Seminar Paper - Advanced Topics
Testing standard load profiles with real world data from Germany

Jan Besler - 5629079

September 30, 2022

Abstract

Real quarter data is used to investigate how comparable the standard load profiles (SLP) (provided by BDEW) are with buildings in the observed area to improve the accuracy of the SLP. The quarter includes households, office buildings, and a factory, to ensure that the SLP $H0, G1$ and $G5$ can be investigated and evaluate how well they match real cases. Various time series clustering methods are used for the evaluation, including k-means and fuzzy c-means as well as dynamic warping distance measures. The results demonstrate that BDEW's rigid models do not stand up to real-world data, such as the insignificance of distinguishing weekends for different seasons and the need for a further breakdown of weekdays across all seasons.

1 Introduction

With an ever increasing share of renewable energies in Germany, both the volatility of the power grid [1] and consumer behaviour [2] are changing. Therefore, the models currently used are no longer as well suited to describe the energy industry and its processes as they were when they were introduced a few decades ago. This also applies to the SLP produced by the Bundersverband der Energie- und Wasserwirtschaft (BDEW) - formerly VDEW - to estimate electricity consumption for consumers with an electricity consumption of less than 100,000 kWh per year. These SLPs were produced for households ($H0$), businesses with working hours from 8am to 6pm on weekdays ($G1$) and bakeries ($G5$) which are relevant for this work[1]. The SLP are intended to be an approximation of the actual electricity consumption profile, yet they may well be inaccurate. The 1999 BDEW report [3] states that variations of -10% up to $+30\%$ can be expected. This variation was confirmed by Stromnetz Berlin GmbH. According to their experience, the differences can fluctuate largely depending on the weather [4].

In this work, the SLPs are examined to see how well they match real data. The results are not expected to match, but an approximation should be apparent. The SLP data are compared with a district in the 'Oststadt' in Karlsruhe, Germany. The Smart East project, that provides the data, monitors the electricity consumption of a quarter. The area includes a multi-generation

[1] the full list of SLPs can be found in Table 8 in the Appendix

house, an office building housing several start-ups, a research institute, and a factory, to ensure a comparison with several SLPs is possible.

The analysis is first performed with descriptive statistics and then extended using time series clusters to identify characteristics for different load profiles. Results are then tested and compared with the BDEW data. These can be used to challenge and improve the SLP, with the aim of having more accurate SLPs and improving forecasts for all stakeholders. Thus reducing the costs associated with balancing the grid and over- and under-production of electricity for varying hours of the day.

2 Related Work

This paper tries to combine two approaches from other papers, namely publications using time series clustering with energy relevant data and those using raw real electricity data to find patterns in electricity consumption. Good basics and overviews of the possibilities of time series clustering by *Liao* [5], provides an overview of commonly used techniques and approaches for time series clustering. Another paper overviewing the use of time series clustering in multiple fields was published by *Aghabozorgi et al.* [6] and explains the use cases for research areas ranging from biology to finance. More specific work,published by *Motlagh et al.* [7], looks at finding load patterns for household electricity consumption. *Košmelj et al.* [8] used time series data over a six-year period to investigate whether European countries are comparable in their per capita electricity consumption. Different approaches to clustering time series and exploring possible energy patterns in buildings were carried out by *Iglesias et al.* [9]. The study used various distance measures to find load profiles for multiple building types.

Yang et al. [10] publishes analyses with raw real electricity data investigated the use of the k-shape algorithm to find electricity consumption patterns and predict consumption more accurately for different government buildings. Another study by *McLoughlin et al.* [11] looks at finding new patterns associated with household electricity consumption over a six-month period using the clustering methods k-means, k-mediod, and self-organising maps.

3 Methodology

The aim of this work is to determine whether the aforementioned SLPs ($H0, G1, G5$) can explain the data from a real urban district. For this purpose, the data are first examined with deterministic statistical methods, which also include the normalisation of the respective time series in order to be able to compare time series of different magnitudes. The normalised time series are based on the average values for each hour over the entire available data. To detect outliers, the same steps are repeated, using the median instead of the average for the Smart East data and then comparing it to the average. If the median and average show no statistically significant differences, it can be assumed that there are no major outliers. The BDEW data are excluded from the search for outliers because the data points are designed to correspond to a specific curve and therefore the differences are negligible.

Further analysis is carried out using clustering methods to identify groups of similar days.

These groups are then examined and similarities are sought for each cluster to determine which, if any, SLP match the clusters. Since the data is a time series, special properties need to be considered and assumptions made. These include the use of the extended Dickey-Fuller test, which checks whether the data exhibit drift, to determine whether the data are stationary or not. If the data are not stationary but show drift, further steps must be taken to eliminate the drift.

The next step is to check the standardised curves for the BDEW data and the Smart East data for matching curves to get a first indication of which load curve belongs to which BDEW SLP. For further investigation, two clustering algorithms are used, k-means and fuzzy c-means clustering. K-means was chosen to have a robust clustering technique as a reference for the fuzzy c-means method, which is expected to perform better than K-means due to its ability to dump non-matching time series into a dumping cluster. The number of centroids is selected using three algorithms commonly mentioned in the literature, the elbow method [12], the average silhouette method [13] and the gap statistic [14]. These three methods are expected to give similar results for the number of optimal centroids. Several metrics were selected to evaluate the models [15, 16], the first being the silhouette score, which is used to measure the distance between clusters by calculating how close each point in a cluster is to the points in its neighbouring clusters. This measure has a range of $[-1,1]$, where a one means perfectly defined clusters and a zero corresponds to random assignment due to overlapping clusters. The mean distance within a cluster is described by $a(i)$, where M_i denotes the individual points. While $b(i)$ is used to calculate the smallest mean distance of M_i to the points of the other clusters.

$$a(i) = \frac{1}{n_k - 1} \sum_{i' \in I_k} d\left(M_i, M_{i'}\right) \tag{1}$$

$$b(i) = \min_{k' \neq k} \frac{1}{n_{k'}} \sum_{i' \in I_{k'}} d\left(M_i, M_{i'}\right) \tag{2}$$

$$s_k = \frac{1}{n_k} \sum_{i \in I_k} \frac{b(i) - a(i)}{max\left(a(i), b(i)\right)} \tag{3}$$

The variable s_k is the mean value of the silhouette widths for a given cluster.

The second metric is the Calinski-Harabasz index, which is defined as the ratio between the dispersion within a cluster and the dispersion between clusters, with a higher index indicating denser and better separated clusters. Here, the numerator of the second fraction describes the dispersion between clusters, where $G^{\{k\}}$ is the barycentre of each cluster and G is the barycentre of the whole data.

$$C = \frac{N-K}{K-1} \cdot \frac{\sum n_k ||G^{\{k\}} - G||^2}{\sum ||M_i^{\{k\}} - G^{\{k\}}||^2} \tag{4}$$

The dispersion within a cluster is given by the denominator of the second fraction as the sum of the squared distances between the points $M_i^{\{k\}}$ and the respective barycentres in the clusters.

The third metric is the Davies-Bouldin index, which is defined as the average similarity measure of each cluster with its most similar cluster to determine the separation between clusters, similar to the silhouette score. The similarity score is measured as a comparison of the distance between clusters and the size of the clusters themselves. Clusters that are denser and further apart will therefore have a lower - in this case better - score, with a minimum score of zero. It is calculated as follows

$$C = \frac{1}{K} \sum_{k=1}^{K} \max_{k' \neq k} \left(\frac{\delta_k + \delta_{k'}}{\Delta_{kk'}} \right) \tag{5}$$

with δ_k as the mean distance between the points belonging to their respective clusters.

The final metric is the Dunn index, which is calculated as the lowest distance within a cluster divided by the largest distance between clusters. The index has a range of $[0, 1]$, with higher values representing better results. It is calculated as follows

$$DI_k = \frac{min\delta\,(C_i, C_j)}{max\Delta_k} \tag{6}$$

where the numerator describes the smallest distance within the cluster and the denominator the largest distance between the clusters, with the advantage that the specific measures can be calculated with arbitrary distance functions.

After determining the best clustering methods according to the metrics introduced, the clusters are manually examined to determine which days are placed in the clusters based on the meta-information introduced earlier. Once the characteristics defining each cluster are determined, the clusters are compared to the existing framework developed by the BDEW to check whether the clusters fit the characteristics defined by the BDEW, including the separation into summer, winter, and transitional seasons.

The clustering techniques chosen for the comparison are the basic k-means algorithm, which acts as a kind of benchmark for the others. In addition, the k-mediods algorithm was chosen, which differs in the handling of the centres by using real data points rather than mean values as in k-means. Another algorithm that uses the Euclidean distance measure is the fuzzy c-means algorithm, which avoids deterministic assignment of time series to clusters but calculates the percentage of data that belong to certain clusters; furthermore, this algorithm can create a dumping cluster to which all data that do not fit into any other cluster are assigned, which is not possible with the previous two algorithms. For the next three clustering techniques, different distance measures were used, including shape-based and dynamic time-warping.

The k-shape [17] algorithm uses a shape-based distance measure and is optimised for clustering time series. The second fuzzy C-means algorithm uses dynamic time warping (DTW) as does the DTW clustering approach. The DTW distance measure is an alternative to the Euclidean distance measure, which solves a minimum cost alignment problem between two time series and is thus more robust to different lengths and shifts of the time series. However, these advantages are accompanied by a higher computational effort[2]. The second fuzzy C-means algorithm works in exactly the same way as the Euclidean algorithm, the only difference being in the calculation of the distances, where the DTW approach is used here to approximate the time series. The DTW clustering also uses the DTW distance measure and its centre evaluation is based on soft-DTW centroids [18].

[2] the computation time increases from three seconds to 3600 seconds on the local computer if DTW is used instead of the Euclidean distance measure.

$$z = \frac{x - \mu}{\sigma} \tag{7}$$

To perform all these clustering algorithms, the R package *dtwclust* with the function *tsclust* was used. In order to make the different time series comparable, the function first applies z-normalisation so that the values in the output do not represent the actual electricity consumption but the normalised values.

4 Evaluation

4.1 Data

The data collected as part of the Smart East project is provided by Stadtwerke Karlsruhe and consists of 15-minute values for each of the collection points. These points include the Mehrgenerationenhaus (MGH), which also houses a bakery and a kindergarten, the Hoepfner area, which mainly consists of a factory and a villa, the Technologiefabrik (TeFak), which houses more than 80 start-ups [19], and the Forschungszentrum Informatik (FZI), which is a research laboratory. Each time series is an aggregation of the electricity consumption of an entire property, so it is not possible to break down the data to individual units, such as households or offices. The data includes measurements from more than 3 years from 2019 to March 2022.

Problem	**Modification Steps**
Mixed time intervals	move all time series to the starting point of midnight $(00:00)$
Shift to German time zone	convert UTC time to German time zone (Berlin)
Incorrect values	identifying using domain knowledge and remove these
Add meta information	add columns for time measures
	like day of the week and heating period

Table 1: Data Engineering Processes.

In order to analyse the data, they first had to be modified and checked for missing or incorrect values. This included shifting the time series depending on whether the timestamp of the measurement was dated to the beginning or the end of the 15-minute interval in order to obtain a coherent representation. In addition, the data had to be standardised with regard to the time zone, which again shifted the time series. When the data sets were examined, it was found that there were both missing and incorrect values. These wrong values were identified and removed. As a result, some days have fewer values, but this impact is considered negligible as the data covers a period of more than 3 years. In addition, new columns have been added to better describe the temporal characteristics of the data, including the day of the week and the seasons according to the standards used by BDEW.

One possible explanation for the missing values on 31.12.2020 is that 2020 was a leap year and therefore only $35,040$ measurement points are stored. And with the 29^{th} February, the 31^{st} December is also no longer stored. For the 29 missing as well as wrong values on the Hoepfner site, a possible explanation is that the electricity meter was replaced during that time and thus

Building	# Values	Explanation
Hoepfner Areal	29	non-existent values
Hoepfner Areal	69	incorrect values
All Buildings	96	whole 31.12.2020 is missing
FZI	288	01.01.2019 until 03.01.2019 is missing

Table 2: Missing and Incorrect Values per Building.

wrong values were produced because the calibration had not yet been adjusted. As a result, the values before the replacement can be questioned overall and are therefore discarded and not used for the analysis. For the missing days at the FZI site, there is no plausible explanation as to why the data points are missing.

To check for outliers in the data, the indexed values are calculated using the median instead of the mean and the t-test is used to check the difference in means. The t-test gives a p-value of 0.9362 and thus the $H0$ must be accepted for any common level of significance, i.e. the difference in means between calculating the mean and the median is zero. It is therefore negligible and it can be assumed that no statistically significant outliers occur. Furthermore, all time series for the buildings were tested for stationarity using an Augmented Dickey-Fuller test. The results show a p-value of 0.01 for all tests, therefore $H0$ can be rejected and the time series are stationary and show no drift.

The data used by the BDEW to calculate the SLP were collected as samples throughout Germany from 1980 to 1997. On the basis of these data, an exemplary load curve was created, which also functions as the SLP. Since the data are publicly available, no further data engineering is required here and the data points are available for use. The data contains the SLP for all building types [20], although only the household type ($H0$), the commercial type 8am - 6pm ($G1$) and the bakery outlet type ($G5$) are of interest for this analysis. The latter is notable for the MGH as it includes a bakery, which strongly influences electricity consumption in the early hours of the day.

The data are divided into three categories: Winter, Summer, Transition (corresponding to the BDEW data to account for seasonal effects). To illustrate the differences in the two data sets, the relative development of the two data sets was standardised and mapped.

In Figure 1, the mismatched SLP are made less visible to focus on the better explanatory SLP. During weekdays, this means that only the $G5$ SLP has a good match with the Smart East data. This can be explained by the influence of the bakery as well as the kindergarten within the building. In addition, due to the Covid 19 pandemic, most people stayed at home, especially during the closures, and therefore consumed more electricity during the day, which explains the difference with the $H0$ profile, especially in the evening hours.

For the Saturday cases, the effect of the bakery can be seen in the early morning hours, although it is not as present as on the weekdays. A combination of $G5$ and $G1$ seems to best explain the real data, while the influence of $H0$ is again negligible, especially in the evening hours. For the Sunday cases, the strong increase in the morning hours cannot be explained by SLP, although one possible explanation is that the bakery is open on Sunday, while the bakeries in the BDEW data are not. Overall, however, no single SLP explains the Smart East data well

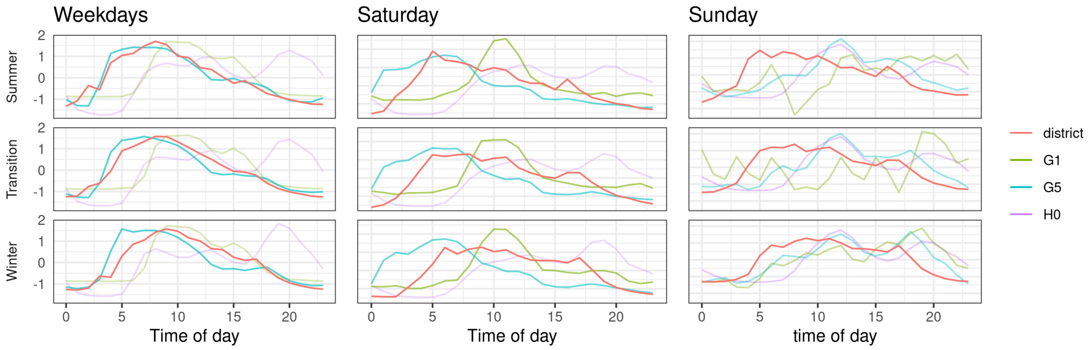

Figure 1: Comparison BDEW and Smart East Data for different SLP

enough, so it is suspected that there is a combination of all three cases.

4.2 Results

4.2.1 Data driven Model Results

Using three methods previously described in Chapter 3, the optimal number of k is chosen for the data at hand. The various methods give $k = 3$, but this can only be considered a recommendation, as the algorithm used is the k-means. It is therefore uncertain whether other algorithms with the chosen methods would lead to the same results. Therefore, $k = 3$ was chosen to narrow down the choice of k, and a range of two to five was chosen to be used for all six clustering algorithms. All of these selected clustering methods are then evaluated internally to determine which k is best for each algorithm. As already described in detail, the internal evaluation methods are:

- Silhouette Score
- Davies - Bouldin Index
- Calinski - Harabasz Index
- Dunn Index

The results for each clustering procedure are displayed in the Table 9, it shows the optimal k for each clustering procedure and thus allows a comparison between the different procedures to ultimately choose the best one for clustering. The results show that the direct comparison of the c-means cluster with DTW and Euclidean measures indicates that Euclidean measures perform better and that the more computationally expensive DTW does not necessarily lead to better cluster accuracy.

The Dunn index differs only slightly across all methods, so its influence is considered negligible. The other three indicators point to the k-mediods and k-shape algorithms, as they achieve the best results with similar values, although k-shape performs best overall. Therefore, the k-shape clustering method is chosen for further analysis to determine which data should be assigned to which cluster. For the comparison part of the different clusters, the values are presented as ratios, as the nominal values do not provide any information due to the different sizes in the clusters.

$$\text{ratio} = \frac{\text{variable} \cap \text{cluster}(i)}{\text{variable}} \tag{8}$$

The ratios offer the advantage that it is possible to work with data sets of different sizes and to compare the variables line by line in order to identify differences more easily.

		BDEW - Season			**Weekend**		
Building	Cluster	Summer	Transition	Winter	Saturday	Sunday	Weekdays
MGH	1^{st}	0.2090	0.1333	0.0103	0.4301	0.3000	0.0301
k-means	2^{nd}	0.0738	0.1000	0.2113	0.1828	0.4556	0.0473
	3^{rd}	0.7090	0.3333	0.0052	0.0753	0.0556	0.4989
	4^{th}	0.0082	0.4333	0.7732	0.3118	0.1889	0.4237
FZI	1^{st}	0.5383	0.6250	0.6985	0.1552	0.0973	0.7829
k-means	2^{nd}	0.4617	0.3750	0.3015	0.8448	0.9027	0.2171
Hoepfner	1^{st}	0.5410	0.1466	0.0296	0.2073	0.2125	0.3186
k-means	2^{nd}	0.4590	0.8534	0.9704	0.7927	0.7875	0.6814
TeFak	1^{st}	0.8443	0.7850	0.8378	0.8182	0.0238	0.9770
c-means	2^{nd}	0.1557	0.2150	0.1622	0.1818	0.9762	0.0230

Table 3: Building-wise Cluster Evaluation

After assigning the data to the selected clusters and evaluating the time based meta-information, the results are summarised in Table 3. The variables *month*, *day of the week* and *calendar week* were omitted due to concerns about multicollinearity with the existing variables *weekend* and *BDEW - season*. Clustering the entire dataset yields the same results for all algorithms used, namely a division into residential and commercial/industrial clusters. On one side the MGH and on the other side FZI, Hoepfner Areal and TeFak, so that a deeper analysis is performed for each building.

For the residential part, the MGH, the clustering technique chosen was the k-means algorithm with four clusters. The first cluster is characterised by Saturdays and also some Sundays, while falling slightly more in the summer time. The second cluster mainly contains Sundays and also some Saturdays, while it focuses on the winter months. The third cluster focuses heavily on the summer months in combination with the weekdays. The fourth cluster is characterised by the fact that it contains winter and transitional days as well as weekdays, but not as clearly as the previous cluster. These clusters show that a combination of summer and winter months, not so much the transitional periods, and the distinction between weekdays and weekends is the driving factor for electricity consumption. This effect is more pronounced in the summer months, as can be seen in clusters three and four, where the differences for the winter period are not as strong as for the summer months. A possible explanation could be that homes are lit more intensively in the darker winter months and people spend more time at home than in the summer months. Figure 2 provides the same information as Table 3, where the clusters three and four are relatively similar and further have a centre line close to the first cluster. Whereas, the second cluster is clearly distinguishable. Overall the lower consumption fits in the narrative of being a

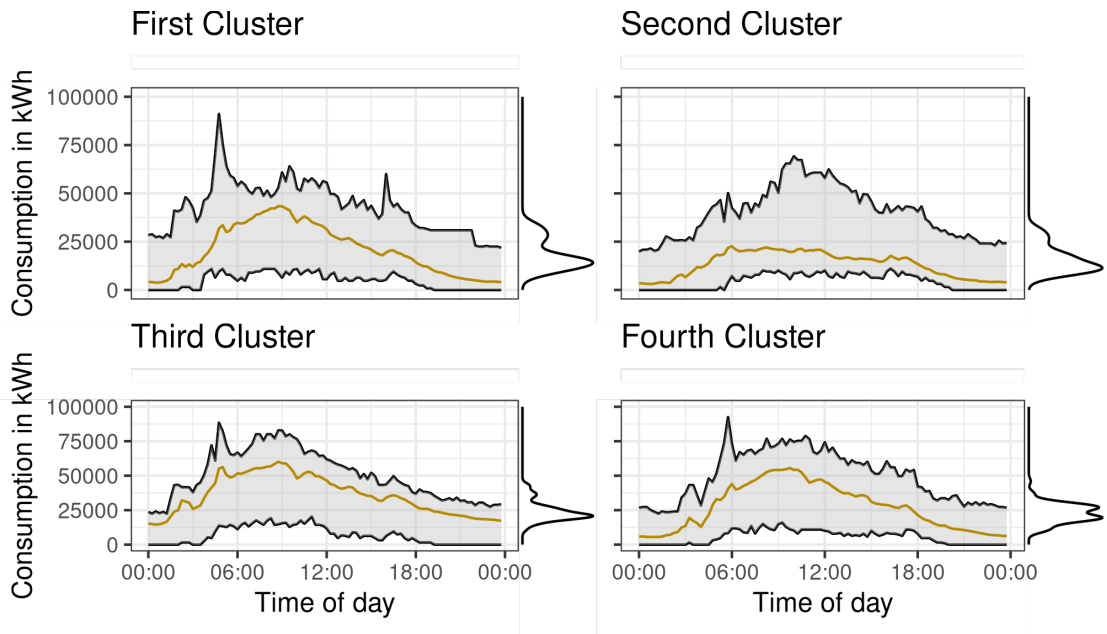

Figure 2: MGH Cluster

cluster characterised by weekends in the winter months.

In the commercial part all buildings are assessed with 2 clusters, the FZI can be mainly characterised by the difference between weekdays and the weekend, the different seasons show a slight correlation between the weekdays and the transitional and winter periods, but the effect is not as clear as the difference in weekdays or as clear as for the MGH. Therefore, the FZI can be described as a building with almost independent electricity consumption from the seasons and a clear distinction between weekdays and weekends, which can be explained by fixed working hours and no special equipment like air conditioners. Figure 3 results in the same conclusion as Table 3, namely that the first cluster describes the weekdays and the second cluster contains the weekend data points. During the weekends the centre line is focused heavily around its average and the extreme values also move roughly in the same range. Hence, it can be assumed a steady electricity usage occurs without much variation.

For TeFak the results were similar, the effects of seasons are balanced and are therefore negligible. However, in cluster one, weekdays and Saturday are almost equal, while cluster two almost exclusively contains Sundays. The difference with FZI and a possible explanation lies in who works in TeFak works. The majority of businesses located here are Start-ups with presumed different work attitudes compared with a university research centre. Therefore working on Saturdays could be one plausible explanation. Another explanation could be the occurrence of conferences and other business-related events that take place on Saturday. Figure 4 accentuates the assumptions made, though there are also presumably events taking place in the second cluster (Sunday cluster). However since the density plot indicates an even steadier electricity usage, it can be assumed that events on a Sunday do not take place regularly compared to the FZI.

The Hoepfner area cannot be described by the days of the week, but only by the BDEW - seasons. The weekday variables do not show a clear distinction between the two clusters, only a slight concentration on the weekdays in the first cluster, although not as clear as in previous cases. A clear feature can be seen in the winter and transition seasons, where the second cluster contains most of these data points. The first cluster could thus be called the summer cluster, although the distinction is not clear-cut, as almost half of the data points are also in the first cluster. A possible explanation for this could be the electricity-intensive winter and transition months as well as hot summer days with a high use of air conditioning, which result in a similar profile. The Figure 5 shows a less clear picture, both clusters seem to include outliers that were not caught in earlier data engineering stages. These can be spotted by the difference of the centre line compared to the maximum values, though the first spike in the second cluster must occur on a frequent basis to influence the centre in such a drastic way. These spikes could be attributed to the usage of air conditioners during the summer time, which would explain the massive occasional spikes in the afternoon. The difference in the first cluster could be explained through potential production stops during Covid-19 in the factory. The production closures likely outweigh the normal production days in the given data, where the latter show extreme values in the graph. In general one could say that the data for the Hoepfner Areal has to be cleaned and investigated further before meaningful results and conclusion can be drawn.

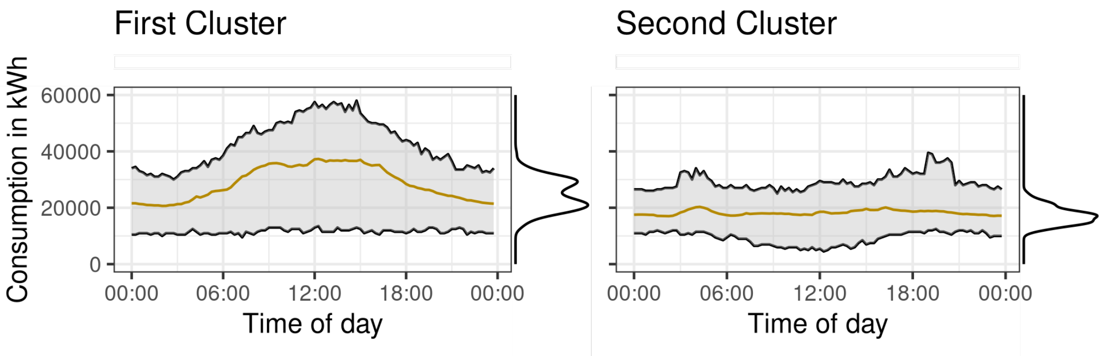

Figure 3: FZI Cluster

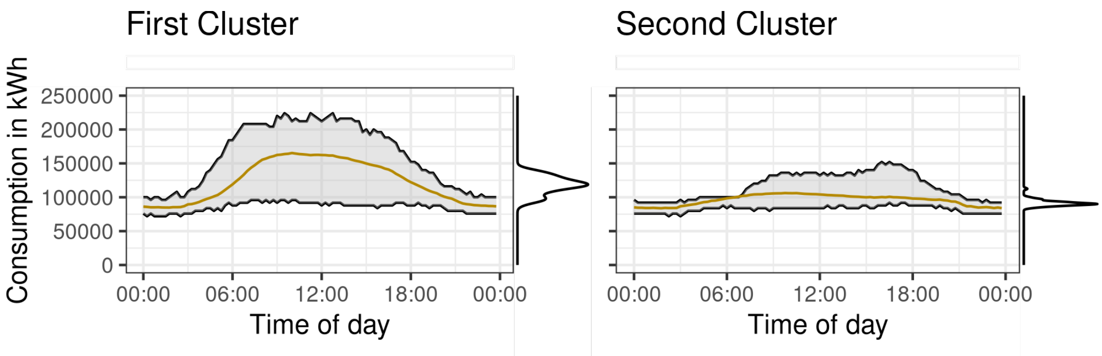

Figure 4: Technologiefabrik Cluster

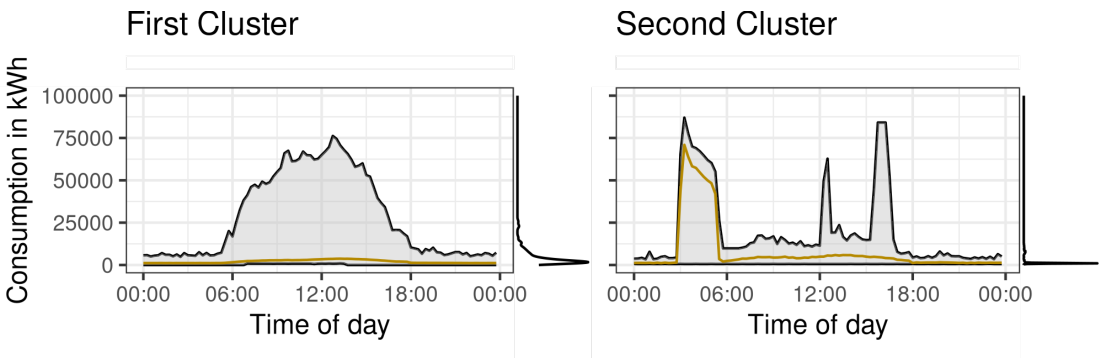

Figure 5: Hoepfner Areal Cluster

4.2.2 BDEW model Results

The second part of the paper deals with the comparison between the model proposed by BDEW, which uses a cluster of $k = 9$, and our own best model. The comparison is done for each building as this approach is close to the SLP approach of BDEW. The results shown in Table 4 demonstrate that our data driven, more dynamic models outperform the fixed 9-cluster models for each building. Furthermore, the negative silhouette score for the Hoepfner area also indicates that a random assignment of the time series to the clusters would provide a better result than the assignment proposed by BDEW.

Due to the poor results of the comparison, the cluster evaluation is only carried out on the TeFak data as an example. This cluster assignment is easiest to present and interpret, however still has the same deficiencies as with the other buildings. The key figures were calculated in the same way as before, so the results must be compared column by column.

For the weekend variables, each cluster except the ninth cluster has a unique value by which it can be classified. The other eight clusters contain four clusters in which the working day dominates (1, 5, 6 and 7), while there are only two clusters for Saturday (3 and 8) and Sunday (2 and 4) respectively. For the BDEW seasonal variables, the clear characterisation is more complicated. Here only five clusters can be described with relatively clear variables. There are two summer clusters (5 and 6), two winter and transition clusters (1 and 7), and one pure transition cluster (9). In particular, clusters 5 and 6 with the combination of working days and summer indicate that a more differentiated distinction is needed to further investigate the differences in these two clusters. The same problem arises for clusters 1 and 7, where the combination of working days, winter, and transition is dominant, leading to the same conclusion that a different model is needed to uncover more relationships. For the ninth cluster, the problem arises that it can be described as a transition cluster, although it has no unique days. The other clusters not mentioned have unique days of the week (Saturday and Sunday) but no assigned seasons, which means that the clusters cannot be compared with the BDEW model, which contains a cluster for each of the six possible combinations of the attributes listed in Table 5.

Therefore, a direct comparison between the two approaches in this paper and in the BDEW paper is not possible. Table 5 raises the question of the need for 9 different clusters when

Data Set	Evaluation Method	BDEW Paper	This Paper
MGH	Number of K	9	4
	Algorithm	k-means	k-means
	Silhouette	0.0236	0.0699
	Calinski - Harabasz	35.0399	75.8356
	Davies - Bouldin	3.3580	2.5736
	Dunn	0.0728	0.0690
FZI	Number of K	9	2
	Algorithm	k-means	k-means
	Silhouette	-0.0319	0.3090
	Calinski - Harabasz	180.7607	506.7903
	Davies - Bouldin	3.5647	1.0578
	Dunn	0.0313	0.0358
Hoepfner	Number of K	9	2
	Algorithm	k-means	k-means
	Silhouette	-0.1203	0.1760
	Calinski - Harabasz	201.6060	62.9782
	Davies - Bouldin	3.1374	1.8935
	Dunn	0.0015	0.0021
TeFak	Number of K	9	2
	Algorithm	k-shape	c-means euclidean
	Silhouette	0.0469	0.5448
	Calinski - Harabasz	108.8538	184.7313
	Davies - Bouldin	2.7800	0.5951
	Dunn	0.0470	0.0476

Table 4: comparison BDEW and own results

there are two cases where the days of analysis need more depth and distinction. The need for a separate transition period can also be questioned if the winter and transition periods mostly occur in the same cluster, as was the case in the previous analysis for the FZI and the Hoepfner area. Furthermore, the weekends are not assigned to a specific season, so that it can be assumed that the electricity consumption on a Saturday or Sunday is largely independent of the season, although here too, since there are two clusters for each day, other undetected factors play an important role.

5 Discussion

The limitations of this work lie primarily in the data used. It cannot be said with certainty that all false values due to damage or errors in the measurement units or infrastructure have been identified and corrected or taken out. A time span of less than 3 years is sufficient to provide

Building	Cluster	BDEW - Season Summer	Transition	Winter	Weekend Saturday	Sunday	Weekdays
TeFak	1^{st}	0.0000	0.1963	0.4189	0.0000	0.0000	0.2396
k-shape	2^{nd}	0.1148	0.1028	0.0541	0.0909	0.5238	0.0138
	3^{rd}	0.1148	0.0654	0.0541	0.5682	0.0000	0.0000
	4^{th}	0.0410	0.0935	0.1081	0.0682	0.4524	0.0046
	5^{th}	0.4426	0.0935	0.0000	0.0000	0.0000	0.2949
	6^{th}	0.2705	0.0748	0.0000	0.0000	0.0238	0.1843
	7^{th}	0.0000	0.1776	0.2838	0.0000	0.0000	0.1843
	8^{th}	0.0164	0.0187	0.0811	0.2273	0.0000	0.0000
	9^{th}	0.0000	0.1776	0.0000	0.0455	0.0000	0.0783

Table 5: Cluster Evaluation for Technologiefabrik

statistically significant results, but not to make general statements in a broader context. This is compounded by the fact that most of the data was recorded during the Covid-19 pandemic, which led to unusual behaviour and electricity consumption [21] as people stayed at home more often. This effect was most pronounced during government-imposed curfews. To normalise the results, these weeks of curfews were pulled out, even if this meant dropping more than half a year of data, especially for the spring and summer months. However, this assumption should be viewed critically, as it assumes that the only skewed days are during the curfew periods and that consumption patterns changed sharply and drastically with the start and end of curfews, which may be considered unrealistic. Another critical point is the limited amount of buildings. There are different buildings with different uses, but only four, which are also all located in the same city. Thus, the load profiles found are only valid for this degree of longitude, as a load profile in the east of Germany is offset in time due to an earlier dawn and sunset[3]. The way data is collected in the buildings also limits the possibilities to filter out individual effects, especially in the Hoepfner area, which hosts different building types with different SLPs. It is also assumed that there are no significant changes in equipment over the measured period, such as new machinery in the factory or the introduction of air conditioning in a whole building, which would drastically change the observed consumption and thus the load profiles. This limits the ability of this paper to make broader and more general assumptions about the electricity consumption of these building types and their respective SLP. When comparing with already determined SLPs from the BDEW paper, the underlying data differ significantly in the time horizons they cover, which is why a direct comparison is only possible to a limited extent. Also, general assumptions and conclusions based on a building with the clearest distinction may be tailored to this specific case and thus may not be meaningful for similar buildings outside the project.

[3]S. Voth (personal communication, 29 June 2022)

6 Conclusion

The evaluated clusters differ greatly from the clusters proposed and used by BDEW. When using real data, it becomes clear that there is no sharp difference between the winter and transition periods, but that they are often assigned to the same clusters. For the commercial buildings, the distinction was mainly in the difference between the days of the week. However, with the exception of the Hoepfner site, this can be partly attributed to the fact that there are several buildings with different uses and SLPs on the site, but due to the fact that only aggregated data points are available, it is impossible to filter individual buildings on the site.

As mentioned earlier, algorithms using the dynamic time-warp distance measure were expected to perform better than algorithms relying on the Euclidean distance measure. Although the results show that all the selected algorithms used the Euclidean distance measure instead of the more complex shape or dynamic time warping based measures, another three of the four buildings were evaluated using the k-means technique. This was also not expected, as this method can be described as a simpler clustering technique. Even though the metrics used to measure the fitness of the clusters cannot compete with other standards, the fixed nine-cluster approach of the BDEW leads to worse results.

For future research, some possibilities to implement the results of this work would be to introduce a data driven number of clusters for different building types instead of nine fixed clusters and time-specific allocations corresponding to SLPs depending on the longitude.

References

[1] K. Schmietendorf, J. Peinke, and O. Kamps, "The impact of turbulent renewable energy production on power grid stability and quality," *The European Physical Journal B*, vol. 90, no. 11, pp. 1–6, 2017.

[2] M. Papież, S. Śmiech, and K. Frodyma, "Effects of renewable energy sector development on electricity consumption–growth nexus in the european union," *Renewable and Sustainable Energy Reviews*, vol. 113, p. 109276, 2019.

[3] B. Schieferdecker, C. Funfgeld, H. Meier, and T. Adam, "Repräsentative vdew-lastprofile," *VDEW-Materialien M-28/99, Frankfurt*, 1999.

[4] M. Santamouris, C. Cartalis, A. Synnefa, and D. Kolokotsa, "On the impact of urban heat island and global warming on the power demand and electricity consumption of buildings—a review," *Energy and buildings*, vol. 98, pp. 119–124, 2015.

[5] T. W. Liao, "Clustering of time series data—a survey," *Pattern recognition*, vol. 38, no. 11, pp. 1857–1874, 2005.

[6] S. Aghabozorgi, A. S. Shirkhorshidi, and T. Y. Wah, "Time-series clustering–a decade review," *Information systems*, vol. 53, pp. 16–38, 2015.

[7] O. Motlagh, A. Berry, and L. O'Neil, "Clustering of residential electricity customers using load time series," *Applied energy*, vol. 237, pp. 11–24, 2019.

[8] K. Košmelj and V. Batagelj, "Cross-sectional approach for clustering time varying data," *Journal of Classification*, vol. 7, no. 1, pp. 99–109, 1990.

[9] F. Iglesias and W. Kastner, "Analysis of similarity measures in times series clustering for the discovery of building energy patterns," *Energies*, vol. 6, no. 2, pp. 579–597, 2013.

[10] J. Yang, C. Ning, C. Deb, F. Zhang, D. Cheong, S. E. Lee, C. Sekhar, and K. W. Tham, "k-shape clustering algorithm for building energy usage patterns analysis and forecasting model accuracy improvement," *Energy and Buildings*, vol. 146, pp. 27–37, 2017.

[11] F. McLoughlin, A. Duffy, and M. Conlon, "A clustering approach to domestic electricity load profile characterisation using smart metering data," *Applied energy*, vol. 141, pp. 190–199, 2015.

[12] L. Kaufman and P. J. Rousseeuw, *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons, 2009.

[13] A. Struyf, M. Hubert, and P. Rousseeuw, "Clustering in an object-oriented environment," *Journal of Statistical Software*, vol. 1, pp. 1–30, 1997.

[14] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters in a data set via the gap statistic," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 2, pp. 411–423, 2001.

[15] Y. Liu, Z. Li, H. Xiong, X. Gao, and J. Wu, "Understanding of internal clustering validation measures," in *2010 IEEE international conference on data mining*, pp. 911–916, IEEE, 2010.

[16] J. C. Dunn, "Well-separated clusters and optimal fuzzy partitions," *Journal of cybernetics*, vol. 4, no. 1, pp. 95–104, 1974.

[17] J. Paparrizos and L. Gravano, "k-shape: Efficient and accurate clustering of time series," in *Proceedings of the 2015 ACM SIGMOD international conference on management of data*, pp. 1855–1870, 2015.

[18] M. Cuturi and M. Blondel, "Soft-dtw: a differentiable loss function for time-series," in *International conference on machine learning*, pp. 894–903, PMLR, 2017.

[19] T. K. GmbH, "Technologie fabrik start-ups." `https://technologiefabrik-ka.de/companies/?lang=en`. last accessed on 17.08.2022.

[20] BDEW, "Standardlastprofile strom." `https://www.bdew.de/energie/standardlastprofile-strom/`. last accessed on 16.08.2022.

[21] A. Bahmanyar, A. Estebsari, and D. Ernst, "The impact of different covid-19 containment measures on electricity consumption in europe," *Energy Research & Social Science*, vol. 68, p. 101683, 2020.

7 Appendix

Name of Test	**Data Input**	**p-value**	**Test Values**
t-test	MGH & H0	$p = 1$	5.8040×10^{-17}
	MGH & G1	$p = 1$	-5.2084×10^{-16}
	MGH & G5	$p = 1$	-3.1392×10^{-16}
	mean & median	$p = 1$	-4.0201×10^{-17}
ADF	FZI	$p = 0.01$	-28.3910
	MGH	$p = 0.01$	-39.9060
	Hoepfner	$p = 0.01$	-29.8890

Table 6: Test Statistics

Algorithm	**Evaluation Method**	**k = 3**	**k = 4**	**k = 5**
k-shape	Silhouette	0.1606	0.1150	0.0535
	Calinski - Harabasz	358.6000	294.3400	176.1000
	Davies - Bouldin	1.8522	1.7932	2.5929
	Dunn	0.0146	0.0146	0.0145
fuzzy c-means dtw	Silhouette	0.0965	0.0066	-0.0612
	Calinski - Harabasz	250.9900	195.8600	128.1100
	Davies - Bouldin	2.3022	7.4439	9.6542
	Dunn	0.0118	0.0118	0.0118

Table 7: DTW evaluation

SLP	Description
G0	average commerce
G1	commerce weekdays 8-18 o'clock
G2	commerce with majority of consumption in evening
G3	commerce continuously running
G4	shops and hairdresser
G5	bakery
G6	commerce running solely on weekends
G7	mobile radio transmitting station
L0	average agriculture
L1	agriculture with milk production and livestock
L2	other agriculture
H0	households

Table 8: all BDEW SLP

Data Set	Evaluation Method	k-means	k-mediods	c-means (euclidean)	k-shape	c-means (DTW)	DTW cluster
whole data	Number of K	2	2	2	2	2	2
	Silhouette	0.2534	0.2547	0.2021	0.1331	0.2297	0.2383
	Calinski - Harabasz	832.7612	725.7763	576.1231	239.5588	698.0843	739.4886
	Davies - Bouldin	1.6345	1.6064	1.8955	2.7686	1.7612	1.7153
	Dunn	0.2372	0.3084	0.1916	0.1472	0.0559	0.2908
MGH	Number of K	4	3	2	2	5	3
	Silhouette	0.0699	0.1366	0.1109	0.1114	−0.0389	0.0913
	Calinski - Harabasz	75.8356	52.2908	66.1333	67.2024	6.9646	39.4109
	Davies - Bouldin	2.5736	2.3033	2.9013	2.8800	13.3910	2.7248
	Dunn	0.0690	0.0656	0.0656	0.0656	0.0676	0.0647
FZI	Number of K	2	2	2	3	4	2
	Silhouette	0.3090	0.2920	0.2767	0.0427	0.1158	0.2909
	Calinski - Harabasz	506.7903	408.9137	490.9277	87.6269	323.2968	443.6366
	Davies - Bouldin	1.0578	1.0995	1.1746	2.3958	1.6524	1.0981
	Dunn	0.0358	0.0348	0.0339	0.0267	0.0348	0.0377
Hoepfner	Number of K	2	2	3	4	3	2
	Silhouette	0.1760	0.0469	0.0470	−0.1791	−0.0878	0.1222
	Calinski - Harabasz	62.9782	51.8790	52.6045	68.4822	44.6026	6.3507
	Davies - Bouldin	1.8935	2.3617	2.3533	2.1006	2.0566	1.9558
	Dunn	0.0021	0.0012	0.0012	0.0018	0.0012	0.0018
TeFak	Number of K	2	2	2	3	2	2
	Silhouette	0.5260	0.5432	0.5448	0.2605	0.0877	0.5374
	Calinski - Harabasz	171.0827	170.2258	184.7313	161.1716	33.4304	161.5311
	Davies - Bouldin	0.6204	0.5964	0.5951	1.3251	2.6397	0.6064
	Dunn	0.0463	0.0424	0.0476	0.0501	0.0321	0.0397

Table 9: Determination of best Clustering Method

Data Set	Evaluation Method	k-means	k-mediods	c-means (euclidean)	k-shape	c-means (DTW)	DTW cluster
MGH	Number of K	9	9	9	9	9	9
	Silhouette	0.0236	0.0187	−0.0572	NaN	−0.0546	−0.0454
	Calinski - Harabasz	35.0399	33.2775	23.4063	33.3092	6.5429	24.4913
	Davies - Bouldin	3.3580	3.8281	3.1878	3.2626	9.0980	3.8055
	Dunn	0.0728	0.0696	0.0650	0.0728	0.0675	0.0701
FZI	Number of K	9	9	9	9	9	9
	Silhouette	−0.0319	−0.1116	−0.1029	−0.0679	−0.0688	−0.0483
	Calinski - Harabasz	180.7607	98.4413	107.2797	144.2442	158.5948	177.3666
	Davies - Bouldin	3.5647	5.1897	5.3765	6.1045	6.6142	4.6388
	Dunn	0.0313	0.0297	0.0297	0.0313	0.0337	0.0334
Hoepfner	Number of K	9	9	9	9	9	9
	Silhouette	−0.1203	−0.2747	−0.2847	−0.2682	−0.2785	−0.3057
	Calinski - Harabasz	201.6060	79.3395	64.7584	41.2544	19.4117	15.6738
	Davies - Bouldin	3.1374	4.1832	4.8814	3.9857	4.7578	3.5349
	Dunn	0.0015	0.0014	0.0012	0.0014	0.0019	0.0019
TeFak	Number of K	9	9	9	9	9	9
	Silhouette	0.0144	−0.0435	−0.0157	0.0469	−0.0565	−0.0526
	Calinski - Harabasz	90.9759	89.3411	71.7619	108.8538	35.2948	69.9425
	Davies - Bouldin	3.6485	3.1199	3.9947	2.7800	8.3431	4.1514
	Dunn	0.0436	0.0515	0.0393	0.0470	0.0361	0.0483

Table 10: Comparison BDEW Results