Université de Neuchâtel
Institut de recherches économiques

IRENE Working Paper 26-03

Residential electricity load profiles and their determinants: A cluster analysis of smart meter data

Valentin Favre-Bulle and Sylvain Weber

Residential electricity load profiles and their determinants: A cluster analysis of smart meter data

Valentin Favre-Bulle[*†] Sylvain Weber[†]

April 2, 2026

Abstract

We analyse two years of hourly electricity consumption data from approximately 4,000 households. Using cluster analyses, we identify three segments with distinct daily load profiles. Using multinomial logit models, we then explore how dwelling characteristics (e.g., heat pumps, photovoltaic panels, appliance usage) and socio-demographic variables (.e.g., employment status and age) influence cluster membership. Dwelling characteristics primarily distinguish low- and high-consumption households, while socio-demographic factors further differentiate among remaining groups. These insights support targeted demand-side policies and enable providers to implement tailored dynamic pricing strategies.

Keywords: Cluster analysis; Demand-side management; Household electricity load profiles; Load curves; Multinomial logit model; Smart meters analytics

JEL Classification: Q41; D12; C38; L94; Q48

[*]University of Neuchâtel, Department of Economics and Business, Switzerland.

[†]University of Applied Sciences Western Switzerland (HES-SO/HEG-Geneva), Switzerland. Contact: valentin.favre-bulle@hesge.ch.

1 Introduction

In the energy sector, households play a key role, as they account for a substantial share of total electricity consumption.[1] Household electricity consumption is highly heterogeneous, and understanding this heterogeneity is crucial for energy sector actors.

Access to high-quality, high-frequency electricity consumption data is essential for analysing household electricity use, as annual data provide only limited insights into usage patterns. Smart meters overcome this limitation by enabling the collection of detailed hourly or quarter-hourly data, which can be used to cluster households into distinct consumption profiles based on their load curves. When combined with socio-demographic characteristics and dwelling features, these profiles offer valuable insights for policymakers and electricity suppliers.

In this paper, we use electricity consumption data from smart meters for approximately 4,000 households in northwestern Switzerland. The dataset spans the years 2023–2024 and includes basic technical information about the dwellings. Additionally, detailed dwelling and socio-demographic data were collected for a subsample of about 150 households. We primarily apply k-means clustering to identify three consumption profiles, and then use a multinomial logit regression model to relate these profiles to the socio-demographic characteristics of households and the technical features of dwellings.

Governments can use such information to design efficient and well-targeted instruments that support demand-side management, encouraging consumers to adjust both the level and timing of their electricity use to better match supply conditions and reduce peak demand, or to implement dynamic pricing, as illustrated by the experience of Hydro-Québec [2]. Electricity utilities can also benefit from a better understanding of customer segments, enabling more targeted strategies and a stronger impact on electricity use, particularly in increasingly deregulated markets. With the expansion of photovoltaic (PV) systems and the associated challenges to grid stability, a detailed understanding of residential electricity consumption becomes even more important.

[1] For example, in 2024, Swiss households accounted for around 35% of total electricity consumption, compared to approximately 30% in 2005, with this share steadily increasing over the past two decades [1].

A substantial body of research has developed on methods for analysing smart meter data[2], with clustering of electricity consumption data being one of the most commonly used approaches to identify groups of households with similar load curves.[3]

[6] apply k-means clustering to identify three distinct electricity load profiles and then use a multinomial logit model to link them to household characteristics (notably the age of household members and the number of electrical appliances). [7] use a finite mixture model and identify ten distinct clusters and confirm the robustness of their results using a bootstrapping procedure. Using k-means clustering on electricity consumption data from approximately 2.5 million households, [8] demonstrate that cluster membership can also be related to spatial characteristics, with distinct consumption patterns across urban, suburban, and rural areas. [9] compare multiple clustering methods (including k-means, k-medoids, and Self-Organising Maps (SOM)) to analyse the consumption data of nearly 4,000 households using daily data over six months. They identify ten profile classes representing typical daily consumption patterns and variation across seasons and days of the week, and, using a multinomial logit model, link these classes to dwelling and household characteristics (such as household size and the presence of a dishwasher).

A different strategy, which can be described as an "inverted approach", consists in inferring household and dwelling characteristics directly from smart meter data. [10] employ various supervised machine learning techniques to analyse smart meter data and, using a separate dataset, infer household characteristics with relatively high accuracy, including household composition (single vs. family) and socio-economic variables such as employment status. [11] employ a Hidden Markov Model framework to group households based on their consumption patterns and show that their method improves the accuracy of inferring specific household characteristics, such as the presence of a dishwasher or children, compared to random guessing.

Our analysis identifies three distinct groups that differ clearly in both the level and timing of electricity consumption. While consumption varies seasonally, typically being higher in winter and lower in summer, the overall load patterns remain largely stable

[2] For a review, see [3].

[3] For comprehensive overviews of clustering techniques in this context, see [4] and [5].

throughout the year. Some clusters also exhibit distinct consumption behaviours across days of the week.

Among the three clusters, one is composed of active flat dwellers with low overall electricity consumption, a distinct peak in the evening hours, limited appliance use, and patterns reflecting high employment rates and the absence of electric heating. The second cluster contains high-consumption households living in houses equipped with heat pumps and electric hot water systems. The last one comprises home-based moderate consumers without electric heating, who consume slightly more due to lower employment and more stable, evenly distributed usage patterns. Overall, dwelling characteristics, in particular the presence of electric heating technologies, play a more decisive role in cluster membership than socio-demographic factors such as household size or age. These three segments provide a useful basis for designing targeted and efficient demand response strategies.

The remainder of the paper is organised as follows. Section 2 describes the data. Section 3 outlines the clustering and multinomial models. Section 4 presents the results, while Section 5 discusses the findings, and Section 6 concludes.

2 Data

Smart meter data was obtained from *La Goule*, a Swiss electricity provider. It contains hourly electricity consumption from approximately 4,000 households located in the Jura Bernois region (northwestern Switzerland), covering the period from January 2023 to December 2024.[4] Only net consumption, i.e., supplied by the provider, is recorded. Self-consumption in households equipped with photovoltaic (PV) systems is not recorded.

[4] We also account for daylight saving time by adjusting the timestamps to the corresponding local time, ensuring that consumption patterns accurately reflect real-life activity. A small share of households also had multiple smart meters, for instance, one for the heat pump and another for the remainder of the household consumption. In such cases, we aggregate consumption from all smart meters at the household level.

2.1 Filtering of the consumption data

First, we drop smart meters with more than 20% missing data or more than 50% of zero hourly consumption values. Since we only observe net consumption, zero values are entirely plausible, for example for photovoltaic (PV) system owners who self-consume during sunny hours or households with battery storage. These thresholds allow us to retain such cases while excluding vacant dwellings with no electricity use. Second, two additional filters were applied to exclude any possible remaining non-residential units with implausible consumption values for residential households. We computed, over the whole period and for each unit, the average consumption for each hour of the day and excluded units with at least one of these exceeding 30 kWh. Then, we compute, for each unit, the normalised average consumption for each hour of the day. It is defined as the share of average consumption in a specific hour relative to the sum of average consumption across all hours of the day.[5] We drop all units that have at least one normalised average hourly consumption value greater than 0.5. These last two filters imply that we do not consider as residential households any units that, on average, consume more than 30 kWh in a single hour or more than 50% of their total daily consumption during a single hour. Table A1 presents the descriptive statistics for the main clustering, which includes 3,304 households.[6]

2.2 Additional household and dwelling characteristics

We have access to information on dwelling characteristics and household types, including the heating system (electrical/heat pump or non-electrical),[7] the presence of photovoltaic (PV) panels and the type of dwelling (house vs. flat) (Table A1).

[5] For example, if a household has an average consumption of 1.5 kWh at noon over the entire period, and the sum of its average consumption across all hours of the day is 10 kWh, then the normalised value for hour 12 is 15%.

[6] Following a primary clustering, we excluded a group of 772 "night-consumer" households characterised by significant high nighttime consumption. All relevant information is provided in B.

[7] More precisely, the heating system variable includes three categories: Heat pump, Storage heating, and Others, with the latter mainly corresponding to oil-based systems. As our analysis focuses on electricity consumption, since both heat pumps and storage heating systems rely on electricity, and while the number of the latter is relatively small, these two categories are combined into a single group. We thus end up with two categories for the heating system: electrical/HP and non-electrical.

We also have detailed information for a subsample of 158 households that participated in the experiment related to this study. For these, we have a large set of variables describing both household composition (such as size, age structure, income, and employment status) and dwelling characteristics (such as size, heating system, and number of electrical appliances, among others) (Table A2).

3 Methods

3.1 Clustering

Even if some studies use hourly or half-hourly data directly for clustering [8], many papers do not rely on raw load profile data but instead use features that represent household consumption patterns. For example, features such as the mean or standard deviation of daily consumption, the load factor, or time-of-day consumption shares can be used. This transformation offers three advantages [12]: (1) it reduces the dimensionality of the dataset, which is often very high with smart meter data. This helps to address the "curse of dimensionality", a common issue when applying algorithms to high-dimensional datasets. While many algorithms perform well on small datasets, they often become intractable when the input is high-dimensional [13]; (2) clustering becomes less sensitive to missing values; and (3) it allows the use of the data even when households datasets vary in length. Following this strategy, our clustering is performed using six electricity consumption features computed for each household.

One feature is the average daily consumption level, which is selected to differentiate households by their consumption levels. We compute four additional features representing each household's electricity consumption across different periods of the day, following [7]. More specifically, these features correspond to normalised electricity consumption shares for each period of the day [6]. In our context, normalised consumption refers to the share of total daily consumption attributed to each period of the day, calculated over the entire duration of the sample. As a result, the four features for each household sum to 1. The periods of the day are night-time (9 p.m. to 7 a.m.), morning (7 a.m. to 11 a.m.), daytime (11 a.m. to 3 p.m.) and evening (3 p.m. to 9 p.m.). These periods are selected to reflect typical daily household consumption patterns. Following [6], a measure of consumption disparity is also used by computing the standard deviation of the mean normalised consumption of the four periods. This sixth feature captures the variability of consumption throughout the periods of the day, with a higher standard deviation indicating greater intra-day variation. All features were standardised (z-scores) to make them comparable in the clustering analysis.

K-means clustering is implemented in Stata software using Euclidean distance as the similarity metric through the "cluster" command. As a robustness check, k-medians clustering is also employed in some analyses. To support the decision regarding the optimal number of clusters, we use the Calinski-Harabasz index, where higher values indicate better clustering performance. As robustness checks, we consider the within-cluster sum of squares (WSS) and its derivates following the approach proposed by [14], as well as the Silhouette index.

An insightful way to examine households' electricity consumption is to explore whether consumption patterns vary across different periods of the year, such as seasons or months, and across days of the week. In the literature, two main approaches are used to analyse this aspect. The first strategy involves performing clustering on the entire dataset and then assessing whether the resulting load curves profiles differ across periods of interest such as seasons or days of the week [e.g. 9, 7]. The second approach segments the data according to the period of interest and then performs clustering separately within each subset [e.g. 15, 16]. The evolution of cluster number and composition can then be analysed overt time. We chose to apply both strategies. The first offers a general understanding of consumption patterns, while the second complements it by allowing for a varying number of optimal clusters across time periods (such as seasons). Thus, we implement the following clustering approaches:

1. A global clustering using the full dataset covering the period 2023–2024;
2. A time-disaggregated clustering, also applied to the full dataset (2023–2024), conducted separately by season and by day of the week.

3.2 Composition of the clusters: Multinomial logit model

To investigate the composition of the clusters, we estimate multinomial logit models. This approach is suitable when the dependent variable consists of more than two unordered categories. Let $y_i \in \{1, \ldots, K\}$ denote the cluster membership of household i, and let x_i be a vector of observed characteristics. The probability that household i belongs to cluster k is then:

$$\Pr(y_i = k \mid x_i) = \frac{\exp(x_i' \beta_k)}{\sum_{j=1}^{K} \exp(x_i' \beta_j)} \tag{1}$$

with one cluster serving as the reference category to ensure identification.

The estimated coefficients of a multinomial logit model cannot be interpreted directly as changes in probabilities. To facilitate interpretation, we therefore focus on marginal effects, which describe how changes in explanatory variables affect the predicted probability of belonging to a given cluster. In particular, marginal effects measure the change in the probability associated with a one-unit increase in a continuous covariate, holding other variables constant as for binary variables, they correspond to the discrete change in probabilities when the variable switches from 0 to 1.

In the main analysis, we estimate the model using the full sample of households and a limited set of variables that are available and relevant for all observations, including dwelling type (house versus flat), the presence of electric heating or a heat pump, and the presence of solar photovoltaic panels. In the complementary analysis, we exploit the subsample of 158 households for which more detailed information is available, allowing for a refined characterization of cluster membership.

4 Results

4.1 Main clustering

After running the k-means clustering, the Calinski-Harabasz index and the within-cluster sum of squares (WSS) clearly support a three-cluster solution ($k = 3$), whereas the Silhouette index is more consistent with a two-cluster solution (Figures C1). Under k-medians clustering, all indices similarly support a three-cluster solution (Figure C2). Taken to-

Table 1: Mean value of the six standardised features - Main clustering - By cluster

	Clusters		
	1	2	3
Sample size	1,207 (36.5%)	521 (15.8%)	1,576 (47.7%)
Night (9PM–7AM)	-0.442	0.256	-0.532
Morning (7AM–11AM)	-0.378	0.355	0.745
Daytime (11AM–3PM)	0.026	-0.423	0.733
Evening (3PM–9PM)	0.907	-0.325	-0.033
Standard deviation	0.093	-0.231	-0.749
Daily mean consumption	-0.469	1.354	-0.350

Notes: This table shows the mean values of the six features used for the main clustering, which are described in detail in Section 3.

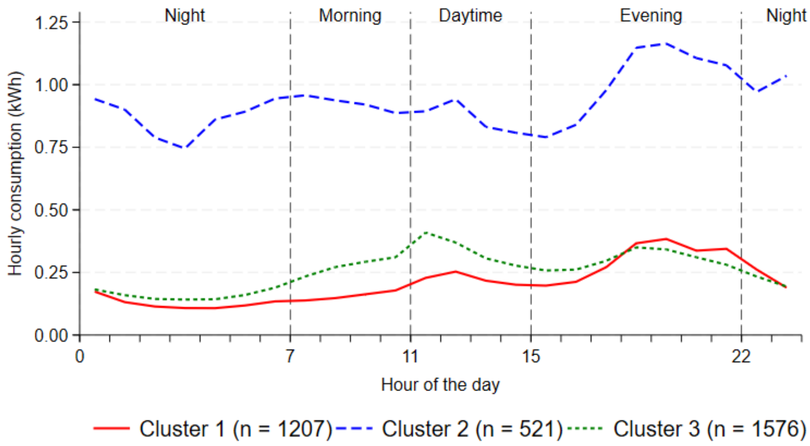
gether, these criteria point to $k = 3$ as the most appropriate number of clusters.[8] Table 1 presents the means of the six standardized features for each of the three clusters, along with the number of households in each cluster. The figures reported in this table show that all six standardized features differ across clusters, indicating that each variable plays a relevant role in distinguishing consumption patterns. The following figures illustrate the clearly distinct consumption levels and load patterns of these three clusters. In particular, the mean consumption profiles in levels (Figure 1a) highlight the differences between high- and low-consuming households, while the mean normalised profiles (Figure 1b) illustrate the intra-day distribution of electricity use. Finally, the plots displaying all individual load curves within each cluster, together with their mean profiles, provide insight into the accuracy of the clustering and the presence of potential outliers (Figures C3).

Cluster 1 is a relatively large group comprising 1,207 households. It is characterised by a low overall consumption level and two daily peaks, occurring around noon and in the evening, with the evening peak being more pronounced.

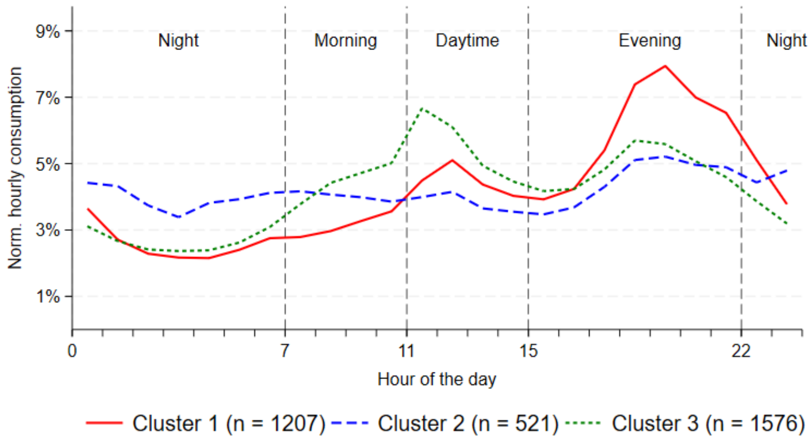
[8] The second approach, which consists in first partitioning the data by season and by day of the week, and then performing the clustering independently for each period, yields approximately the same results. Based on the clustering validity indices, the optimal number of clusters does not vary significantly across different seasons or days of the week, and a three-cluster solution is selected in most cases. These results indicate that clustering by season or by day of the week does not lead to the emergence or disappearance of clearly distinct consumption groups, but rather suggests that the underlying structure of consumption patterns remains stable over time. Details of these results are available upon request.

Figure 1: Consumption by cluster - Main clustering

(a) Levels (kWh)



(b) Normalised



Cluster 2 is a smaller cluster (521 households) exhibiting a substantially higher consumption level, while remaining relatively stable throughout the day. Electricity use remains high during the night, with a small peak in the morning and a larger peak in the evening. **Cluster 3** is the largest cluster, with 1,576 households. It displays a relatively low consumption level and two daily peaks at noon and in the evening, with the midday peak being slightly more pronounced than the evening one.

Overall, consumption across the three clusters differs significantly in terms of levels: Cluster 2 appears to consume three to four times more than the other two, and in a more stable manner, with fewer peaks in consumption. Cluster 1 and Cluster 3 differ more significantly in terms of the timing of their consumption peaks.

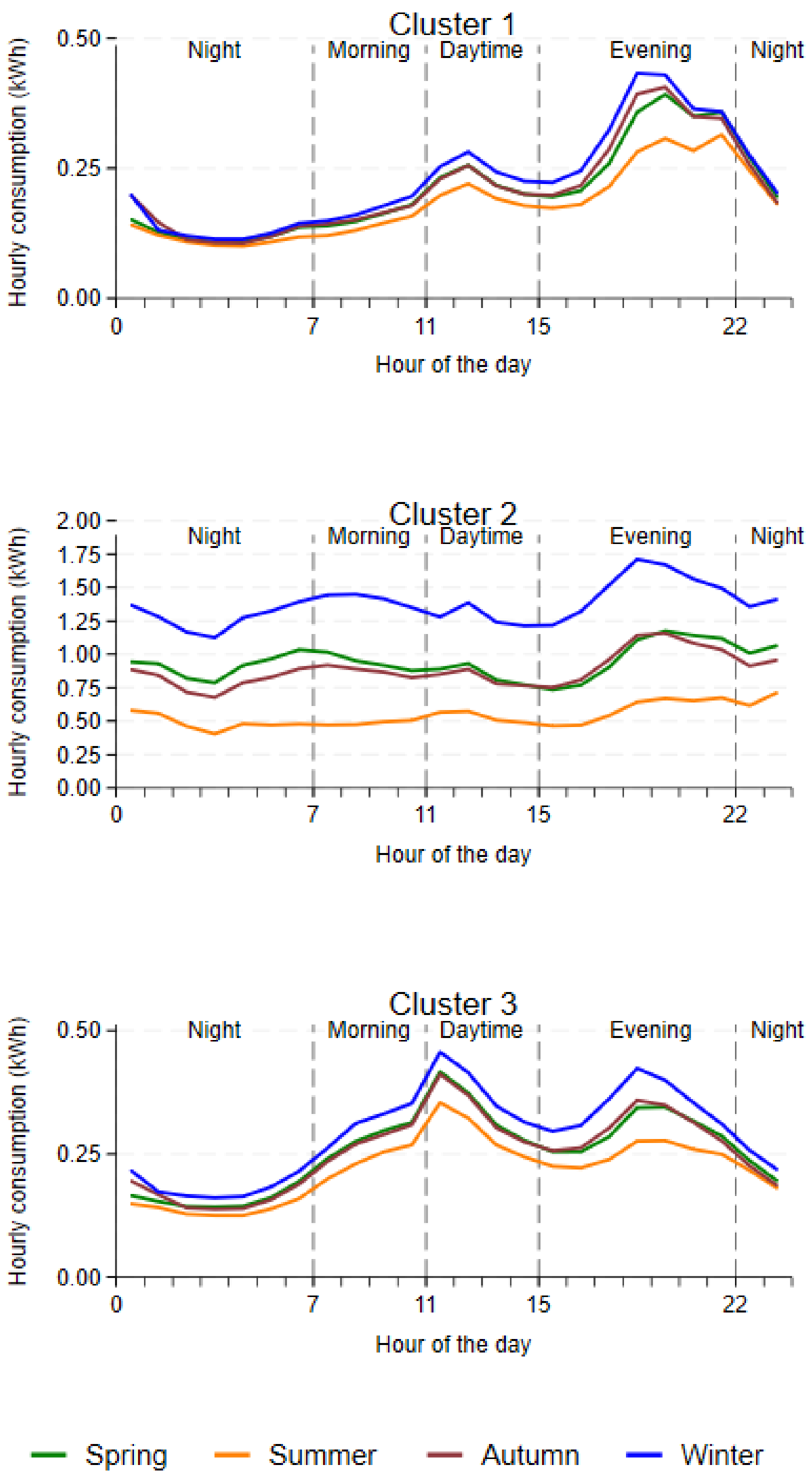
The three identified clusters are then analysed in greater detail by examining their electricity consumption, decomposed by season and day of the week, to assess potential differences in both the level and the shape of load profiles throughout the year and across the week.

Consumption levels vary significantly throughout the year for the three clusters, as illustrated by the seasonal load profile graphs (Figures 2). In particular, household consumption tends to be higher in winter and lower in summer, with similar intermediate levels observed during both autumn and spring. For Cluster 1, this difference in consumption is more pronounced during the evening hours; for Cluster 2, the variation spans the full 24-hour period; and for Cluster 3, it is noticeable throughout the entire daylight period. The magnitude of variability between the season with the lowest and the highest consumption reaches up to approximately 33% for Clusters 1 and 3, but is even greater for Cluster 3, with consumption levels up to three times higher in winter compared to summer, for example.

Contrary to consumption levels, usage patterns remain relatively stable, as confirmed by the normalised load profiles (Figures C4). The only notable changes are a slight reduction in the normalised evening consumption in summer for Clusters 1 and 3, and a decrease during daylight hours for Cluster 3, compensated by an increase in the evening period.

Clear differences can be observed between weekdays and weekend days, particularly for Cluster 1, whose households consume more during the daytime on weekends (up to 30%), while evening consumption remains relatively stable, except on Friday and Saturday evenings, where it appears to be slightly lower. For Clusters 2 and 3, differences across the days of the week remain limited (Figure C5).

Figure 2: Load profiles (kWh) by season - Main clustering



4.2 Multinomial logit model - Main clustering

We then apply a multinomial logit model to assess whether the three main variables available for all households help explain cluster membership by estimating the probability of belonging to each cluster.

Table 2 presents the results. Households living in flats, without electric heating or heat pumps but equipped with photovoltaic (PV) systems, are significantly more likely to belong to Cluster 1. In contrast, households living in houses with electric heating or heat pumps and PV systems are significantly more likely to belong to Cluster 2. Finally, households living in houses without electric heating, heat pumps, or PV systems are significantly more likely to be part of Cluster 3. All estimated coefficients are statistically significant at the 1% level, except for the coefficient for PV in Cluster 1, which is only significant at the 10% level.

4.3 Multinomial logit model - Main clustering subgroup

We then apply the same multinomial logit model to the subsample of 158 households for which a broader set of dwelling and household characteristics is available. The results are reported in Table 3.

Table 2: Multinomial-logit model - Averaged Marginal Effect - Main clustering

	Cluster 1	Cluster 2	Cluster 3
Dwelling (1 = House)	-0.301***	0.201***	0.100***
	(0.017)	(0.015)	(0.020)
Elec. Heating/HP	-0.254***	0.505***	-0.251***
	(0.025)	(0.030)	(0.030)
Presence of PV	0.343***	0.055*	-0.398***
	(0.051)	(0.031)	(0.039)
Cluster size	1207	520	1575

Notes: Standard errors in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Marginal effects computed at the sample means (discrete change from the base level for binary variables).

Table 3: Multinomial logit model - Average marginal effects - Main clustering subgroup

	Cluster 1	Cluster 2	Cluster 3
Dwelling (1 = House)	-0.226*	0.260**	-0.034
	(0.126)	(0.131)	(0.146)
HH size: 2	0.122	-0.079	-0.043
	(0.131)	(0.209)	(0.208)
HH size: 3+	0.081	0.048	-0.130
	(0.171)	(0.303)	(0.251)
Child under 18	0.139	-0.029	-0.111
	(0.167)	(0.196)	(0.160)
Presence of PV	0.187	0.042	-0.229***
	(0.121)	(0.127)	(0.086)
Elec. heating/HP	-0.273***	0.338**	-0.065
	(0.096)	(0.158)	(0.136)
Hot Water (Elec. Source)	-0.122	0.269*	-0.148
	(0.106)	(0.141)	(0.108)
Electrical car	-0.150	0.244	-0.094
	(0.111)	(0.179)	(0.150)
High-energy appliance use	-0.036***	0.012	0.023*
	(0.014)	(0.013)	(0.014)
Age (40-59)	-0.022	-0.187	0.209*
	(0.112)	(0.130)	(0.115)
Age (60+)	-0.081	-0.051	0.133
	(0.147)	(0.179)	(0.139)
Employed	0.296***	-0.041	-0.255*
	(0.085)	(0.158)	(0.151)
Education = Middle	-0.029	-0.045	0.074
	(0.120)	(0.155)	(0.121)
High Education	-0.015	-0.099	0.114
	(0.109)	(0.122)	(0.105)
High trust in gov.	0.035	0.153	-0.188*
	(0.093)	(0.111)	(0.096)
High biospherism	0.002	-0.198*	0.196*
	(0.095)	(0.107)	(0.110)
Cluster size	52	59	47

Notes: Standard errors in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Marginal effects computed at the sample means (discrete change from the base level for binary variables).

Residing in a dwelling without electric heating or heat pumps, exhibiting lower usage of high-energy appliances, and a high share of employed individuals are strongly increase the likelihood of belonging to Cluster 1. Living in a flat (i.e., not residing in a house), is also positively associated with Cluster 1 membership with weaker statistical significance (10% level, $p - value = 0.076$). The coefficient associated with owning photovoltaic

(PV) panels is sizeable, but narrowly fails to reach statistical significance at the 10% level ($p - value = 0.12$). Nevertheless, the sign and direction of the coefficient are consistent with the results obtained for the full sample and remain positive.

Residing in a house, having electric heating or heat pumps, having an electric hot water system, and exhibiting lower biospherism scores are all factors that are associated with a relatively higher likelihood of belonging to Cluster 2.

Finally, the absence of photovoltaic (PV) systems is strongly associated with belonging to Cluster 3. In addition, higher use of high-energy appliances, being middle-aged (40–59), being non-employed, exhibiting lower trust in government, and having higher biospherism scores are associated with higher likelihood of belonging to Cluster 3, with coefficients statistically significant at the 10% level.

5 Discussion

This cluster analysis of electricity consumption patterns among approximately 4,000 households allows us to draw several noteworthy conclusions.

First, electricity consumption varies significantly across households, both in terms of overall levels and usage patterns. This heterogeneity enables the identification of distinct and meaningful consumption clusters that can be used to better design demand-side tools.

Second, we identify three distinct types of electricity-consuming households based on their consumption profiles. Cluster 1 consists mainly of households living in flats and highly engaged in professional activities. These households are typically absent during the day and rely less on high-consumption appliances. This is reflected in their low overall electricity use and limited consumption peaks, mostly occurring in the evening. This consumption pattern, characterised by notably lower electricity use during the afternoon, is consistent with the higher likelihood of photovoltaic (PV) ownership in this group, as on-site generation reduces net electricity consumption during production hours. We refer to this group as *active flat dwellers*. By contrast, Cluster 2 includes households living in large houses equipped with energy-intensive technologies, such as heat pumps or electric hot water systems. These households exhibit the highest levels of electricity consumption, with sustained demand throughout the day. They also display greater seasonal variation in electricity use, which is consistent with the higher prevalence of heat pumps in this group. We label this group as *high-consumption households*. Cluster 3 consists of households that are less likely to have heat pumps or electric water heating, a pattern similar to that observed in Cluster 1. However, they tend to consume slightly more electricity, likely due to being less engaged in employment and spending more time at home. They also use appliances more frequently. This last group appears to exhibit more stable consumption patterns over time, with limited variation across both seasons and days of the week, at least less than what is observed for Cluster 1. We refer to this group as *home-based moderate consumers*.

Third, dwelling characteristics appear to play a more significant role in determining cluster membership than socio-demographic variables. Thus, the presence of heat pumps, electric hot water systems, or photovoltaic panels strongly influences cluster membership and clearly distinguishes high-consumption households from others.

Employment status, and potentially also time spent at home (despite the absence of a variable related to remote work), appears to be the key socio-demographic factor in predicting cluster membership that is not directly related to dwelling characteristics. Although only the age category 40–59 is statistically significant in predicting membership in Cluster 3 (and not the 60+ category), the sign and magnitude of the coefficients suggest that older households are more likely to belong to Cluster 3, while younger households tend to be in Clusters 1 and 2. The limited number of observations may reduce the statistical power needed to detect these effects with significance. This pattern could also help explain the results for the employment status variable, with a higher proportion of retired individuals in Cluster 3, who are likely to spend more time at home, compared to more active, working individuals in Cluster 1.

Finally, variables such as dwelling size and the presence of children do not have statistically significant effects on cluster membership. However, the sign and magnitude of the coefficients suggest that households with children under 18 and larger household sizes are more likely to belong to Cluster 1 but not to Cluster 3, while no clear pattern is observed for Cluster 2. This suggests that household size and related demographic factors may still play a role in differentiating consumption behaviours among lower-consumption households (Cluster 1 versus Cluster 3), once major structural drivers of energy demand - such as the presence of heat pumps - have already separated Cluster 2 from Clusters 1 and 3.

Fourth, normative orientations appear to play a limited role in explaining cluster membership, even if trust in government and biospheric values seem to be associated with membership in Clusters 2 and 3. Indeed, these results should be interpreted with caution, given how these variables are self-reported, and further analysis would be needed to better understand the underlying mechanisms. Education level does not appear to be associated with cluster membership in any case.

Fifth, although there are substantial differences in electricity consumption levels across seasons and times of the year, these variations do not reflect major changes in behavioural patterns. The overall structure of daily and weekly consumption remains relatively stable, suggesting that seasonality primarily affects intensity rather than usage habits.

Overall, these findings highlight the importance of accounting for heterogeneity in residential electricity demand when designing demand-side policies, as different household segments may respond differently to targeted interventions.

6 Conclusion

The clustering of approximately 4,000 households reveals clear differences in electricity consumption levels and usage patterns across three clusters, suggesting that demand-side measures and tariff structures could be more effective if adapted to specific household profiles.

For instance, households in Cluster 2 exhibit substantial seasonal variation in electricity consumption, should be taken into account when considering incentives or dynamic pricing schemes. By contrast, households in Cluster 3 display more stable consumption patterns over time, which may indicate that changes in their electricity use require more sustained or longer-term efforts.

Another example is Cluster 1, which appears to consist of externally active individuals. Therefore, asking them to reduce consumption during midday peak hours or to shift their usage to coincide with photovoltaic (PV) production hours is more challenging, as they are often not at home during those times. Identifying ways to assist them in shifting or scheduling the consumption of certain appliances or activities could offer a viable solution.

Across clusters, dwelling equipment appears to play a central role in shaping electricity consumption patterns. This suggests that a stronger focus on households equipped with energy-intensive or electricity-producing technologies, potentially including photovoltaic systems, could be relevant. In this context, arrangements encouraging local electricity sharing between households producing electricity and non-producing households with higher electricity needs represents a promising avenue, although their effectiveness would require further assessment.

Despite these insights, several limitations and avenues for further research remain. First, capturing differences in electricity consumption patterns could be further improved, particularly for households exhibiting pronounced night-time consumption. While the current approach allows for the identification of night consumers, more detailed information and the use of more refined clustering techniques could help better distinguish and characterise these consumption profiles. Fuzzy k-means could also be explored to capture households with mixed consumption patterns or uncertain cluster membership.

Second, the analysis linking consumption clusters to socio demographic characteristics relies on a relatively small subsample, which limits statistical power. To strengthen external validity, future research would benefit from larger samples and from including households located in different regions of Switzerland.

Finally, information on the level of self-consumption is not available in the dataset. While self-consumed electricity plays a different role for electricity providers and public authorities than grid-supplied consumption, it remains highly relevant for understanding when and how much electricity households consume, and at what times. Access to such information would make it possible to better assess consumption behaviour under varying weather conditions, such as sunny versus non-sunny days, and to better anticipate demand patterns when photovoltaic production is low or absent.

CRediT authorship contribution statement

Valentin Favre-Bulle: Writing – original draft, Writing – review & editing, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization
Sylvain Weber: Writing – review & editing, Validation, Supervision, Project administration, Funding acquisition, Conceptualization.

Data availability statement

The data that support the findings of this study are not publicly available due to confidentiality agreements with the electricity provider but are available from the authors upon reasonable request and with permission from the data provider.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process.

During the preparation of this work, the author(s) used ChatGPT (OpenAI) to improve the clarity, coherence, grammar, and overall fluency of the text. After using this tool, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

Acknowledgements

This work was funded by the Swiss Federal Office of Energy under contract number Sl/502530-01 (project INFINEED).

Appendix

A Descriptive statistics

Table A1: Descriptive statistics (Main clustering) - Total

	Overall
Sample size	3,304
Presence of PV	
No	3,169 (95.9%)
Yes	135 (4.1%)
Type of heating	
Other	2,933 (88.8%)
Elec. Heat./HP	371 (11.2%)
Type dwelling	
Flat	2,084 (63.1%)
House	1,220 (36.9%)

Table A2: Descriptive statistics (Main clustering) - Subgroup

	Subgroup
Sample size	158
Type of dwelling	
Flat	51 (32.3%)
House	107 (67.7%)
Household size	
1	23 (14.6%)
2	60 (38.0%)
3+	75 (47.5%)
Child under 18	
No	102 (64.6%)
Yes	56 (35.4%)
Presence of PV	
No	118 (74.7%)
Yes	40 (25.3%)
HP or elect. heating	
No	120 (75.9%)
Yes	38 (24.1%)
Hot water by elec.	
No	100 (63.3%)
Yes	58 (36.7%)
Presence of electric car	
No	139 (88.0%)
Yes	19 (12.0%)
Use of high-energy appliances	15.0
Age of the resp.	
18-39	44 (27.8%)
40-59	72 (45.6%)
60+	42 (26.6%)
Employment status	
No	37 (23.4%)
Yes	121 (76.6%)
Education level	
Low	47 (29.9%)
Middle	33 (21.0%)
High	77 (49.0%)
Trust in government	
Low	99 (62.7%)
High	59 (37.3%)
Biospherism	
Low	45 (28.5%)
High	113 (71.5%)

B Primary clustering

The first clustering is performed on the whole data set and according to the two methods of clustering (k-means and k-medians as robustness check). Based on clustering validity criteria for selecting the number of clusters, the optimal solution consists of two groups (see Figure B1). One of the groups consists of 772 households and exhibits high consumption levels and percentages at night but a very similar one as the other bigger group during the day. Especially this cluster exhibits clear increase in consumption at some specific time of the day (9 p.m. or 11 p.m.) (see Figure B2). A large share of these households own a hot water heater that works during the night; some others have an electric heating system that also works during the night, while others have a heat pump that, depending on the settings made by the households, works during the night. The logit model seems to confirm this hypothesis as having an electrical heating and having PVs is significantly correlated with membership to Cluster 1 (see Table B1).

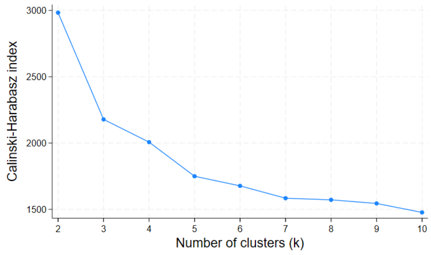
These pronounced consumption peaks during the night appear to capture most of the differences between households. As a result, households without such peaks cannot be distinguished from one another, which keeps the optimal number of clusters at two. To improve this situation, we tested additional specifications by introducing a seventh feature specifically designed to capture nighttime consumption peaks. The aim was to better group these specific consumers and to strengthen the differentiation among the remaining households. These tests were not conclusive and did not improve the clustering results, so we decided to discard the seventh feature

Since clustering results are more informative with more than two groups, we treat these 772 households as a distinct "Night Consumers" (NC) group. We then perform a separate clustering on the remaining 3,304 households, excluding NCs [9].

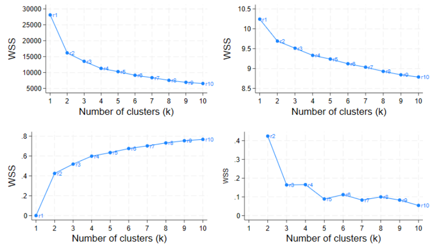
[9] To the 3,276 households in Cluster 1, we also add 28 households from Cluster 2 that belong to the subgroup for which more detailed data are available, in order to increase the statistical power of the multinomial model analysis.

Figure B1: Optimal Number of Clusters – Primary clustering (K-means)

(a) Calinski index



(b) WSS



(c) Silhouette index

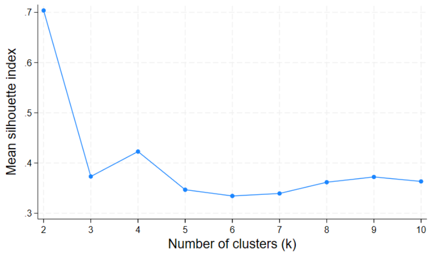
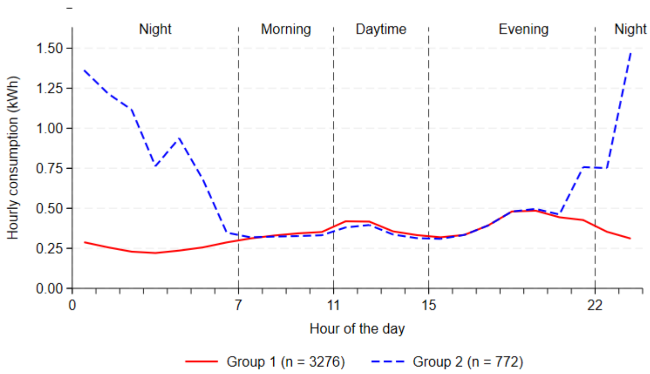


Figure B2: Consumption by cluster - Primary clustering

(a) Levels (kWh)



(b) Normalised

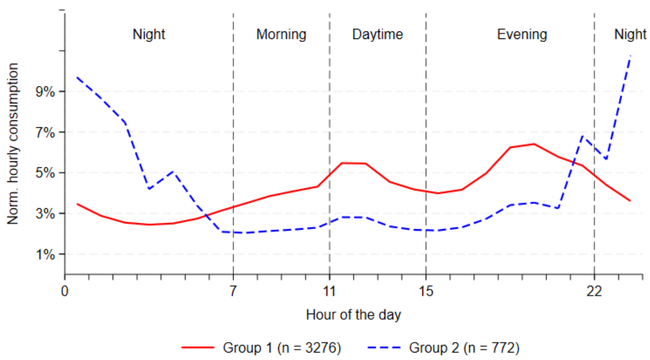
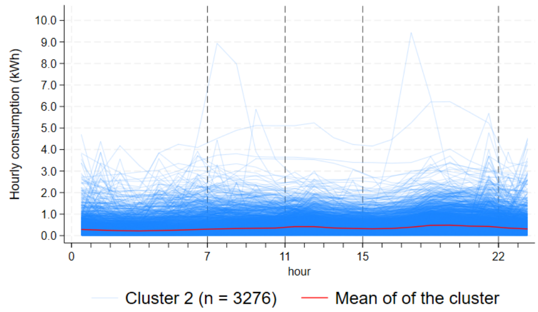
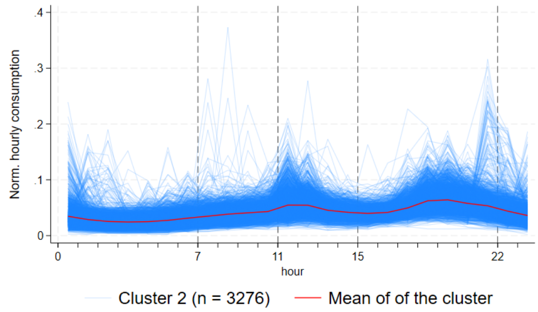


Figure B3: Individual load curves - Primary clustering

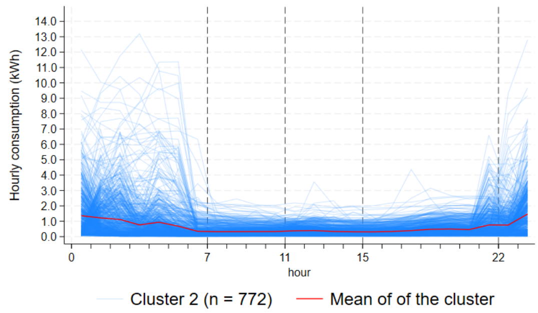
(a) Cluster 1 (in kWh)



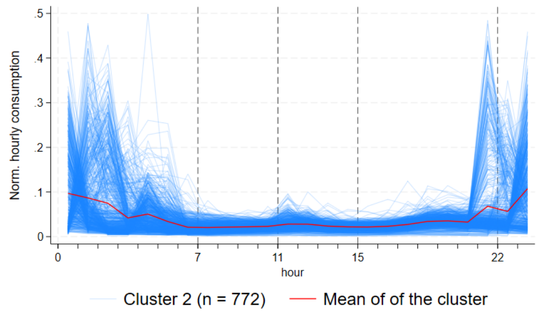
(b) Cluster 1 (normalised)



(c) Cluster 2 (in kWh)



(d) Cluster 2 (normalised)



Note: To improve visual clarity, households with any normalised hourly consumption values in the top 0.5% were excluded.

Figure B4: Multinomial-logit model - Averaged Marginal Effect - Primary clustering

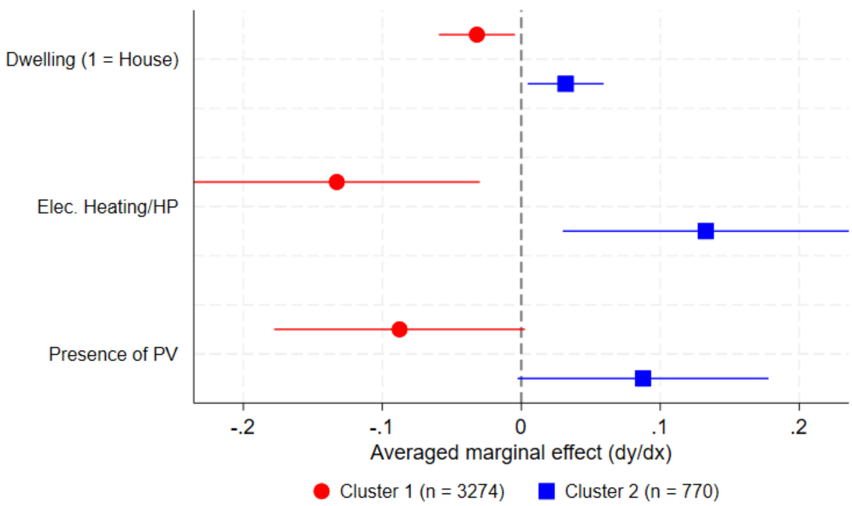


Table B1: Multinomial-logit model - Averaged Marginal Effect - Primary clustering

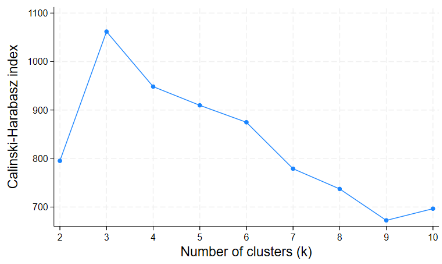
	Cluster 1	Cluster 2
Dwelling (1 = House)	-0.032**	0.032**
	(0.014)	(0.014)
Elec. Heating/HP	-0.133***	0.133***
	(0.022)	(0.022)
Presence of PV	-0.088***	0.088***
	(0.033)	(0.033)
Cluster size	3274	770

Notes: Standard errors in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Marginal effects computed at the sample means (discrete change from the base level for binary variables).

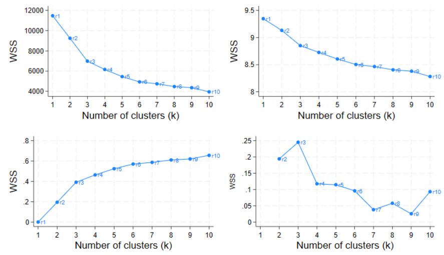
C Main clustering

Figure C1: Optimal Number of Clusters – Main clustering (K-means)

(a) Calinski index



(b) WSS



(c) Silhouette index

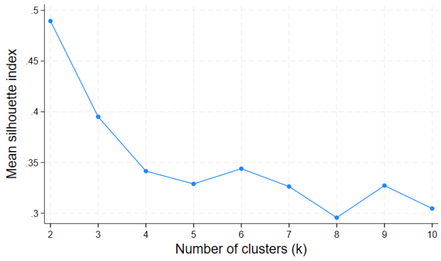
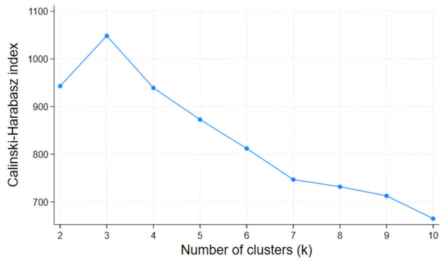
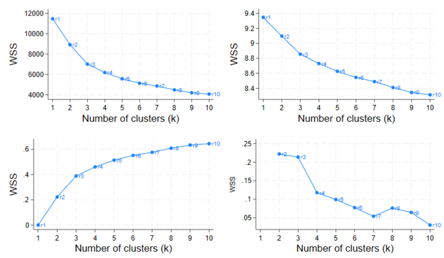


Figure C2: Optimal Number of Clusters – Main clustering (K-medians)

(a) Calinski index



(b) WSS



(c) Silhouette index

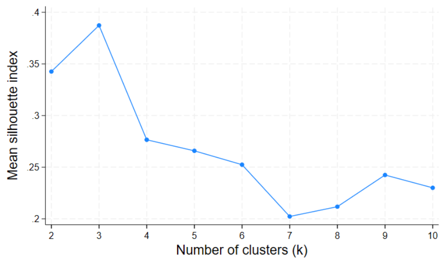
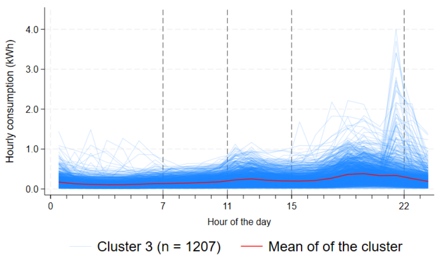
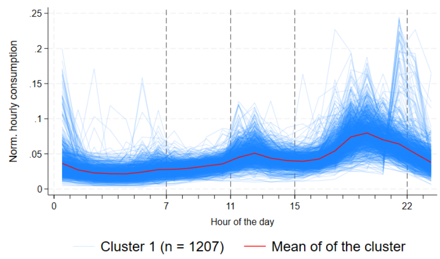


Figure C3: Individual load profiles - Main clustering

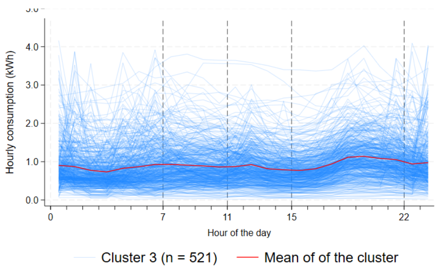
(a) Cluster 1 (in kWh)



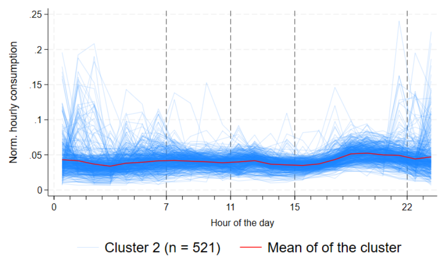
(b) Cluster 1 (normalised)



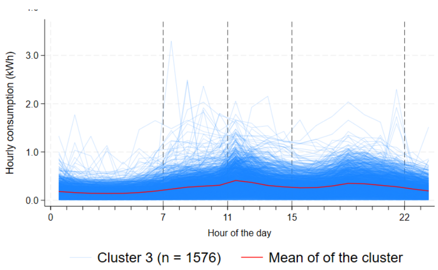
(c) Cluster 2 (in kWh)



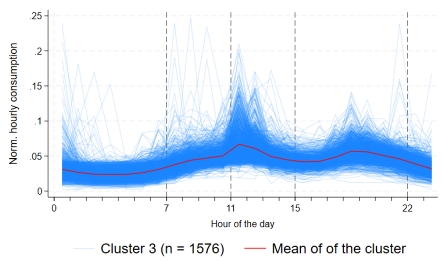
(d) Cluster 2 (normalised)



(e) Cluster 3 (in kWh)



(f) Cluster 3 (normalised)



Note: To improve visual clarity, households with any normalised hourly consumption values in the top 0.5% were excluded.

Figure C4: Normalised load profiles by season - Main clustering

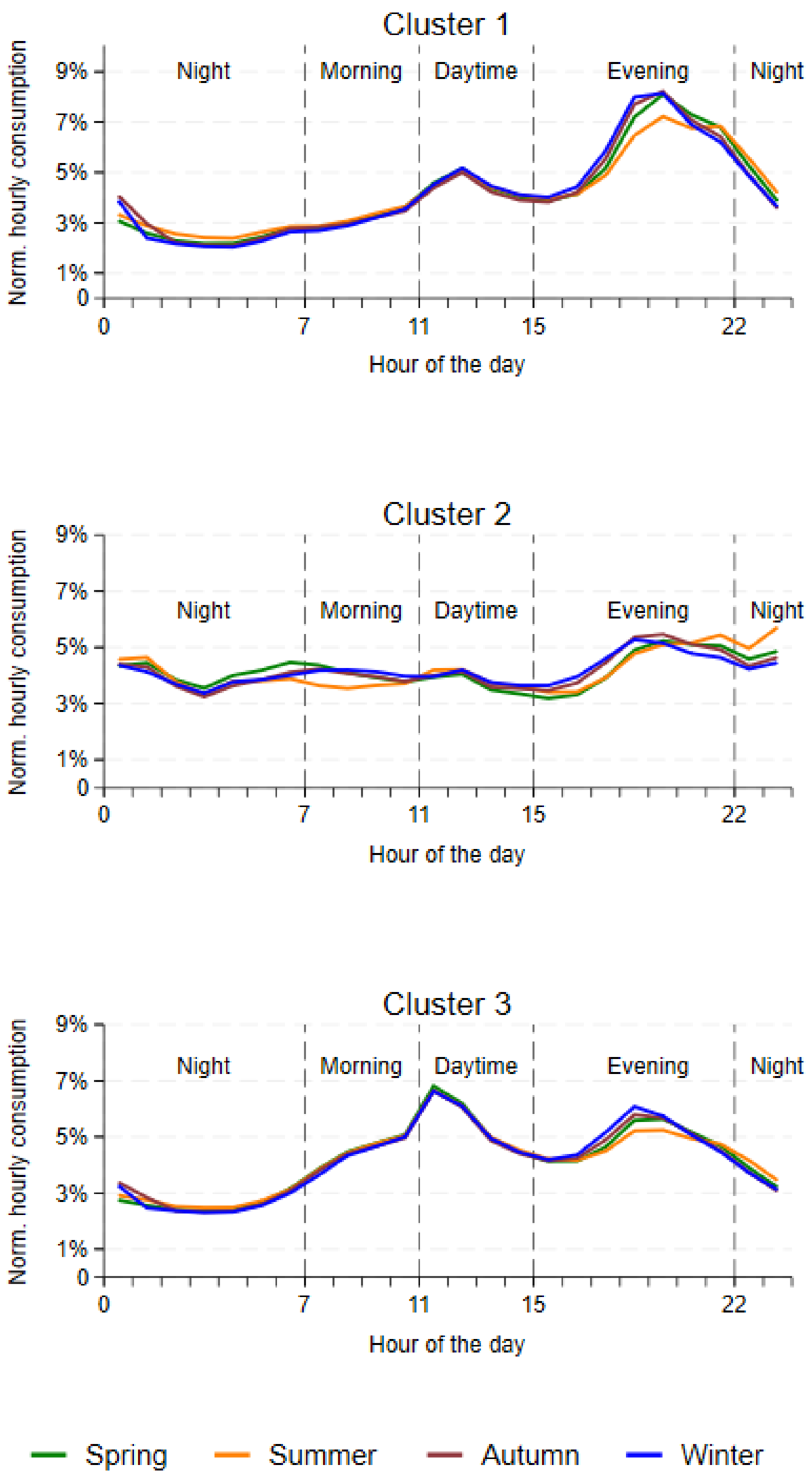
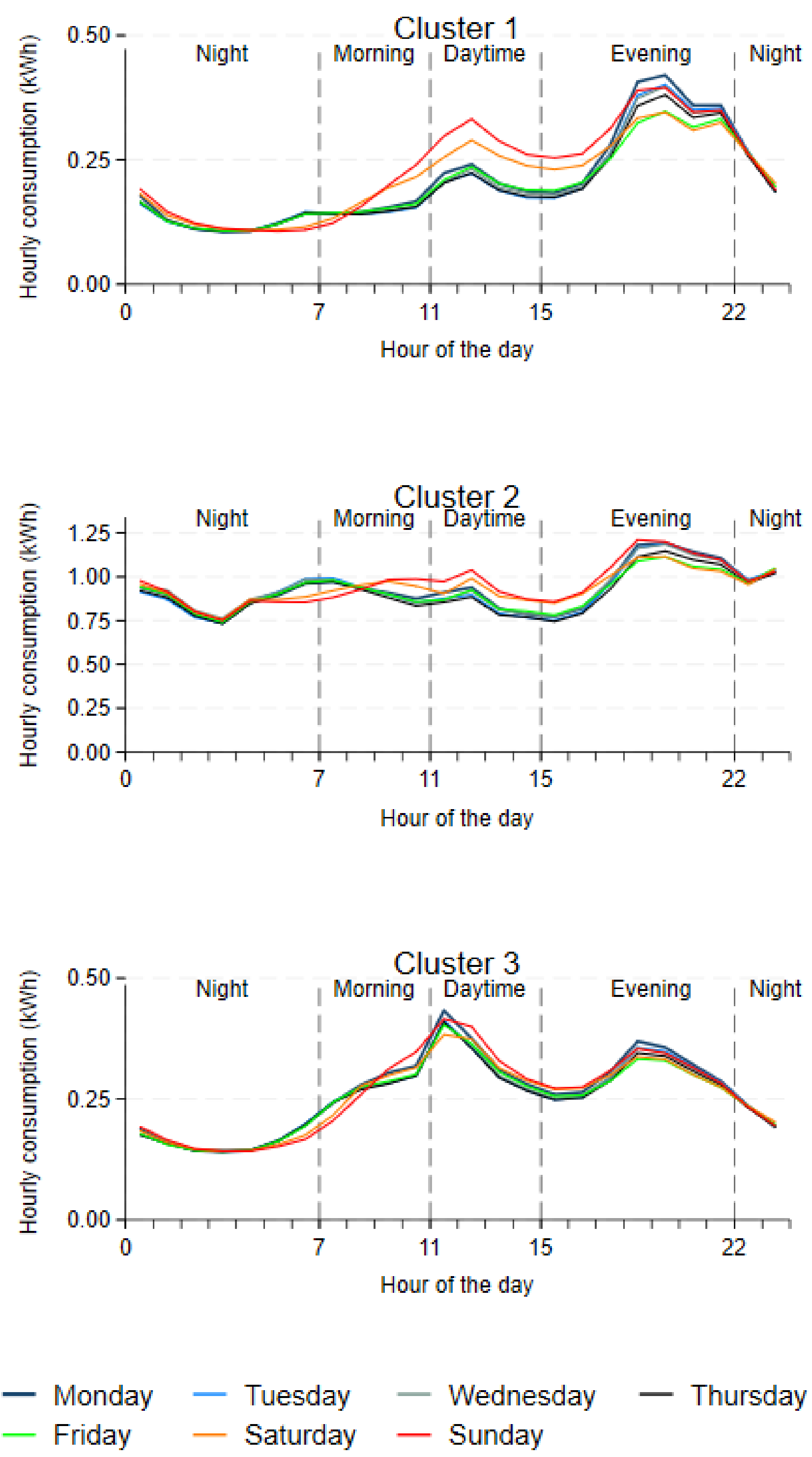


Figure C5: Load profiles (kWh) by day of the week - Main clustering



References

[1] SFOE, Schweizerische Elektrizitätsstatistik 2024 / Statistique suisse de l'électricité 2024, Electricity statistics, Swiss Federal Office of Energy (SFOE) (2024).
URL https://www.bfe.admin.ch/bfe/en/home/supply/statistics-and-geodata/energy-statistics/electricity-statistics.html/#tab__content_bfe_en_home_versorgung_statistik-und-geodaten_energiestatistiken_elektrizitaetsstatistik_jcr_content_par_tabs

[2] F. Pelletier, A. Faruqui, Does dynamic pricing work in a winter-peaking climate? A case study of hydro Quebec, The Electricity Journal 35 (2) (2022) 107080. doi:https://doi.org/10.1016/j.tej.2022.107080.
URL https://www.sciencedirect.com/science/article/pii/S1040619022000069

[3] B. Yildiz, J. Bilbao, J. Dore, A. B. Sproul, Recent advances in the analysis of residential electricity consumption and applications of smart meter data, Applied Energy 208 (2017) 402–427. doi:https://doi.org/10.1016/j.apenergy.2017.10.014.
URL https://www.sciencedirect.com/science/article/pii/S0306261917314265

[4] G. Chicco, Overview and performance assessment of the clustering methods for electrical load pattern grouping, Energy 42 (2012) 68–80. doi:https://doi.org/10.1016/j.energy.2011.12.031.
URL https://www.sciencedirect.com/science/article/pii/S0360544211008565

[5] A. Rajabi, M. Eskandari, M. J. Ghadi, L. Li, J. Zhang, P. Siano, A comparative study of clustering techniques for electrical load pattern segmentation, Renewable and Sustainable Energy Reviews 120 (2020) 109628. doi:https://doi.org/10.1016/j.rser.2019.109628.
URL https://www.sciencedirect.com/science/article/pii/S1364032119308354

[6] S. Yilmaz, S. Weber, M. K. Patel, Who is sensitive to DSM? Understanding the determinants of the shape of electricity load curves and demand shifting: Socio-demographic characteristics, appliance use and attitudes, Energy Policy 133 (2019) 110909. doi:https://doi.org/10.1016/j.enpol.2019.110909.
URL https://www.sciencedirect.com/science/article/pii/S0301421519304872

[7] S. Haben, C. Singleton, P. Grindrod, Analysis and clustering of residential customers energy behavioral demand using smart meter data, IEEE Transactions on Smart Grid 7 (2016) 136–144. doi:https://doi.org/10.1109/TSG.2015.2409786.
URL https://ieeexplore.ieee.org/document/7063233

[8] J. Zethmayr, R. S. Makhija, Six unique load shapes: A segmentation analysis of Illinois residential electricity consumers, The Electricity Journal 32 (9) (2019) 106643. doi:https://doi.org/10.1016/j.tej.2019.106643.
URL https://www.sciencedirect.com/science/article/pii/S1040619019301800?via%3Dihub

[9] F. McLoughlin, A. Duffy, M. Conlon, A clustering approach to domestic electricity load profile characterisation using smart metering data, Applied Energy 141 (2015) 190–199. doi:https://doi.org/10.1016/j.apenergy.2014.12.039.
URL https://www.sciencedirect.com/science/article/pii/S0306261914012963

[10] C. Beckel, L. Sadamori, T. Staake, S. Santini, Revealing household characteristics from smart meter data, Energy 78 (2014) 397–410. doi:https://doi.org/10.1016/j.energy.2014.10.025.
URL https://www.sciencedirect.com/science/article/pii/S0360544214011748

[11] A. Albert, R. Rajagopal, Smart meter driven segmentation: What your consumption says about you, IEEE Transactions on Power Systems 28 (2013) 4019–4030. doi:https://doi.org/10.1109/TPWRS.2013.2266122.
URL https://ieeexplore.ieee.org/document/6545387

[12] T. Räsänen, M. Kolehmainen, Feature-based clustering for electricity use time series data, in: M. Kolehmainen, P. Toivanen, B. Beliczynski (Eds.), Adaptive and Natural Computing Algorithms, Vol. 5495, Springer, Berlin, Heidelberg, 2009, pp. 401–412. doi:https://doi.org/10.1007/978-3-642-04921-7_41.
URL https://link.springer.com/chapter/10.1007/978-3-642-04921-7_41

[13] S. Yilmaz, J. Chambers, M. K. Patel, Comparison of clustering approaches for domestic electricity load profile characterisation - Implications for demand side management, Energy 180 (2019) 665–677. doi:https://doi.org/10.1016/j.energy.2019.05.124.
URL https://www.sciencedirect.com/science/article/pii/S0360544219310060

[14] A. Makles, Stata tip 110: How to get the optimal k-means cluster solution, The Stata Journal 12 (2012) 347–351. doi:https://doi.org/10.1177/1536867X1201200213.
URL https://doi.org/10.1177/1536867X1201200213

[15] J. D. Rhodes, W. J. Cole, C. R. Upshaw, T. F. Edgar, M. E. Webber, Clustering analysis of residential electricity demand profiles, Applied Energy 135 (2014) 461–471. doi:https://doi.org/10.1016/j.apenergy.2014.08.111.
URL https://www.sciencedirect.com/science/article/pii/S0306261914009349

[16] J. L. Viegas, S. M. Viera, R. Melicio, V. M. F. Mendes, J. M. C. Sousa, Classification of new electricity customers based on surveys and smart metering data, Energy 107 (2016) 804–817. doi:https://doi.org/10.1016/j.energy.2016.04.065.
URL https://www.sciencedirect.com/science/article/pii/S0360544216304789