World Academy of Science, Engineering and Technology
International Journal of Energy and Power Engineering
Vol:17, No:12, 2023

Two-Level Identification of HVAC Consumers for Demand Response Potential Estimation Based on Setpoint Change

M. Naserian, M. Jooshaki, M. Fotuhi-Firuzabad, M. Hossein Mohammadi Sanjani, A. Oraee

***Abstract*—**In recent years, the development of communication infrastructure and smart meters have facilitated the utilization of demand-side resources which can enhance stability and economic efficiency of power systems. Direct load control programs can play an important role in the utilization of demand-side resources in the residential sector. However, investments required for installing control equipment can be a limiting factor in the development of such demand response programs. Thus, selection of consumers with higher potentials is crucial to the success of a direct load control program. Heating, ventilation, and air conditioning (HVAC) systems, which due to the heat capacity of buildings feature relatively high flexibility, make up a major part of household consumption. Considering that the consumption of HVAC systems depends highly on the ambient temperature and bearing in mind the high investments required for control systems enabling direct load control demand response programs, in this paper, a solution is presented to uncover consumers with high air conditioner demand among a large number of consumers and to measure the demand response potential of such consumers. This can pave the way for estimating the investments needed for the implementation of direct load control programs for residential HVAC systems and for estimating the demand response potentials in a distribution system. In doing so, we first cluster consumers into several groups based on the correlation coefficients between hourly consumption data and hourly temperature data using K-means algorithm. Then, by applying a recent algorithm to the hourly consumption and temperature data, consumers with high air conditioner consumption are identified. Finally, demand response potential of such consumers is estimated based on the equivalent desired temperature setpoint changes.

***Keywords*—**Data-driven analysis, demand response, direct load control, HVAC system.

I. Introduction

THE use of renewable energy has been on the rise due to the increase in demand for electricity consumption. At the other hand increasing electrical energy consumption [1], the proliferation of renewable energy sources aiming at elevating energy security and reducing the production of greenhouse gases [2], and maintaining the balance between consumption and production of electric power in the modern power grids [3] are among the main challenges facing power system planners and operators. Addressing these issues calls for the provision of higher levels of flexibility for which demand response plays a key role [4]. Demand response refers to the adjustments that electricity consumers make to their normal usage patterns based on the fluctuations in the electricity price over time, or based on the incentive payments that encourage them to reduce their electricity consumption when the wholesale market prices are high or the system is at risk [5].

Direct load control belongs to a category of demand response known as incentive-based programs. In a direct load control program, customers' equipment is directly controlled by the utility. When the demand response program is invoked, each equipment is either turned off or its time of operation is slightly shifted to fall in an adjacent period of lower demand. Direct load control is typically used for small commercial and residential customers. In exchange for taking the control of customers' equipment, utilities grant incentive payments or credits to customers. Also, utilities often provide customers with several options such as the ability to cancel the program and to limit its annual frequency and duration of execution. In general, direct load control programs are relatively simple, reliable, and popular among residential customers [5].

Direct load control programs can exploit the flexibility of household equipment, thereby increasing the overall flexibility of a power system [6]. HVAC systems are high consumption residential equipment featuring high flexibility potentials thanks to the heat capacity of buildings, and thus, being regarded as popular equipment for direct load control programs [7]. Numerous research works have focused on proposing mathematical models for flexile equipment such as HVAC systems to investigate their demand response potentials. In this respect, a model for HVAC systems is proposed in [8] to investigate the impact of different temperature dead band for HVAC system in direct load control program. The authors in [9] proposed a model for various flexible loads such as residential HVAC systems under incentive-based demand response programs considering customers' comfort to evaluate the demand response potentials. One of the significant challenges in adopting these mathematical-modeling-based methods is that the aggregator or system operator must have sufficient information about several parameters of the customers' equipment. The computational cost of solving the

Mehdi Naserian is with Sharif University of Technology, Tehran, Iran.
Mohammad Hossein Mohammadi Sanjani is with CITCEA-UPC, Polytechnic University of Catalonia, Barcelona, Spain (e-mail: mohammad.hossein.mohammadi@upc.edu).
Ashknaz Oraee is with University of Cambridge, Cambridge, United Kingdom, (IEEE Senior).
Mohammad Jooshaki is with Geologian Tutkimuskeskus, Espoo, Finland, (IEEE Senior).
Mahmud Fotuhi-Firuzabad is with Sharif University of Technology, Tehran, Iran, (IEEE Fellow).

models for a large set of consumers can be regarded as another limiting factor.

In view of the challenges mentioned and the difficulties in extracting some of the parameters or achieving accurate outcomes when simplified models are used, such techniques may not be applicable for large number of customers.

Accordingly, many studies leveraged data-driven approaches to detect customers with high temperature-dependent power consumption and extracting HVAC consumption based on weather conditions in order to evaluate demand response potential of the air conditioning systems. In this respect, the authors in [10] proposed an unsupervised load analysis method for large-scale customers by assessing seasonal demand variations to divide the total energy consumption into the amount of HVAC consumption and baseline load. Then, based on the thermal dynamic model of the HVAC system, an approach to estimate parameters of the HVAC system and demand response potential is presented. The authors in [11] proposed a method where the total consumption of the household and the outdoor temperature are used as input data to disentangle HVAC consumption from baseline load with the goal of measuring expected reduction in HVAC consumption in response to temperature setpoint changes. Authors in [12] proposed a method for using consumption pattern of every customer to identify those with high potentials for adjusting the set point of HVAC when participated in demand response programs. This method also divides the total consumption into baseline load and HVAC consumption. In [13], regression-based techniques are utilized to detect consumers with the highest temperature dependent consumption pattern.

While the previous studies have provided valuable insights into the techniques for separating baseline load from the HVAC system consumption and identifying high consumption customers, such research works either use data screening methods to partition customers or do not provide an appropriate algorithm for separating the baseline load from the HVAC consumption. In addition, the methods utilized for identifying high consumption customers are typically computationally expensive, thereby not applicable for a large number of customers. Some studies only focused on the separation of baseline load from the HVAC system consumption, and did not address the estimation of the demand response potential of the customers. Motivated by these points, this paper proposes a two-stage method for detecting consumers with high HVAC consumption among large number of customers. The outcomes will then be employed to estimate demand response potential of the consumers.

II. Data Analysis

In this paper, data from the Low Carbon London Project have been used [14]. This dataset includes electricity consumption for a set of 5,567 households in London that participated in the project between November 2011 and February 2014. The customers' consumption data have been collected at half-hourly intervals using smart meters. There are two groups of customers in this dataset: 1) a subset of approximately 1100 customers who were exposed to time-varying energy prices in 2013, and 2) the rest of the customers being exposed to a fixed rate tariff (14.228 pence/kWh). For the first group, prices had three levels, namely high (67.20 pence/kWh), medium (11.76 pence/kWh), and low (3.99 pence/kWh).

Based upon the initial preprocessing of the dataset, the consumption data from January 1, 2013 to the end of December 2013 were selected as the information was found to be more consistent. Even the selected period includes missing values during some time intervals which are eliminated in the analyses. Considering that the data collected by the smart meters are available with higher resolutions compared to the accessible weather data, two methods can be considered for adapting the time scales:

1) Sampling the consumption data at the frequency of weather data points, taking into account the higher resolution of the consumption data, results in losing some information. However, this approach can maintain the accuracy as no synthetic data are generated. In addition, as the size of the dataset is reduced, the computational complexities decrease.
2) Curve fitting is used in weather data to fill in missing measurements in the sampling frequency of consumption data. This method might reduce the accuracy of the model considering the injection of synthetic data, yet avoids losing high-resolution consumption data.

In this research, the first method is used to harmonize the time scale of different datasets in order to maintain the model accuracy.

HVAC systems can be operated in cooling and heating modes. Depending on the regional weather condition, the use of cooling and heating systems and the utilization ratio of each mode are different. Therefore, in order to determine whether the customers of a region use cooling, heating, or both, firstly, the aggregated profile of the hourly consumption of the customers (half-hourly sampled consumption data) is obtained. Then, the aggregated consumption of customers is plotted as a function of ambient temperature and a piece-wise linear function is fitted to the data as can be seen in Fig. 1. According to this figure, as the temperature decreases, the consumption increases. Therefore, based on the ambient temperature range and the relation between the power consumption and temperature, it can be implied that in the case under study, heating is the dominant mode of HVAC operation.

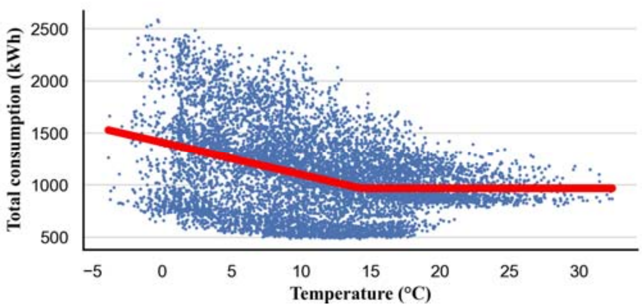

Fig. 1 Aggregated consumption data of customers versus outdoor temperature, and the piece-wise linear curve fitted on the data

III. Methodology

As concluded from the data pre-processing, the consumption and weather data show the notable use of heating systems. However, it can be inferred from Fig. 1 that many customers do not use electric heating systems, or the use of such systems by these customers is negligible. One of the major challenges for a load aggregator is to find customers with high HVAC consumption in order to select the customers with high flexibility, thus, of potential interest for direct load control demand response programs. In this regard, at the first stage, the correlation (r) between hourly consumption and temperature during a year is calculated separately for each customer according to (1), which represents the correlation coefficient r between two variables x and y [15]. In this equation, $\bar{x}$ and $\bar{y}$ respectively indicate the mean values of x and y.

$$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2}} \quad (1)$$

Then, K-means clustering method [16] is applied to the obtained correlation coefficients to divide customers into different groups. The objective function of this clustering algorithm is to minimize the summation of the square errors between the members of each cluster and the associated cluster center, which is expressed in (2). In this equation, K is the number of clusters, C_k is the set of customers in cluster k, $L_{m,d}$ represents the set of features for customer m, and $c_{k.d}$ indicates the center of cluster k, and d represents number of features.

$$minimize\ OF = \sum_{k=1}^{K}\sum_{m\in C_k}\left\|L_{m,d} - c_{k,d}\right\|^2 \quad (2)$$

In order to choose the best value for K, two indices, namely inertia and average Silhouette score, are used in this paper. The inertia expresses the summation of the square values of the distances between data points and their associated cluster centers, thus, the lower the value of the inertia parameter the more coherent the clusters. Equation (3) expresses the Silhouette scores for each point and their average value.

$$S_i = \begin{cases} 1-\frac{a_i}{b_i} & a_i < b_i \\ 0 & a_i = b_i \\ \frac{b_i}{a_i}-1 & a_i > b_i \end{cases} \rightarrow S_{avg} = \frac{1}{N}\sum_{i=1}^{N} S_i \quad (3)$$

where, N represents number of data points, a_i represents the average distance between data point i and all the other data points being in the same cluster, and b_i indicates the minimum value of the average distance between data point i and all the data points falling in a cluster to which data point i does not belong. The average value of the Silhouette index is between -1 and 1, and the closer this number is to 1, the better the data clustering is.

As consumers with high HVAC demand are identified in the first stage, an algorithm is presented in the second stage to separate baseline load and HVAC consumption for each of the customers identified in the first stage. This is to evaluate HVAC demand response potential based on set point changes. The algorithm comprises of the following steps:

1. The break point for all customers is considered to be 15 °C because based on the data analyzed before, the use of heating system at temperatures above 15 °C is unlikely. Also, if it is in use, the system consumption and therefore the system flexibility are relatively negligible. Thus, the customer's hourly consumption for temperatures above 15 °C are eliminated.
2. The hourly consumption data for ambient temperatures below 15 °C are divided into two categories based on the 70 percentile of consumption, in such a way that consumption data greater than the 70 percentile are assumed to be related to HVAC consumption and consumption data less than the 70 percentile are considered as the baseline load.
3. A line is fitted for HVAC consumption and another one with zero slope is fitted for baseline load according to (4). In this equation, P_h represents the customer's consumption in timestep h, a represents the slope of the line fitted for the HVAC consumption, b represents the intercept of the line fitted for HVAC, and c represents the baseline load.

$$\begin{cases} P_h = aT_h + b & HVAC \\ P_h = c & Baseline \end{cases};\forall h \in \{1,\ldots,24\} \quad (4)$$

4. After fitting the above lines, the data are re-categorized based on the line in between the two lines in such a way that the hourly consumption data above this line are related to HVAC consumption and the hourly consumption data below this line are related to the baseline load. The function of this line is according to (5):

$$P_h = \frac{a}{2}T_h + \frac{b+c}{2};\forall h \in \{1,\ldots,24\} \quad (5)$$

5. Step 3 and step 4 are repeated until the results do not change significantly.

As the algorithm is applied to all customers, those with negative slope a are chosen as those with highest HVAC consumption. The more negative the value of a, the more flexible the consumption of the HVAC system and the greater the customer’s demand dependence on ambient temperature.

The reason for using a two-stage technique to identify the customers with high HVAC consumption is that the algorithm of the second stage is significantly slower than that of the first stage and it is substantially time-consuming for a large number of customers.

In order to measure the demand response potential of a HVAC system, it is assumed that a shift in the setpoint of the HVAC system is equivalent to a similar change in the ambient air temperature with an identical consumer comfort. This simplified assumption ignores the effects of the heat capacity of buildings, and therefore the demand response potential obtained from this simplified method is lower than the actual case, because due to the heat capacity of a building, it takes a longer time for the air temperature inside the house to be equal to the outdoor air temperature. In this paper, a 4-degree decrease in the set point is considered for estimating the demand response potential of a HVAC system. This set point change, as explained before, is equivalent to a 4-degree increase in the outdoor air temperature. This amount of set point change is considered as it does not significantly endanger the thermal comfort of the customers. The demand response potential of each customer is calculated separately according to (6), where a_i represents the slope of the line fitted for the HVAC system, and T is the outdoor air temperature during the demand response program implementation hour. It is worth mentioning that the break point $X0_i$ is considered 15 °C in our proposed method.

$$DR_i = min\big(4a_i . a_i(X0_i - T)\big) \qquad T \leq X0_i \tag{6}$$

IV. Study Results

The proposed method is implemented for 5,526 residential customers participated in the Low Carbon London Project. In the first stage, correlation coefficient between a customer's hourly consumption and hourly outdoor temperature for one-year data is calculated for all customers. The coefficient is then used as the customer's feature to cluster customers into different groups. Running K-means clustering algorithm with different K values results in the inertia indices presented in Fig. 2. According to this figure, the higher the number of groups or the value of K, the lower the inertia. To decide the best value of K, we need the Silhouette index as well. The average silhouette indices for different K values are provided in Table I. According to this table, the best number of groups is 2, nonetheless, because the inertia index is high for this K value. Considering both inertia and Silhouette indices the number of groups is selected to be 6 for which the average Silhouette index is the second highest and the inertia index is low.

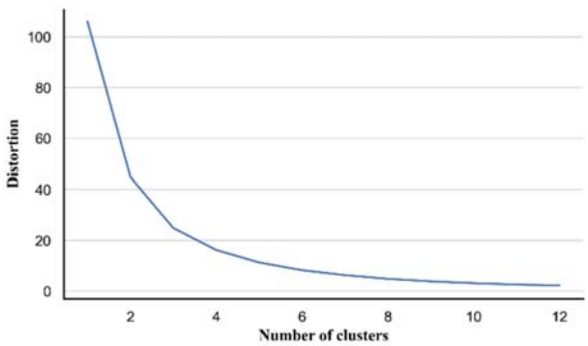

Fig. 2 Inertia index for different number of clusters

TABLE I
Average Silhouette Index for Different Number of Groups

Different number of groups	Average silhouette index
2	0.6226
3	0.5204
4	0.5166
5	0.5173
6	0.5259
7	0.5155
8	0.5220

Accordingly, the K-means algorithm divides customers into six groups. Values of the Silhouette index for all members of the groups are illustrated in Fig. 3. In this figure, the average value of the Silhouette index is shown with a dashed red line and the distribution of silhouette indices for each group is depicted with distinct colors. The numbers of customers in each group are given in Table II. As the next step, all the customers in each group are aggregated, i.e., summation of hourly consumption of the customers is determined. Finally, for visualization, the algorithm proposed in Stage 2 is applied, which results in the graphs depicted in Fig. 4. According to this figure, Group 2 is the high-HVAC-consumption group, Group 1, Group 3, and Group 4 are among the groups with medium HVAC consumption and Group 5 and Group 6 do not noticeably use the HVAC system. After the approximate selection of customers in the first stage, groups of interest based on the temperature dependence of the demand can be chosen for the next stage. We choose Group 2, which is the group with the highest demand-temperature dependence. In Stage 2, the algorithm proposed in this stage is applied to hourly consumption and hourly temperature data for each customer. Then, customers with negative slope a are chosen as consumers with high HVAC consumption. For these customers, the estimated value for demand response potential of the HVAC systems based on a 4-degree decrease in the temperature setpoint are given in Table III. As can be seen in this table, average demand response potential of the members in this group is about 113 Watt.

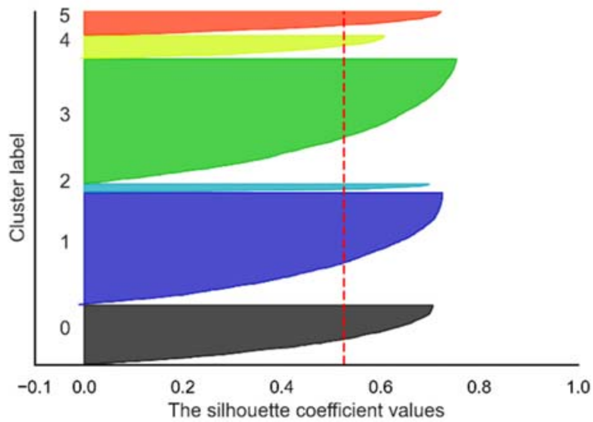

Fig. 3 Distribution of Silhouette index for all clusters

TABLE II
Number of Members in Each Group

Groups	Number of members
1	937
2	128
3	386
4	1738
5	1980
6	357

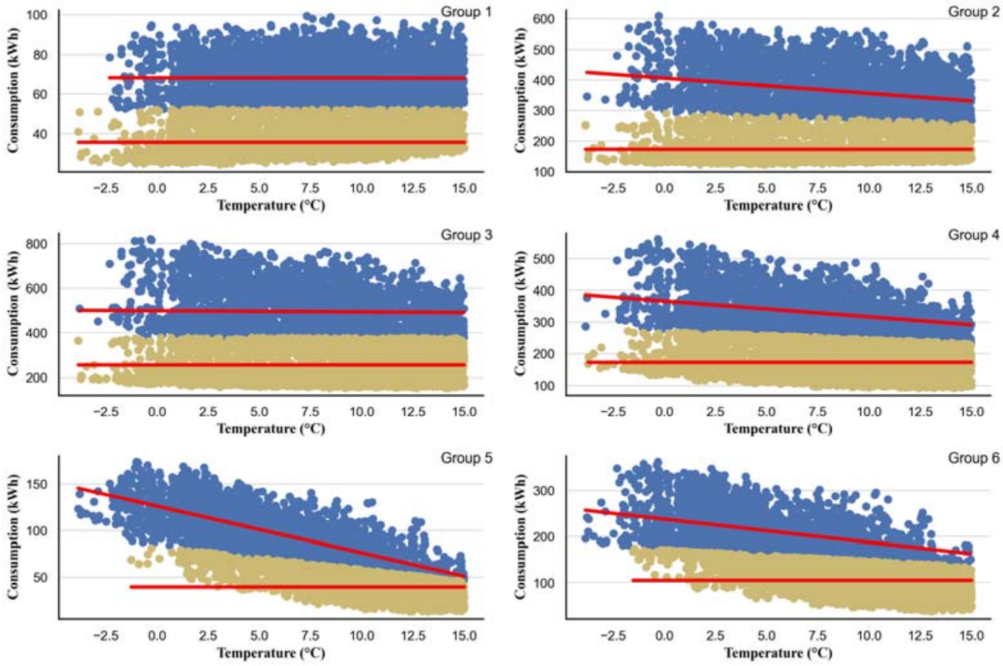

Fig. 4 Proposed algorithm in Stage 2 for aggregated customers in each group

TABLE III
STATISTICAL DATA FOR THE DEMAND RESPONSE POTENTIAL OF THE MEMBERS IN GROUP 6

	Demand response potential (W)
Mean	113.07
Std	103.816
Min	0.052
25 percentile	35.805
50 percentile	84.63
75 percentile	157.17
Max	573.29

V. CONCLUSIONS

Considering high costs of controlling equipment in direct load control programs on one hand, and considerable flexibility of HVAC systems on the other hand, in this paper, a two-stage method was proposed to identify consumers with high HVAC demand among large number of customers. In the first stage, consumers are clustered based on the correlation coefficient between hourly consumption and hourly temperature data extracted for each consumer. Then, in the second stage, by choosing consumers with high-temperature-dependent demand as identified in the first stage, an algorithm is proposed to detect consumers with excessive HVAC consumptions. In this algorithm, baseline load and HVAC consumption are separated. Finally, demand response potential of the HVAC system is measured based on the setpoint change. The reason for using two stages to separate customers is that the algorithm of the second stage is much slower than that of the first stage and it is specifically time-consuming for a large number of customers. The numerical study showed the applicability of the proposed algorithm for identifying consumers with high HVAC demand and separating HVAC consumption from baseline load in order to estimate the demand response potential.

ACKNOWLEDGMENT

The third author would like to acknowledge financial support from the Iran National Science Foundation (INSF).

REFERENCES

[1] Lee, J., et al., “Improvements to the customer baseline load (CBL) using standard energy consumption considering energy efficiency and demand response.” Energy, 2018. 144: p. 1052-1063.
[2] Lacommare, K.S., “Review of the Recent Frequency Performance of the Eastern, Western and ERCOT Interconnections.” 2010.
[3] Paterakis, N.G., O. Erdinç, and J.P. Catalão, “An overview of Demand Response: Key-elements and international experience.” Renewable and Sustainable Energy Reviews, 2017. 69: p. 871-891.
[4] Adams, J., M. O’Malley, and K. Hanson, “Flexibility requirements and potential metrics for variable generation: Implications for system planning studies.” Princeton, NJ: NERC, 2010: p. 14-17.
[5] Qdr, Q., “Benefits of demand response in electricity markets and recommendations for achieving them.” US Dept. Energy, Washington, DC, USA, Tech. Rep, 2006.
[6] Haider, H.T., O.H. See, and W. Elmenreich, “A review of residential demand response of smart grid.” Renewable and Sustainable Energy Reviews, 2016. 59: p. 166-178.
[7] Hu, M., F. Xiao, and L. Wang, “Investigation of demand response potentials of residential air conditioners in smart grids using grey-box room thermal model.” Applied energy, 2017. 207: p. 324-335.
[8] Ali, M., A. Safdarian, and M. Lehtonen. “Demand response potential of residential HVAC loads considering users preferences.” in IEEE PES innovative smart grid technologies, Europe. 2014. IEEE.
[9] Ni, Z. and A. Das, “A new incentive-based optimization scheme for residential community with financial trade-offs.” IEEE Access, 2018. 6: p. 57802-57813.
[10] Qi, N., et al., “Smart meter data-driven evaluation of operational demand response potential of residential air conditioning loads.” Applied Energy, 2020. 279: p. 115708.
[11] Dyson, M.E., et al., “Using smart meter data to estimate demand response potential, with application to solar energy integration.” Energy Policy, 2014. 73: p. 607-619.
[12] Jazaeri, J., T. Alpcan, and R.L. Gordon. “Customer selection for residential demand response with thermostatically controlled loads.” in 2019 7th International Istanbul Smart Grids and Cities Congress and Fair (ICSG). 2019. IEEE.

[13] Grabner, M., A. Souvent, and N. Suljanovič. "Evaluating Temperature-Dependent Consumers in a Demand Response Program using Machine Learning." in 2020 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe). 2020. IEEE.

[14] Data Available from: https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households.

[15] Wu, W.-J. and Y. Xu. "Correlation analysis of visual verbs' subcategorization based on Pearson's correlation coefficient." in 2010 International Conference on Machine Learning and Cybernetics. 2010. IEEE.

[16] Singh, A., A. Yadav, and A. Rana. "K-means with Three different Distance Metrics." International Journal of Computer Applications, 2013. 67(10).