 232 feeders. These 232 feeders were used for the clustering process.

4 Feeder features

The most important topology features and monitored variables for the 232 feeders are presented in Figure 4. It can clearly be seen that some of these features, such as the power factor, are similar for most feeders. However, significant diversity appears for other features, as it is the case of the neutral current.

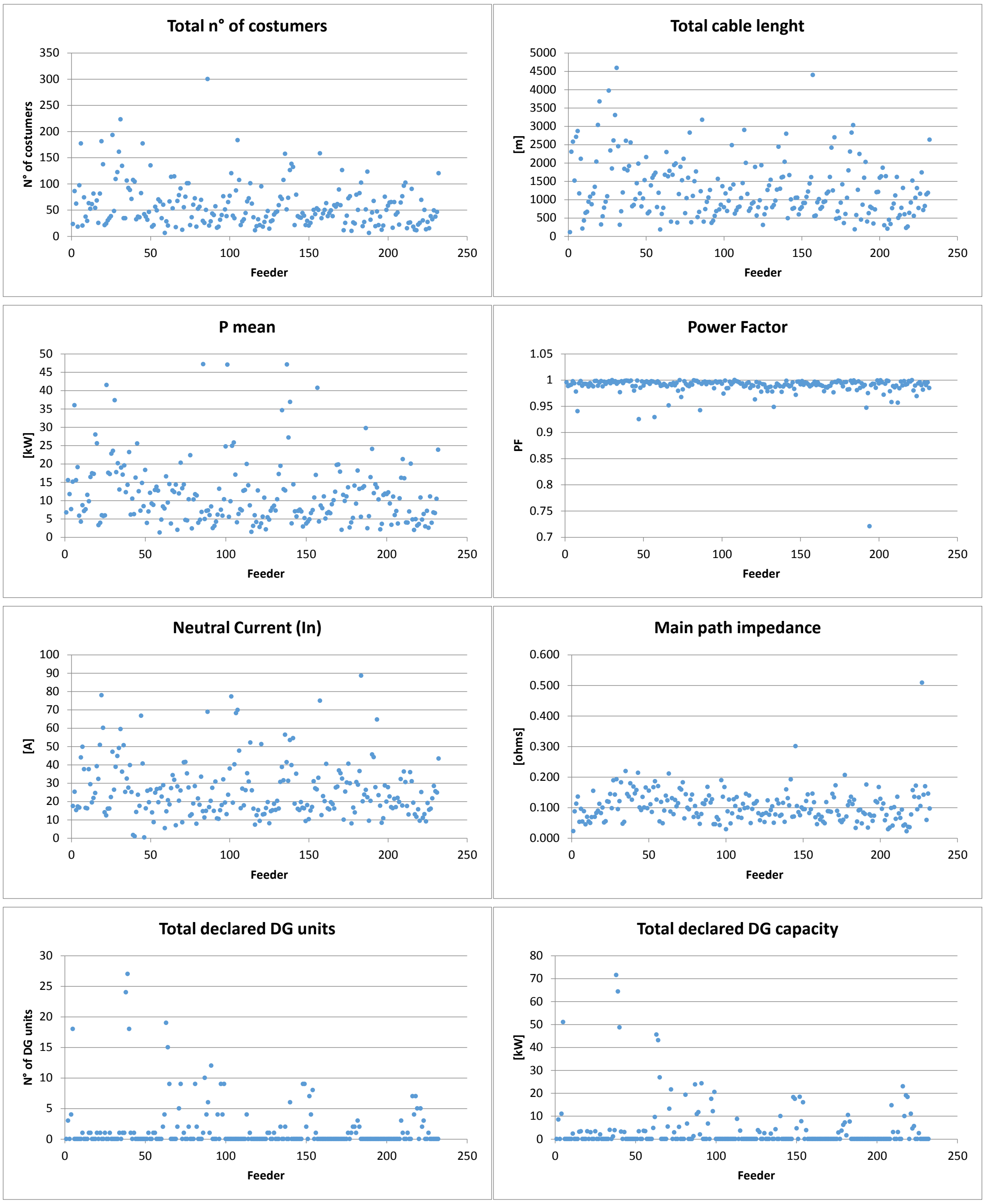


Figure 4. Feeder features

A series of observations can be done:

- **PF**. In terms of planning and operation, power factor is usually considered to be around 0.97. As it can see in Figure 4, in most of cases the real value is over the traditional assumption.
- **In**. The neutral current can reach average values of 90 A during the day.

4.1 Simplification and Selection of Features

Given that a compromise has to be found in order to take into account diversity and the potential identification of representative families, the selection and simplification of the features (or attributes) used for the description of the 232 feeders was made based on the literature review [1]-[3].

The data from the monitors describes the daily tendency of active and reactive power, voltage, current, etc. in all of the feeders considering a 10 minutes resolution throughout a day. This means 144 available values to describe each parameter per phase per feeder. The direct use of this volume of data would not only make the clustering technique slow but also distorts the balance between topological characteristics and monitoring data.

Following the adopted approach in [1], the monitoring related features were processed using mean values and associated standard deviations. Using mean values and the corresponding standard deviations allows reducing three-phase and time-series values into a single representative value.

In addition to the features that provide a general picture of the characteristics of a feeder (e.g., number of customers, customer type, active and reactive power, etc.) parameters such as impedance and neutral current were also considered as important [1], [2]. The selection of the features required a trial and error approach by which the set of features with the best performance was selected. Table 2 shows the final features used for the description of the feeders.

The main path distance is associated to the path connecting the farthest customer to the head of the feeder. The average path distance is the mean value of each customer's path to the head of the feeder. The total path impedance is calculated as the total sum of the impedances of each customer's path to the head of the feeder.

Table 3. Adopted Features

1- Number of domestic unrestricted customers	9- Total path impedance [ohms]
2- Number of domestic TWO RATE customers	10- Neutral current [A]
3- Number of small non domestic OFF PEAK customers	11- Mean 3ϕ daily active power [kW]
4- Number of small non domestic unrestricted and TWO RATE customers	12- Daily mean standard deviation of 3ϕ active power [kW]
5- Number of LV medium non domestic customers	13- Daily mean standard deviation of 1ϕ active power [kW]
6- Total conductor length [m]	14- Mean 3ϕ daily reactive power [kvar]
7- Main path distance [m]	15- Power Factor (PF)
8- Average path impedance [ohms]	

5 Clustering Techniques

The representative feeders can be obtained by applying what is known as clustering techniques or grouping algorithms. This methodology consists in grouping similar samples of data in function of the features used for their characterization. The result is a set of K clusters (groups) each one defined by a centroid (i.e., representative feeder) [4] [5].

The high degree of complexity and the large volume of information require the clustering process to proceed with different levels of detail. Given that distributed generation (DG) was in some feeders significant and hence producing noticeable impacts on the monitored features, it was decided to divide the feeders into two macro-categories: with and without DG penetration. Without this separation, any clustering technique would result in groups of feeders without a robust correlation within them.

Mathematically, each macro-category has M patterns (feeders) forming the set $\mathbf{L} = \{\mathbf{l}^{(m)}, m = 1, \ldots, M\}$. Each pattern is characterized by H features forming the set $\mathbf{l} = \{l_h, h = 1, \ldots, H\}$.

In this report, a set of different clustering algorithms were considered and compared. Indifferently of the applied algorithm, the individuals (from the total population) must be grouped according to the features selected to characterise them. The obtained set of centroids can be represented as $\mathbf{R} = \{\mathbf{r}^{(k)}, k = 1, \ldots, K\}$.

The clustering process can be summarised in a series of steps as follows:

1- Gather the network and monitored data for the M feeders (patterns).

2- Data cleansing.

3- A set of H representative features is selected.

4- The data is standardised and processed in order to form an M x H matrix which is the main input for the clustering algorithm.

5- The clustering algorithms are run and the centroids created.

6- The solutions are validated and the most suitable algorithm is selected.

7- The optimal number of clusters K is defined.

8- The results are analysed and the representative feeders described.

were steps 3 to 4 are repeated for each macro-category.

5.1 Clustering Algorithms

Data clustering algorithms can be generally separated in partition and hierarchical algorithms [6]. In the case of partition algorithms, clusters are determined all at once by allocating the K centroids and associating the rest of the elements to them. They have the peculiarity that one of the input parameters is the number of clusters in which data wants to be partitioned. Hierarchical clustering algorithms can be agglomerative or divisive. In the first case, each one of the elements is initially an isolated cluster (composed by only one element). The algorithm stars joining these isolated clusters in function of their similarity until a unique and final one is obtained. The divisive mode works in the opposite way as it starts with only one cluster composed by the total population and then divides it progressively.

This report presents two different partition clustering algorithms that are: improved k-means++ and k-medoids++. One hierarchical algorithm is also presented.

5.1.1 Improved k-means++

K-means groups the set of M patterns into the desire number of clusters K by minimizing the total Sum of Square Distance (SSE) between the data and the associated clusters, as show in the equation below.

$$SSE = \sum_{k=1}^{K} \sum_{\mathbf{l} \in \mathbf{L}^{(k)}} d(\mathbf{r}^{(k)}, \mathbf{l})$$

Where $d(\mathbf{r}^{(k)}, \mathbf{l})$ is the Euclidean distance between the element $\mathbf{l}$ and centroid $\mathbf{r}^{(k)}$. It means that the algorithm tends to obtain the most compact structure possible to minimise the distance between each element and the centroid of its cluster. The algorithm starts by initializing (locating) the K set of centroids to then associate the closer elements so as to minimise the SSE. Once all the elements have been assigned, the algorithm recalculates the new centroids as the average of the elements belonging to each cluster. The whole procedure is repeated until there is no further variation. As the centroids are H-dimensional points calculated as an average they are not necessarily an element of the population. The representative feeder is defined as the closest element to the centroid.

The k-means clustering algorithm was improved by utilizing and “smart” initialisation process to avoid obtaining low quality structures. The new k-means++ algorithm starts the process by randomly allocating only the first centroid and associating a probability to the location of the second one. The farthest elements to the first cluster have a higher probability to become the next centroid [7]. The application of this modification showed a noticeable improvement of the results.

A further modification of the k-means algorithm was applied. The k-means++ algorithm was run several times as in each initialisation the new location of the starting centroids could lead to different solutions. In each one of these internal runs the objective function of this process was to originally minimize the SSE. The modification consisted in using the Global Silhouette Coefficient (GS) (see section 5.2), calculated considering the isolated patterns, to define the optimal solution after each iteration [8], [9]. This modification tended to isolate uncommon feeders thus leading to better results.

5.1.2 K-medoids++

The k-medoids algorithm is strongly related to k-means but with the particular difference that the initial centroids are elements of the population. It was modified as well by utilizing the smart initialisation process from k-means++.

5.1.3 Hierarchical clustering

The agglomerative hierarchical algorithm was run using the Ward’s Variance Method (WVM) in which the distance between a pair of clusters is equal to how much will the SSE increase by merging them [9]. Different methods of the hierarchical algorithm were applied but always leading to less compact and spherical clusters than WVM.

The basic algorithm is very simple. The steps are listed below.

1- Each element is its own cluster.

2- As long as there are more than one cluster

   a. Find the closest pair of clusters.

   b. Fuse them together.

3- Return as an output the dendrogram that contains the information of all the process.

5.2 Normalisation of data

Each one of the feeders was characterised by an H-dimensional vector. In all of the algorithms presented, these vectors are grouped together in function of their similarity based on the Euclidean distance between them. The Euclidean distance is sensible to differences in the scale of magnitude of the features used to characterise the feeders. In the context of this project, the nature of features was very diverse given that network and monitoring data were combined. A normalisation (also called standardisation) process was needed in order to make these features vary in comparable ranges [11], [12].

The max-min standardisation formula consisting on the use of minimum and maximal values was used for normalising each one of the selected features.

$$l_h^{(\mathrm{m})} = \frac{l_h^{(\mathrm{m})^*} - Min\{l_h\}}{Max\{l_h\} - Min\{l_h\}}$$

Where Min $\{l_h\}$ and Max $\{l_h\}$ correspond to the minimum and maximum values of the feature *h* and $l_h^{(\mathrm{m})^*}$ is the non-normalised value of the feature *h* for the element *m*. If there are no negative values, the variation range of each one of the features is going to be limited between 0 and 1. Given that this normalisation process uses minimum and maximum values, it is sensitive to the presence of noise and uncommon cases.

A different normalisation process known as z-score proposed in [3] and [11] was also considered. It normalises data by using mean values and standard deviations. Both values depend on the whole data and are also sensitive to the presence of noise.

After an analysis of the minimum and maximum values for each feature considered in this report, it was found that noise or outliers were not present. The best performance was found to be adopting the max-min standardisation process.

6 Clustering validation

The validation process of the clustering structures allows us to determine two important things: 1) the most appropriate clustering algorithm to be used, and 2) the optimal number of clusters to partition the population of M feeder.

A set of distances are defined [14]:

a) *Feeder-to-feeder* distance: the distance between two members $\mathbf{l}^{(i)}$, $\mathbf{l}^{(j)}$ of a group where each one of the members corresponds to a h-dimensional vector representing a feeder.

$$d(\mathbf{l}^{(i)},\mathbf{l}^{(j)}) = \sqrt{\frac{1}{H}\sum_{h=1}^{H}(l_h^{(i)} - l_h^{(j)})^2}$$

b) *Feeder-to-cluster* distance: the distance between a representative feeder (centroid) $\mathbf{r}^{(k)}$ and the subset $\mathbf{L}^{(k)}$, calculated as the geometric mean of the distances between the centroid and each one of the feeders belonging to the group $\mathbf{L}^{(k)}$.

$$d(\mathbf{r}^{(k)},\mathbf{L}^{(k)}) = \sqrt{\frac{1}{n^{(k)}}\sum_{m=1}^{n^{(k)}} d^2(\mathbf{r}^{(k)},\mathbf{l}^{(m)})^2}$$

c) *Intra-class* distance: calculated by using the *Feeder-to-cluster* distance for all the members $\mathbf{n}^{(k)}$ of a cluster or group $\mathbf{L}^{(k)}$.

$$\hat{d}(\mathbf{L}^{(k)}) = \sqrt{\frac{1}{2n^{(k)}}\sum_{v=1}^{n^{(k)}} d^2(\mathbf{l}^{(v)},\mathbf{L}^{(k)})^2}$$

6.1 Clustering Validity indicators

None of the clustering algorithms presented above (or found in the literature) provide the optimal number of clusters. This has to be determined by assessing the quality of the clusters resulting from different values of K. To perform this cluster assessment in a comprehensive way, four indices are adopted. In addition, these indices will also be used to assess the relative performance of the three clustering algorithms. Their formulation has been uniformed in such a way that higher values imply better results. The values for each indicator will depend on the algorithm and the number K of clusters.

Two of the adopted indices were presented in [15]. They are based on the concept of the Euclidean distance, structure compactness, distance among different clusters, etc.

1. Variance Ratio Criterion

$$VRC(\mathbf{Y},K,\mu) = M\left(1+\frac{W}{K-1}\right)^{-1}\left(1-\frac{W}{M-K}\right) \qquad W = \sum_{k=1}^{K}(\mathrm{n}^{(k)}-1)\left(1-\frac{\mathrm{n}^{(k)}}{M}\frac{\hat{d}^2(\mathbf{L}^{(k)})}{\hat{d}^2(\mathbf{L})}\right)$$

2. Similarity Matrix Indicator

$$SMI(\mathbf{Y},K,\mu) = \left\{\max_{i>j}\left[\left(1-\frac{1}{\ln[d(\mathbf{r}^{(i)},\mathbf{r}^{(j)})]}\right)^{-1}\right]\right\}^{-1}$$

Two further indices were considered from [8] and [9]. They are based on the silhouette width index which represents how strongly related is a feeder to the cluster it has been associated to.

3. Silhouette coefficient

The silhouette global coefficient (GS) gives us an idea of how strongly related are the elements (feeders) to the group they form part of. Each one of the elements is represented by its silhouette width index. The elements forming part of the same group are joined together and can be represented in a figure similar to a silhouette. The width of each independent silhouette is related to the number of elements forming part of the same group. The corresponding height (equal to the silhouette width index) represents how strongly related is each one of the feeders to the group it belongs to [8], [9].

$$s_i = \frac{b_i - a_i}{\max\{b_i - a_i\}} \quad \text{, silhouette width index for the i} - \text{object , with } -1 \leq s_i \leq 1$$

$$S_j = \frac{1}{r_j}\sum_{i=1}^{r_j} s_i \quad \text{, local coefficient}$$

$$GS = \frac{1}{N_c}\sum_{j=1}^{N_c} S_j \quad \text{, local coefficient}$$

Where r_j is the number of feeders by cluster; a_i is the mean distance between the i-object and other objects of the same group j; and, b_i is the minimum mean distance between the i-object and the objects of the closest group to j.

4. Average Silhouette Coefficient (AvgSC)

The mean value of the silhouette (AvgSC) width index for the whole population was also used as a validity index. It is calculated as:

$$AvgSC = \frac{1}{M}\sum_{i=1}^{M} s_i$$

7 Clustering Results

The following section presents the clustering results in terms of the representative feeders for each of the macro-categories, i.e., with and without PV panels.

7.1 Feeders without declared DG units

A set of 156 feeders without declared PV panels characterised by the features described in section 4.1 (Table 3) was separately clustered.

7.1.1 Algorithm selection and determination of K

The macro-category without PV panels was clustered applying the 3 proposed clustering algorithms and results were compared using the 4 indices. All indices were calculated without considering isolated clusters (only composed by one feeder) as they tend to falsely increase their values making any comparison unrealistic. Results, calculated varying the number *K* of clusters, are presented in Figure 6 to 10. The Total Sum of Square Errors (SSE) is as well presented for each case.

Figure 6 shows the total Sum of Square Errors (SSE) for the 3 proposed algorithms. It can be seen that Hierarchical algorithm presents the lower values. This was expected as Ward’s method merges elements by incrementing the SSE’s value as less as possible. The value of the SSE gives an idea of the compactness of clusters but it does not give a notion of between clusters relationship. The figure shows that there is not significant variation of total intra-cluster distances after the number *K* of clusters exceeds ~10. This means that it is most possible that the optimal number of clusters is around this value.

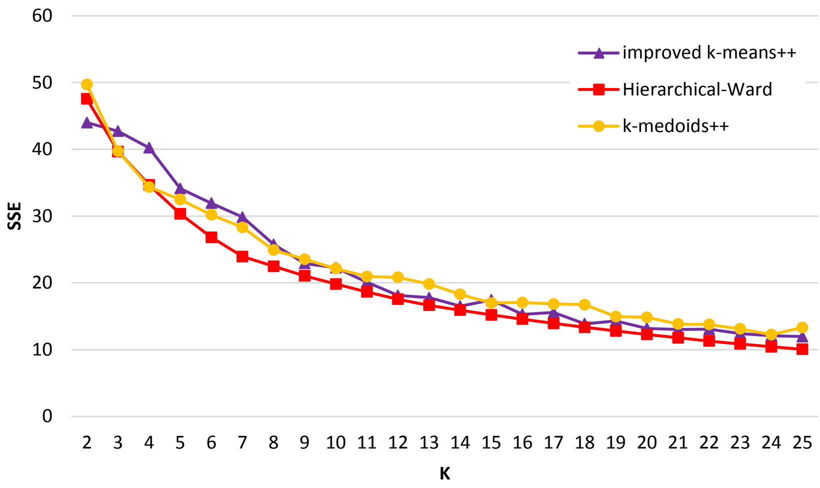


Figure 6. SSE for K=2 to 25

The VRC index is not directly suitable to determine the optimal number of clusters as the values obtained for different K cannot be compared. Therefore, the formulation presented in [19] was considered:

$$wVRC_k = (VRC_k - VRC_{k-1}) - (VRC_{k+1} - VRC_k)$$

The set of 4 indices are presented from Figure 7 to 10. They have all been normalised with respect to the results obtained with the k-means++ algorithm so they could be compared (with the exemption of the *w*VRC index). All indicators were calculated without considering isolated clusters as they tend to falsely increase their values making any comparison unrealistic.

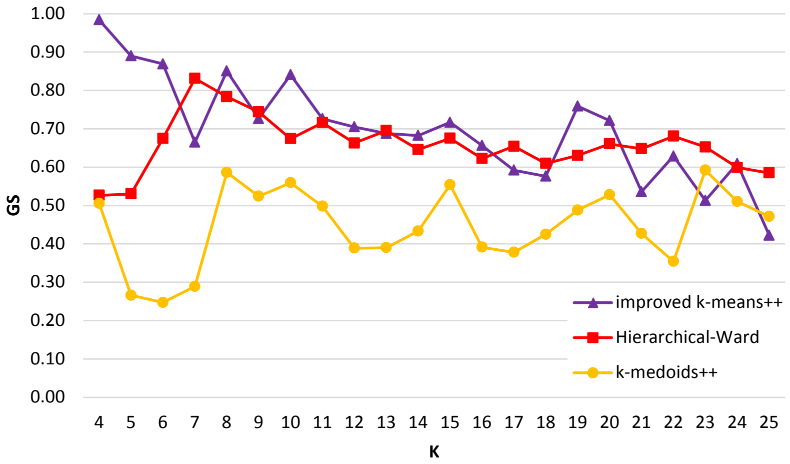


Figure 7. GS for K=4 to 25

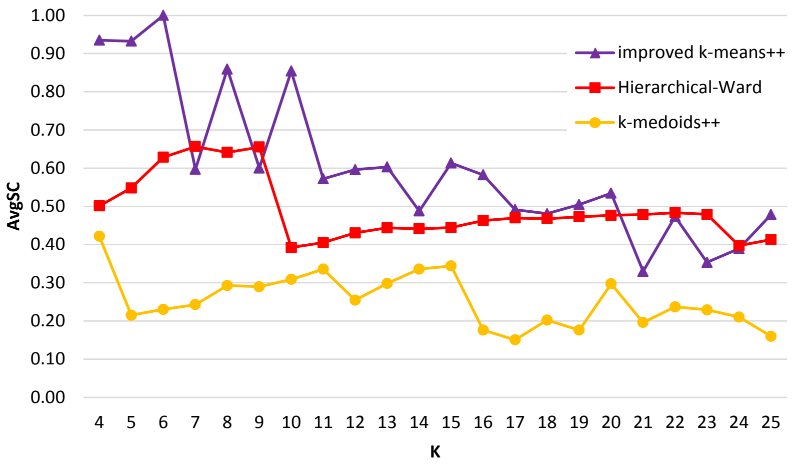


Figure 8. AvgSC for K=4 to 25

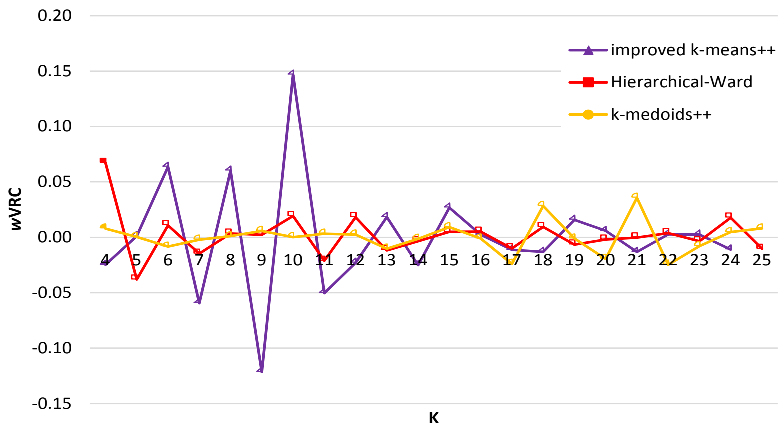


Figure 9. VRC for K=4 to 25

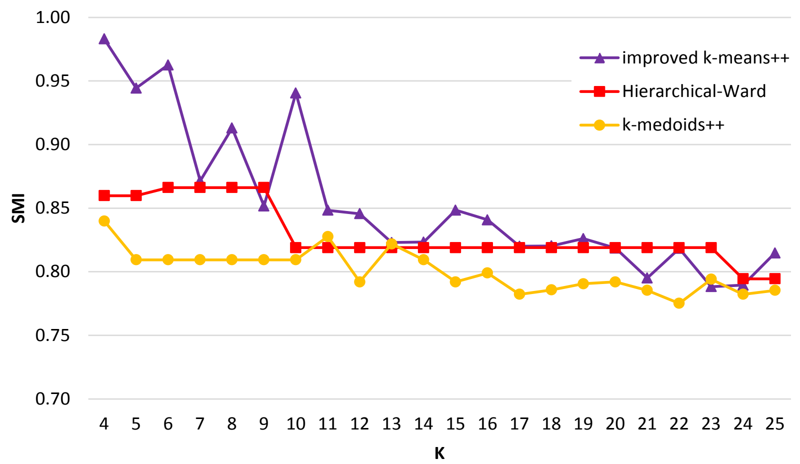


Figure 10. SMI for K=4 to 25

The selection of the optimal number of clusters is not always straight forward. The optimal K needs to come from a compromise between a high performance of the different indices and an adequate number of clusters. For instance, even if most indices present high values for low values of K there would not be any practical interest in obtaining a very low number of clusters as the level of detail for each one would be limited. Therefore, the evolution of the indices has to be studied in order to obtain a reasonable number of clusters able to properly characterize the data set under study.

For each clustering algorithm, the best solution was identified. This solution results from a compromise between a logical number of clusters and high values for the considered indices. Table 4 shows the optimal K number of clusters for each algorithm and the corresponding indices.

Table 4. Cluster assessment (without PV panels)

Algorithm	Improved k-means++	Hierarchical	k-medoids++
*w*VRC	0.15	0.02	0.01
SMI	1.58	1.46	1.36
GS	0.37	0.33	0.26
AvgSC	0.55	0.42	0.19
Optimal *K*	10	9	8

According to the analysis of results, *K*=10 for the improved k-means++ algorithm corresponded to the optimal solution, i.e., best values in most cases and coherent number of clusters. It has to be noticed that 10 is not necessary the final number of representative feeders as some of the groups may consist of isolated clusters composed of only one feeder (e.g., excluded for having uncommon characteristics).

In order to statistically compare the performance of the improved k-means++ with the hierarchical algorithm, the results from 100 executions of the improved k-means++ (1000 initializations of the centroids) for K=10 are presented in Figure 11 as boxplots. To facilitate the comparison, values have been normalized with respect to the results from the hierarchical algorithm in Table III as in the above formula.

$$index_i^{(e)} = \frac{index_i^{(e)^*}}{index_{i_{hierarchical}}}$$

where $index_i^{(e)^*}$ is the non-normalized value of index *i* for the *e*th execution and $index_{i_{hierarchical}}$ the value obtained with hierarchical clustering for index *i*.

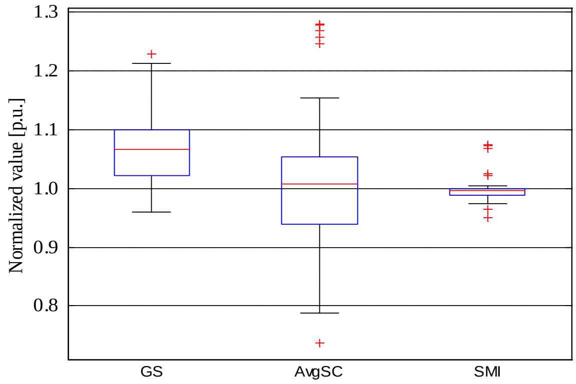


Figure 11. Improved k-means++ vs. hierarchical clustering for K=10

The boxplots reveal median values for the GS, AvgSC and SMI indices (medium horizontal bar), the inter-quartile range (rectangle, 50% of a normal distribution probability), the typical extremes (slim black lines, 99.3%) and outlier values (small crosses). It is clear that the improved k-means++ can statistically lead to better results than hierarchical clustering. For instance, in the case of the GS index, 91% of the cumulative curve was above 1 per unit.

Most of indices values depend on the dataset characteristics so cannot be compared with previous works. However, GS and AvgSC should be independent of the type of data as they result from a standardized index. Taking this into account, the AvgSC showed considerable better performance than results from [2] (about 60% higher).

7.1.2 Set of representative feeders for K=10

The silhouette plot in Figure 12 represents the final set of clusters. Only 8 of the 10 obtained clusters were considered. The 2 remaining ones (corresponding to isolated clusters, i.e., only one feeder) were considered as particular cases could not be representative.

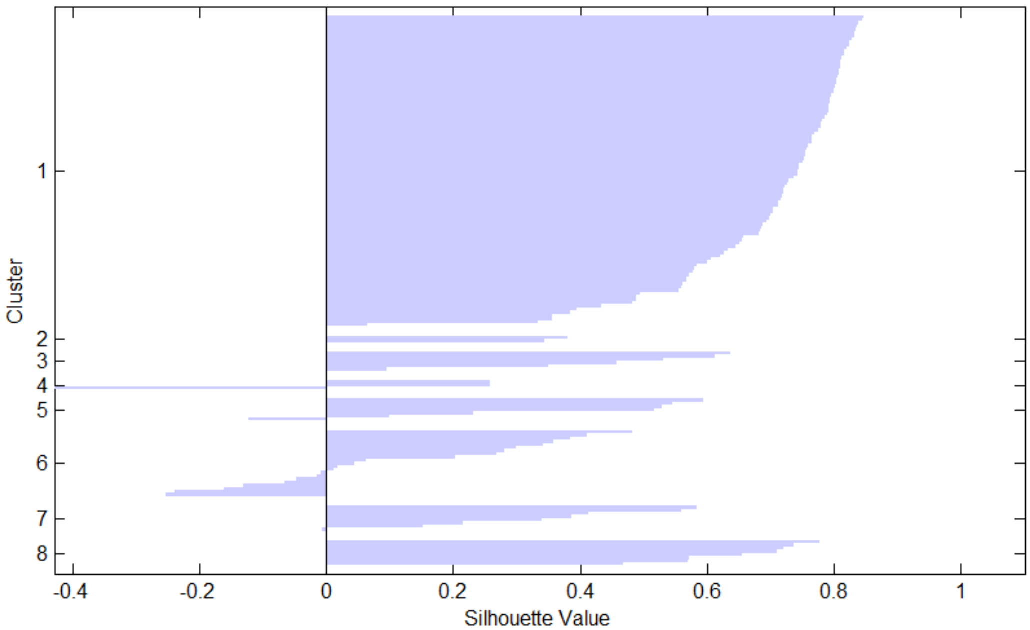


Figure 12– Silhouette plot for the set of 11 final clusters without DG

The final step is to check whether the feeders associated to each of the 8 clusters present similar characteristics. In addition, the representative feeders have to be clearly distinguishable between them. This was done manually by comparing the relative similarity of the different characteristics per feeder. It was found that the 8 clusters can indeed be considered as a statistical representation of the analysed population.

7.2 Feeders with declared DG units

The presence of DG in each feeder can affect some of the attributes previously considered for the characterization. For instance, it is expected that a progressive reduction on the active power will appear as we increase the local generation. The possibilities of understanding the composition in terms of DG in the UK or any country LV network not only increase the understanding in terms of the actual penetration state, it can provide useful information related to the present impacts that those units are having on network metrics.

A set of 76 feeders with declared PV panels was considered. A set of 4 new features representing the presence of DG was introduced (15-17) giving the final set of features presented in Table 5. Because of this, only some of the attributes from section 4.1 (Table 3) were considered in order reduce the complexity of the problem.

Table 5. Considered features

1- Total number of customers	10- Neutral current [A]
2- Number of domestic TWO RATE customers	11- Mean 3ϕ daily active power [kW]
3- Number of small non domestic OFF PEAK customers	12- Mean 3ϕ daily reactive power [kvar]
4- Number of small non domestic unrestricted and TWO RATE customers	13- Power Factor (PF)
5- Number of LV medium non domestic customers	14- Total PV declared capacity [kW]
6- Total conductor length [m]	15- Number of declared PV panels
7- Main path distance [m]	16- Penetration level (N° PV panels/N° customers)
8- Average path impedance [ohms]	17- Total PV declared capacity [kW]/ N° of PV panels
9- Total path impedance [ohms]	

7.2.1 Algorithm selection and determination of K

Similarly to the methodology applied in section 7.1, data was clustered applying the 3 proposed clustering algorithms and results were compared using the 4 validation indices. Results, calculated varying the number *K* of clusters, are presented in Figure 13 to 16. All indices were again normalized respect to the Improved k-means++ values and calculated without considering isolated clusters (with the exemption of the *w*VRC index).

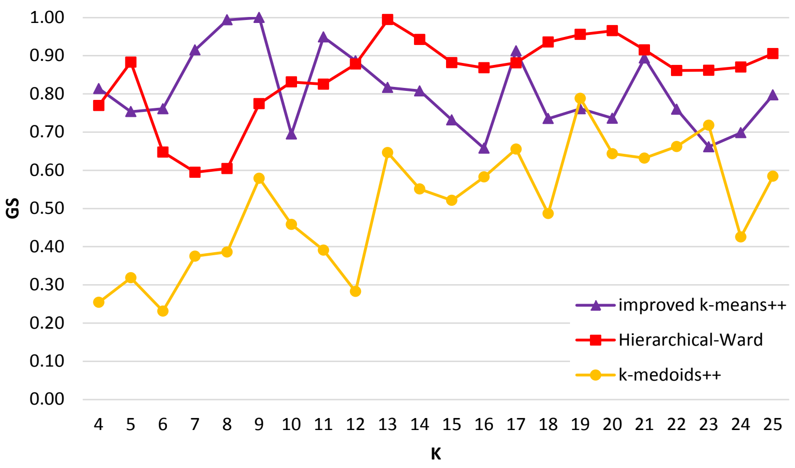


Figure 13. GS for K=4 to 25

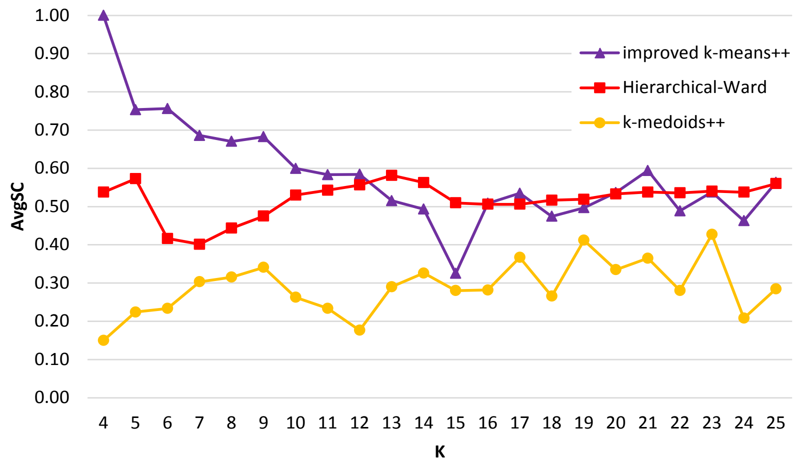


Figure 14. AvgSC for K=4 to 25

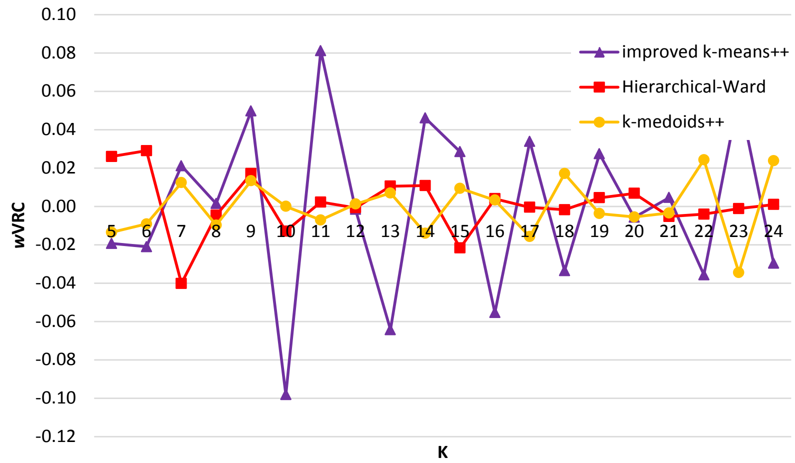


Figure 15. *w*VRC for K=4 to 24

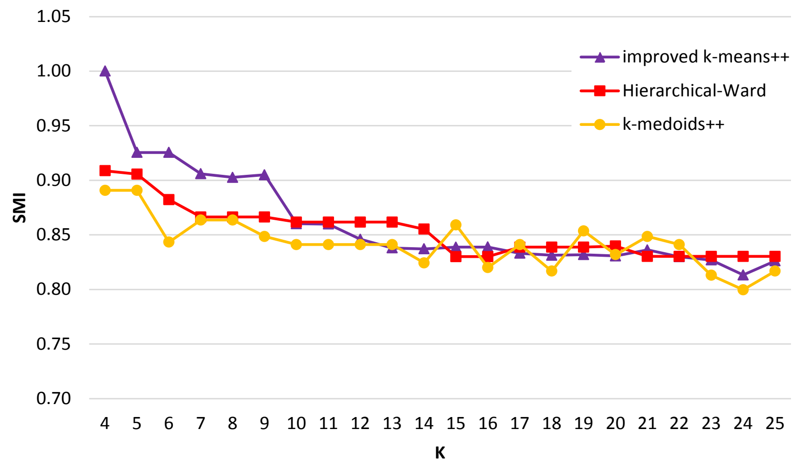


Figure 16. SMI for K=4 to 25

For each clustering algorithm, the best solution was identified. This solution results from a compromise between a logical number of clusters and high values for the considered indices. Table 6 shows the optimal *K* number of clusters for each algorithm and the corresponding indices.

Table 6. Cluster assessment (with PV panels)

Algorithm	Improved k-means++	Hierarchical	k-medoids++
*w*VRC	0.05	0.01	0.01
SMI	1.54	1.45	1.43
GS	0.50	0.44	0.26
AvgSC	0.42	0.38	0.22
Optimal *K*	9	13	9

According to the analysis of results, *K*=9 for the improved k-means++ algorithm corresponded to the optimal solution, i.e., best values in most cases and coherent number of clusters. It has to be noticed that 9 is not necessary the final number of representative feeders as some of the groups may consist of isolated clusters composed of only one feeder (e.g., excluded for having uncommon characteristics).

7.2.2 Set of representative feeders with declared DG

The optimal K number obtained was equal to 9. The k-means++ was found again to be the most suitable algorithm. This, however, resulted in 3 isolated clusters. The silhouette plot of the final 6 clusters considered is presented in Figure 17.

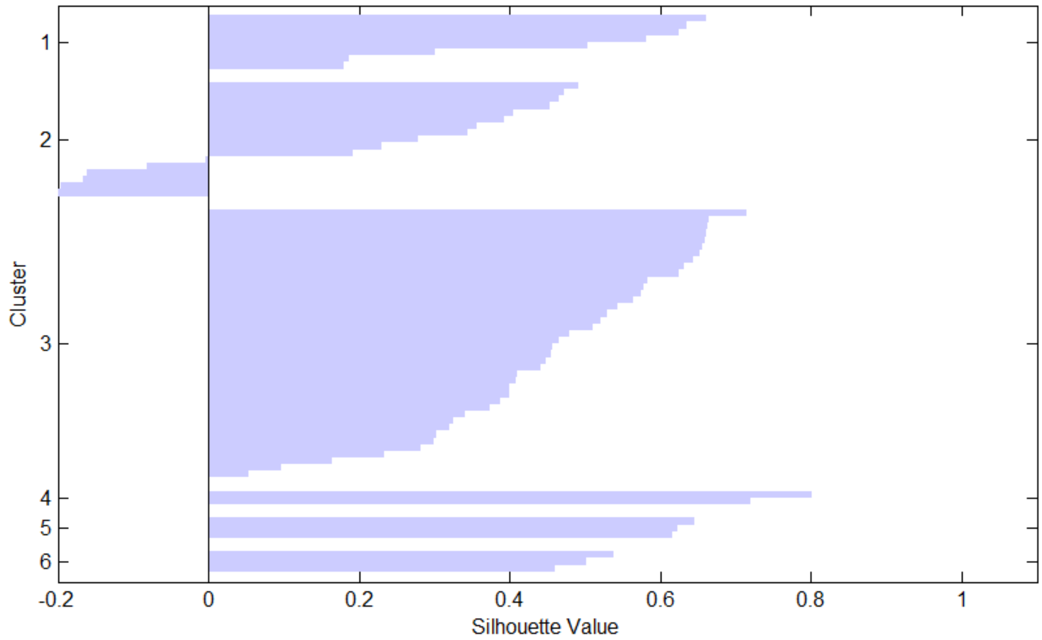


Figure 17. Silhouette plot for K=6

After a thorough manual check of the 6 obtained clusters, only centroids 1, 4 and 6 where considered as representative feeders with PV panels. Clusters 2, 3 and 5, presented small penetrations of PV panels (smaller than 5% of the customers) and hence required further analysis. In fact, they were presenting similar characteristics to the ones of clusters 6, 1 and 3 respectively.

By using the SMI index, modified to be calculated only between a pair of clusters, a symmetric "similarity matrix" was obtained were the value of SMI* for the ij element gives an idea of between clusters proximity. This matrix is presented in Figure 18.

CLUSTER	1	2	3	4	5	6	7	8
1	0.00	2.22	1.64	2.09	1.63	1.61	1.93	1.53
2	2.22	0.00	2.01	2.24	2.15	1.88	1.78	2.19
3	1.64	2.01	0.00	1.81	1.67	1.59	1.83	1.73
4	2.09	2.24	1.81	0.00	2.12	1.98	1.95	2.05
5	1.63	2.15	1.67	2.12	0.00	1.69	2.04	1.74
6	1.61	1.88	1.59	1.98	1.69	0.00	1.63	1.70
7	1.93	1.78	1.83	1.95	2.04	1.63	0.00	1.95
8	1.53	2.19	1.73	2.05	1.74	1.70	1.95	0.00
2*	1.61	2.00	1.64	1.92	1.80	**1.50**	1.60	1.69
3*	**1.37**	2.16	1.61	2.06	1.66	1.54	1.87	1.57
5*	1.59	2.07	**1.53**	1.79	1.77	1.61	1.81	1.72

Figure 18. SMI matrix

The index depends on feeder-to-feeder distance between centroids. It means that higher values are related to pairs of clusters with closed centroids.

By analysing the matrix, it was found that clusters 2*, 3* and 5* (the "*" correspond to the partition with DG penetration) could be actually included in clusters 6, 1 and 3 respectively. In all of three cases, apart from a non-significant PV panel penetration, there was no other metric that could encourage the creation a unique cluster (e.g. reverse power flow, PF, etc.).

Clusters 1*, 4* and 6* were renamed to 9, 10 and 11 respectively, in order to follow the numeration in section 7.1.2.

7.3 Final set of representative feeders

After the comprehensive review and the validation stage a final set of 11 representative feeders was obtained. The set of representative feeders is summarised in table 7 showing some of the main properties qualitatively compared. Figure 19 shows their distribution in the LV network.

Table 7 – Final set of representative feeders

k	Total cable length	N° of customers	Type of customers	Power consumption	Observations
1	Small	Low	Domestic (mainly domestic unrestricted)	Low	N/A
2	Small-medium	Medium-high	Domestic (presence of some low consumption non-domestic)	Hihgest	Highly density area - High neutral current
3	Low	Low	Domestic (presence of some low consumption non-domestic and LV medium non domestic customers)	Medium	High neutral current
4	Medium	Medium	Non-domestic and domestic (considerable presence of LV medium non-domestic customers)	Medium-high	N/A
5	Low	Low	Domestic and non-domestic (61% small non-domestic customers)	Medium	High neutral current
6	Medium	Medium	Domestic (mainly domestic unrestricted)	Medium	N/A
7	High	High	Domestic (mainly domestic unrestricted)	High	Low neutral current
8	Low	Low	Domestic (high presence of domestic two rate customers)	Low	Main cable path represents 50% of the total cable length
9	Small	Low	Domestic (mainly domestic unrestricted)	Lowest	High PV panels penetration level (~40%)
10	Medium	Medium	Domestic (presence of LV medium non domestic customers)	Low	Medium PV panels penetration level (~30%)
11	Large	Medium-high	Domestic (mainly domestic unrestricted)	High-medium	Low PV panels penetration level (~20%)

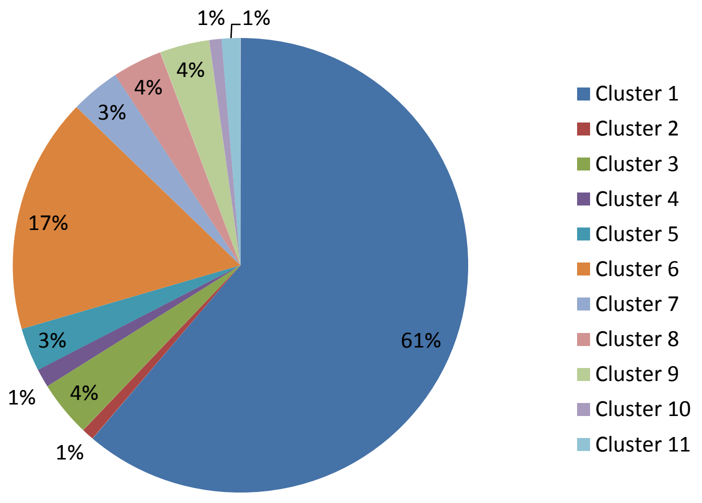


Figure 19. Distribution of representative feeders

The final set of representative feeders with detailed information per feeder is presented in the Appendix.

8 Conclusions and Future Work

This report described a methodology to obtain a set of representative feeders from a sample of monitored LV networks in the North West of England by applying clustering techniques.

The following are the key aspects of the report:

- The proposed methodology based the initial considerations such as selection of features, validity indicators and potential clustering algorithms on previous works.
- The proposed methodology improved most of previous works by 1) ensuring a higher quality of the clusters (using more validity indicators), and 2) ensuring a significant diversity of feeder characteristics. Furthermore, here a clear justification of the most adequate clustering technique is provided.
- A set of initial 232 LV feeders was divided into those with PV panels (76) and those without (156). A final set of 11 clusters (groups) and their representative feeders were obtained. Only 3 of them represented those with PV panels.
- The proposed methodology is generic and can be easily applied to a much larger set of LV networks within ENWL or to other regions or countries (if similar levels of data are available).
- An improved k-means++ algorithm is proposed (better performance have been proved).

The following are key findings from this work:

- Clusters 1, 6 and 7 correspond to pure domestic feeders of different lengths. These representative feeders correspond to ~80% of the whole population under analysis.
- Cluster 1 corresponds to more than half of the population and consists of domestic customers (mainly “domestic unrestricted”) with a small total cable length (qualitatively speaking).
- The mean silhouette coefficient, one of the parameter used for determining the optimal number of clusters, was found to be higher (i.e., better) than the one from the Australian report (0.55 against 0.23).
- The customer classification (i.e., CDMC tariff code) proved to be a fundamental parameter as it is the strongest link between network and monitored data. A simpler customer classification (e.g., residential, commercial and industrial) could have led to lower quality of the clusters and hide the interaction between LCT and different profile classes (traditional loads).
- In terms of planning and operation, power factor is usually considered by DNO or Transmission System Operators (TSO) to be around 0.97. In most of cases the real value is over the traditional assumption.
- The neutral current (In) was found to reach average values of 90A during the day. It was found that in domestic clusters the n° of customers reduces the diversified In. At the same time, clusters with high presence of non-domestic customers presented higher In level.
- The presence of DG clearly reduces active and reactive power consumption and increment the PF level.

9 REFERENCES

[1] Schneider, KP; Chen, Y; Chassin, DP; Pratt, RG; Engel, DW; Thompson, SE; “Modern Grid Initiative - Distribution Taxonomy Final Report”; Pacific Northwest National Laboratory; November 2008.
[2] Berry, A.M; Moore, T.; Ward, K.K.; Lindsay, S.A; Proctor, K; “National Feeder Taxonomy – Describing a Representative Feeder Set for Australian Electricity Distribution Networks”; CSIRO; June 2013.
[3] Yingliang, L ; Wolfs, P.J.; “Taxonomic description for western Australian distribution medium-voltage and low-voltage feeders”; Generation, Transmission & Distribution, IET; vol. 8; Issue 1; pp. 104 – 113; January 2014.
[4] Pascual, D.; Pla, D.; Sanchez, S.; “Algoritmos de Agrupamiento”; Métodos informàticos avanzados; pp. 163-175.
[5] A. Mèffe, C. Oliveira, “Classification techniques applied to Electrical Energy Distribution Systems”, Electricity Distribution, CIRED, 18th International Conference and Exhibition on, pages 1-5, June 2005.
[6] Periklis, A.; “Data Clustering Techniques”; Qualifying Oral Examination Paper; University of Toronto - Department of Computer Science; March 2002.
[7] Arthur, D.; Vassilvitskii, S.; “k-means++: The Advantages of Careful Seeding”; Technical Report 2006-13; Stanford InfoLab; 2006.
[8] Rousseeuw, P.J.; “Silhouettes: a graphical aid to the interpretation and validation of cluster analysis”; Journal of Computational and Applied Mathematics; vol. 20; pp. 53–65; November 1987.
[9] Scarlatache, F.; Grigoraş, G.; Chicco, G.; Cârţină, G.; “Using k-Means Clustering Method in Determination of the Optimal Placement of Distributed Generation Sources in Electrical Distribution Systems”; Optimization of Electrical and Electronic Equipment (OPTIM), 2012 13th International Conference on; pp. 953 – 958; May 2012.
[10] Shalizi, C.; “Distances between Clustering, Hierarchical Clustering”; Carnegie Mellon University – Department of Statistics; September 2009.
[11] de Souto, M.C.P. ; de Araujo, D.S.A. ; Costa, I.G. ; Soares, R.G.F. ; Ludermir, T.B. ; Schliep, A.; “Comparative study on normalization procedures for cluster analysis of gene expression datasets”; Neural Networks, 2008. IJCNN 2008. (IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on; pp. 2792 – 2798; June 2008.
[12] Milligan, G.W.; Cooper, M.C.; “A Study of Standardization of Variables in Cluster Analysis”; Journal of Classification; vol. 5; pp. 181 – 204; 1988.
[13] Kaur, A. ; Kumar, P. ; Kumar, P.; “Effect of noise on the performance of clustering techniques”; Networking and Information Technology (ICNIT), 2010 International Conference on; pp. 504 - 506; June 2010.
[14] Chicco, G.; Napoli, R.; Postolache, P.; Scutariu, M.; Toader, C.; “Customer characterization options for improving the tariff offer”; Power Systems IEEE Transactions; vol. 18; pp. 381 – 387; Feb 2003.
[15] Chicco, G.; Napoli, R.; Piglione, F.; Postolache, P.; Scutariu, M.; Toader, C.; “Application of clustering techniques to load pattern-based electricity customer classification”; Electricity Distribution, 2005. CIRED 2005. 18th International Conference and Exhibition on; pp. 1-5; June 2005.
[16] Jain, A.K.; Dubes, R.C; “Algorithms for clustering data”; Prentice-Hall; 1988.
[17] Photovoltaic Geographical Information System (PVGIS). Joint Research Centre website. [online]. Available from: http://re.jrc.ec.europa.eu/pvgis/ [Accessed 04th May 2014].
[18] Massey, F. J.; "The Kolmogorov-Smirnov Test for Goodness of Fit."; Journal of the American Statistical Association; Vol. 46, No. 253; pp. 68–78; March 1951.
[19] E. Mooi, M. Sarstedt, “A Concise Guide to Market Research”, chapter 9: “Cluster analysis”, Ed. Springer, 2011.

APPENDIX

Detailed Results for the Final Clusters

Looking at the clusters set in more detail, Figures 21 to 27 provide box-plots for each feature considered in the analysis. The box-plots reveal median values for each intra-class attribute (medium horizontal bar), the inter-quartile range (rectangle), the typical extremes (slim black lines) and outlier values (small crosses). In the case of normal distribution probability, the relation between the boxplot and the PDF is shown in Figure 20.

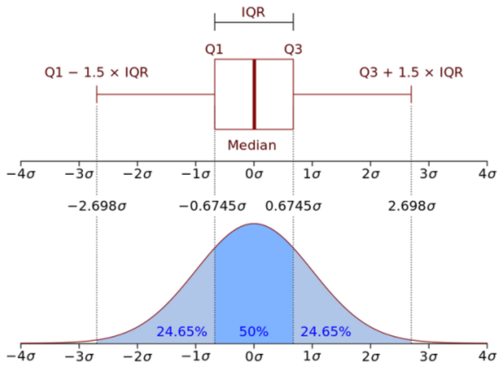


Figure 20. Box-plot vs normal PDF

By examining the box-plots, key differences and equivalencies between clusters were found. The set of box-plots gives useful information related to the characteristics of England's North West feeders.

Figures 21 to 27 show as well a good compactness for each box-plot. At the same time it is possible to identify marked differences between clusters with consolidates the structure quality clamed after the validation process.

Some remarks:

- The median of power factor is in most cases over 0.98
- All feeders analysed are mostly residential
- Neutral current presents significant values (up to 70A average). The diversified neutral current (In/n° of customers) was found to decrease as the number of customers go up in the case of pure domestic clusters.
- The presence of DG clearly reduces active and reactive power consumption and increment the PF level.

Detailed Results for the Final Clusters (without DG)

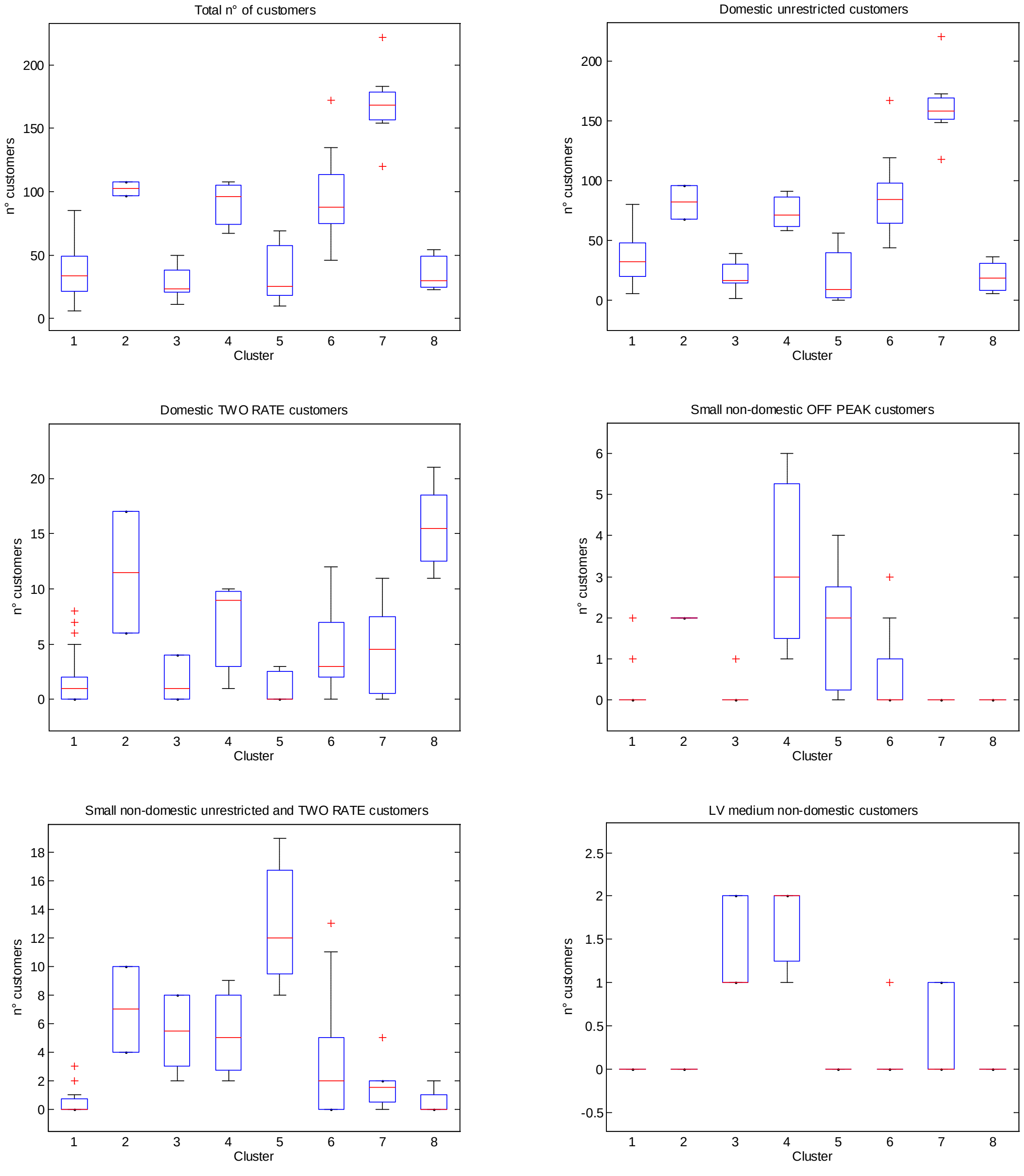


Figure 21. N° and customer type (without DG)

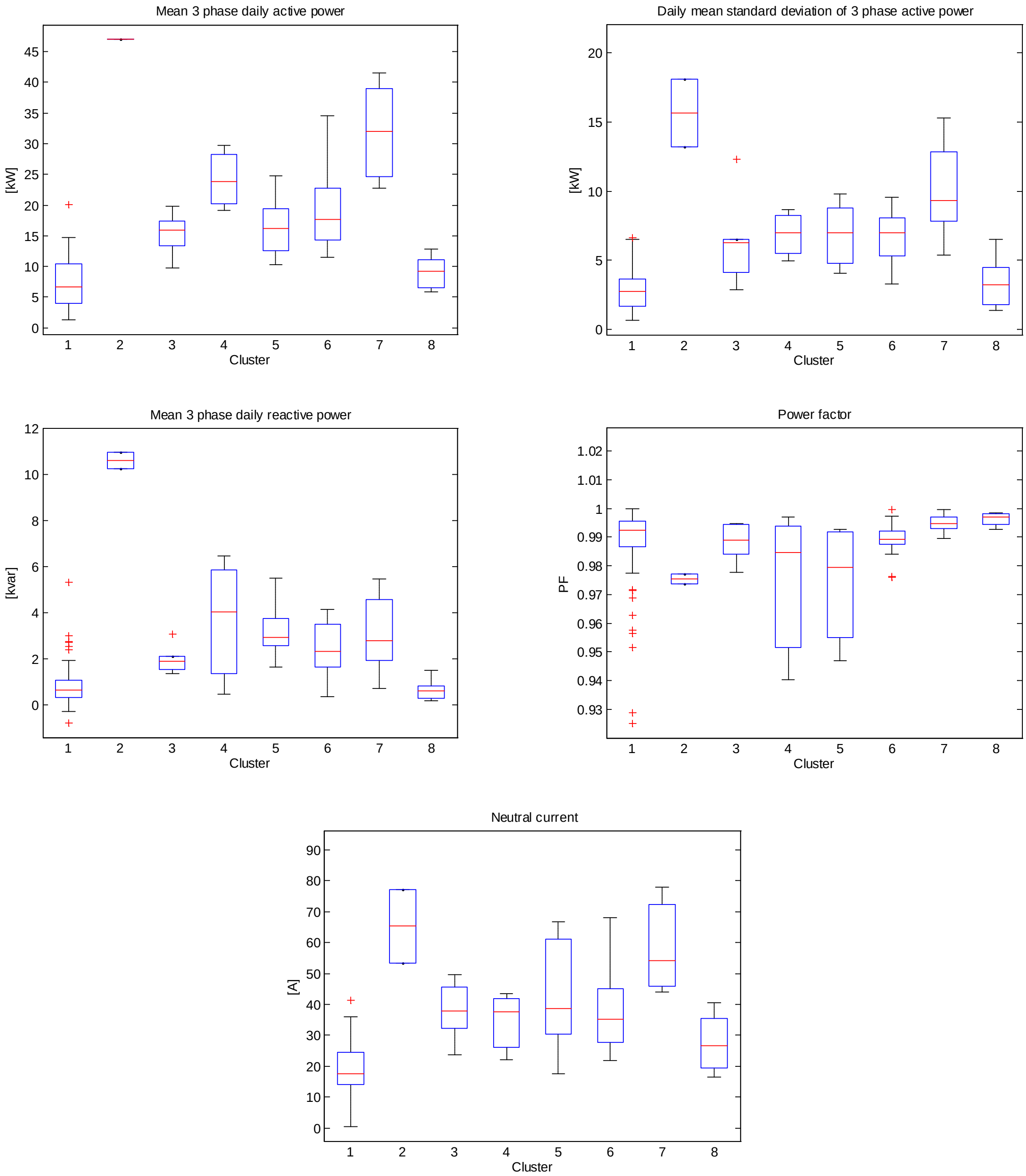


Figure 22. Monitoring data (without DG)

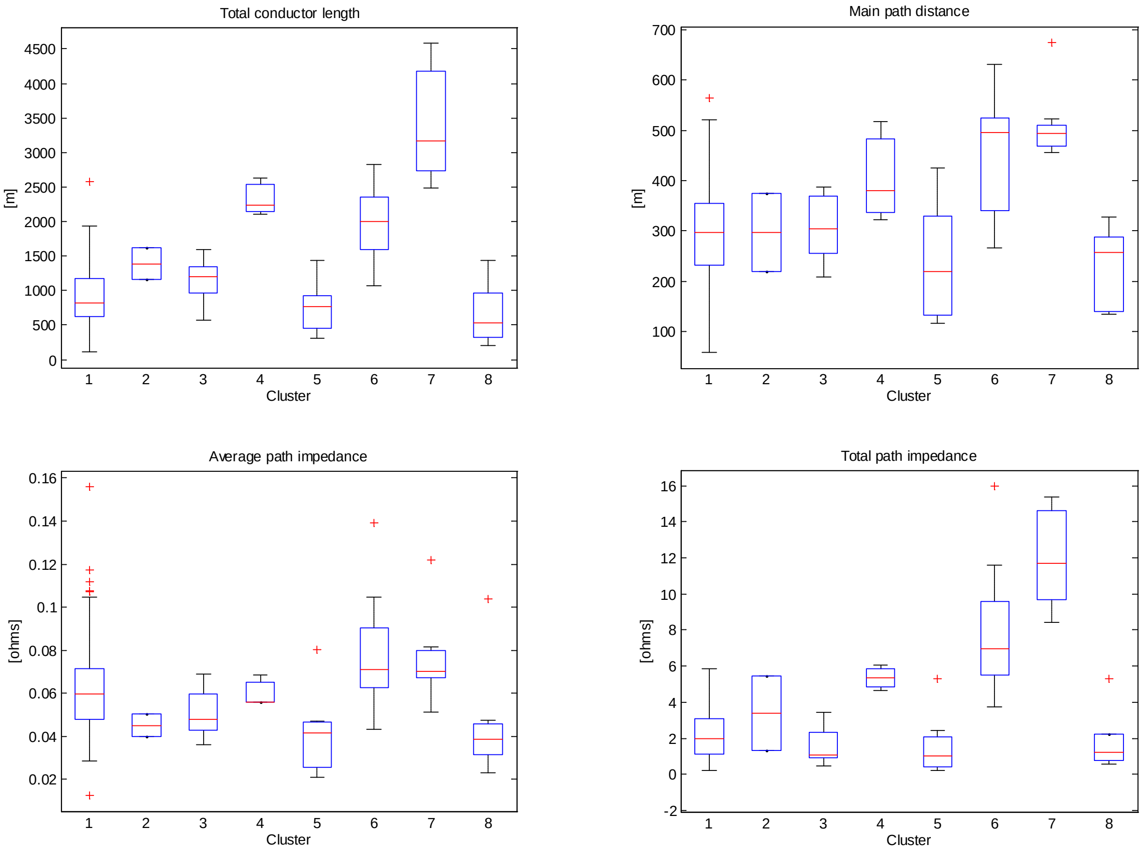


Figure 23. Conductor characteristics (without DG)

Detailed Results for the Final Clusters (with DG)

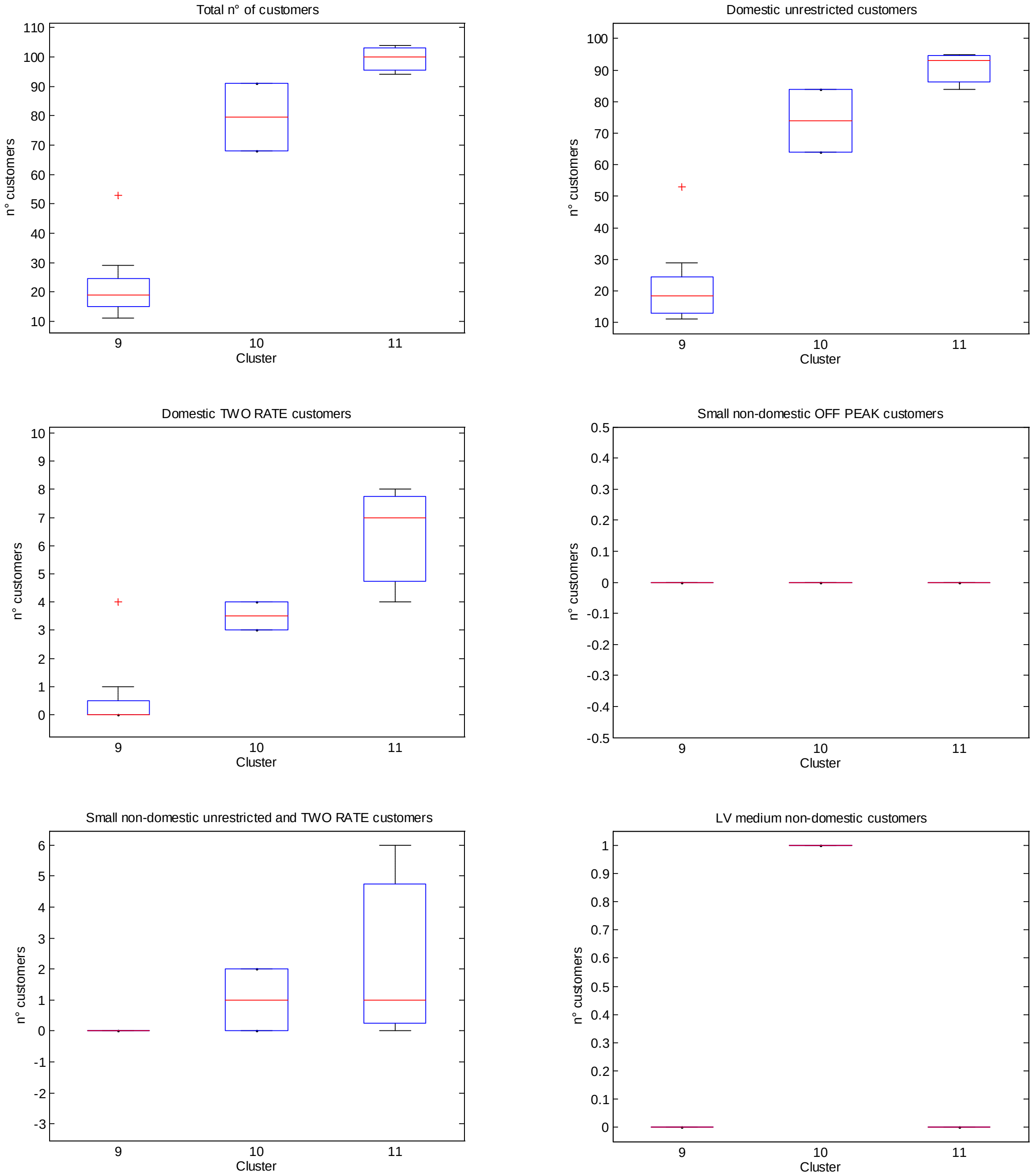


Figure 24. N° and customer type (with DG)

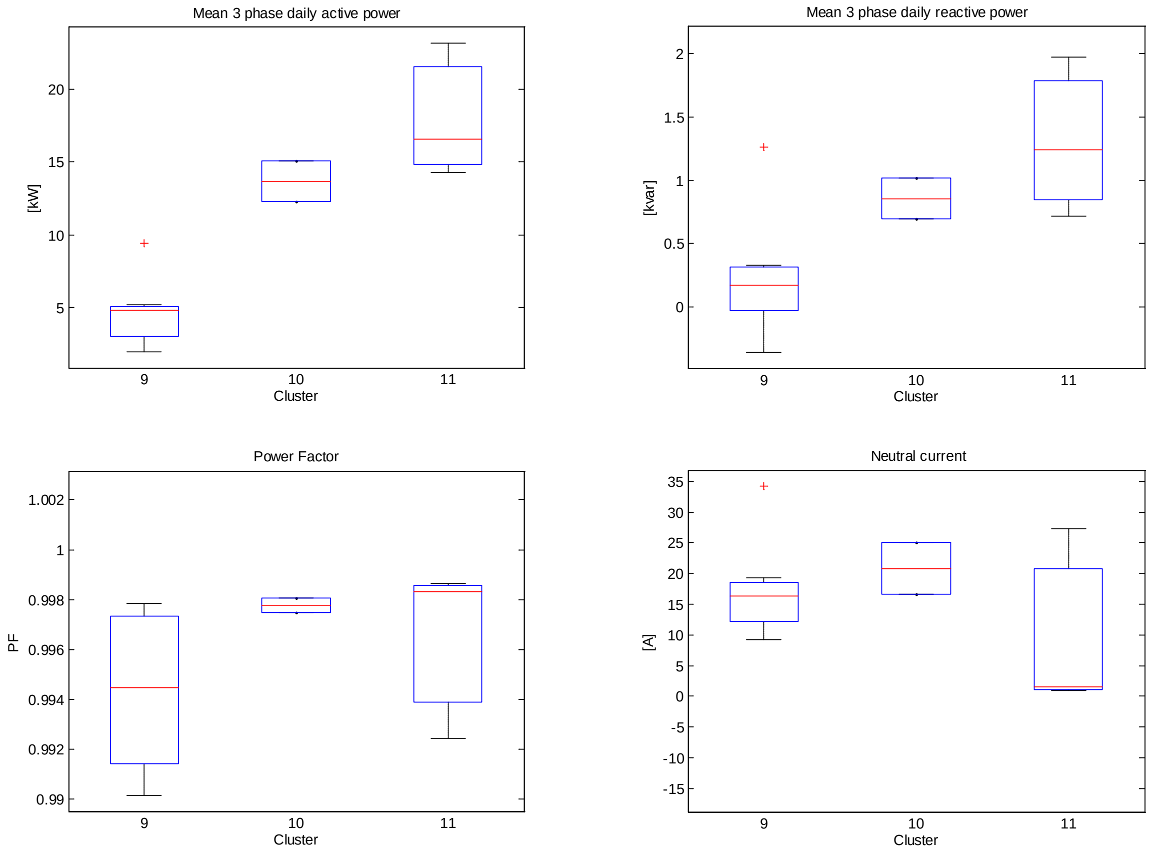


Figure 25. Monitoring data (with DG)

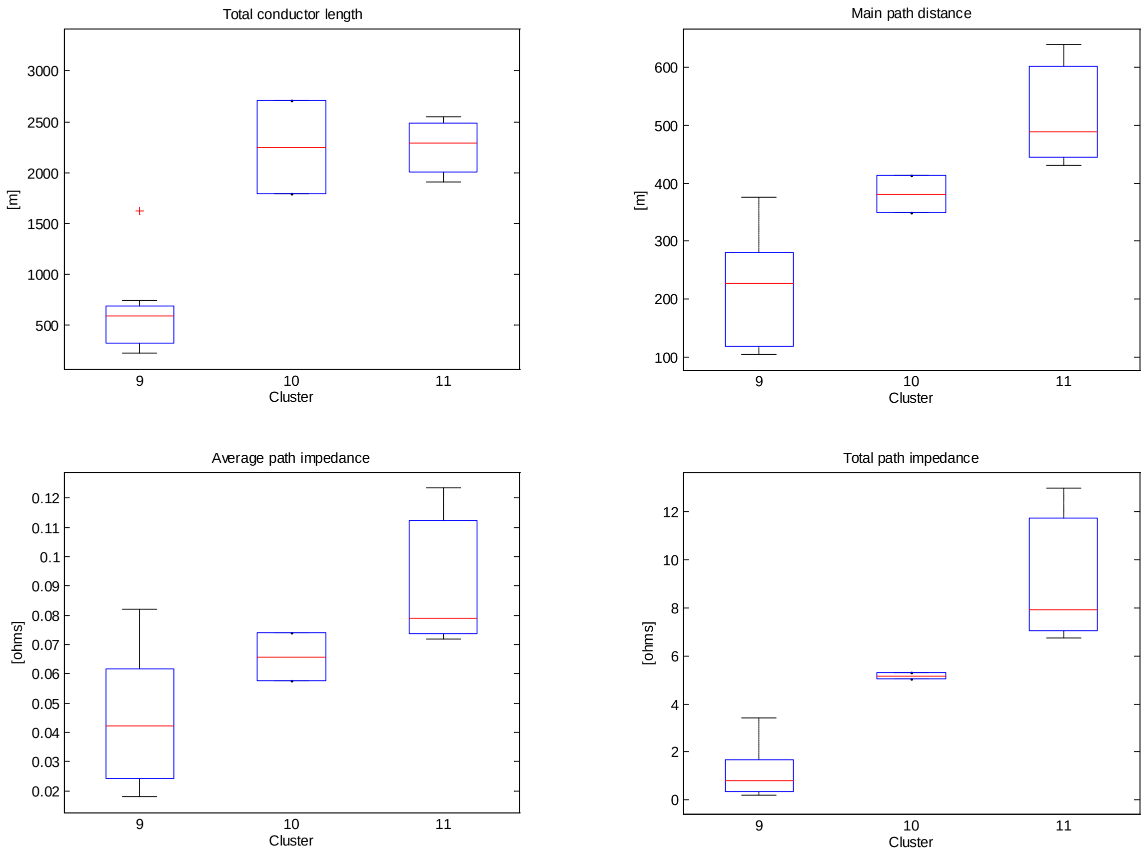


Figure 26 Conductor characteristics (with DG)

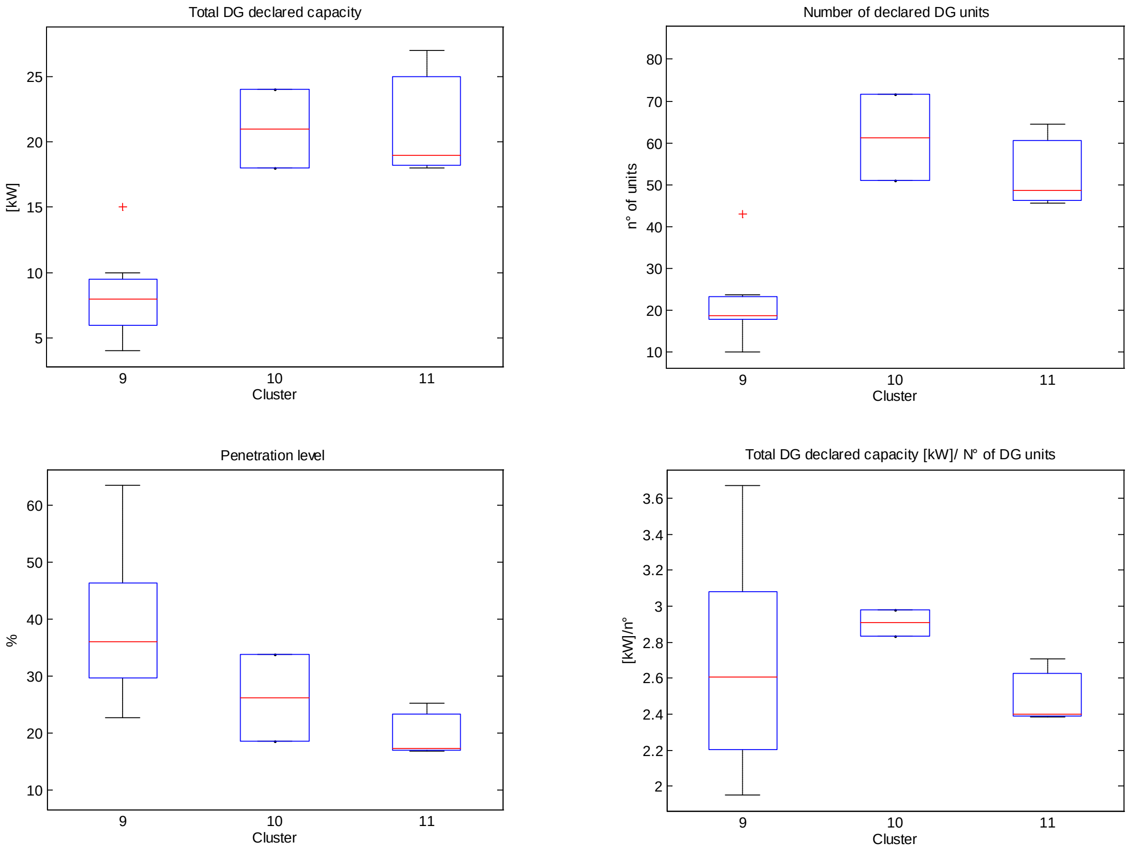


Figure 27. DG penetration characteristics

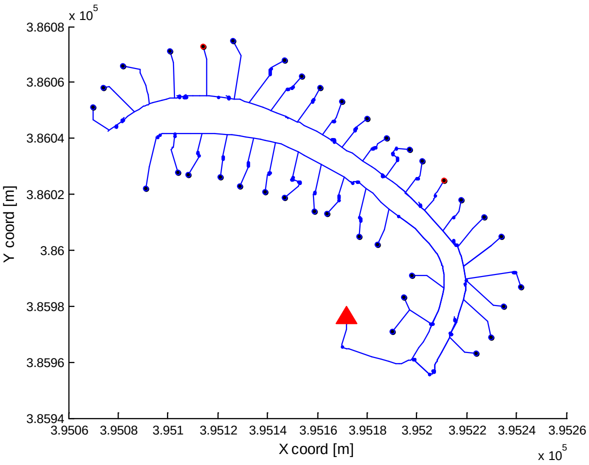
Representative Feeder 1

The main characteristics of this representative feeder are listed below.

1. Main customer composition: Domestic (mainly domestic unrestricted)
2. (Qualitative) Length: Small
3. (Qualitative) Daily mean three-phase active power consumption: Low
4. (Qualitative) Power factor: High (inductive)
5. PV Panels: No
6. Other comments: N/A

The following table and figure present further features of the feeder as well as its topology.

	SUBSTATION	333008
	FEEDER	275033193
CDCM TARIFF MAPPING CODE	Total n° of customers	36
	511	34
	531	2
	591	0
	631-661	0
	751	0
Conductor	Total length [m]	1002
	Main path [m]	270
	Average path impedance [ohm]	0.05
	Total impedance path [ohm]	1.85
	In [A]	15.91
	Mean 3ϕ daily active power [kW]	7.01
	Mean 3ϕ daily reactive power [kvar]	0.99
	Power Factor (PF)	0.988



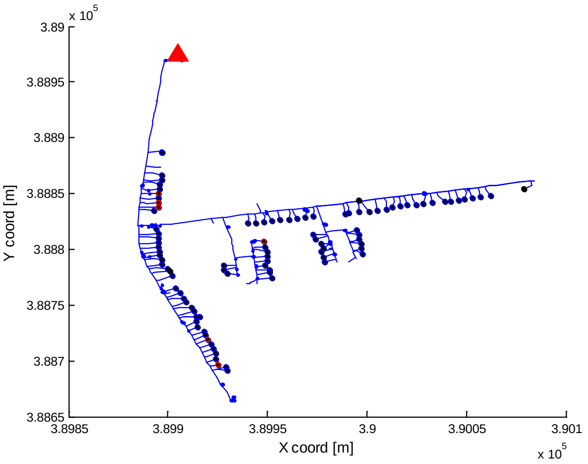
Representative Feeder 2

The main characteristics of this representative feeder are listed below.

1. Main customer composition: Domestic (mainly domestic unrestricted but with presence of some low consumption non-domestic)
2. (Qualitative) Length: Small-medium
3. (Qualitative) Daily mean three-phase active power consumption: Highest
4. (Qualitative) Power factor: Lower (inductive)
5. PV Panels: No
6. Other comments: Highly density area (there’s a customer per each 10m of cable). There is possibly the presence of apartments. It presents the highest values of neutral current.

The following table and figure present further features of the feeder as well as its topology.

	SUBSTATION	330269
	FEEDER	266026370
	Total n° of customers	108
CDCM TARIFF MAPPING CODE	511	96
	531	6
	591	2
	631-661	4
	751	0
Conductor	Total length [m]	1164
	Main path [m]	374
	Average path impedance [ohm]	0.05
	Total impedance path [ohm]	5.43
	In [A]	77.20
	Mean 3ϕ daily active power [kW]	47.02
	Mean 3ϕ daily reactive power [kvar]	10.96
	Power Factor (PF)	0.974



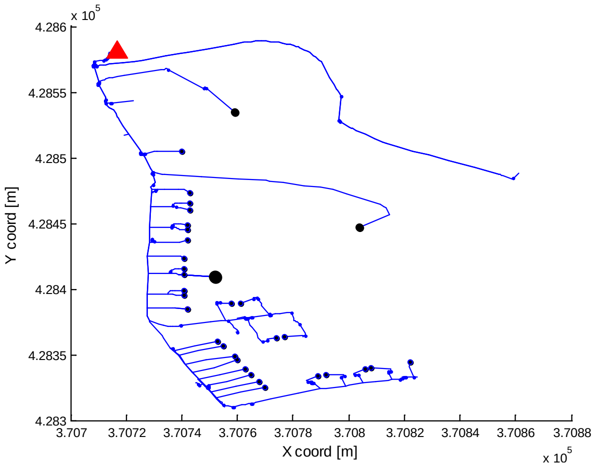
Representative Feeder 3

The main characteristics of this representative feeder are listed below.

1. Main customer composition: Domestic (domestic unrestricted with presence of some low consumption non-domestic and LV medium non domestic customers)
2. (Qualitative) Length: Small
3. (Qualitative) Daily mean three-phase active power consumption: Medium
4. (Qualitative) Power factor: High (inductive)
5. PV Panels: No
6. Other comments: It presents high values of neutral current even if the number of customers is low.

The following table and figure present further features of the feeder as well as its topology.

	SUBSTATION	450028
	FEEDER	113038446
CDCM TARIFF MAPPING CODE	Total n° of customers	38
	511	30
	531	1
	591	0
	631-661	6
	751	1
Conductor	Total length [m]	1591
	Main path [m]	370
	Average path impedance [ohm]	0.06
	Total impedance path [ohm]	2.33
	In [A]	45.54
	Mean 3ϕ daily active power [kW]	16.34
	Mean 3ϕ daily reactive power [kvar]	2.11
	Power Factor (PF)	0.988



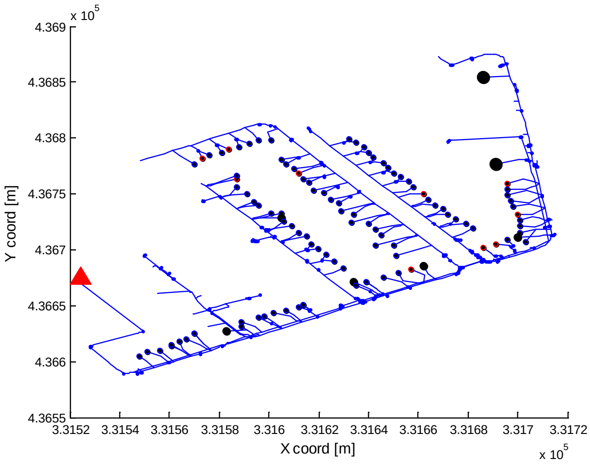
Representative Feeder 4

The main characteristics of this representative feeder are listed below.

1. Main customer composition: Domestic-non and domestic (mainly domestic unrestricted but with presence of some low consumption non-domestic and LV medium non domestic customers)
2. (Qualitative) Length: Large
3. (Qualitative) Daily mean three-phase active power consumption: Medium-high
4. (Qualitative) Power factor: Medium (inductive)
5. PV Panels: No
6. Other comments: N/A

The following table and figure present further features of the feeder as well as its topology.

	SUBSTATION	423108
	FEEDER	158024677
CDCM TARIFF MAPPING CODE	Total n° of customers	108
	511	91
	531	10
	591	3
	631-661	2
	751	2
Conductor	Total length [m]	2241
	Main path [m]	517
	Average path impedance [ohm]	0.06
	Total impedance path [ohm]	6.04
	In [A]	22.22
	Mean 3ϕ daily active power [kW]	29.73
	Mean 3ϕ daily reactive power [kvar]	0.49
	Power Factor (PF)	0.997



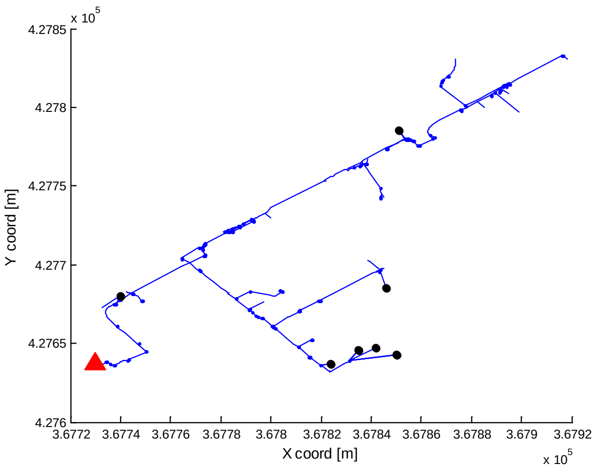
Representative Feeder 5

The main characteristics of this representative feeder are listed below.

1. Main customer composition: Domestic and non-domestic (2/3 domestic and 1/3 of mainly “631-661” customers)
2. (Qualitative) Length: Small
3. (Qualitative) Daily mean three-phase active power consumption: Medium
4. (Qualitative) Power factor: Medium (inductive)
5. PV Panels: No
6. Other comments: It presents high values of neutral current even if the number of customers is low.

The following table and figure present further features of the feeder as well as its topology.

	SUBSTATION	450092
	FEEDER	118030759
CDCM TARIFF MAPPING CODE	Total n° of customers	23
	511	9
	531	0
	591	2
	631-661	12
	751	0
Conductor	Total length [m]	764
	Main path [m]	253
	Average path impedance [ohm]	0.05
	Total impedance path [ohm]	1.08
	In [A]	17.63
	Mean 3ϕ daily active power [kW]	10.28
	Mean 3ϕ daily reactive power [kvar]	1.64
	Power Factor (PF)	0.990



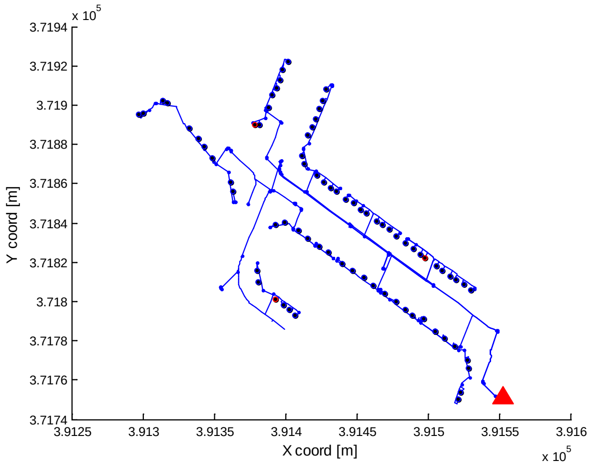
Representative Feeder 6

The main characteristics of this representative feeder are listed below.

1. Main customer composition: Domestic (mainly domestic unrestricted)
2. (Qualitative) Length: Large
3. (Qualitative) Daily mean three-phase active power consumption: Medium
4. (Qualitative) Power factor: High (inductive)
5. PV Panels: No
6. Other comments: N/A

The following table and figure present further features of the feeder as well as its topology.

	SUBSTATION	332927
	FEEDER	281061517
CDCM TARIFF MAPPING CODE	Total n° of customers	76
	511	73
	531	3
	591	0
	631-661	0
	751	0
Conductor	Total length [m]	1664
	Main path [m]	360
	Average path impedance [ohm]	0.09
	Total impedance path [ohm]	6.97
	In [A]	35.05
	Mean 3ϕ daily active power [kW]	14.38
	Mean 3ϕ daily reactive power [kvar]	1.68
	Power Factor (PF)	0.990



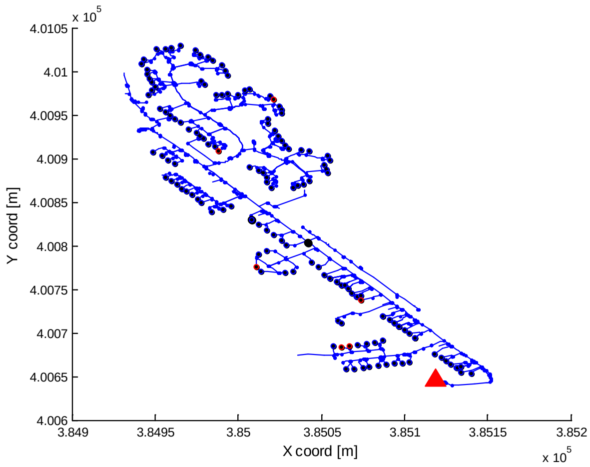
Representative Feeder 7

The main characteristics of this representative feeder are listed below.

1. Main customer composition: Domestic (mainly domestic unrestricted)
2. (Qualitative) Length: Largest
3. (Qualitative) Daily mean three-phase active power consumption: High
4. (Qualitative) Power factor: High (inductive)
5. PV Panels: No
6. Other comments: The neutral current is lower than expected even if the number of customers is the highest one. This contradicts the tendency in other cases (possibly related to a better balance of loads).

The following table and figure present further features of the feeder as well as its topology.

	SUBSTATION	165198
	FEEDER	193051280
CDCM TARIFF MAPPING CODE	Total n° of customers	169
	511	161
	531	6
	591	0
	631-661	2
	751	0
Conductor	Total length [m]	2865
	Main path [m]	522
	Average path impedance [ohm]	0.05
	Total impedance path [ohm]	8.69
	In [A]	43.91
	Mean 3ϕ daily active power [kW]	35.95
	Mean 3ϕ daily reactive power [kvar]	4.01
	Power Factor (PF)	0.993



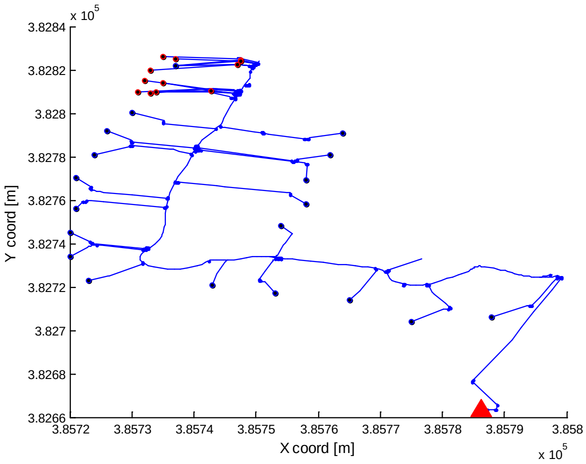
Representative Feeder 8

The main characteristics of this representative feeder are listed below.

1. Main customer composition: Domestic (big presence of domestic two rate customers)
2. (Qualitative) Length: Small
3. (Qualitative) Daily mean three-phase active power consumption: Low
4. (Qualitative) Power factor: High (inductive)
5. PV Panels: No
6. Other comments: The main cable path represents 50% of the total cable length (feeder slightly branched)

The following table and figure present further features of the feeder as well as its topology.

	SUBSTATION	333940
	FEEDER	272040852
CDCM TARIFF MAPPING CODE	Total n° of customers	31
	511	19
	531	12
	591	0
	631-661	0
	751	0
Conductor	Total length [m]	561
	Main path [m]	264
	Average path impedance [ohm]	0.04
	Total impedance path [ohm]	1.11
	In [A]	22.80
	Mean 3ϕ daily active power [kW]	8.58
	Mean 3ϕ daily reactive power [kvar]	0.62
	Power Factor (PF)	0.996



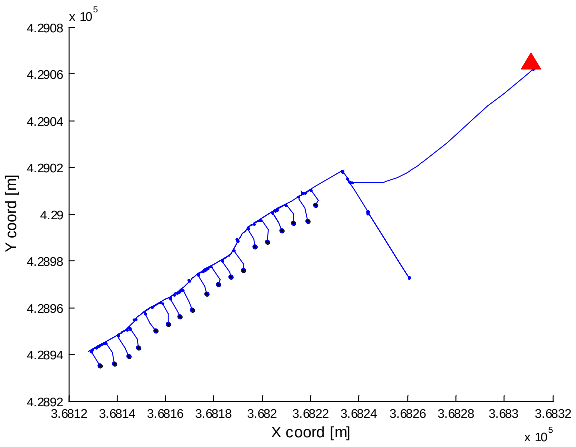
Representative Feeder 9

The main characteristics of this representative feeder are listed below.

1. Main customer composition: Domestic (mainly domestic unrestricted)
2. (Qualitative) Length: Small
3. (Qualitative) Daily mean three-phase active power consumption: Low
4. (Qualitative) Power factor: Very high (inductive)
5. PV Panels: Yes. High penetration level (~40%).
6. Other comments: This feeder is similar to “representative feeder 1” in terms of conductor and customers characteristics. However, there is a noted reduction of active and reactive power consumption. There is an increment of neutral current in relation with similar feeders.

The following table and figure present further features of the feeder as well as its topology.

	SUBSTATION	451004
	FEEDER	119066451
CDCM TARIFF MAPPING CODE	Total n° of customers	18
	511	18
	531	0
	591	0
	631-661	0
	751	0
DG characteristics	Declared DG units	7
	Total DG capacity [kW]	23
	Penetration level	38.9%
	Mean DG capacity [kW]	3.3
Conductor	Total length [m]	593
	Main path [m]	242
	Average path impedance [ohm]	0.03
	Total impedance path [ohm]	0.63
	In [A]	19.25
	Mean 3ϕ daily active power [kW]	4.77
	Mean 3ϕ daily reactive power [kvar]	0.29
	Power Factor (PF)	0.998



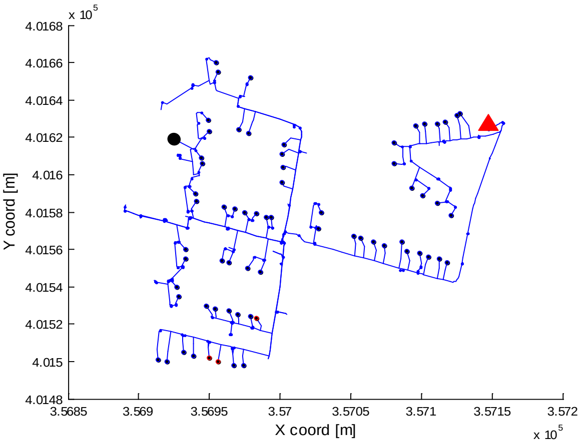
Representative Feeder 10

The main characteristics of this representative feeder are listed below.

1. Main customer composition: Domestic (mainly domestic unrestricted but with presence of LV medium non-domestic customers)
2. (Qualitative) Length: Medium
3. (Qualitative) Daily mean three-phase active power consumption: Low
4. (Qualitative) Power factor: Very high (inductive)
5. PV Panels: Yes. Medium penetration level (~30%)
6. Other comments: There is a low neutral current level in function of the customers' composition.

The following table and figure present further features of the feeder as well as its topology.

	SUBSTATION	211951
	FEEDER	63057173
CDCM TARIFF MAPPING CODE	Total n° of customers	68
	511	64
	531	3
	591	0
	631-661	0
	751	1
DG characteristics	Declared DG units	24
	Total DG capacity [kW]	71.5
	Penetration level	33.8%
	Mean DG capacity [kW]	3.0
Conductor	Total length [m]	1793
	Main path [m]	414
	Average path impedance [ohm]	0.07
	Total impedance path [ohm]	5.03
	In [A]	25.07
	Mean 3ϕ daily active power [kW]	12.25
	Mean 3ϕ daily reactive power [kvar]	0.69
	Power Factor (PF)	0.998



The University of Manchester

Representative Feeder 11

The main characteristics of this representative feeder are listed below.

1. Main customer composition: Domestic (mainly domestic unrestricted)
2. (Qualitative) Length: Large
3. (Qualitative) Daily mean three-phase active power consumption: High-medium
4. (Qualitative) Power factor: Highest (inductive)
5. PV Panels: Yes. Low penetration level (~20% considering all cluster)
6. Other comments: Similar to cluster 6 in terms of conductor and customer's characteristics. However, there is a reduction in terms of active and reactive power consumption and an increment of power factor. The neutral current is practically zero.

The following table and figure present further features of the feeder as well as its topology.

	SUBSTATION	211951
	FEEDER	63057172
CDCM TARIFF MAPPING CODE	Total n° of customers	100
	511	93
	531	7
	591	0
	631-661	0
	751	0
DG characteristics	Declared DG units	27
	Total DG capacity [kW]	64.4
	Penetration level	25.2%
	Mean DG capacity [kW]	2.4
Conductor	Total length [m]	1912
	Main path [m]	640
	Average path impedance [ohm]	0.08
	Total impedance path [ohm]	7.90
	In [A]	1.59
	Mean 3ϕ daily active power [kW]	14.24
	Mean 3ϕ daily reactive power [kvar]	0.72
	Power Factor (PF)	0.999