Application of data analytics for advanced demand profiling of residential load using smart meter data

Application of Data Analytics for Advanced Demand Profiling of Residential Load using Smart Meter Data

Jelena Ponoćko, *Student Member IEEE*, and Jovica V. Milanović, *Fellow IEEE*

School of Electrical and Electronic Engineering
University of Manchester
Manchester, UK
jelena.ponocko@manchester.ac.uk; milanovic@manchester.ac.uk

Abstract—**This paper proposes a methodology for demand profiling, namely load decomposition, of aggregated residential load based on smart meter (SM) data. The methodology is applicable to both active and reactive load, following an assumption that SMs can monitor real-time active power consumption of individual appliances. Only a number of households in the aggregation are equipped with SMs in this study. The non-monitored users' load is decomposed using artificial neural network (ANN) trained with the available SM data. Information about load composition, in terms of load categories or load controllability, can be highly beneficial for various demand response (DR) applications. Different levels of SM coverage are considered in the study to illustrate the effect of the level of SM coverage on the accuracy of total aggregated load decomposition. The results show that the consumption of some load categories can be estimated with high confidence, even at lower levels of SM coverage.**

Index Terms-- **Demand-side management, distribution network, smart meter, neural networks**

I. INTRODUCTION

Following the roll-out of smart meters (SMs) in residential districts around the world, the end-users will gain better observability of their consumption, as well as higher potential to participate in the power network daily operation. Higher granularity of low-level consumption data in the future distribution grid will bring benefits to both consumers and the distribution system operator (DSO). On the one hand, smart metering will facilitate awareness of consumers about their daily consumption and enable them to make savings by reacting to price signals or various types of incentives triggered by their electricity supplier. On the other, SM data will provide information to the DSO about individual load profiles, enabling more advanced profiling of consumers in different areas and at different levels of aggregation.

Load profiling has shown crucial role in the studies of direct load control (DLC), demand response (DR) programs, design of tariffs and involvement of local generation [1]. An important part of load profiling is flexibility profiling, i.e. assessment of the size of controllable (shiftable) load within the total load. The assessment can be performed in two dimensions:

- *Time*: observing the change in the size of controllable load within the total load over a day or a season;
- *Space*: observing the size of controllable load over a distribution network. In this case, different network buses will have different flexibility potential, depending on their load mix (namely residential, industrial or commercial users).

Furthermore, load can be disaggregated (decomposed) into load categories, such as resistive loads, induction motors, lighting, etc., in order to obtain a more detailed insight into the types of load utilized on a daily or seasonal basis. Following the DLC scenario, where certain load categories are equipped with smart controllers, load disaggregation would provide information on the amount and profile (category) of the disposable controllable load. Also, for any kind of incentive-based DR, estimation/prediction of load composition at aggregation point could show if the DR potential (flexibility) of the aggregated consumption is big enough for the needs of the program, or some additional measures are necessary.

This paper introduces load profiling methodology relying on minute-based consumption data coming from SMs. The main assumption is that SMs measure consumption of individual domestic appliances, which enables detailed time-based analysis of DR potential of the residential end-users. This approach is motivated by the existence of small trial sites performing load measurements on appliance level, mainly in residential sector [2-4]. Assuming this kind of smart metering system, a bottom-up approach can be used to assess load flexibility using SM data. For the non-monitored load, i.e. the end-users without an SM, artificial neural network (ANN) based approach is taken for load decomposition, as described in the following chapters.

II. DEMAND PROFILING OF RESIDENTIAL LOAD

Demand profiling in this paper refers to load decomposition (disaggregation), which results in time-varying load composition, i.e. contribution (in per unit or percentage) of different load categories to the total active or reactive load.

Following methodology discussed in [5], load categories are defined as groups of appliances with similar voltage-dependent steady-state and dynamic load characteristics. Furthermore, load categories are divided into controllable and uncontrollable based on their potential to be shifted in time. According to the most commonly used appliances in residential sector in the UK, seven categories are recognized in this analysis and presented in Table I as: single-phase constant torque induction motors (CTIM1), three-phase constant torque induction motors (CTIM3), single-phase quadratic torque induction motors (QTIM1), controllable resistive loads (R_C), uncontrollable resistive loads (R_{UC}), switch-mode power supply (SMPS) loads and Lighting.

The full list of appliances was adopted from the residential load model given in [3]. Due to the lack of larger datasets coming from the existing pilot sites, this model was adopted to generate synthetic SM data for the case studies in this paper.

TABLE I. LOAD CATEGORIES AND CORRESPONDING TYPES OF APPLIANCES

Load controllability	Load categories	Residential appliances
Controllable	CTIM1	Dish washer, tumble dryer, washing machine, washer-dryer, vacuum cleaner
	CTIM3	Electrical space heating
	QTIM1	Chest freezer, fridge-freezer, fridge, upright freezer
	R_C	Water heater, electrical shower, storage heater
Uncontrollable	R_{UC}	Iron, hob, oven
	SMPS	Answer machine, CD player, Clock, telephone, high fidelity (HiFi) appliances, Fax machine, PC, printer, TV, VCR-DVD, receiver, microwave
	Lighting	Lighting

III. SMART METERING

In the two-way communication between electricity suppliers and end-users provided by the SMs, the end-users may receive daily and seasonal change in electricity price as a signal to change their load patterns and make savings for both themselves and suppliers [6]. The users will be provided with close to real-time information about their energy consumption, as well as the cost of the kWh and the amount of CO_2 emission [7, 8]. In the UK, customers will also choose whether they want their measurement data to be sent every 30 minutes, daily or monthly.

Next to the benefits provided to the end-users, there are several advantages SMs could bring to the DSO, including:

- More accurate load forecasting and distribution system state estimation, at different aggregation levels
- Improved fault detection and faster fault restoration

- Monitoring of distributed generation, and thus avoiding appearance of hidden load in the network

At this point, the transmission system operator (TSO) in the UK uses DR only from large industrial users and estimates it in advance, so that the balancing between generation and load can be done 2-60 minutes in advance of real time (the data is received every 30 minutes from generation units and large loads) [9]. On the other hand, higher level of load aggregation in residential districts can also have a significant potential for DR – only in UK residential sector forms around 30% of the overall consumption. A potential hurdle for practical implementation might be the discrepancy between the time steps at which data is collected from SMs and time granularity of system operation activities.

Aggregation of SM data is proposed as a way of enhancing the DR potential of end-users, and protecting privacy of the users. Although an aggregator might include customers which are not necessarily supplied from the same distribution substation, it is assumed that the consumption data from all the households in this study are aggregated at the common substation (bulk) point. This type of communication architecture also enables more efficient data handling, i.e., reduction of the size of data sent to the upstream network [9].

IV. METHODOLOGY

The diagram in Fig. 1 presents methodology for load decomposition in case of a smart metering system with partial coverage (i.e. only some consumers in the aggregation area have an SM). Two assumptions are made in this respect:

1) SMs can measure consumption (real and reactive power) of individual appliances (block {1}); if not measured, reactive load can be derived probabilistically, using a range of possible values of power factor for each appliance; this approach was taken in this paper
2) Measurement of total consumption (real and reactive power) at the substation point (block {5}) is available every minute.

Following the first assumption, part of the consumption which is monitored with SMs can be decomposed into categories or controllable/uncontrollable load by simply summing consumption of appliances belonging to the same category (block {2}) according to the classification in Table I. Load composition is given as time-changing participation (in per unit - p.u.) of each of the seven categories within the total load. Since there are non-monitored households, ANN based approach is adopted to estimate the demand composition of the total aggregated load based on the available SM data. A two-layer feed-forward ANN with Bayesian Regulation Backpropagation is used, following its good performance reported in [10]. In this case, available SM data is used for training the ANN, using total (aggregated from SMs) active and reactive power as input data and participation of the seven categories (in p.u.) as the target data (block {3}).
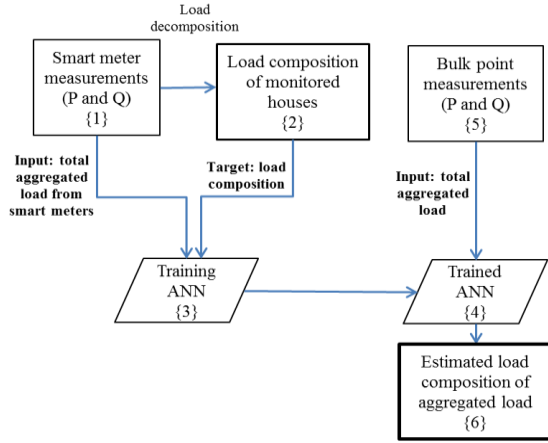
Figure 1. Flowchart of load decomposition

The training process is performed over 7 days data, which includes minute-based real power (P) and reactive power (Q) measurements, so there are 7*1440 samples for each of the variables, presented in a matrix form as follows:

$$Input = \begin{bmatrix} P_1 & \cdots & P_i & \cdots & P_{7*1440} \\ Q_1 & \cdots & Q_i & \cdots & Q_{7*1440} \end{bmatrix}. \qquad (1)$$

Target data, defined as W, presents the participation of each load category in p.u. If in a time step i, active load of category j equals P_{ji}, then the participation w_{ji}^P of that category is given as:

$$w_{ji}^P = \frac{P_{ji}}{P_i}, \qquad (2)$$

where P_i is the total active load in a time step i. Target data can then be presented in matrix form as follows:

$$Target = \begin{bmatrix} w_{1,1}^P & \cdots & w_{1,7*1440}^P \\ w_{2,1}^P & \cdots & w_{2,7*1440}^P \\ \vdots & \cdots & \vdots \\ w_{7,1}^P & \cdots & w_{7,7*1440}^P \end{bmatrix}. \qquad (3)$$

In case of reactive power, the participation of each category is given as:

$$w_{ji}^Q = \frac{Q_{ji}}{Q_i} = \frac{P_{ji}\tan(\varphi_{ji})}{P_i\tan(\varphi_i)} = w_{ji}^P \frac{\tan(\varphi_{ji})}{\tan(\varphi_i)}, \qquad (4)$$

where φ_{ji} and φ_i are phase angles of category j and total load in time step i, respectively.

Once trained, the ANN (block {4}) uses measurements of total active and reactive load at the substation point as the input, giving its load composition as the output (block {6}). It is worth mentioning that in case of smart metering system with full coverage (base case), the load decomposition process would consist only of parts described by blocks {1} and {2} and it would be trivial. Load composition of the base case is given in Fig. 2 and used as a reference for the calculation of accuracy in the case studies.

V. CASE STUDIES

In order to prevent high influence of individual load profiles on the total load shape and to reflect a relatively large number of households supplied by the same distribution substation, aggregation of 1000 houses is analyzed in this paper. Following this, residential occupancy statistics for the UK is adopted from [11], where it is stated that 29% of houses in the residential district accommodate single resident, 35% accommodate two, 16% have three residents and 20% have four.

As suggested in [3], in order to prevent under-estimation of electrical consumption and provide sufficient data for detailed modelling of distribution networks, one minute-based load profile is chosen as the reference. Considering that balancing between generation and demand is performed every minute, the available flexibility should be reported to the network operator at this rate as well [9]. It was reported in [12] that the active and reactive consumption could be measured by SMs over periods from 1 to 60 minutes. Assuming the freedom of end-users to choose how often their consumption data will be sent, the SM data streams are taken to be with one, 10, 30 or 60 minutes rate, randomly. According to [13], up to 20% of active load measurements at substation points are inaccurate. Therefore, 20% of the overall data coning from SMs is assumed to be missing ("NaN" values). The case studies shown here analyze data after it has been processed, restored and aggregated, though the data preprocessing and restoration is out of the scope of this paper.

Objective of the examples described in the case studies is to illustrate the effect of smart metering system coverage on the accuracy of demand decomposition in an aggregation of 1000 households. Therefore, three cases will be investigated: Case I: 20% coverage, i.e. 200 households out of 1000 have smart meters; Case II: 50% coverage; Case III: 80% coverage.
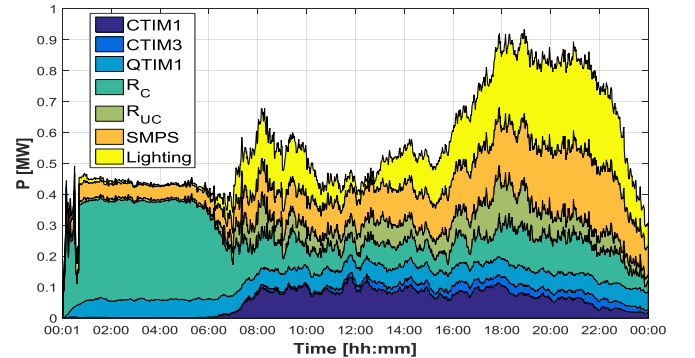

Figure 2. Load composition of 1000 households with 100% SM coverage (base case)

VI. RESULTS

The accuracy of the load decomposition is assessed using the load share error (WFE), calculated for each time step (minute) as follows:

$$WFE_{cat} = WF_{cat,ANN} - WF_{cat,real} \text{ ,} \tag{5}$$

where $WF_{cat,ANN}$ is the share (contribution) of the load category obtained as the result of the ANN, and $WF_{cat,real}$ is the actual share of the category (from the base case), both given in p.u. based on the monthly peak load. Fig. 3 presents the way errors are accounted for on the example of two load categories, namely controllable and uncontrollable load. Shares of controllable and uncontrollable load are obtained by summing up the shares of corresponding load categories, following Table I.
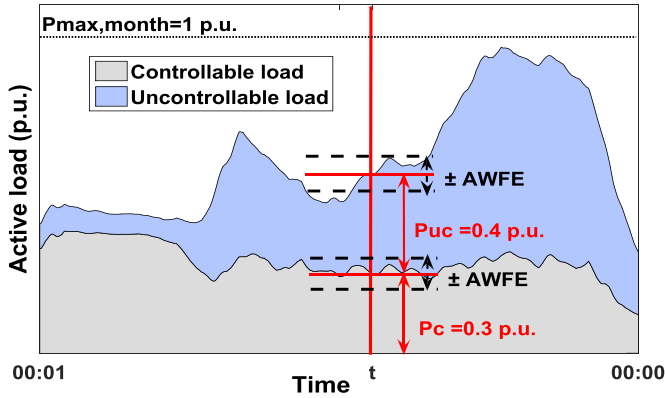

Figure 3. Presentation of the confidence level on the example of controllable/uncontrollable load over one day (24 hours)

If the total load at time t equals 0.7 p.u. (where 1 p.u. refers to the maximum monthly load at the aggregation point), and the estimated load shares of controllable and uncontrollable load are 0.3 p.u. and 0.4 p.u., respectively, then the real values of the shares are within the following ranges:

$$P_C = 0.3 \, p.u. \pm AWFE, \tag{6}$$
$$P_{UC} = 0.4 \, p.u. \pm AWFE. \tag{7}$$

Here, $AWFE$ refers to the absolute value of the WFE. In each time step, the WFE for controllable load has the same absolute value as the uncontrollable load, but with the opposite sign, as the sum of these two shares always gives the same total load for that time step.

Figures 4-10 present the probability density function (PDF) of the WFE and cumulative distribution function (CDF) of the AWFE for individual load categories. Figures compare the distribution of errors based on the SM coverage level, including the base case (100% SM coverage). The 90th percentile of the errors (read from the CDF plot) is chosen for

comparison, as it shows the maximum error in 90% of the observed time steps (here, 1440 time steps over a 24h period).

As seen in Fig. 4, shares of the CTIM1 category are estimated with high accuracy at 80% SM coverage level, with 90th percentile of AWFE around 0.06 p.u. At the same time, the accuracy at 50% and 20% coverage is almost the same, with 90th percentile AWFE around 0.11 p.u. Looking at the PDF plot, it can be seen that the most probable WFE for 80% and 50% coverage is zero, same as the base case. At 20% coverage level, the WFE distribution is bimodal, having peaks around 0 and -0.08 p.u.
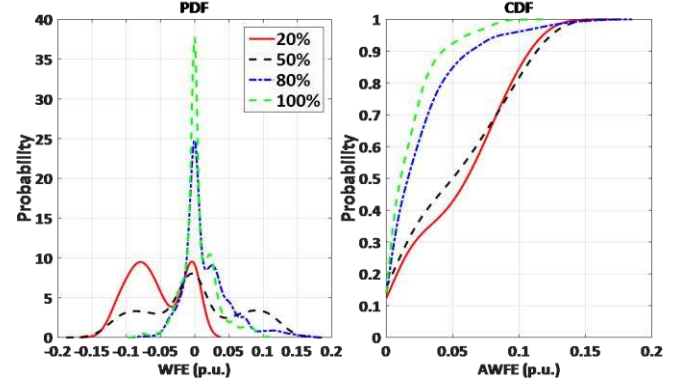

Figure 4. WFE and AWFE of the estimated share of the CTIM1 category (e.g. dish washer, washing machine, clothes dryer, etc.)

Estimation of CTIM3 shares shows very high accuracy, as seen in Fig. 5. The 90th percentile of AWFE is smaller than 0.05 p.u. for all the SM coverage levels. The most probable WFE is different from zero only for 20% SM coverage, still being very small, around -0.02 p.u.
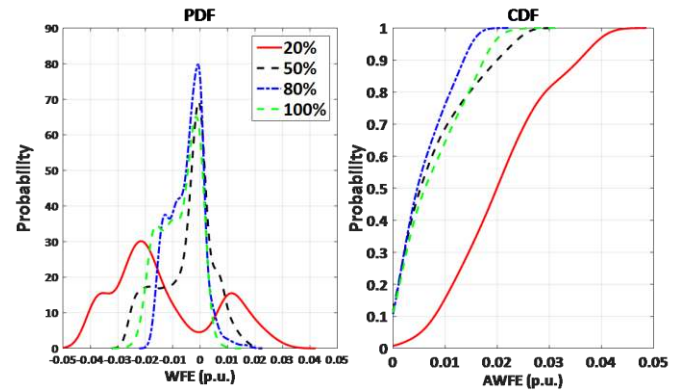

Figure 5. WFE and AWFE of the estimated share of the CTIM3 category (e.g. electrical space heating)

Cold appliances' shares in the overall load can also be very accurately estimated even at 20% SM coverage, where the 90th percentile of AWFE is around 0.06 p.u. (Fig. 6). The most

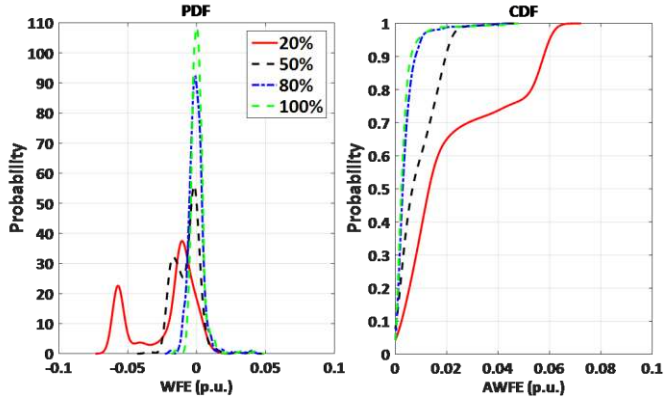probable WFE in all cases is between 0 and -0.01 p.u.

Figure 6.   WFE and AWFE of the estimated share of the QTIM1 category
(e.g. cold appliances)

Controllable resistive loads are estimated with visibly lower accuracy at SM coverage levels under 80%. As seen in Fig. 7, the 90th percentile of AWFE is around 0.25 p.u. at 50% SM coverage, and 0.65 p.u. at 20% SM coverage. Thus this category needs higher SM coverage for confident estimation/prediction of its shares during the day. Uncontrollable resistive loads' contribution to the total load can be estimated with higher accuracy, having 90th percentile of the AWFE less than 0.05 p.u. at 80% coverage and around 0.1 p.u. at 50 and 20% coverage, as seen in Fig. 8.

Electronic devices' shares can be assessed with relatively high accuracy at 80% and 50% SM coverage, while at 20% coverage the 90th percentile of the error is over 0.15 p.u. (Fig. 9). Regarding the Lighting category (Fig. 10), its share can be estimated with 90th percentile of error up to 0.1 p.u. for SM coverage of 50% and higher. At 20% coverage, majority of errors is up to 0.27 p.u.
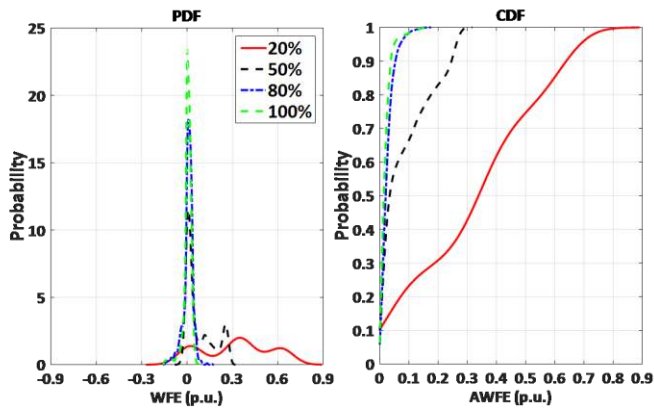

Figure 7.   WFE and AWFE of the estimated share of the R_C category (e.g. water heater, storage heater, etc.)
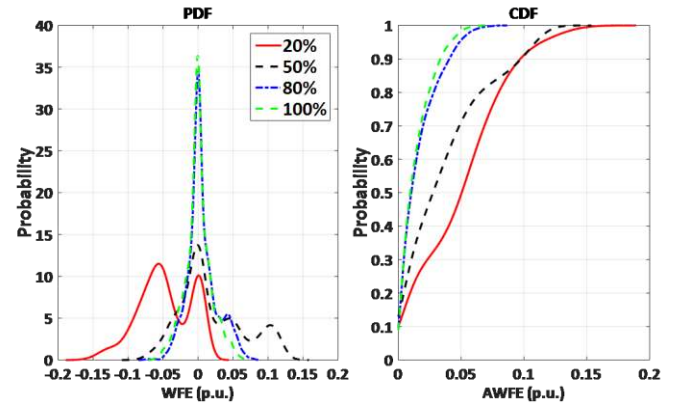

Figure 8.   WFE and AWFE of the estimated share of the R_{UC} category (e.g. hob, oven, iron, etc.)
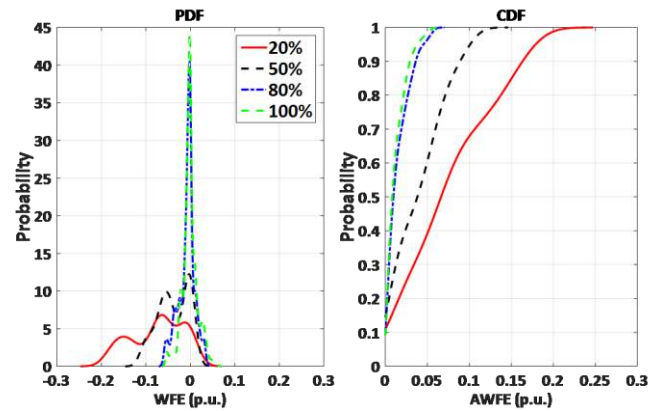

Figure 9.   WFE and AWFE of the estimated share of the SMPS category (e.g. electronic devices)

Finally, looking at the distribution of errors for controllable load in Fig. 11, the 90th percentile of AWFE is around 0.05 p.u. at 80% SM coverage, and around 0.15 p.u. at 50% coverage. The majority of errors is over 0.5 p.u. at 20% coverage, presenting very low accuracy.

Table II presents the 90th percentile of the confidence level for AWFE for different load categories and controllable load, over a range of SM coverage levels. Highlighted cells are those presenting AWFE not higher than 0.1 p.u., reflecting high accuracy. As seen from the accuracy analysis, there are load categories whose share can be estimated with high confidence even with lower levels of SM coverage (e.g. 20%) – these are CTIM3, QTIM1 and R_{UC}. At the same time, SMPS and Lighting require 50% coverage for confident share estimation, while CTIM1 and R_C need 80% coverage.
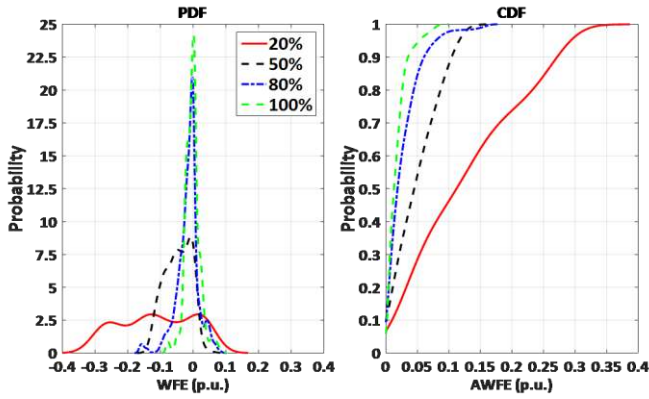
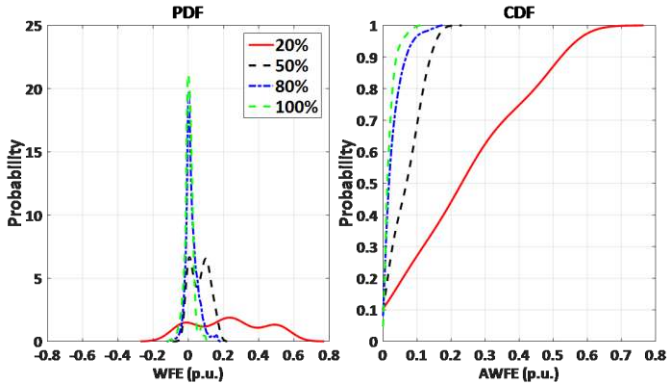Figure 10. WFE and AWFE of the estimated share of the Lighting category

Figure 11. WFE and AWFE of the estimated share of the controllable load

It can be concluded that for the DR programs relying on cold appliances (QTIM1 category) or induction motor-based space heating appliances (CTIM3), even lower SM coverage (20%) allows very accurate estimation of their share and flexibility potential using this methodology. Load flexibility estimation for DR programs based on washing/drying machines and similar appliances (CTIM1) can also be fairly accurate with 20% and 50% SM coverage levels. As for the voltage-based DR, relying on resistive loads (water heaters, for example), 80% SM coverage is required for R_C, which is why in this case there are more SMs needed to be installed for more confident prediction/estimation of the DR potential.

VII. CONCLUSION

This paper presented methodology for load decomposition based on SM data and ANN. Results showed that the level of SM coverage influences the accuracy of load decomposition only with respect to some load categories. The methodology proposed in this paper can be straightforwardly extended to network buses with mixed load sectors, i.e. residential, commercial and industrial. Also, obtaining the optimal aggregation level and data granularity, as means of saving expenditures for expanding data servers and communication line capacity, will be one of the objectives of future research.

Further studies should also reveal what is the required accuracy of load decomposition for different types of system studies.

TABLE II.   AWFE WITH 90TH PERCENTILE CONFIDENCE LEVEL

Load category	SM coverage level			
	100%	80%	50%	20%
CTIM1	0.04	0.06	0.12	0.11
CTIM3	0.02	0.01	0.02	0.04
QTIM1	0.01	0.01	0.02	0.06
R_C	0.04	0.05	0.25	0.64
R_{UC}	0.04	0.04	0.10	0.10
SMPS	0.03	0.04	0.09	0.16
Lighting	0.03	0.06	0.10	0.27
Controllable	0.04	0.07	0.13	0.52

REFERENCES

[1]   E. Carpaneto and G. Chicco, "Probabilistic characterisation of the aggregated residential load patterns," *Generation, Transmission & Distribution, IET,* vol. 2, pp. 373-382, 2008.

[2]   J. Z. Kolter and M. J. Johnson, "REDD: A public data set for energy disaggregation research," in *Workshop on Data Mining Applications in Sustainability (SIGKDD), San Diego, CA,* 2011, pp. 59-62.

[3]   I. Richardson, M. Thomson, D. Infield, and C. Clifford, "Domestic electricity use: A high-resolution energy demand model," *Energy and Buildings,* vol. 42, pp. 1878-1887, 2010.

[4]   M. Pipattanasomporn, M. Kuzlu, S. Rahman, and Y. Teklu, "Load Profiles of Selected Major Household Appliances and Their Demand Response Opportunities," *IEEE Transactions on Smart Grid,* vol. 5, pp. 742-750, 2014.

[5]   Y. Xu and J. V. Milanović, "Day-Ahead Prediction and Shaping of Dynamic Response of Demand at Bulk Supply Points," *IEEE Transactions on Power Systems,* vol. 31, pp. 3100-3108, 2016.

[6]   D. S. Kirschen, A. Rosso, M. Juan, and L. F. Ochoa, "Flexibility from the demand side," in *Power and Energy Society General Meeting, 2012 IEEE*, 2012, pp. 1-6.

[7]   *Smart Energy GB, [Online].Available:* https://www.smartenergygb.org/en/faqs.

[8]   *Smart meter data - A guide to your rights and choices, [Online].Available:* http://www.energy-uk.org.uk/policy/smart-meters.html.

[9]   K. Samarakoon, J. Ekanayake, and N. Jenkins, "Reporting Available Demand Response," *Smart Grid, IEEE Transactions on,* vol. 4, pp. 1842-1851, 2013.

[10]   Y. Xu and J. V. Milanović, "Accuracy of ANN based methodology for load composition forecasting at bulk supply buses," in *Probabilistic Methods Applied to Power Systems (PMAPS), 2014 International Conference on*, 2014, pp. 1-6.

[11]   F. Lamberti, D. Cuicai, V. Calderaro, and L. F. Ochoa, "Estimating the load response to voltage changes at UK primary substations," in *IEEE PES ISGT Europe 2013*, 2013, pp. 1-5.

[12]   "D3.4 Smart meters architecture and data model analysis ", NOBEL GRID project, [Online].Available: http://nobelgrid.eu/deliverables/ , 2016.

[13]   X. Chen, C. Kang, X. Tong, Q. Xia, and J. Yang, "Improving the Accuracy of Bus Load Forecasting by a Two-Stage Bad Data Identification Method," *IEEE Transactions on Power Systems,* vol. 29, pp. 1634-1641, 2014.