Life characterization of power distribution transformers using clustering techniques

Edouard Mulliez[1], Miguel A. Sanz-Bobi[1], Álvaro Sanchez[1], Peyman Mazidi[1], Ana González[2], Raúl Bachiller[2]

[1]*Comillas Pontifical University, Engineering School, Institute for Research in Technology, Madrid, Spain*
masanz@comillas.edu

[2]*IBERDROLA, Distribution Division, Dept. of Control Systems, Bilbao, Spain*
ana.gb@iberdrola.es, rbpr@iberdrola.es

ABSTRACT

The current development of the smart grids has considerably increased the amount of research studies around new exploitation paradigms focused on the electrical distribution systems. One of the key elements of electrical distribution networks is the distribution transformer that supports the load that has to feed the consumer needs. This paper aims at characterizing the life of the distribution transformers using clustering techniques. This will make it possible to focus the attention of the Distribution System Operator on particular groups of distribution transformers reducing the amount of information to be analyzed. Also this classification combined with the study of stress indicators for each distribution transformer can be used to complement the criteria used for the network planning and operation.

1. INTRODUCTION

Recent years have been marked by important development and investments in the field of smart grids. They have fostered research which is oriented to a better analysis and study of the distribution network operation by the Distribution Systems Operators (DSOs). Smart grids are characterized by a two-way flow of power and information (Wang et al., 2015). The Advanced Metering Infrastructure (AMI), which is installed in a distribution network, is a key element of the smart grids and the number of installations is growing. As an example, in Europe it is expected that in 2020 (European Commission, 2014) 200 million smart meters will be installed for electricity measurement. Similar effort is being carried out in the rest of the world. This growing flow of up-going information provides interesting indications about the patterns of electricity consumption. These indications, when correctly analyzed, can lead to significant improvements in network operation. The Smart World concept can be defined as an integrated environment with smart grids, homes, buildings and with new infrastructure and tools (Hernandez et al., 2012). Its goals are efficient maintenance and operation, security, mobility, energy savings and quality of life. In order to reach these objectives, the load at the distribution network has to be carefully studied. This can be done from several perspectives that range from the particular consumer load profile till a high level, such as the points of electricity distribution or secondary power substations. The contributions of this paper are not concerned with theoretic aspects, but rather practical ones. Now the DSOs are receiving more and more information from smart meters that are located in customers and secondary power substations. Such a huge amount of information is difficult to analyze without convenient tools. The contributions of this paper are on the DSO practical perspective and are focused on the proposition of analysis methods able to give support in the process of taking decisions in operation and maintenance of a smart distribution network. These methods are based on the characterization of the load in the distribution transformers (DTs) and the possible health conditions of them. Also, this information can be used in other aspects of energy management such as Demand Side Management (DSM) by different agents. It is possible that the DSO has a general idea of the load profiles than the DTs have, but now using the information coming from the smart meters, the DSO could obtain a more accurate view. The knowledge of the characteristic profiles in the low voltage side of the DTs can contribute to a better balance of load at the medium level, to a more accurate view of the relationships between load profiles of consumers as observed in the DT, to a better detection of energy losses due to different causes, and, in general, to improving the capability of operation and management of the low voltage network.

Extended literature about methods characterizing the life observed in a DT, one of the objectives of this paper, does not exist. However load forecasting is a well known area of interest in the electrical field and the methods used there will be taken into account and extended to the context of this paper. Two main families of them can be found in the scientific literature that has been applied to characterizing the electrical consumption in DTs: the methods based on load forecasting and the methods based on load profiling. Forecasting the electrical load at the distribution level can often be a challenge due to possible different requested scenarios of load. In (Amjady et al. 2010), a bilevel strategy with an Artificial Neural Network is used to forecast the hourly load of a microgrid, and (Chen et al. 2014) forecasts the electrical load of a secondary power substation of the distribution network in China. In this reference, as measurement errors are common at the substation level, the authors propose a two-stage bad data identification method to improve the forecast accuracy. Load profiling is a key tool used to characterize the electrical consumption and the behavior of the electrical customers. It is necessary for establishing efficient demand-response programmes and for tariff setting. Load profiling can be applied to households (McLoughlin et all, 2015, Mahmoudi et al. 2010, Kwac et al. 2014 and Albert et al. 2015) or to industrial consumers (Mutanen et al., 2011 and Panapakidis et al. 2012) . In (Verdú et al., 2006) the electrical customers are sorted according to three load shape indices by using a classification tree. The authors of (Panapakidis et al., 2012) focus their contribution on the design of Real Time Pricing options for industrial customers. Yearly profiling and seasonal profiling are compared. The results show that considering the seasonal variations, the potential benefits increase.

Thus, in the area of distribution networks, the behavior of the final electrical consumer has been extensively studied, however, very few studies can be found concerning the load profiles in one of their key elements such as the DTs and for this reason this paper tries to cover this gap. In order to reach this objective, characterization of the load in DTs will be based on similar techniques used for the analysis of the final electrical consumer consumption. The analysis of the DT load can provide the DSO with guidelines for the operation, maintenance and management of the distribution network. The data used in this paper and the methods and results described are part of a project titled “Real proven solutions to enable active demand and distributed generation flexible integration, through a fully controllable low voltage and medium voltage distribution grid”, in short UPGRID, which received funding from the European Union’s Horizon 2020 research and innovation programme under the grant agreement No 646.531.

This paper is organized around the following sections. Section II focuses on the clustering of the real daily load profiles of the DTs used as a data set. Section III proposes guidelines for the DSO. Finally, the conclusions are drawn in Section IV and ideas for future research are suggested.

2. CLUSTERING DISTRIBUTION TRANSFORMERS ACCORDING TO THEIR DAILY LIFE LOAD PROFILES

A distribution network contains an important number of DTs. Usually if the distribution network consists of a total or partial Smart Grid, the amount of data received is huge even when SCADA systems and other tools are in charge of processing and showing the most relevant information in an efficient way. Clustering the DTs according to their daily load profiles makes it possible to reduce an important amount of information available for the DSO in groups of DTs with similar load demand from the consumers. This data mining technique permits the DSO to have a quick view of those DTs that carry out similar work. The categorization of a DT by its particular contribution to the distribution network in terms of typical load profile managed is knowledge that is useful in the cases of fault solutions, better balance of load at the upper level of medium voltage, planning or new investments required as simple examples of its benefit. This section presents an application of a clustering method using the daily load profiles from a set of DTs of a real distribution network. This is only an application to a small part of an actual low voltage network in order to present the potential of the use of traditional clustering techniques for the DSO. In subsection II-A, the data set and its preparation process are introduced. In subsection II-B, different clustering approaches are presented. The daily load profiles are clustered using as a first step the k-means algorithm (Hartigan, 1975) and then hierarchical clustering (Kaufman, 1990) to reduce the number of clusters previously obtained. Observe that the load profile of a DT is not extensively studied in scientific literature due to the lack of availability of data taken in real time till now.

2.1. Characteristics of the data set used and data preparation

This paper will use a data set that consists of hourly load measurements of 59 DTs. They are located in a distribution network in several secondary substations and they have a nominal power rating of 630 kVA. For each DT, the hourly active and reactive powers are known for a period spanning over one to three years. These measurements were collected by supervision meters located in the low voltage side of the DTs. The hourly apparent power was computed since it is directly linked to the capacity limit of the DTs. As a first step, the load data was prepared. The main outliers were detected using basic statistical and visual methods and then eliminated. Also, for each DT, the days with missing values were eliminated.

2.2. Clustering the load daily profiles per DT

The daily load profiles of the DTs were then clustered. Two main types of clustering approaches can be distinguished: direct clustering and indirect clustering. In indirect clustering, data-reduction techniques, such as Principal Component Analysis (PCA) (Jolliffe, 1986) or Curvilinear Component Analysis (CCA) (Demartines, 1997) are applied prior to the clustering process. These methods reduce the data set size and thus the computational burden of the clustering process (Chicco et al., 2006). Since all DTs have the same nominal power rating, no normalization is applied to the load measurements prior to clustering. The main algorithms are: k-means, fuzzy k-means, follow-the-leader, Self-Organizing Maps (SOM) and hierarchical clustering. In (Granell et all, 2015), the performances of hierarchical clustering, k-means and the Dirichlet Process Mixture Model are compared using different known validity indices: Mean Index Adequacy (MIA), Davies Bouldin Index (DBI) and Clustering Dispersion Indicator (CDI). The hierarchical clustering is implemented with different distances. The best scores are obtained with single link hierarchical clustering. Nevertheless, the authors must point out that hierarchical clustering leads to unbalanced clusters, which is often not a recommendable feature for this type of process. K-means leads to more balanced clusters and it is the fastest technique.

In (Kwac et al., 2014), a load shape dictionary is developed for the classification of household load profiles. The load shapes are obtained through adaptive k-means, the number of elements being then reduced by hierarchical clustering. Taking this procedure as a reference, a similar approach has been applied in this paper. The daily load curves from all DTs were first clustered using the k-means algorithm. The number of clusters was decided using the Elbow method (Kodinariya et al., 2013) as starting point and testing several optional values around trying to keep the most important characteristics of the resulting profiles in the clusters obtained. Finally, the number of clusters selected was 15 which corresponds to a reduction of 75% with respect to the original number of DTs analyzed showing a reasonable grouping capability. The clusters provide information about how similar the DT load profiles are, and if the information coming from the 59 DTs can be reduced to a few significant patterns, to 25% or 15 groups in this case. This grouping of DTs allows the DSO to reduce the information available and to better know at a quick glance the details of the different characteristics of the electrical distribution in the region that is managed. This is a practical contribution for helping the DSO. Of course this information is also useful from a different perspective such as DSM because it is possible to observe DTs where the variation of load could be important during the day and where this is not occurring, permitting or not the feasibility of a DSM programme.

As mentioned, the daily load curves from all DTs were first clustered using the k-means algorithm with the Euclidean distance. In order to check the credibility of the results, the same daily load curves were also clustered using a Self-Organizing Map (SOM) algorithm (Kohonen, 2001, Kohonen, 2014) which is an alternative method of clustering. The results after applying SOM were almost identical to those obtained with k-means. That is, 99.6% of the days distributed among the clusters were the same for the SOM and k-means algorithms confirming the results of grouping. The centers of the clusters obtained with k-means are represented in Figure 1 for the 24 hours of the day. Many clusters with low load values are very close together at the bottom of Figure 1. It is possible that the experienced DSO already has an idea about the results obtained in this reduction of information, but it is difficult to keep in mind its accuracy, and for this reason this method contributes to a practical explanation of the real observed load patterns in the DTs.

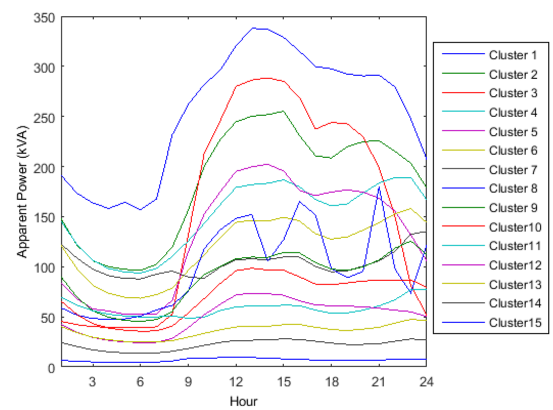

Figure 1. Centers of the clusters obtained with the k-means algorithm

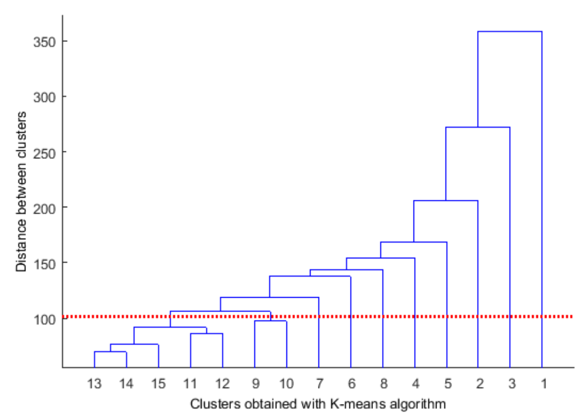

Figure 2. Dendrogram of the hierarchical clustering of the centers obtained with the k-means algorithm

Furthermore, observing that very close patterns were obtained with 15 clusters, there was an attempt to reduce the number of clusters even more, however still explaining the main core of information. In order to do this, a hierarchical clustering based on Euclidean distance, was carried out using the centers of the clusters previously obtained with the k-means algorithm. The result of the hierarchical clustering is shown by the dendrogram in Figure 2 where the closest centers or similar daily load patterns are those with the lowest load mean values (clusters 9 to 15). Observing Figure 2 the original 15 clusters were reduced to 10. This corresponds to cutting the hierarchical tree at the level of the red dotted line (see Figure 2) which was the threshold of similarity selected. The first eight clusters (1 to 8) on the right side of the x-axis of Figure 2 remain unaffected. The new cluster 9 is the result after merging the former clusters 9 and 10, whereas the new cluster 10 is the result after merging the former clusters 11 to 15.

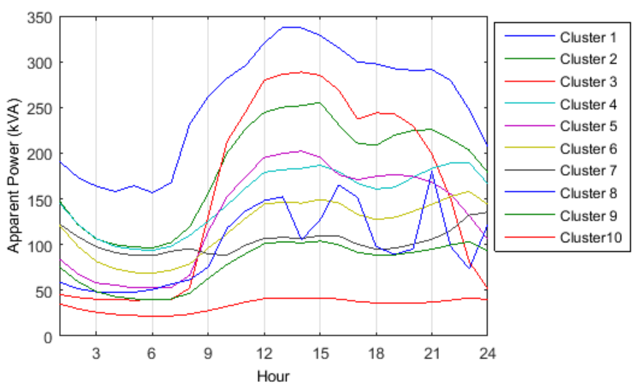

Figure 3. Centers of the clusters obtained with k-means and hierarchical clustering

The centers of the clusters obtained by first applying k-means and then reduced by hierarchical clustering, are represented in Figure 3. The clusters are sorted according to the decreasing value of their load mean value. Figure 3 shows different types of daily load profiles in the DTs studied that are briefly commented about in the next lines.

The DTs in clusters 7 and 10 have a load profile relatively constant all day long, but at different levels of load. However, for most clusters there is a significant difference between day and night hours. This difference is strongly marked for the DTs in clusters 1, 2, 3 and 5. The highest difference is observed for the DTs in cluster 3, whose night load is particularly low compared to its day time load.

The cluster centers generally have two peaks: the first one near 2 pm and the second one, less pronounced, near 9 pm. The latter is slightly pronounced for clusters 1, 5 and 9 whereas it is strongly pronounced for clusters 2, 4, 6 and 8. In cluster 8, there are three different peaks: one at 1 pm, one at 4 pm and another at 9 pm. All these characteristics are very valuable for increasing the knowledge about the characteristics of load profiles demanded and this can be used for a better distribution network management because it helps to anticipate the evolution of the load at each DTs.

The patterns (centers of the clusters) represented in Figure 3 for the different load profiles in the DTs and the area around them covered by the real individual load profiles belonging to each cluster are represented in Figure 4. The deviations between the patterns and their associated individual load profiles of the DTs are generally more pronounced during day time hours than during night time hours. This difference is emphasized for clusters 1, 3, 5 and 8. The maximum difference between individual load profiles in DTs belonging to a same cluster ranges from 100 kVA for cluster 6 (16 % of the total power of the DT) to nearly 160 kVA for cluster 3 (25 % of the total power of the DT).

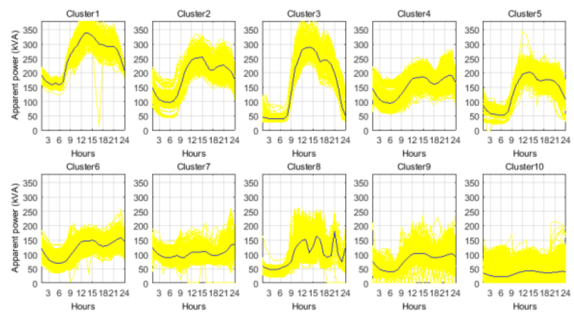

Figure 4. Patterns in Figure 3 (blue lines) and area of the individual load profiles of the DTs that are covered (yellow area)

An alternative approach would have been to directly apply the k-means algorithm with a smaller number of clusters (10 for example) instead of a combination of k-means and hierarchical clustering methods. This option leads to more balanced clusters. However, in this case, the variability inside the clusters with a high load is very important. Distinguishing high daily load profiles is more important than distinguishing low daily load profiles, since the first ones put a higher stress on the distribution network. When successively applying k-means and hierarchical clustering, the intra-cluster variability is more important for clusters of a low load and less important for high load clusters. Since the objective is to distinguish high values clusters, the second approach was finally chosen.

3. PRACTICAL INFORMATION FOR DSO NETWORK OPERATION

The analysis of the load profiles at the distribution transformer level provides interesting recommendations or guidelines for the DSO network operation. In subsection III-A, the relation between the groups of DTs previously established and the aging of the DTs due to their workload is studied. In subsection III-B, an analysis is developed showing how the DTs are distributed among the clusters obtained for similar load profiles. Finally, in subsection III-C, the potential to coordinate the network operation with some DSM actions is analyzed.

3.1. Aging of the DTs due to their workload

One of the main aging factors for distribution transformers is their temperature due to their workload. The mechanical breakdown of the insulation materials, due to thermal aging, is responsible for an important part of the transformer faults (Feng et al., 2012). In general, distribution transformers undergoing the highest load are those enduring the highest temperatures and thus are the ones with the lowest life span.

The study of the load and time of DTs gives some insight about their life span and the stress they have to bear. This is important from a network operation point of view because this information can help to map, a priori, strongest and weakest points within the distribution network in the area managed by the DSO. The following life span indicators were proposed in order to be compared and later selected as the most convenient: the maximum load observed, the 99% quantile value, the 97% quantile, the 90% quantile, the load median value and the percentage of hours above a given threshold of load. After a detailed study, the quantile values are more robust and reliable indices than the maximum value, which is highly sensitive to the outliers. The percentage of hours above a given threshold of load is also an interesting indicator.

In Figure 5, the values of the life span indicators for each DT are represented. The DTs were sorted by increasing mean yearly load. The values for the life span indicators were normalized by the nominal capacity of each DT, i.e. 630 kVA. Generally, these indicators are coherent with one another. Nonetheless, for some DTs (the cases of DTs number 48 and 56 for example), the load median values have an evolution different from their own quantile values. The maximum value for the most stressed DT remains under 70% of its maximum capacity.

The percentages of hours above two given thresholds (30% and 40% of the maximum capacity of the DTs) were computed for each DT as examples of a sensibility analysis. As shown in Figure 6, the majority of the DTs never reach the lowest threshold.

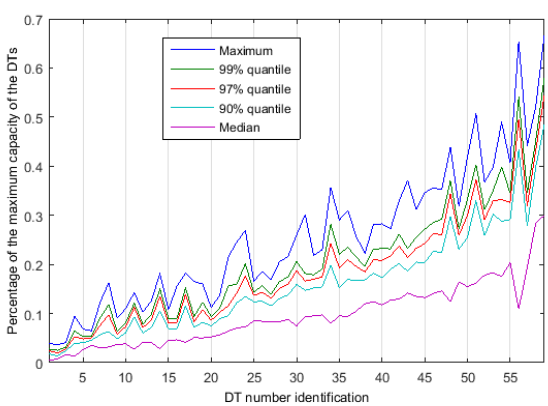

Figure 5. Normalized life indicators tested for the DTs.

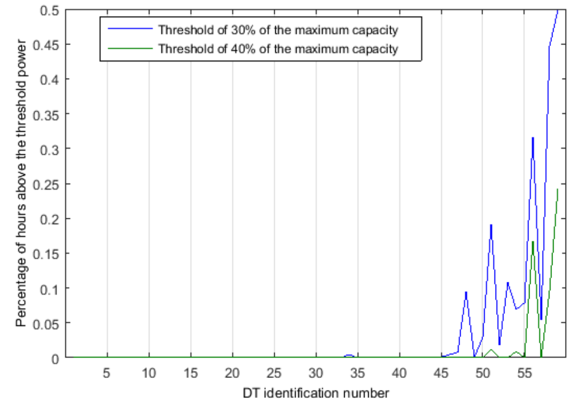

Figure 6. Percentage of hours above a given threshold by the DTs.

The DTs in Figure 6 were sorted by increasing the mean yearly load profile. There are two DTs (the cases of the DTs numbers 56 and 59) which are above the 40% of the load threshold more than 15% of the time observed. Six other DTs are between 30% and 40% for nearly 20% of the time. This means that the DTs analyzed are not suffering a significant degradation due to temperature because the load managed is usually far away from the maximum level for what they are designed for.

3.2. Analysis about how the DTs are distributed among the clusters obtained for similar load profiles

In order to give the DSO information about the robustness of a particular DT to a cluster, the distribution of the clusters among the DTs was studied. The objective is to classify the DTs with similar behavior by group according to the load profiles observed in the DTs. In the previous section, a definition was provided explaining how a DT belongs to a particular cluster. However, it is possible that during some days the load profile of a DT could be closer to the pattern of other clusters and this is convenient to take into account in order to have an idea as to the reliability of the cluster for representing a DT. For each DT, the percentage of days belonging to each cluster was computed. The DTs were sorted by increasing daily mean load observed. The results show that then DT number 1 has the lowest mean load observed whereas DT number 59 has the highest one.

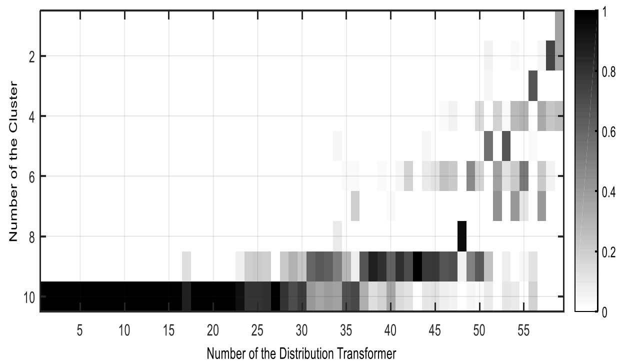

Figure 7. Percentage of days observed that belong to a particular cluster for each DT.

The obtained heat map in Figure 7 highlights some interesting features. The scale of grey colors on the right side corresponds to the percentage of days belonging to each cluster (1 is 100%). As expected, in Figure 7 DTs with low mean load profile always belong to low load clusters, whereas DTs with high mean load profiles belong to high load clusters. The dispersion among the different clusters seems to be more important for DTs that usually have high load levels. There is an important number of DTs whose days only belong to cluster 10, which is a cluster characterized by a relatively constant load throughout the day. This is a special contribution of the paper identifying the DTs where the profile of load is more stable than others, and in this case, what the changes expected on theirs loads are.

Table 1. DTs per group.

Composition and number of DTs in each group		
Group of DT	Clusters	Number of DTs
1	10	24
2	9, 10	15
3	9	4
4	8	1
5	7,10	1
6	6,9	4
7	5,9	1
8	5	1
9	4,6,9	1
10	4,6,7	3
11	4,6	1
12	3,10	1
13	2,4	1
14	1,2,4	1

A methodology was developed to regroup the DTs in a limited number of groups, building on the repartition of the DTs between the different clusters. For this purpose, only the main clusters, which each DT belongs to, were considered. A particular cluster was labeled as the main cluster of a DT if there were more than 1 day out of 7 of the DT belonging to that cluster. It is then possible to classify the DTs according to their main clusters: a group of DTs is constituted by the DTs which have the same main clusters. Table 1 shows the composition of the groups of DTs and the number of DTs each one contains.

The first group contains DTs which belong to cluster 10; the second group contains DTs whose days belong to cluster 10 and 9, and so on. Fourteen groups of DTs are obtained. Table 1 suggest the types that are possible to find for a particular DT. It can be used for load balance and/or DSM strategies. The biggest group is the group with only cluster 10, followed by the groups with clusters 9 and 10. There are several groups with a single DT. The DTs were then sorted according to the group they belong to, from groups 1 to 14.

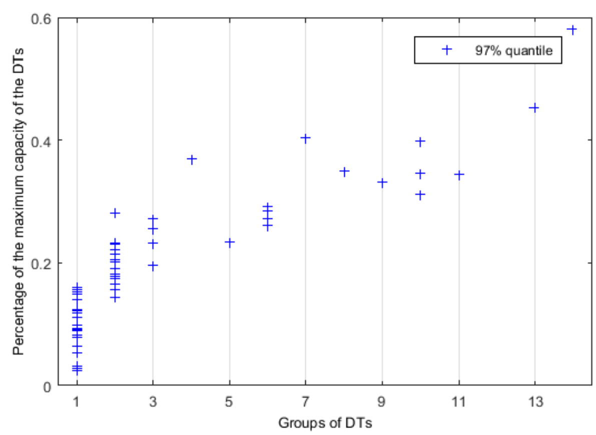

Figure 8. Normalized life span indicators (97% quantile) regrouped by group of DTs.

In Figure 8, one of the life span indicator presented in the previous sub-section, the 97% load quantile, was plotted based on the group of DTs considered. The values of this indicator are coherent inside the same group of DTs. Thus, DTs from the same group are undergoing a similar life degradation. This also means that the DTs from the same group are aging in a similar manner due to their load profiles. The classification of the DTs emphasizes the different level of stress that the DTs are submitted to throughout their lives.

3.3. Potential to coordinate the network operation with some DSM actions

If the peak load of some DTs is too important and sustained over time, some new investments may be necessary to reinforce the network in order to give a good quality of service. These expenses can be avoided or reduced if the peak loads are reduced by coordinating the network operation and some DSM actions oriented to the consumers. Therefore, as is well known, DSM could be applied to diminish or shift the peak load of some DTs which are close to their capacity limit. The DTs analyzed in this paper have a maximum load far below their capacity limit. Therefore, applying here DSM actions coordinated with network operation does not make any sense and this would not reduce the investments required by the current distribution network. Just as a simple example, if the capacity limit of the DTs was lower (smaller transformers), the first DT to be considered for a possible attempt of load reduction, would be the DT whose main cluster is cluster 1, which also corresponds to the DT with the highest stress scores (see Figure 5 and 6).

Consumers with low entropy and large relative load peaks during peak hours have greater potential for DSM programs (Kwac et al., 2013). Taking this fact into account, consumers fed by DTs belonging to clusters 1, 2 and 3 seem to be good candidates for DSM actions.. The centers of these clusters are indeed characterized by a sharp difference in load level between daytime and night time. Their peak load is well-defined and localized near 2 pm. These clusters correspond to the DTs 57 to 59 of Figure 7. As has been pointed out previously, the application of DSM in the study case would nevertheless not help in avoiding current investments in the distribution network as the DTs studied are far below their capacity limit.

4. CONCLUSION

In this paper, part of the distribution transformers of a low voltage electrical distribution network was studied. It belongs to one of the four demonstration areas of the UPGRID project funded from the European Union's Horizon 2020 research and innovation programme. The first contribution of the paper is the practical analysis of load profiles in DTs using measurements coming from smart meters. First, the daily load profiles from all DTs under study representing their lives were clustered in order to reduce the huge amount of information and obtain the most representative load patterns (clusters). Then, the DTs were regrouped depending on their main clusters, that is to say, the clusters to which a significant part of their days belong to. This methodology is a contribution that presents advantages for network operation and management of the DTs in several areas such as better balance of load at the upper level of medium voltage network, improvement in the accuracy of the characteristic load profile of DTs, better knowledge for DSM actions and the possibility to detect energy losses due to technical and fraudulent causes in contrast to the DT load profiles with the aggregated measurements coming from the customers meters.

The classification of the DTs can be used by the DSO to get a quick visualization as to how the load is distributed in its managed area and their typical life profiles reduced to a few patterns. Another contribution of the paper was the analysis of the robustness of the cluster assigned to each DT identifying possible close different load profiles that sometimes could exist due to a different use of energy by the customers supplied. Two other important added values were studied from the knowledge included in these load patterns. First, some life indicators for the DTs were proposed and tested in this paper in order to know how their workload can affect their life expectancy. A second added value obtained from the patterns of load identified is that it can be used to target consumers and geographical areas for possible DSM actions. In the case analyzed, the benefits from a possible application of DSM programs are very limited given that the load of the DTs is far below their limit of capacity.

ACKNOWLEDGEMENT

The paper includes partial results of the UPGRID project. The authors are grateful to all the other partners cooperating in the project and the support received from them. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 646.531, for which the authors wish to express their gratitude. For further information check the website: http://upgrid.eu/

REFERENCES

Amjady, N., Keynia, F., Zareipour, H. (2010) Short-term load forecast of microgrids by a new bilevel prediction strategy. *IEEE Transactions on Smart Grid,* vol. 1, no. 3, pp. 286-294.

Albert, A., Rajagopal, R. (2015) Thermal profiling of residential energy consumption using smart meter data. *IEEE Transactions on Power System*, vol.30, no. 2, pp. 1-10.

Bishop, C. (2006) *Pattern Recognition and Machine Learning*, Springer

Chen, X., Kang, C., Tong, X., Xia, Q, Yang, J. (2014) Improving the accuracy of bus load forecasting by a two-stage bad data identification method. *IEEE Transactions on Power Systems*, vol 29, no. 4, pp. 1634-1641.

Chicco, G., Napoli, R., Piglione, F. (2006) Comparisons among clustering techniques for electricity customer classification. *IEEE Transactions on Power Systems*, vol. 21, no. 2, pp. 933-940.

Demartines, P., Hérault, J. (1997) Curvilinear Component Analysis: A Self-Organizing Neural Network for Nonlinear Mapping of Data Sets. *IEEE Transactions on Neural Networks*, vol. 8, no. 1, pp. 148–154.

European Commission report (2014). Benchmarking smart metering deployment in the EU-27 with a focus on electricity. European Commission report, 2014. http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52014DC0356&from=EN (accessed February 2018).

Feng, D.Y., Wang, Z.D., Jarman, P. (2012) Modeling thermal life expectancy of the UK transmission power transformers. *Proc. 2012 IEEE International Conference on High Voltage Engineering and Application*, pp. 540-543.

Granell, R., Axon, C. J., Wallom, D. C. H. (2015) Impacts of Raw Data Temporal Resolution Using Selected Clustering Methods on Residential Electricity Load

Profiles. *IEEE Transactions on Power Systems,* vol. 30, no. 6, pp. 3217-3224.

Hartigan, J.A. (1975) *Clustering algorithm,* John Wiley & Sons, Inc.

Hernández, L. Baladrón, C., Aguiar, J.M., Calavia, L., Carro, B., Sánchez-Esguevillas, A., Cook, D.J., Chinarro, D., Gómez, J. (2012). A study of the relationship between weather variables and electric power demand inside a smart grid/smart world framework. *Sensors,* vol. 12, no. 9, pp. 11571-11591.

Jolliffe, I.T. (1986) *Principal Component Analysis*, Springer-Verlag.

Kaufman, L., Roussew, P.J. (1990) *Finding Groups in Data - An Introduction to Cluster Analysis,* A Wiley-Science Publication John Wiley & Sons.

Kodinariya, T.M, Makwana, P.R. (2013) Review on determining number of Cluster in K-Means Clustering. *International Journal of Advance Research in Computer Science and Management Studies*, vol. 1, no.6, pp. 90-95.

Kohonen,T. (2001) *Self-Organizing Maps*, Springer

Kohonen,T. (2014). *MATLAB Implementations and Applications of the Self-Organizing Map*. Unigrafia Oy, Helsinki, Finland

Kwac, J., Tan, C.W., Sintov, N., Flora, J., Rajagopal, R. (2013) Utility customer segmentation based on smart meter data: Empirical study. *2013 IEEE International Conference on Smart Grid Communications*, pp. 720-725.

Kwac, J., Flora, J., Rajagopal, R. (2014). Household energy consumption segmentation using hourly data. *IEEE Transactions on Smart Grid,* vol. 5, no. 1, pp. 420-430.

Mahmoudi-Kohan, N., Moghaddam, M.P., Sheikh-El-Eslami, M.K, Shayesteh, E. (2010) A three-stage strategy for optimal price offering by a retailer based on clustering techniques. *International Journal of Electrical Power & Energy Systems,* vol. 32, no. 10, pp. 1135-1142.

McLoughlin, F., Duffy, A., Conlon, M. (2015) A clustering approach to domestic electricity load profile characterisation using smart metering data. *Applied Energy* vol. 141, no. March, pp. 190-199.

Mutanen, A., Ruska, M., Repo, S., Järventausta, P. (2011) Customer classification and load profiling method for distribution systems. *IEEE Transactions on Power Delivery*, vol. 26, no.3, pp. 1755-1763.

Panapakidis, IP., Simoglou, CK, Alexiadis, M.C., Papagiannis, G.K. (2012) Determination of the optimal electricity selling price of a retailer via load profiling. in *Proc. 2012 of the 47th International Universities Power Engineering Conference (UPEC),* pp. 1-6.

Shalev-Shwartz, S., Shai Ben-David, S. (2014) *Understanding Machine Learning From Theory to Algorithms,* Cambridge University Press.

Verdú, S.V., García, M.O., Senabre, C., Marín, A.G., Franco, F.J.G (2006) Classification, filtering, and identification of electrical customer load patterns through the use of self-organizing maps. *IEEE Transactions on Power Systems,* vol.21, no. 4, pp. 1672-1682.

Wang, Y., Chen, Q., Kang, C., Zhang, M., Wang, K., Zhao, Y. (2015) Load profiling and its application to demand response: A review. *Tsinghua Science and Technology* vol. 20, no. 2, 2015, pp. 117-129.