An empirical analysis of domestic electricity load profiles: Who consumes how much and when?

Trotta, Gianluca

Published in:
Applied Energy

DOI (link to publication from Publisher):
10.1016/j.apenergy.2020.115399

Publication date:
2020

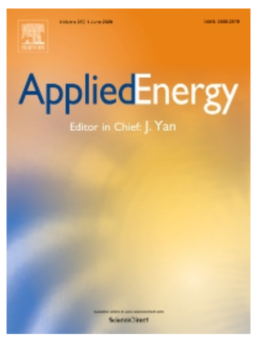
Document Version
Accepted author manuscript, peer reviewed version

Citation for published version (APA):
Trotta, G. (2020). An empirical analysis of domestic electricity load profiles: Who consumes how much and when? *Applied Energy*, *275*(1 October), 1-14. Article 115399. https://doi.org/10.1016/j.apenergy.2020.115399

An empirical analysis of domestic electricity load profiles: Who consumes how much and when?

First and corresponding author:

Gianluca Trotta
Aalborg University Copenhagen
The Faculty of Engineering and Science
Department of the Built Environment
Address: A.C. Meyers Vænge 15, 2450 Copenhagen - Denmark
E-mail: gianluca.trotta@seri.de



Please cite as follows:

Trotta, G. (2020). An empirical analysis of domestic electricity load profiles: Who consumes how much and when? Applied Energy, 275, 115399. DOI: 10.1016/j.apenergy.2020.115399.

An empirical analysis of domestic electricity load profiles: Who consumes how much and when?

Abstract

With the increased share of renewables in power generation, end users play a key role in keeping the demand at levels that better match variable supply, maintaining lower overall system costs, and reducing carbon dioxide emissions. To increase the potential for demand-side flexibility, a deeper understanding of domestic electricity load profiles is needed. Informed by customer grouping based on similar consumption patterns and drivers, targeted interventions can be better designed to time-shift peak loads and reduce overall demand. Thus, the objectives of this study are (i) to segment households in relation to their electricity load patterns using K-means clustering and (ii) to investigate household characteristics that have an influence on electricity load patterns by employing multinomial probit regression. This study uses hourly electricity consumption for 2017, combined with population-based register data for a large sample of Danish households. The results indicate that four distinct Danish household groups are characterized by different timing and magnitudes of electricity consumption, which are influenced by specific sociodemographics and dwelling characteristics. Similarities between the groups emerge with respect to the evening peak consumption, seasonal variation in electricity demand, and weekend morning demand ramp-up. Challenges and opportunities for domestic load profiling in the power industry and policymaking are discussed.

Keywords: domestic electricity load profiles; hourly electricity consumption; household segmentation; cluster analysis; multinomial probit regression; Denmark.

1. Introduction

As the share of intermittent renewables in the supply mix grows, and more electric vehicles and electric heat pumps hit the market, demand-side flexibility is needed to aid system balancing. To make demand more flexible and to reduce peak pressure on the grid, a deeper understanding of the daily domestic electricity consumption patterns is required [1], [2], [3]. The widespread smart electric metering deployment with two-way communication capabilities has made a wealth of information available that can reveal significant differences about the way domestic consumers use electricity, thus accommodating targeted demand-side response strategies, energy efficiency programs, and energy-saving recommendations and improving the individual or aggregate load forecast accuracy and the reliability of the power grid [4], [5], [6].

Clustering is the core data-mining technique used in residential smart-meter data for discovering homogeneous household groups sharing a similar magnitude and timing of electricity demand. When groups of electricity load patterns are combined with information about household influences on specific electricity demand patterns, valuable insight can be derived for utilities and policymakers. In the last decade, clustering techniques have been employed to segment domestic electricity consumption (using data from smart meters) in Finland [7], California [8], [9], Ireland [3], [6], [10], [11], Taiwan [5], Texas [12], Portugal [13], [14], [15], France [16], Japan [17], Spain [18], UK [19], [20], Denmark [21], and Illinois [22]. Despite being very valuable, most of these studies are limited by the sample size, time period, and geographical scale of the analysis and, when investigated, by the information about the factors influencing electricity load patterns.

Against this background, using a large dataset, this study aims to cluster Danish households with a similar magnitude and timing of electricity demand and to investigate the sociodemographics and dwelling characteristics influencing electricity demand. Unlike previous studies, this article uses hourly electricity consumption data combined with population-based registered data. The dataset well represents the entire population of Denmark and offers a unique opportunity to segment Danish households in relation to their hourly electricity load patterns using K-means clustering [23] and to investigate the determinants of electricity demand among different household groups by applying a multinomial probit regression. The results can support utilities and policymakers in the design of tailored demand-side management and energy efficiency strategies for load shifting and domestic electricity demand (and associated carbon dioxide emission) reduction.

Denmark is an interesting case study to analyze because the domestic electricity consumption and share of variable renewable energy in electricity production are expected to increase in the next few years [24]. In 2017, domestic electricity consumption accounted for 31.6% of the total electricity consumption, and 70.6% of the total electricity production was generated by renewable energies, including wind (67.4%), biomass (25.9%), solar (3.4%), biogas (3.1%), and hydro (0.1%) [25], [26]. By 2030, 100% of electricity will be produced by renewable energy [24]. According to the Danish Energy Agency [26], the number and consumption of domestic electrical appliances are expected to increase by 2.3% and 0.3% annually from 2017 to 2030, respectively. In addition, the growth of electric vehicle sales, which is expected to account for at least 22% of the total sales of new cars in 2030, and the electrification of the heating sector, combined with a larger heated floor area, will further stress the grid. To cope with this and other energy and emission mitigation challenges, on June 26, 2019, under the agreement "A Fair Direction for Denmark" [27], the new government committed to reducing greenhouse gas (GHG) emissions by 70% by 2030, relative to 1990 levels (the previous target was 40%). Meeting the Danish government's climate target will imply significant changes in domestic electric usage to keep the demand at levels that better match the higher share of renewable energy in electricity generation.

The paper is structured as follows. Section 2 briefly reviews the literature on domestic electricity consumption clustering. Section 3 presents the data used for the Danish case study. Section 4 describes the methodology employed in the study. The results are presented in Section 5 and are discussed in Section 6. Section 7 provides concluding remarks and future research directions.

2. Literature Review

Driven by the widespread application in many areas of science that involve the extraction of similar groups of objects or individuals from data, several clustering techniques have been proposed in the literature [23], [28], [29], [30], [31], [32]. Generally, the choice of the clustering algorithm depends on the nature of the research and available data. Clustering methods include K-means, K-medoids, hierarchical, self-organized maps, or other clustering approaches, with K-means being one of the most widely used due to its stability, efficiency, and ability to work well with a large dataset [3], [23], [33].

In the analysis of smart electrical data, several studies have classified customers in relation to the similar volume and timing of electricity consumption [5], [6], [8], [9], [13], [17], [21], [22], [29], while a few others, in addition to the different electricity load profiles, have investigated their underlying patterns [3], [10], [11], [12], [14], [15]. Identifying the main drivers of specific groups of electricity load profiles is crucial for the design of any tailored demand-side strategy and for creating an incentive to save energy. This valuable knowledge for utilities and policymakers regarding who consumes how much energy and when they consume it can only be derived using the combination of smart electrical data with other data containing information about households.

For example, Rhodes et al. [12] investigated the temporal variation using K-means clustering in electricity use for 103 homes in Austin (Texas) and assessed the correlations between household and dwelling characteristics (collected through surveys), identifying clusters through binary probit regression. Moreover, working at home, the number of hours of television watched per week, and education were significant determinants of inclusion in a given cluster. In Ireland, McLoughlin et al. [11] clustered 3,941 households over six months using K-means, K medoids, and self-organizing maps. The electricity load profile classes were then linked to information on the dwelling, occupant, and appliance characteristics to determine the likelihood of a customer with specific characteristics belonging to a particular profile class through multinomial logistic regression. Similarly, Viegas et al. [3] employed K-means clustering to data from 4,232 Irish households and used multinomial logistic regression to investigate the main drivers of electricity consumption using survey data. In Portugal, Gouveia and Seixas [14] and Gouveia et al. [15] combined a smart-meter cluster analysis with door-to-door surveys covering information about socioeconomic, dwelling, and appliance characteristics for a sample of 265 and comprising data on 19 households in the Évora municipality, respectively, using hierarchical clustering. Gouveia and Seixas [14] primarily distinguished between fuel poverty, standard comfort, and "fat energy" households, while Gouveia et al. [15] evaluated the influence of different external air temperature thresholds on daily electricity consumption profiles.

The contribution of this article to the literature is three-fold. First, this is the first study that uses hourly electricity consumption data to segment a large sample of Danish households. Second, unlike previous studies that used survey data to investigate the drivers of electricity consumption, this study combines hourly electricity consumption data with information on households and dwellings taken from administrative register data. Administrative records have more abundant, detailed, and reliable information than self-reported surveys and are not subject to sample, response, and non-response biases. Third, to better inform utilities and policymakers on domestic electric use during different periods of the year, the electricity load profiles of each identified cluster are broken down into seasons, weekdays, and weekends.

3. Data

The original sample size consisted of 19,734 households with 191,716,879 hourly observations of electricity consumption data covering the period from January 1, 2017, to December 31, 2017. Several pre-processing steps including data cleaning, data integration, data selection, and data transformation were applied. At first, about 16% of the observations consisting in 15-minute resolution were aggregated on an hourly basis for comparability with other consumption data. The hourly electricity consumption data were then combined with information regarding dwelling and sociodemographic characteristics of the households (e.g., income, number of people, type of dwelling, household size, etc.) provided by Statistics Denmark. The hourly electricity consumption data are collected by Energinet, which is an independent public enterprise owned by the Danish Ministry of Climate, Energy and Utilities[1], and then sent to Statistics Denmark.

The link between the electricity consumption data and the rich information from the administrative records was possible because all people living in Denmark are assigned a unique personal identification number ("Det Centrale Person Register": CPR), which is used by public authorities to store specific personal information, such as income, family composition, etc. Access to microdata can be granted to researchers and analysts from Danish research environments according to the "need-to-know" principle, which implies that researchers and analysts can access the data required for a specified purpose. The data are confidential and are accessed through a secure server at Statistics Denmark [34], [35]. The combination of the electricity consumption data with the population-based register data allowed to remove from the sample households living in a farmhouse that, compared to other families, typically have a significantly higher consumption due to, for example, dairy production. The final sample obtained is composed of 15,433 households, and the number of hourly observations of electricity consumption is 134,271,332 encompassing the period from January 1, 2017, to December 31, 2017, and accounting for data-entry errors (missing period indicating the decimal place) and missing information. Since there was nothing systematic going on that made some data more likely to be missing than others ("Missing Completely at Random": MCAR), the few missing hourly consumption data were ignored, and no imputation was required [36].

The sample descriptive statistics compared to the whole population are provided in Table 1.

Table 1. Descriptive statistics.

	Sample					Population
Variables	**N**	**Mean**	**Std Dev**	**Min**	**Max**	-
Hourly electricity consumption (kWh) (January 1, 2017 – December 31, 2017)	134,271,332	0.36	0.45	0	3.109	-
Normalized hourly electricity consumption (kWh) (January 1, 2017 – December 31, 2017)	134,271,332	0.12	0.14	0	1	-
Variables	**N**	**Mean (%)**	**Std Dev**	**Min**	**Max**	**Mean (%)**

[1] https://en.energinet.dk/

Building year[2]						
Before 1963	6,698	43.5%	0.5	0	1	46.6%
1963-1972	2,657	17.3%	0.38	0	1	17.3%
1973-1978	1,623	10.5%	0.31	0	1	9.6%
1979-1998	2,664	17.3%	0.38	0	1	15.2%
1999-2006	921	6%	0.24	0	1	5.5%
After 2006	827	5.4%	0.26	0	1	5.7%
Dwelling type						
Single-family detached house	8,583	56.9%	0.5	0	1	42.9%
Terraced, linked or semi-detached house	2,585	17.1%	0.38	0	1	15.9%
Multi-dwelling house	3,932	26%	0.44	0	1	41.2%
Floor area (m^2)						
≤ 75 m^2	3,156	20.5%	0.4	0	1	26.8%
76 m^2 – 95 m^2	2,744	17.8%	0.38	0	1	19.8%
96 m^2 – 120 m^2	2,787	18.1%	0.39	0	1	18%
121 m^2 – 150 m^2	2,935	19.1%	0.39	0	1	16.6%
≥ 151	3,764	24.5%	0.43	0	1	18.8%
Dwelling tenure						
Owned	8,680	57.8%	0.49	0	1	48.9%
Rented	6,346	42.2%	0.49	0	1	51.1%
Heating system						
Heating system powered by natural gas, liquid/solid fuels, or straw	13,759	89.3%	0.31	0	1	93.4%
Heating system powered by electricity	1,652	10.7%	0.31	0	1	6.6%
Government office region						
Region of Southern Denmark	7,673	49.8%	0.5	0	1	21.2%
Central Denmark Region	1,209	7.8%	0.27	0	1	22%
North Denmark Region	808	5.2%	0.22	0	1	10.2%
Zealand Region	5,557	36%	0.48	0	1	14.2%
Capital Region of Denmark	164	1%	0.1	0	1	32.3%
Household size						
1 member	6,077	40.3%	0.49	0	1	43%
2 members	5,342	35.4%	0.48	0	1	33.2%
3 members	1,455	9.7%	0.3	0	1	10.1%
4 members	1,575	10.4%	0.31	0	1	10%
5 members or more	627	4.2%	0.2	0	1	3.7%
Household disposable income						
Lowest 20% (0 – 182,325 DKK)	2,781	20%	0.39	0	1	20.2%
Quintile 2 (182,330 – 264,629 DKK)	2,781	20%	0.39	0	1	20.1%
Quintile 3 (264,650 – 380,071 DKK)	2,781	20%	0.39	0	1	19.5%
Quintile 4 (380,073 – 554,405 DKK)	2,781	20%	0.39	0	1	18.5%
Highest 20% (554,531 – 8,024,383 DKK)	2,781	20%	0.39	0	1	20.9%
Head of household[3] working status						
Outside workforce	6,621	46.3%	0.49	0	1	46.6%
Worker (skilled/unskilled)	3,850	26.9%	0.42	0	1	23.4%
Self-employed	437	3%	0.16	0	1	3%
Routine nonmanual	1,657	11.6%	0.32	0	1	11.7%

[2] The "building year" categorization follows the Danish Building Regulation and the energy labels typically linked to the year of construction. https://sparenergi.dk/

[3] "Head of household" is the member of the household who has the highest individual income.

Manager	1,739	12.2%	0.12	0	1	15.3%
Head of household educational level						
Primary	4,657	28.4%	0.47	0	1	25.2%
Secondary	5,878	44.6%	0.5	0	1	42.3%
Tertiary	3,082	26.9%	0.42	0	1	32.5%

Compared to the whole population, the sample is not representative in terms of regional distribution. Households living in the capital region of Denmark are underrepresented due to the delayed roll out of smart meters in the municipality of Copenhagen compared to other areas of Denmark. In addition, owner-occupied, large single-family detached houses, and dwellings equipped with a heating system powered by electricity are slightly overrepresented. All the other sociodemographic and dwelling variables represent the overall Danish household characteristics well.

4. Methodology

The method of K-means clustering was used to segment households in relation to their hourly electricity load patterns. The goal of K-means clustering, which is one of the most popular unsupervised machine learning algorithms [23], [33], [37], is to partition *n* data points into homogeneous *k* clusters in such a way that households exhibiting similar magnitudes and time of use of electricity demand are grouped. In particular, K-means clustering minimizes the sum of the squared error over all clusters as follows:

$$J = \sum_{k=1}^{K} \sum_{x_i \in c_k} ||x_i - \mu_k||^2 \tag{1}$$

where μ_k is the mean value of the points assigned to the cluster c_k and $||x_i - \mu_k||^2$ is the standard Euclidian distance function between the data point x_i and its closest centroid μ_k. A smaller *J* results in more similar within-group data. To improve the accuracy and efficiency of the K-means clustering algorithm and to reduce the differences in consumption volume across households while retaining the consumption patterns, the electricity consumption data were scaled to fit in the range of (0, 1) by using minimum-maximum normalization [38], [39].

The main advantage of K-means lies in its speed of convergence, relative ease of implementation, and efficiency. In addition, unlike other clustering techniques, such as hierarchical clustering, K-median, and k-medoids, which require quadratic computation and are therefore restricted to small datasets, K-means clustering works for linear boundaries and is particularly suitable for large datasets. Despite its desirable properties, some caveats associated with the K-means method exist, such as the selection of the correct number of clusters, and random initialization of the algorithm (resulting in locally optimal solutions). To estimate the optimum number of clusters, in addition to the scree plot of the curve of the within sum of squares (*WSS*) and its logarithm form (*log WSS*) according to the number of clusters *k*, the η^2 coefficient and the proportional reduction of error (*PRE*) coefficient was used, as proposed by Makles [40], as follows:

$$\eta_k^2 = 1 - \frac{WSS(K)}{WSS(1)} = 1 - \frac{WSS(K)}{TSS} \ \forall k \in K \tag{2}$$

$$PRE_k = \frac{WSS(k-1) - WSS(k)}{WSS(k-1)} \ \forall k \geq 2 \tag{3}$$

where *WSS(K) [WSS (k-1)]* is the *WSS* for cluster solution *k* (*k-1*), and *WSS(1)* is the *WSS* for cluster solution *k=1* (for the nonclustered data), and η_k^2 and PRE_k measure the proportional reduction of the *WSS* for cluster solution *k* compared to the total sum of squares (*TSS*) and the previous solution with *k-1* clusters, respectively [40]. To find a better local minimum, 20 cluster solutions, *k=1,…, 20*, with 50 multiple random starting points were performed, and the one with the smallest *WSS* was selected.

For comparison purposes only, Ward's agglomerative hierarchical clustering method [41] was also employed to segment households in relation to the similar volume and timing of electricity consumption. Agglomerative hierarchical clustering methods start with each point in a cluster of its own, and successively merge pairs of clusters until there is only one cluster. Unlike K-means, no random initial conditions are imposed [42], [43], [44]. The distance between the new cluster and the rest of the clusters is determined by the linkage method. Ward's method, also known as incremental sum of squares method or Ward's minimum variance method, attempts to maximize the between-cluster distances and minimize the within-cluster distances [42], [43], [44], [45]. The results of Ward's method representing the clustering structure of the data were displayed as a dendrogram. In addition to the Ward's method dendrogram, the Calinski and Harabasz Index was used for choosing the optimal number of clusters [46]. This criterion combines the within and between cluster sum of squares to evaluate the cluster validity [47]; the larger the value of this index (pseudo-F statistic) the more distinct the clusters. Compared to K-means, the Ward's method does not typically scale well to large datasets and tends to produce more even-sized clusters [42], [43], [44], [45].

In the second part of the analysis, a multinomial probit regression was employed to investigate the probability of a household with specific characteristics, controlling for geographical and climate areas, of having a specific electricity load profile (in terms of magnitude and time of use).

The dependent variable y is categorical and unordered and represents different clusters or electricity load profiles *k=1,…, K*. The multinomial probit regression is an extension of the binomial probit regression that allows for a dependent variable with more than two categories. For each observation, only one of the clusters is non-zero. By belonging to a specific cluster k, the household derives the "utility level" of $U_{i,k}$, which depends on a set of observable sociodemographic and dwelling characteristic variables $V_{i,k}$ such as income, number of people, type of dwelling, and so on and unobservable factors denoted by $\varepsilon_{i,k}$.

$$U_{i,k} = V_{i,k} + \varepsilon_{i,k}, \qquad k = 1, \dots K, \qquad \varepsilon \sim N[0, \Sigma] \tag{4}$$

where $\varepsilon_{i,k}$ is normally distributed. Therefore, the probability of a household belonging to a specific cluster *k* over alternatives *K* can be expressed as follows:

$$Pr_{i,k} = Pr\left(U_{i,k} > U_{i,K}\right), \qquad k = 1, \dots K, \qquad \forall k \neq K \tag{5}$$

$$= Pr\left(U_{i,k} - U_{i,K} > 0\right) = Pr(V_{i,k} - V_{i,K} > \varepsilon_{i,K} - \varepsilon_{i,k})$$

In particular, the likelihood of a household belonging to cluster *k = 1* over four clusters, for example, $Pr_{i,1}$, can be expressed as follows:

$$Pr_{i,1} = Pr\left(U_{i,1} > U_{i,2}, and\ U_{i,1} > U_{i,3}, and\ U_{i,1} > U_{i,4}\right) \tag{6}$$

$$= Pr(V_{i,1} - V_{i,2} > \varepsilon_{i,2} - \varepsilon_{i,1}, and\ V_{i,1} - V_{i,3} > \varepsilon_{i,3} - \varepsilon_{i,1}, and\ V_{i,1} - V_{i,4} > \varepsilon_{i,4} - \varepsilon_{i,1})$$

$$= Pr(\widetilde{\varepsilon_{i,21}} < \widetilde{V_{i,12}}, and\ \widetilde{\varepsilon_{i,31}} < \widetilde{V_{i,13}}, and\ \widetilde{\varepsilon_{i,41}} < \widetilde{V_{i,14}})$$

$$= \int_{-\infty}^{V_{i,12}} \int_{-\infty}^{V_{i,13}} \int_{-\infty}^{V_{i,14}} f_i\left(\widetilde{\varepsilon_{i,21}}, \widetilde{\varepsilon_{i,31}}, \widetilde{\varepsilon_{i,41}}\right) \widetilde{d\varepsilon_{i,21}}, \widetilde{d\varepsilon_{i,31}}, \widetilde{d\varepsilon_{i,41}}$$

where $\widetilde{\varepsilon_{i,21}} = \varepsilon_{i,2} - \varepsilon_{i,1}$; $\widetilde{\varepsilon_{i,31}} = \varepsilon_{i,3} - \varepsilon_{i,1}$; $\widetilde{\varepsilon_{i,41}} = \varepsilon_{i,4} - \varepsilon_{i,1}$; $\widetilde{V_{i,12}} = V_{i,1} - V_{i,2}$; $\widetilde{V_{i,13}} = V_{i,1} - V_{i,3}$; $\widetilde{V_{i,14}} = V_{i,1} - V_{i,4}$ and $f_i\left(\widetilde{\varepsilon_{i,21}}, \widetilde{\varepsilon_{i,31}}, \widetilde{\varepsilon_{i,41}}\right)$ is the joint probability density function of $\widetilde{\varepsilon_{i,21}}, \widetilde{\varepsilon_{i,31}}, \widetilde{\varepsilon_{i,41}}$.

To identify the parameters of the model, the normalization $\beta_0=0$ (reference category) is imposed for the group exhibiting an average electricity consumption closer to that of the entire sample. Given the non-linear nature of the multinomial probit model, to provide a direct interpretation of the coefficients, the average marginal effects (AMEs) are estimated [48]. All analyses were performed using STATA MP 15.1.

5. Results

5.1 Electricity Load Profiles

The results indicate clustering with k = 4 to be the optimal solution. At k = 4, a kink exists in both the WSS and *log(WSS)* figures (Figure 1A, 1B). η_4^2 points to a reduction in the WSS by 63%, and PRE_4 to a reduction of about 12% compared to the k = 3 solution PRE_4 (Appendix, Table 2); however, the reduction in WSS for k > 4 is negligible (Figure 1; Appendix, Table 2).

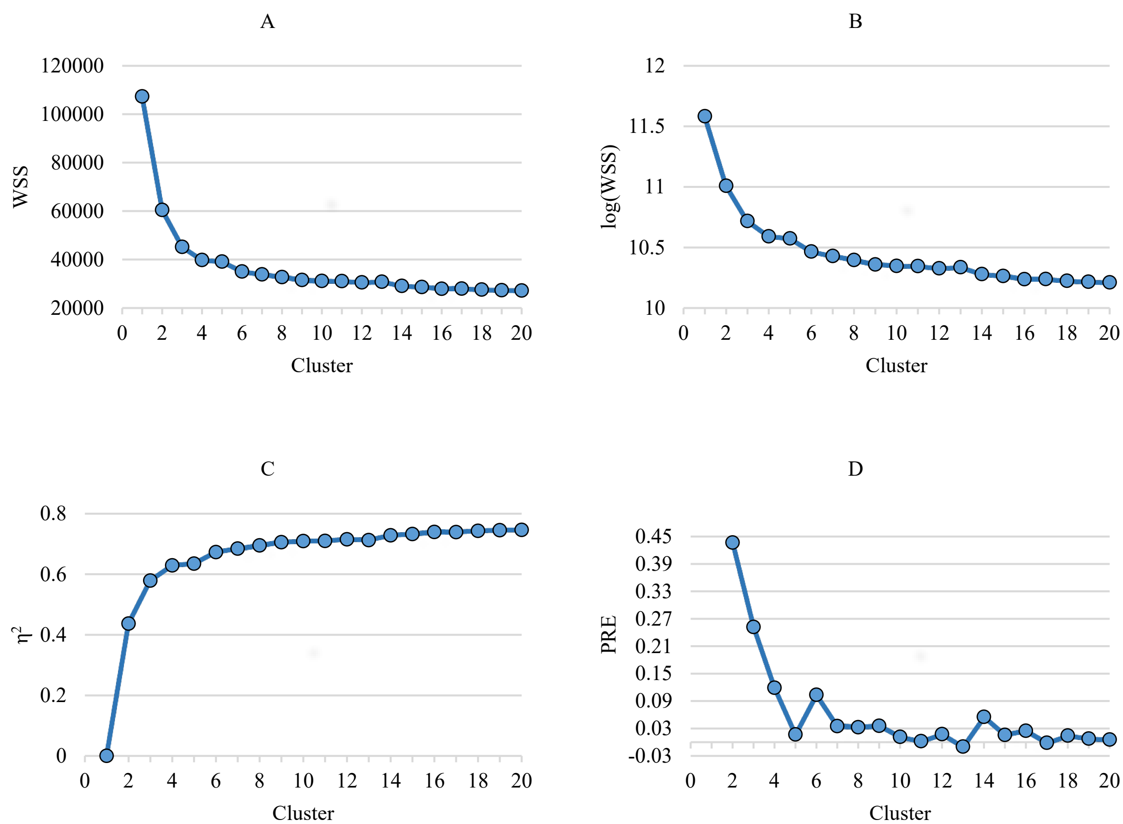


Figure 1. (A) *WSS,* (B) *log(WSS),* (C) η^2, and (D) *PRE* for all *K* cluster solutions.

Figure 2 illustrates the dendrogram obtained using Ward linkage criterion. Both the dendrogram and the Calinski and Harabasz Index (Appendix, Table 3) indicate the presence of three groups among the cluster solutions. Unlike K-means clustering, Ward's algorithm tends to merge clusters with a small number of observations and to produce relatively equally sized clusters [42], [43], [44], [45].

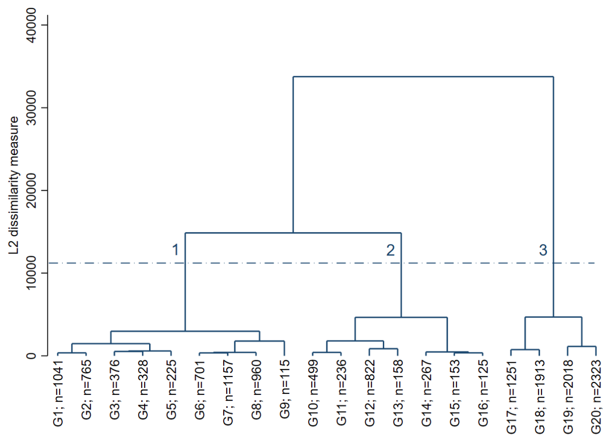


Figure 2. Cluster dendrogram using *Ward linkage* for all *K* cluster solutions.

Next, the results refer to K-means clustering, which is the preferred approach for this particular dataset. Figure 3 depicts the box and whisker plot, which represents the hourly electricity consumption distribution and skewness of the four clusters.

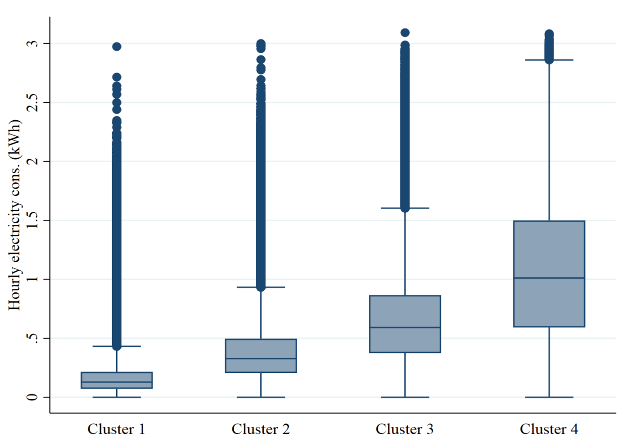


Figure 3. Box and whisker plot describing the hourly electricity consumption data (kWh) of four clusters.

The key values of the box and whisker plot are a five-number summary, which consists of the minimum value, first quartile (25th percentile), median (50th percentile), third quartile (75th percentile), and maximum value of the data. The median (50th percentile), which is the solid line in the middle of the boxes, represents the median value of the hourly electricity consumption (kWh) for each cluster. The lower and upper end of the boxes shows the 25th and 75th percentiles of the data (interquartile range), respectively. The whiskers are the two lines outside the boxes that represent data outside the interquartile range. At the end of the whiskers, the minimum and maximum values represent the lowest and highest data points (excluding outliers), respectively, and provide information about the dispersion of the data. Lastly, the circles at the end of the upper whisker represent the outliers (single data points). The median line of each cluster lies outside the other box, indicating four distinct groups. In addition, with the median slightly closer to the bottom of each box and with the whisker shorter on the lower end of the boxes, the distribution for each cluster is positively skewed. As the hourly electricity consumption increases from Clusters 1 to 4, the distribution increases. Cluster 1 demonstrates the lowest data variability (standard deviation of 0.13 kWh) followed by Clusters 2 (standard deviation of 0.24 kWh), 3 (standard deviation of 0.37 kWh), and 4 (standard deviation of 0.6 kWh).

The first cluster of households comprises 43% of the sample. The second, third, and fourth clusters represent 34.7%, 16.2%, and 5.7% of the sample, respectively. Figure 4 depicts the annual average electricity load patterns of the four clusters.

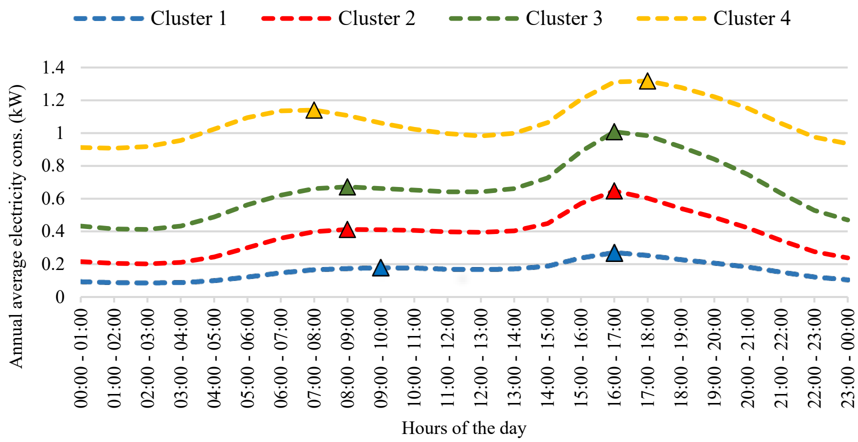


Figure 4. Annual average electricity load patterns (kW) of the four clusters.

Cluster 1 represents 6,667 households with a flat load pattern. For Cluster 1, the annual average baseload is 0.09 kW, the average peak load is 0.27 kW, and the average load is 0.16 kW. Cluster 2 contains 5,379 households. For Cluster 2, the annual average baseload is 0.2 kW, the average peak load is 0.65 kW, and the average load is 0.39 kW. Cluster 3, which represents 2,505 households, is characterized by relatively high electricity consumption. For Cluster 3, the annual average baseload is 0.41 kW, the average peak load is 1 kW, and the average load is 0.66 kW. Cluster 4 is composed of 882 households and reflects the heaviest users of electricity across the year. For Cluster 4, the annual average baseload is 0.9 kW, the average peak load is 1.3 kW, and the average load is 1.08 kW. The annual average daily electricity consumption varies considerably among clusters. Households belonging to Cluster 1 have an annual average daily electricity consumption of 3.9 kW, followed by Clusters 2 (9.3 kW), 3 (15.9 kW), and 4 (25.9 kW).

The triangles highlight the morning and evening peak hours. Cluster 1 is characterized by average morning peak hours between 9 and 10 and average evening peak hours between 16 and 17. Households belonging to Cluster 2 or 3 have average morning peak hours between 8 and 9 and average evening peak hours between 16 and 17, whereas households associated with Cluster 4 have average morning peak hours between 7 and 8 and average evening peak hours between 17 and 18. For all four clusters, the average evening peaks are higher than the average morning peaks. Figure 5 shows the seasonal average electricity load patterns of the four clusters.

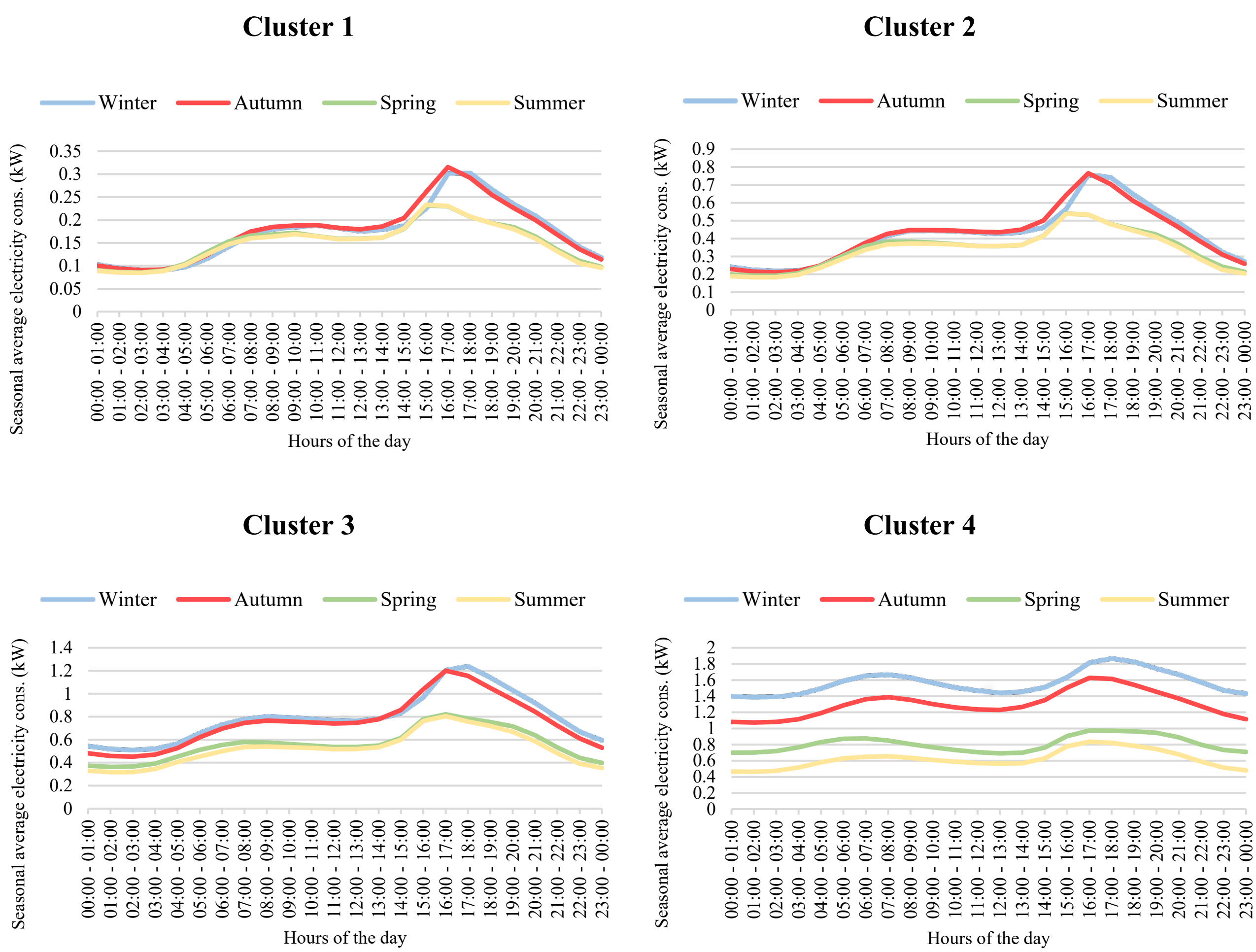
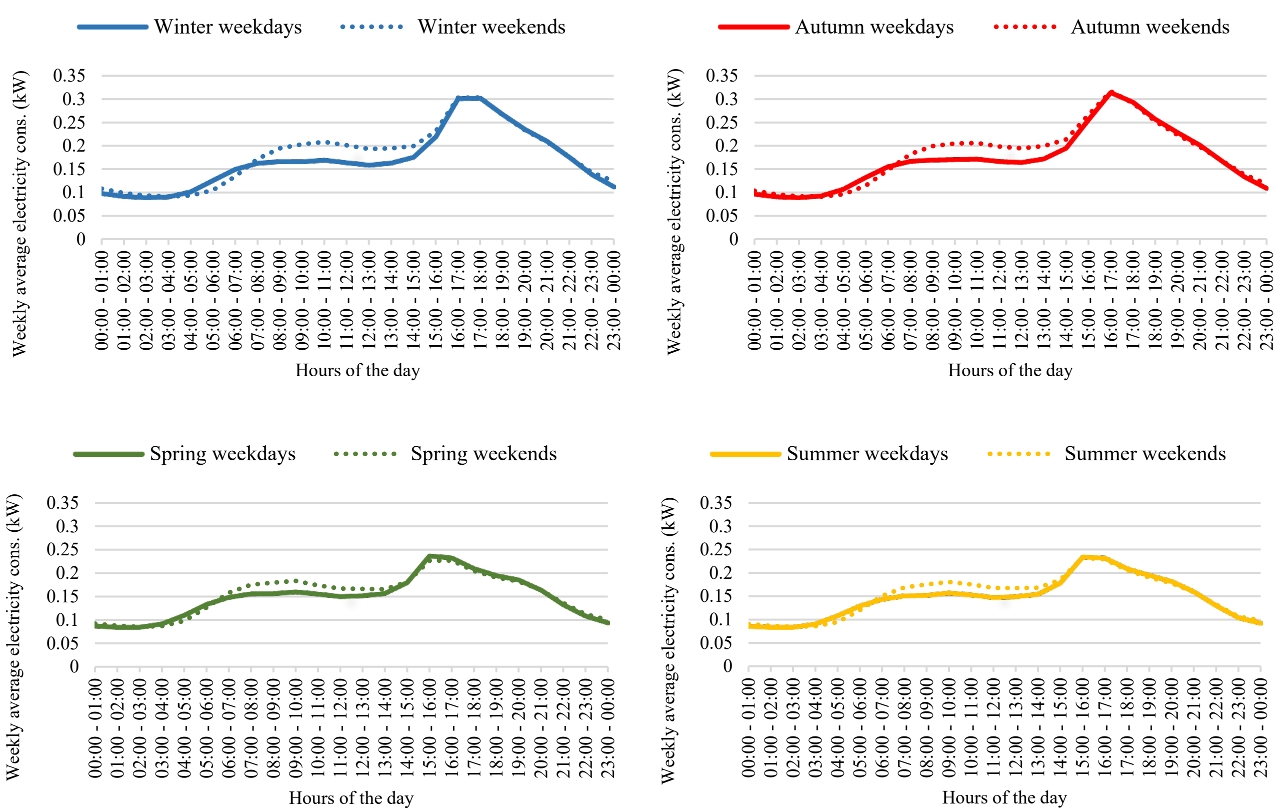


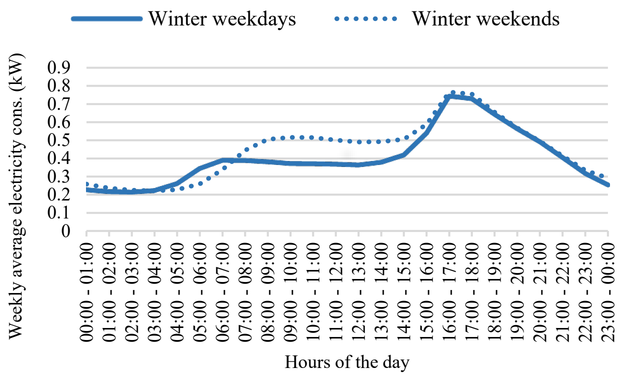
Figure 5. Seasonal average electricity load shapes (kW) of the four clusters.

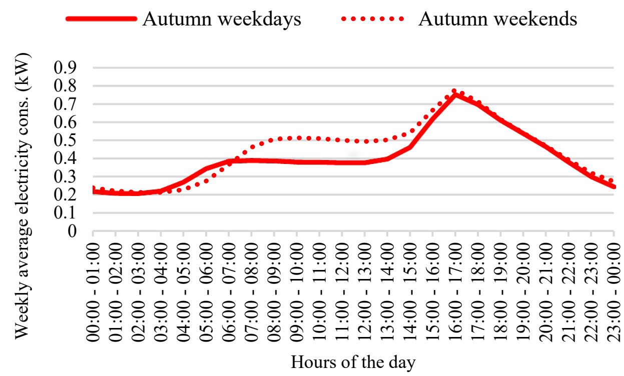
Common to the four clusters is the seasonal variation in electricity demand. In cold seasons, the likelihood that the household spends more time at home and, therefore, more often uses appliances increases. In addition, for those households with a heating system powered by electricity, the seasonal gap can primarily be attributed to the electricity demand for space heating. Generally, in both absolute and relative terms, the difference in electricity consumption within seasons becomes more extensive as the electricity consumption increases. For example, the average winter load of Cluster 1 is 16.4% higher than the average summer load, whereas the average winter load of Cluster 4 is 153.9% higher than the average summer load. Peak usage times vary slightly within season and cluster. Compared to winter demand, summer peak demand typically occurs earlier in the morning and earlier in the evening. Except for Cluster 4, the average morning and evening peak hours during spring and summer are similar. Compared to the other clusters, Cluster 4 exhibits earlier morning peak hours and later evening peak hours during summer and spring. Figure 6 compares the average electricity load patterns (kW) of the four clusters between weekdays and weekends during the four seasons.

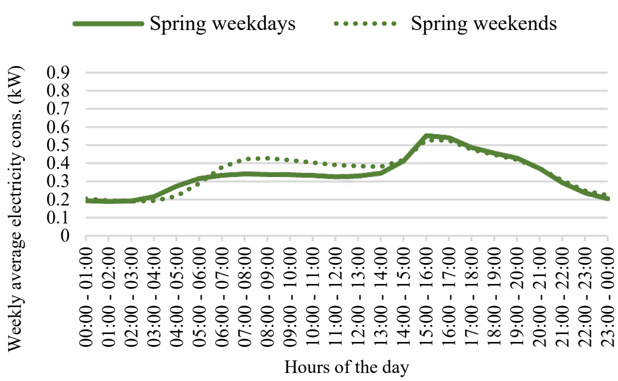
Cluster 1

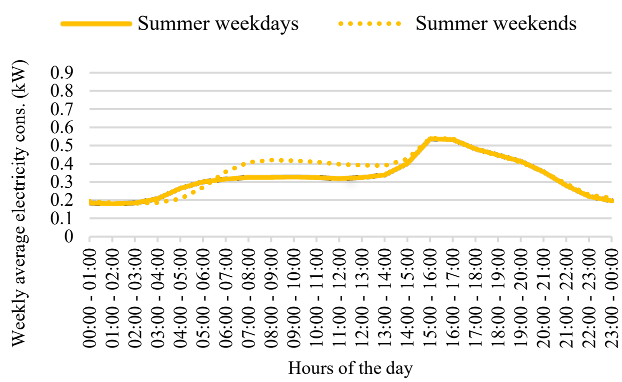


Cluster 2

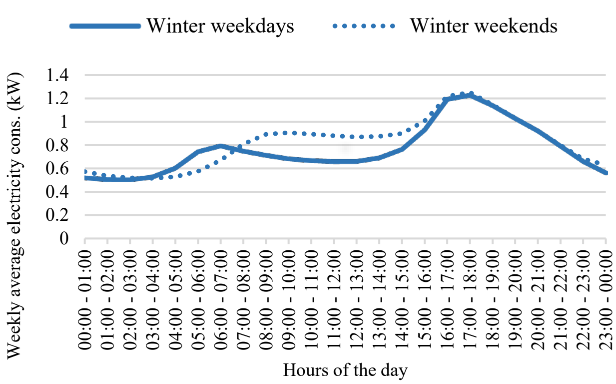


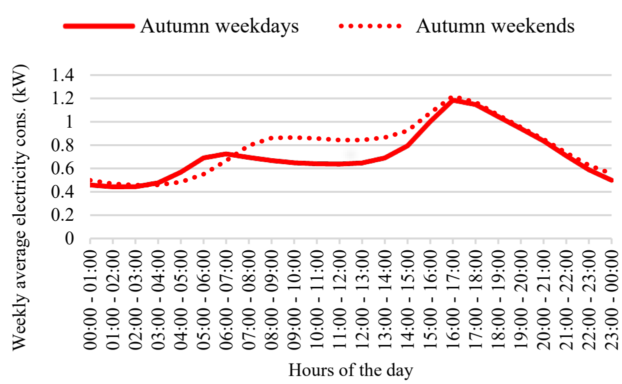


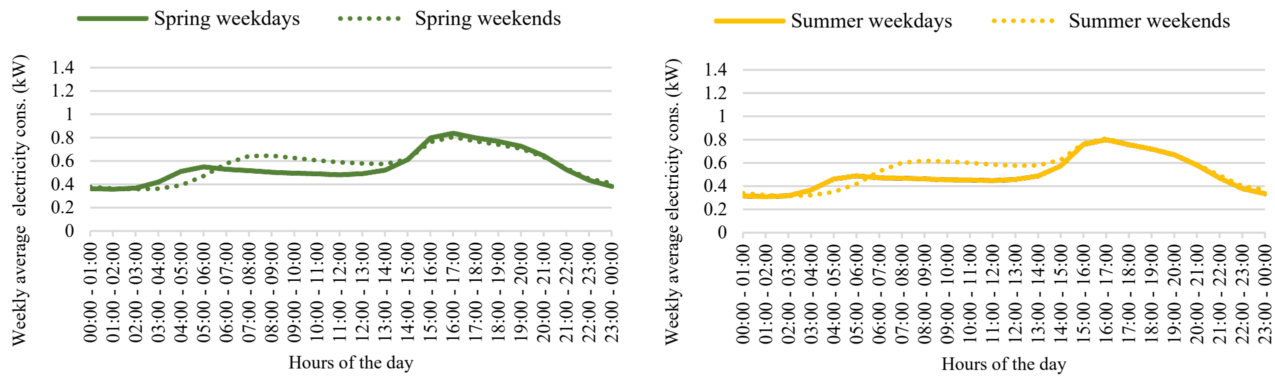




Cluster 3







Cluster 4

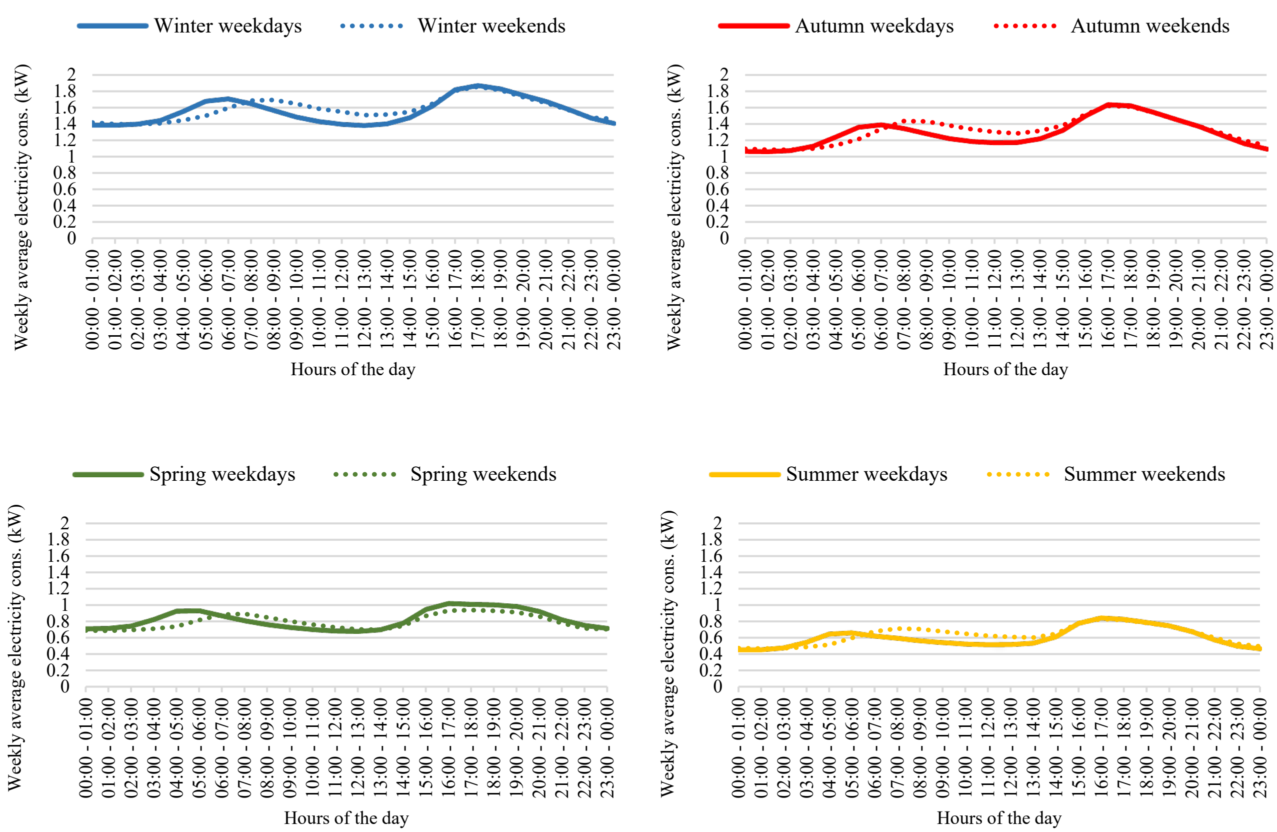


Figure 6. Weekly average electricity load patterns (kW) of the four clusters.

The electricity load shapes vary with respect to the days of the week, reflecting various (or the lack of) energy activities of the households. During weekdays, late morning and mid-day are often characterized by off-peak hours due to work activities, while during weekends, households can start the day later and stay home for electricity-consuming activities, such as cooking, laundry, and so on. The late-morning and mid-day gap of electricity consumption between weekdays and weekends is smaller in warm seasons and higher in cold seasons. Clearly, as the outdoor temperature increases, the likelihood to participate in outdoor activity increases. However, these well-defined patterns of electricity use with respect to weekdays and weekends differ among clusters. In relative terms, between weekdays and weekends, a larger late-morning and mid-day gap of electricity consumption exist in Cluster 3, followed by Clusters 2, 1, and 4. Nonetheless, for all four clusters, the weekday and weekend evening peak hours overlap.

5.2 Determinants of Household Electricity Use

Table 4 depicts the AMEs for multinomial probit estimations of the four clusters. The results show the main factors (sociodemographic and dwelling characteristics) influencing households grouped by different electricity load profiles (clusters). The multinomial probit estimations are represented in the appendix (Table 5). For dummy variables that indicate different categories of a single underlying variable, the AME indicates how the predicted probability of observing that a household belongs to a specific cluster *k* over alternatives *K* changes as the dummy variables change from 0 to 1 relative to the reference category [49].

Table 4. Average marginal effects (AMEs) of multinomial probit estimations for the four clusters.

Variables	Cluster 1 ($\beta_0=0$)	Cluster 2	Cluster 3	Cluster 4
	AMEs	AMEs	AMEs	AMEs
Building year (Ref = Before 1963)				
1963-1972	0.03** (0.01)	0.04*** (0.01)	-0.05*** (0.01)	-0.01 (0.00)
1973-1978	0.01 (0.01)	0.04*** (0.01)	-0.04*** (0.01)	-0.01* (0.00)
1979-1998	0.02 (0.01)	0.02 (0.01)	-0.02** (0.01)	-0.01 (0.00)
1999-2006	0.02 (0.01)	0.02 (0.02)	-0.04*** (0.01)	0.00 (0.01)
After 2006	0.03* (0.02)	0.02 (0.02)	-0.03** (0.01)	-0.02** (0.01)
Dwelling type (Ref = Single-family detached house)				
Terraced, linked or semi-detached house	0.09*** (0.01)	-0.02 (0.01)	-0.05*** (0.01)	-0.02* (0.01)
Multi-dwelling house	0.24*** (0.01)	-0.1*** (0.02)	-0.13*** (0.02)	-0.00 (0.01)
Floor area (Ref = ≤ 75 m^2)				

76 m^2 – 95 m^2	-0.07*** (0.01)	0.09*** (0.02)	-0.02 (0.02)	-0.00 (0.01)
96 m^2 – 120 m^2	-0.13*** (0.01)	0.14*** (0.02)	-0.03* (0.01)	0.02 (0.01)
121 m^2 – 150 m^2	-0.15*** (0.01)	0.13*** (0.02)	0.01 (0.02)	0.01 (0.01)
≥ 151	-0.21*** (0.01)	0.13*** (0.02)	0.05** (0.02)	0.03*** (0.01)
Dwelling tenure (Ref= Owned)				
Rented	0.02 (0.01)	0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)
Heating system (Ref = Heating system powered by natural gas, liquid/solid fuels, or straw)				
Heating system powered by electricity	-0.22*** (0.02)	-0.07*** (0.02)	0.16*** (0.01)	0.13*** (0.00)
Government office region (Ref = Region of Southern Denmark)				
Central Denmark Region	0.04** (0.01)	-0.05*** (0.02)	0.01 (0.01)	0.01 (0.01)
North Denmark Region	0.02 (0.01)	0.03 (0.02)	-0.04** (0.01)	-0.02 (0.01)
Zealand Region	-0.05*** (0.01)	0.01 (0.01)	0.04*** (0.01)	0.00 (0.00)
Capital Region of Denmark	-0.06 (0.04)	-0.04 (0.04)	0.08*** (0.02)	0.02 (0.01)
Household size (Ref =1 member)				
2 members	-0.16*** (0.01)	0.12*** (0.01)	0.03*** (0.01)	0.00 (0.01)
3 members	-0.22*** (0.01)	0.14*** (0.01)	0.09*** (0.01)	0.00 (0.01)
4 members	-0.29*** (0.01)	0.16*** (0.02)	0.12*** (0.01)	0.01 (0.01)
5 members or more	-0.33*** (0.02)	0.12*** (0.03)	0.19*** (0.01)	0.02* (0.01)
Household disposable income (Ref = Lowest 20%)				
Quintile 2	0.02 (0.01)	0.01 (0.01)	-0.02 (0.01)	-0.01 (0.01)
Quintile 3	0.01 (0.01)	0.03 (0.01)	-0.02* (0.01)	-0.01 (0.01)
Quintile 4	-0.03** (0.01)	0.06*** (0.01)	-0.02 (0.01)	-0.01 (0.01)
Highest 20%	-0.05** (0.02)	0.04** (0.02)	0.01 (0.01)	-0.00 (0.01)
Head of household working status (Ref = Outside workforce)				
Worker (skilled/unskilled)	-0.04*** (0.01)	-0.00 (0.01)	0.04*** (0.01)	0.01 (0.01)
Self-employed	-0.04 (0.02)	-0.02 (0.02)	0.04** (0.02)	0.02* (0.01)
Routine nonmanual	-0.02 (0.01)	-0.00 (0.01)	0.01 (0.01)	0.01 (0.01)
Manager	-0.05*** (0.01)	0.01 (0.01)	0.02* (0.01)	0.02** (0.01)
Head of household education level (Ref = Primary)				
Secondary	0.02* (0.01)	-0.00 (0.01)	-0.01 (0.01)	-0.01 (0.00)
Tertiary	0.03*** (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01* (0.01)
Number of observations	13,602	13,602	13,602	13,602

Robust standard errors in parentheses. * p<.05, ** p<.01, *** p<.001

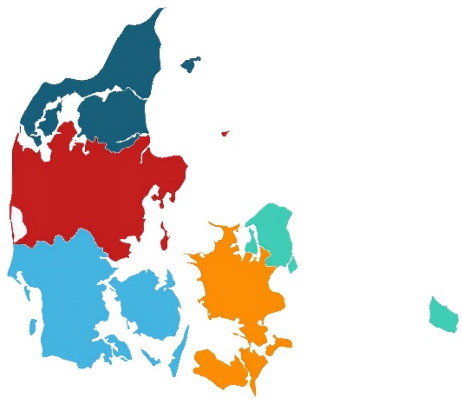
Households belonging to Clusters 1 and 2 are more likely to live in a relatively newer building, whereas the opposite is true for households belonging to Clusters 3 and 4. Compared to households living in a single-family detached house, households belonging to Cluster 1 are 9% more likely to live in a terraced, linked, or semi-detached house and are 24% more likely to live in a multi-dwelling house. In contrast, households belonging to Clusters 2 and 3 are 10% and 13% less likely to live in a multi-dwelling house, respectively, compared to households living in a single-family detached house. The larger the dwelling, the smaller the probability is that a household belongs to Cluster 1, whereas a reverse effect is observed in Cluster 2, and partially in Clusters 3 and

4. Similarly, a household with many members is less likely to belong to Cluster 1, while it is positively associated with Clusters 2, 4, and especially 3. A large number of members in the household, which can be associated with the presence of children, might contribute to explaining the earlier morning peaks [50] in households belonging to Clusters 2, 3, and 4 compared to Cluster 1. Continuing this line of inquiry at the household income level, high levels of income are negatively associated with Cluster 1 but positively associated with Cluster 2. However, no linear correlation between higher levels of household income and electricity consumption exists (Clusters 3 and 4).

While the type of dwelling tenure does not play any role in any electricity profile, clear results emerge with respect to the heating system type in the dwelling tenure. Households belonging to Clusters 1 and 2 are 22% and 7%, respectively, less likely to have a heating system powered by electricity relative to households with a heating system powered by other sources of energy. In contrast, households belonging to Clusters 3 and 4 are 16% and 13%, respectively, more likely to have a heating system powered by electricity relative to households with a heating system powered by other sources of energy. This result contributes to explaining the differences in the magnitude of electricity consumption among clusters and the higher morning peaks in Cluster 4.

The results also indicate the geographical and climate areas in which different household groups are more likely to be found. As shown in Figure 7, no significant variation exists in the average yearly maximum and minimum temperature and sunlight hours across regions in 2017.[4] Unlike other countries, the variation in domestic electricity demand within geographical areas due to different weather conditions is likely to be small in Denmark.

[4] https://www.dmi.dk/



Region	Av. temp. (°C)	Av. max temp. (°C)	Av. min temp. (°C)	Sunlight hours ☼
■ North Denmark Region	8.5	25.2	-11.6	1539.6
■ Central Denmark Region	8.8	26.1	-11.7	1483.8
■ Region of Southern Denmark	9.2	25.9	-10.5	1538.5
■ Zealand Region	9.1	25.5	-11.1	1548.9
■ Capital Region of Denmark	9.2	25.8	-11.4	1540.5

Figure 7. 2017 regional mean annual, minimum, and maximum temperature and sunlight hours.

Households belonging to Cluster 1 are more likely to live in the central Denmark region, while households belonging to Cluster 2 are 5% less likely to live in the same region relative to households living in the region of southern Denmark. In contrast, households belonging to Cluster 3 are 4% and 8% more likely to reside in the Zealand region and the capital region of Denmark, respectively, compared to households that reside in the southern Denmark region. However, it is important to interpret this result with caution, given the small representation of households living in the capital region of Denmark. Concerning working status, compared to the heads of households that are outside the workforce, managers and self-employed persons are more likely to belong to Clusters 3 and 4, thus partially contributing to explaining the higher levels of electricity consumption, especially during the morning. No clear results emerge for Cluster 2, whereas skilled or unskilled workers and managers are less likely to belong to Cluster 1 compared to the heads of households that are outside the workforce. Lastly, the educational level of the heads of households appears to be a predictor of different electricity load profiles. A higher educational level is positively associated with heads of households belonging to Cluster 1, while negatively correlated with heads of households belonging to Cluster 4.

6. Discussion

The empirical analysis has offered insight into the characteristics of households with similar electricity load patterns, the amount of electricity they consume, and at what time they consume it. In summary, the results illustrate four different profiles of Danish households regarding electricity use. Cluster 1, which represents the largest group of the sample, is characterized by low electricity consumption. Households belonging to this group are more likely to be single-person households with a high level of education living in a relatively newer and small multi-dwelling house. Households belonging to Cluster 2 consume, on average, more than twice the consumption of households belonging to Cluster 1. Households in Cluster 2 are more likely to comprise more than two members possessing a high level of household income, living in a large dwelling. Both Clusters 1 and 2 are less likely to be associated with a heating system powered by electricity. Households belonging to Cluster 3 consume, on average, more than three times the consumption of households belonging to Cluster 1. Households in Cluster 3 are more likely to comprise many members who live in old, large single-family detached houses. Cluster 4 has the highest electricity consumption, and households in this cluster are more likely to live in large houses. Compared to the other clusters, households belonging to Cluster 4 have earlier morning peaks. The heads of household for both Clusters 4 and 3 are more likely to be self-employed. Unlike Clusters 1 and 2, households belonging to Clusters 3 and 4 are more likely to have a heating system powered by electricity.

The results indicate that the sociodemographic and dwelling characteristics of the households are crucial indicators explaining and forecasting electricity load patterns and can support electric utilities and policymakers in personalizing interventions designed to time-shift peak loads, including time-of-use pricing, critical peak pricing, real-time pricing, and overall demand reduction. For example, households belonging to Clusters 3 and 4, who are more likely to have high occupancy rates and live in a large (and potentially inefficient) dwelling that uses electricity for space heating, might have a high potential to shift demand and participate in energy efficiency programs that reduce space heating[5].

Segment-specific rate design based on the four household load profile clusters allows consideration of consumption heterogeneity but also avoids the complexity of rate design on the individual customer level [52]. When designing segment-specific time-variable electricity rates, distribution and retail market players should define some key parameters such as the number of time and price zones, the price level in each period, and the start and end times of each time zones [52], [53]. Simulation scenarios may then assess and compare the potential impacts of different time-variable electricity rates on each representative load profile and the whole power system. Informed by customer grouping, time-varying rates could also improve the economic attractiveness and uptake of solar PV, by allowing owners to self-consume electricity during higher-priced peak hours [53], [54]. However, before deploying time-variable rates at large scale, well-defined pilots would be crucial to provide "real world" experience about what works and what does not [53].

In addition to household characteristics, several other factors play a significant role in shaping electricity demand and must be considered when designing and implementing demand response (DR). Exposing (targeted) consumers to the volatility of wholesale market prices is efficient from a neo-classical economic viewpoint but is prevented by systematic behavioral biases in consumer decision-making, such as heuristics, risk aversion, lack of attention to energy cost, and status quo bias [55], [56], [57]. In fact, a trade-off seems to exist between the complexity of the tariff model and the engagement of households in DR programs [58]. To further complicate things, the variable charge that depends on electricity consumed represents only 18% of the total electricity price[6] in addition to the

[5] Difficulties may arise in practice, for example, in households with numerous members (including children) who might be less flexible in incorporating time-shifting in their daily routines [51].

[6] The price Danish households pay comprises 18% in energy component payments, 17% in grid payments, and 65% in taxes, including Public Service Obligation (PSO) and VAT payments. By 1 January 2022, the PSO subsidizing renewable energy production and development will be phased out completely [59].

reluctant switching contract behavior of Danish customers [60]. Therefore, even under the assumption of entirely rational and informed individuals, a weak price signal would likely not encourage consumers to shift demand from peak to off-peak time due to the small financial gains associated with an active response to the program. Given the current unfavorable pricing structure for DR, it may be rational for households not to invest the time and effort to shift demand.

To help activate flexible demand and minimize distortions, Katz et al. [60] suggested introducing a dynamic approach to levies and taxes to increase the incentive to respond by a factor of three in a Danish setting. The debate about the introduction of a dynamic electricity tax and an overall reduction of the electricity taxes has recently moved into the agenda of the Danish Ministry of Climate, Energy and Utilities [24].

Consistent exposure to information feedback can augment an effective DR program implementation (e.g., in-home display) that can enhance consumer learning and awareness of price events and consumption by making electricity more visible [61], [62], [63], [64]. Also, shadow bills that inform customers about how their actual bill compares to what they would have paid under alternative tariffs could better educate them about the opportunities of specific time-varying rates and increase the switching rates [65]. In turn, this could facilitate the transition from simple to more complex tariffs [58].

However, to understand and influence electricity demand, the focus should not only be placed on price signals, household characteristics, and behavioral biases but also on the practices that people perform in their everyday lives. This is because similar households can have very different magnitudes and timing of electricity consumption due to the heterogeneity of domestic social practices and their order in time [66], [67], [68], [69], [70], [71]. The weekend morning ramp-up in demand provides a simple but effective illustration of how social practices shape electricity demand and its variation between seasons (Figure 6). Therefore, any DR strategy should be informed by the level of discretionary load that households are able and willing to shift and curtail, which is determined by the complex interaction between their characteristics, behavioral biases, responses to different price signals, and everyday practices.

7. Conclusions

Denmark has committed to an ambitious GHG reduction target of 70% by 2030, relative to 1990 levels, and to 100% of electricity consumption from renewables. Demand-side flexibility can play a crucial role in facilitating the integration of higher shares of renewables in the supply mix while reducing grid investments needed for the electrification of the heating and transport sectors. To improve the design of DR programs and incentivize flexibility, enhanced knowledge on domestic electricity load profiles is necessary. Thus, using hourly electricity consumption data combined with population-based register data, this study set out to segment a large sample of Danish households in relation to the time, magnitude, and drivers of electricity consumption. Both K-means clustering and multinomial probit regression are employed to determine the annual, seasonal, and weekly domestic electricity load profiles and their drivers. The results reveal four distinct household groups characterized by different timing and magnitudes of electricity use, which are influenced by specific sociodemographic and dwelling characteristics. In all four clusters, the average evening peak consumption, seasonal variation in electricity demand (especially for those with a higher likelihood to have an electrical heating system), and weekend morning ramp-up in demand are similar. This finding reveals some patterns of synchronisation of practices across the society. Valuable insight for utilities and policymakers aiming at reducing the peak demand and overall electricity consumption are derived. Moreover, the results support domestic electricity load and peak demand forecasting needed to improve the operational efficiency and reliability of the power system.

While providing a refined residential electricity consumption segmentation, the results cannot account for the habitual practices explaining electricity load patterns. For this reason, as a part of the eCAPE "New Energy Consumer Roles and Smart Technologies – Actors, Practices and Equality" project, two representative survey questionnaires focusing on everyday household energy-consuming practices will be conducted. The first survey will investigate the time of use and sequences of everyday practices with an emphasis on gender (in)equality. The second survey will closely examine households with low electricity consumption and the extent (if any) that ethical issues related to energy and the environment affect everyday practices. By integrating quantitative and qualitative research techniques aimed at investigating not only the timing, amount, and drivers of electricity demand but also the underlying practices, new DR programs can be tailored to suit household lifestyles.

Acknowledgements

This work was financed by the European Research Council (ERC) under project grant Gram-Hanssen/786643, eCAPE "New Energy Consumer Roles and Smart Technologies – Actors, Practices and Equality". The author would like to thank Anders Rhiger Hansen, Kirsten Gram-Hanssen, Toke Haunstrup Christensen, the anonymous reviewers, and the participants of the "13th International Workshop on Empirical Methods in Energy Economics" (January 13-14, 2020) in Zurich for their valuable comments on the earlier version of this paper.

References

[1] Chicco, G. (2012). Overview and performance assessment of the clustering methods for electrical load pattern grouping. *Energy*, *42*(1), 68–80. doi:10.1016/j.energy.2011.12.031.

[2] Hayn, M., Bertsch, V., & Fichtner, W. (2014). Electricity load profiles in Europe: The importance of household segmentation. *Energy Research & Social Science*, *3*, 30–45. doi:10.1016/j.erss.2014.07.002.

[3] Viegas, J. L., Vieira, S. M., Melício, R., Mendes, V. M. F., & Sousa, J. M. (2016). Classification of new electricity customers based on surveys and smart metering data. *Energy*, *107*, 804–817. doi:10.1016/j.energy.2016.04.065.

[4] Pina, A., Silva, C., & Ferrão, P. (2012). The impact of demand side management strategies in the penetration of renewable electricity. *Energy*, *41*(1), 128–137. doi:10.1016/j.energy.2011.06.013.

[5] Hsiao, Y. H. (2014). Household electricity demand forecast based on context information and user daily schedule analysis from meter data. *IEEE Transactions on Industrial Informatics*, *11*(1), 33–43. doi:10.1109/TII.2014.2363584.

[6] Haben, S., Singleton, C., & Grindrod, P. (2015). Analysis and clustering of residential customers energy behavioral demand using smart meter data. *IEEE Transactions on Smart Grid*, *7*(1), 136–144. doi:10.1109/TSG.2015.2409786.

[7] Räsänen, T., Voukantsis, D., Niska, H., Karatzas, K., & Kolehmainen, M. (2010). Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data. *Applied Energy*, *87*(11), 3538–3545. doi:10.1016/j.apenergy.2010.05.015.

[8] Kwac, J., Flora, J., & Rajagopal, R. (2014). Household energy consumption segmentation using hourly data. *IEEE Transactions on Smart Grid*, *5*(1), 420–430. doi:10.1109/TSG.2013.2278477.

[9] Kwac, J., Flora, J., & Rajagopal, R. (2016). Lifestyle segmentation based on energy consumption data. *IEEE Transactions on Smart Grid*, *9*(4), 2409–2418. doi:10.1109/TSG.2016.2611600.

[10] Beckel, C., Sadamori, L., Staake, T., & Santini, S. (2014). Revealing household characteristics from smart meter data. *Energy*, *78*, 397–410. doi:10.1016/j.energy.2014.10.025.

[11] McLoughlin, F., Duffy, A., & Conlon, M. (2015). A clustering approach to domestic electricity load profile characterisation using smart metering data. *Applied Energy*, *141*, 190–199. doi:10.1016/j.apenergy.2014.12.039.

[12] Rhodes, J. D., Cole, W. J., Upshaw, C. R., Edgar, T. F., & Webber, M. E. (2014). Clustering analysis of residential electricity demand profiles. *Applied Energy*, *135*, 461–471. doi:10.1016/j.apenergy.2014.08.111.

[13] Ramos, S., Duarte, J. M., Duarte, F. J., & Vale, Z. (2015). A data-mining-based methodology to support MV electricity customers' characterization. *Energy and Buildings*, *91*, 16–25. doi:10.1016/j.enbuild.2015.01.035.

[14] Gouveia, J. P., & Seixas, J. (2016). Unraveling electricity consumption profiles in households through clusters: Combining smart meters and door-to-door surveys. *Energy and Buildings*, *116*, 666–676. doi:10.1016/j.enbuild.2016.01.043.

[15] Gouveia, J. P., Seixas, J., & Mestre, A. (2017). Daily electricity consumption profiles from smart meters-Proxies of behavior for space heating and cooling. *Energy*, *141*, 108–122. doi:10.1016/j.energy.2017.09.049.

[16] Belaïd, F. (2016). Understanding the spectrum of domestic energy consumption: Empirical evidence from France. *Energy Policy*, *92*, 220–233. doi:10.1016/j.enpol.2016.02.015.

[17] Ozawa, A., Furusato, R., & Yoshida, Y. (2016). Determining the relationship between a household's lifestyle and its electricity consumption in Japan by analyzing measured electric load profiles. *Energy and Buildings*, *119*, 200–210. doi:10.1016/j.enbuild.2016.03.047.

[18] Benítez, I., Díez, J. L., Quijano, A., & Delgado, I. (2016). Dynamic clustering of residential electricity consumption time series data based on Hausdorff distance. *Electric Power Systems Research*, *140*, 517–526. doi:10.1016/j.epsr.2016.05.023.

[19] Anderson, B., Lin, S., Newing, A., Bahaj, A., & James, P. (2017). Electricity consumption and household characteristics: Implications for census-taking in a smart metered future. *Computers, Environment and Urban Systems*, *63*, 58–67. doi:10.1016/j.compenvurbsys.2016.06.003.

[20] Satre-Meloy, A., Diakonova, M., Grünewald, P. (2019). What makes you peak? Cluster analysis of household activities and electricity demand. *ECEEE 2019 Summer Study on Energy Efficiency in Buildings*. https://bit.ly/36kOEx5

[21] Tureczek, A., Nielsen, P., & Madsen, H. (2018). Electricity consumption clustering using smart meter data. *Energies*, *11*(4), 859. doi:10.3390/en11040859.

[22] Zethmayr, J., & Makhija, R. S. (2019). Six unique load shapes: A segmentation analysis of Illinois residential electricity consumers. *The Electricity Journal*, *32*(9), 106643. doi:10.1016/j.tej.2019.106643.

[23] Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, *31*(8), 651–666. DOI: 10.1016/j.patrec.2009.09.011.

[24] Danish Ministry of Climate, Energy and Utilities. (2018). *Energy Agreement, 2018*. https://en.efkm.dk/media/12307/energy-agreement-2018.pdf

[25] Danish Energy Agency. (2017). *Energy in Denmark 2017*. https://ens.dk/sites/ens.dk/files/Statistik/energyindenmark2017.pdf

[26] Danish Energy Agency. (2019). *Denmark's Energy and Climate Outlook 2019*. https://ens.dk/sites/ens.dk/files/Analyser/deco19.pdf

[27] Denmark Government. (2019). *A fair direction for Denmark*. https://www.socialdemokratiet.dk/media/8602/a-fair-direction-for-denmark.pdf

[28] Chicco, G., Napoli, R., & Piglione, F. (2006). Comparisons among clustering techniques for electricity customer classification. *IEEE Transactions on Power Systems*, *21*(2), 933–940. doi:10.1109/TPWRS.2006.873122.

[29] Hsu, D. (2015). Comparison of integrated clustering methods for accurate and stable prediction of building energy consumption data. *Applied Energy*, *160*, 153–163. doi:10.1016/j.apenergy.2015.08.126.

[30] Yildiz, B., Bilbao, J. I., Dore, J., & Sproul, A. B. (2017). Recent advances in the analysis of residential electricity consumption and applications of smart meter data. *Applied Energy*, *208*, 402–427. doi:10.1016/j.apenergy.2017.10.014.

[31] Adolfsson, A., Ackerman, M., & Brownstein, N. C. (2019). To cluster, or not to cluster: An analysis of clusterability methods. *Pattern Recognition*, *88*, 13–26. doi:10.1016/j.patcog.2018.10.026.

[32] Rodriguez, M. Z., Comin, C. H., Casanova, D., Bruno, O. M., Amancio, D. R., Costa, L. D. F., & Rodrigues, F. A. (2019). Clustering algorithms: A comparative approach. *PloS One*, *14*(1), e0210236. doi:10.1371/journal.pone.0210236.

[33] Cembranel, S. S., Lezama, F., Soares, J., Ramos, S., Gomes, A., & Vale, Z. (2019). A short review on data mining techniques for electricity customers characterization. *2019 IEEE PES GTD Grand International Conference and Exposition Asia* (GTD Asia) (pp. 194–199). IEEE. doi:10.1109/GTDAsia.2019.8715891.

[34] Statistics Denmark. (2014). *The Danish system for access to micro data*. https://www.dst.dk/ext/5452354440/0/israel2016/Annex-B5-15-Introduction-to-the-Danish-system-for-access-to-microdata--pdf

[35] Statistics Denmark. (2017). *Data confidentiality*. https://www.dst.dk/ext/502998790/0/formid/Data-Confidentiality-Policy-at-Statistics-Denmark--pdf

[36] Wilson, S. E. (2015). Methods for clustering data with missing values. https://www.math.leidenuniv.nl/scripties/MasterWilson.pdf

[37] Jin, X., & Han, J. (2017). K-means clustering. In *Encyclopedia of machine learning and data mining* (695–697). doi:10.1007/978-1-4899-7687-1.

[38] Piao, M., Shon, H. S., Lee, J. Y., & Ryu, K. H. (2014). Subspace projection method based clustering analysis in load profiling. *IEEE Transactions on Power Systems*, *29*(6), 2628–2635. doi:10.1109/TPWRS.2014.2309697.

[39] Tureczek, A. M., Nielsen, P. S., Madsen, H., & Brun, A. (2019). Clustering district heat exchange stations using smart meter consumption data. *Energy and Buildings*, *182*, 144–158. doi:10.1016/j.enbuild.2018.10.009.

[40] Makles, A. (2012). Stata tip 110: How to get the optimal k-means cluster solution. *The Stata Journal*, *12*(2), 347–351. doi:10.1177/1536867x1201200213.

[41] Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American statistical association, 58*(301), 236-244. doi: 10.1080/01621459.1963.10500845.

[42] Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). Cluster analysis. John Wiley & Sons.

[43] Xu, G., Zong, Y., & Yang, Z. (2013). Applied data mining. CRC Press.

[44] Zaki, M. J., & Meira, W. (2014). Data mining and analysis: fundamental concepts and algorithms. Cambridge University Press.

[45] Strauss, T., & von Maltitz, M. J. (2017). Generalising Ward's method for use with Manhattan distances. *PloS one, 12*(1), e0168288. doi: 10.1371/journal.pone.0168288.

[46] Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods, 3*(1), 1-27. doi: 10.1080/03610927408827101.

[47] Liu, Y., Li, Z., Xiong, H., Gao, X., & Wu, J. (2010). Understanding of internal clustering validation measures. In 2010 *IEEE International Conference on Data Mining* (pp. 911-916). IEEE. doi: 10.1109/ICDM.2010.35.

[48] Bartus, T. (2005). Estimation of marginal effects using margeff. *The Stata Journal*, *5*(3), 309–329. doi:10.1177/1536867X0500500303.

[49] Trotta, G. (2018). The determinants of energy efficient retrofit investments in the English residential sector. *Energy Policy*, 120, 175-182. doi: 10.1016/j.enpol.2018.05.024.

[50] Nicholls, L., & Strengers, Y. (2015). Peak demand and the 'family peak' period in Australia: Understanding practice (in) flexibility in households with children. *Energy Research & Social Science*, *9*, 116–124. doi:10.1016/j.erss.2015.08.018.

[51] Friis, F., & Christensen, T. H. (2016). The challenge of time shifting energy demand practices: Insights from Denmark. *Energy Research & Social Science*, *19*, 124–133. doi:10.1016/j.erss.2016.05.017.

[52] Flath, C., Nicolay, D., Conte, T., van Dinther, C., & Filipova-Neumann, L. (2012). Cluster analysis of smart metering data. *Business & Information Systems Engineering*, 4(1), 31-39. doi: 10.1007/s12599-011-0201-5.

[53] Faruqui, A., Hledik, R., & Palmer, J. (2012). *Time-varying and dynamic rate design*. Regulatory Assistance Project. https://www.raponline.org/wp-content/uploads/2016/05/rap-faruquihledikpalmer-timevaryingdynamicratedesign-2012-jul-23.pdf

[54] Hledik, R., Gorman, W., Irwin, N., Fell, M., Nicolson, M., & Huebner, G. (2017). The value of TOU tariffs in Great Britain: Insights for decision-makers. *Citizen Advice, Final Report, 1*. https://www.citizensadvice.org.uk/Global/CitizensAdvice/Energy/The%20Value%20of%20ToU%20Tariffs%20in%20GB%20-%20Volume%20I.pdf

[55] Pollitt, M. G., & Shaorshadze, I. (2013). The role of behavioural economics in energy and climate policy. In *Handbook on energy and climate change*. doi:10.4337/9780857933683.00035.

[56] Frederiks, E. R., Stenner, K., & Hobman, E. V. (2015). Household energy use: Applying behavioural economics to understand consumer decision-making and behaviour. *Renewable and Sustainable Energy Reviews*, *41*, 1385–1394. doi:10.1016/j.rser.2014.09.026.

[57] Hobman, E. V., Frederiks, E. R., Stenner, K., & Meikle, S. (2016). Uptake and usage of cost-reflective electricity pricing: Insights from psychology and behavioural economics. *Renewable and Sustainable Energy Reviews*, *57*, 455–467. doi:10.1016/j.rser.2015.12.144.

[58] Grünewald, P., McKenna, E., & Thomson, M. (2015). Keep it simple: time-of-use tariffs in high-wind scenarios. *IET Renewable Power Generation*, *9*(2), 176–183. doi:10.1049/iet-rpg.2014.0031.

[59] Danish Utility Regulator. (2019). *National Report 2019 for Denmark. Developments in 2018*. https://forsyningstilsynet.dk/media/6616/national-report-2019-for-denmark.pdf

[60] Katz, J., Kitzing, L., Schröder, S. T., Andersen, F. M., Morthorst, P. E., & Stryg, M. (2018). Household electricity consumers' incentive to choose dynamic pricing under different taxation schemes. *Wiley Interdisciplinary Reviews: Energy and Environment*, *7*(1), e270. doi:10.1002/wene.270.

[61] Faruqui, A., Sergici, S., & Sharif, A. (2010). The impact of informational feedback on energy consumption—A survey of the experimental evidence. *Energy*, *35*(4), 1598–1608. doi:10.1016/j.energy.2009.07.042.

[62] Jessoe, K., & Rapson, D. (2014). Knowledge is (less) power: Experimental evidence from residential energy use. *American Economic Review*, *104*(4), 1417–1438. doi:10.1257/aer.104.4.1417.

[63] Lynham, J., Nitta, K., Saijo, T., & Tarui, N. (2016). Why does real-time information reduce energy consumption? *Energy Economics*, *54*, 173–181. doi:10.1016/j.eneco.2015.11.007.

[64] Aydin, E., Brounen, D., & Kok, N. (2018). Information provision and energy consumption: Evidence from a field experiment. *Energy Economics*, *71*, 403–410. doi:10.1016/j.eneco.2018.03.008.

[65] Borenstein, S. (2013). Effective and equitable adoption of opt-in residential dynamic electricity pricing. *Review of Industrial Organization*, 42(2), 127-160. doi: 10.1007/s11151-012-9367-3.

[66] Gram-Hanssen, K. (2011). Understanding change and continuity in residential energy consumption. *Journal of Consumer Culture*, *11*(1), 61–78. doi:10.1177/1469540510391725.

[67] Gram-Hanssen, K. (2013). Efficient technologies or user behaviour, which is the more important when reducing households' energy consumption? *Energy Efficiency*, *6*(3), 447–457. doi:10.1007/s12053-012-9184-4.

[68] Gram-Hanssen, K. (2014). New needs for better understanding of household's energy consumption–behaviour, lifestyle or practices? *Architectural Engineering and Design Management*, *10*(1-2), 91–107. doi:10.1080/17452007.2013.837251.

[69] Torriti, J., Hanna, R., Anderson, B., Yeboah, G., & Druckman, A. (2015). Peak residential electricity demand and social practices: Deriving flexibility and greenhouse gas intensities from time use and locational data. *Indoor and Built Environment*, *24*(7), 891–912. doi:10.1177/1420326X15600776.

[70] Torriti, J. (2017). Understanding the timing of energy demand through time use data: Time of the day dependence of social practices. *Energy Research & Social Science*, *25*, 37–47. doi:10.1016/j.erss.2016.12.004.

[71] Yunusov, T., Lorincz, M. J., & Torriti, J. (2018). *Role of household activities in peak electricity demand and distributional effects of time-of-use tariffs*. http://centaur.reading.ac.uk/79505/

Appendix

Table 2. Descriptive statistics of *WSS, log(WSS)*, η^2, and *PRE* for all *K* cluster solutions.

Cluster (K)	WSS	log(WSS)	η^2	PRE
1	107349.3	11.584	0.000	.
2	60460.5	11.010	0.437	0.437
3	45216.8	10.719	0.579	0.252
4	39829.2	10.592	0.629	0.119
5	39158.6	10.575	0.635	0.017
6	35100.1	10.466	0.673	0.104
7	33858.4	10.430	0.685	0.035
8	32745.1	10.397	0.695	0.033
9	31569.9	10.360	0.706	0.036
10	31196.8	10.348	0.709	0.012
11	31124.0	10.346	0.710	0.002
12	30571.8	10.328	0.715	0.018
13	30868.3	10.337	0.712	-0.010
14	29152.5	10.280	0.728	0.056
15	28682.3	10.264	0.733	0.016
16	27962.3	10.239	0.740	0.025
17	27993.0	10.240	0.739	-0.001
18	27580.6	10.225	0.743	0.015
19	27359.7	10.217	0.745	0.008
20	27203.7	10.211	0.747	0.006

Table 3. Calinski-Harabasz Index.

Cluster (K)	Pseudo-F statistic
2	8044.38
3	8882.92
4	6363.49
5	6218.91
6	5486.21
7	4962.98
8	4577.42
9	4224.87
10	3827.93
11	3598.41
12	3296.81
13	3104.47
14	2916.17
15	2780.99
16	2645.49
17	2508.45
18	2383.21
19	2291.27
20	2190.99

Table 5. Results of the multinomial probit (MNP) estimations for the four clusters.

Variables	Cluster 1 (β_0=0)	Cluster 2	Cluster 3	Cluster 4
	MNP	MNP	MNP	MNP
Building year (Ref = Before 1963)				
1963-1972	-	-0.05 (0.05)	-0.46*** (0.07)	-0.39*** (0.1)
1973-1978	-	0.03 (0.07)	-0.32*** (0.08)	-0.33** (0.11)
1979-1998	-	-0.04 (0.06)	-0.24** (0.07)	-0.27* (0.11)
1999-2006	-	-0.06 (0.08)	-0.35** (0.1)	-0.16 (0.14)
After 2006	-	-0.12 (0.1)	-0.41** (0.12)	-0.63*** (0.18)
Dwelling type (Ref = Single-family detached house)				
Terraced, linked or semi-detached house	-	-0.44*** (0.07)	-0.77*** (0.09)	-0.83*** (0.14)
Multi-dwelling house	-	-1.31*** (0.07)	-1.91*** (0.11)	-1.4*** (0.19)
Floor area (Ref = ≤ 75 m^2)				
76 m^2 – 95 m^2	-	0.5*** (0.07)	0.19 (0.11)	0.27 (0.19)
96 m^2 – 120 m^2	-	0.86*** (0.07)	0.43*** (0.11)	0.84*** (0.19)
121 m^2 – 150 m^2	-	0.97*** (0.08)	0.8*** (0.12)	0.96*** (0.2)
≥ 151	-	1.24*** (0.09)	1.34*** (0.12)	1.71*** (0.2)
Dwelling tenure (Ref= Owned)				
Rented	-	-0.06 (0.06)	-0.15 (0.08)	-0.28* (0.13)
Heating system (Ref = Heating system powered by natural gas, liquid/solid fuels, or straw)				
Heating system powered by electricity	-	0.91*** (0.12)	2.24*** (0.12)	3.8*** (0.14)
Government office region (Ref = Region of Southern Denmark)				
Central Denmark Region	-	-0.26** (0.08)	-0.09 (0.1)	-0.03 (0.15)
North Denmark Region	-	-0.06 (0.09)	-0.36*** (0.12)	-0.46** (0.18)
Zealand Region	-	0.24*** (0.04)	0.47*** (0.05)	0.39*** (0.07)
Capital Region of Denmark	-	0.19 (0.24)	0.79** (0.25)	0.79** (0.3)
Household size (Ref =1 member)				
2 members	-	0.95*** (0.05)	0.9*** (0.07)	0.85*** (0.11)
3 members	-	1.29*** (0.08)	1.56*** (0.1)	1.23*** (0.15)
4 members	-	1.66*** (0.1)	2.11*** (0.11)	1.72*** (0.15)
5 members or more	-	1.75*** (0.15)	2.72*** (0.15)	2.22*** (0.2)
Household disposable income (Ref = Lowest 20%)				
Quintile 2	-	-0.06 (0.06)	-0.19* (0.08)	-0.26* (0.13)
Quintile 3	-	0.03 (0.07)	-0.16 (0.08)	-0.17 (0.13)
Quintile 4	-	0.27*** (0.08)	0.04 (0.1)	0.01 (0.14)
Highest 20%	-	0.32*** (0.1)	0.29** (0.11)	0.2 (0.15)
Head of househoold working status (Ref = Outside workforce)				
Worker (skilled/unskilled)	-	0.2*** (0.05)	0.44*** (0.07)	0.4*** (0.1)
Self-employed	-	0.12 (0.13)	0.45** (0.15)	0.6** (0.2)
Routine nonmanual	-	0.08 (0.07)	0.19* (0.09)	0.33** (0.13)
Manager	-	0.23** (0.08)	0.38*** (0.1)	0.59*** (0.13)
Head of household education level (Ref = Primary)				
Secondary	-	-0.09 (0.05)	-0.14* (0.06)	-0.23* (0.09)
Tertiary	-	-0.16** (0.06)	-0.25** (0.08)	-0.41*** (0.11)
Constant		-1.24*** (0.09)	-1.85*** (0.12)	-3.16*** (0.22)
Number of observations	13,602	13,602	13,602	13,602

Robust standard errors in parentheses. * p<.05, ** p<.01, *** p<.001