

Generating Load Profiles Using Smart Metering Time Series

Christian Bock^{1,2}(✉)

¹ Institute of Computer Science, Heinrich-Heine-University,
40225 Düsseldorf, Germany
`bock@cs.uni-duesseldorf.de`

² BTU EVU Beratung GmbH, 40545 Düsseldorf, Germany

Abstract. In this work we present a practice-oriented approach for generating *load profiles* as a means to forecast energy demand by using *smart metering time series*. The general idea is to apply *fuzzy clustering* on historic consumption time series. The segmentation yielded helps electricity companies to identify customers with similar consumption behavior. This knowledge can be used to plan available energy capacities in advance. What makes this approach special is that this approach segments consumption time series by time in addition to identifying customer groups. This is done not only to accommodate for customers potentially behaving completely different on working days than on local holidays for example, but also to build the resulting load profiles in a way the electricity companies can adapt with minimal adjustments. We also evaluate our approach using two real world smart metering datasets and discuss potential improvements.

Keywords: Big data · Data mining · Knowledge discovery · Clustering · Time series · Smart metering · Load profiles

1 Introduction

When electricity companies are confronted with the task to ensure the availability of energy for every customer, they are faced with the important obligation to announce the demand to the energy producers in advance. This is due to the fact that energy producers require lead time to ensure the availability of necessary capacities. Thus, in order to be able to announce the demand in advance, electricity companies require a means to forecast the total energy demand of their customers.

In general however, the electricity demand is not known in advance. Among other things, this is because not every customer behaves the same on every day; for example, it seems reasonable that during working days energy consumption is different from weekends. In addition, customers like families, singles, agricultural organizations and businesses typically each have different daily routines and thus might genuinely differ in their consumer behavior. To forecast the energy demand

for all customers, electricity companies use so called *load profiles*. In the case of Germany, the specific profiles used are usually the *standard load profiles* released by the “*Bundesverband der Energie- und Wasserwirtschaft*” (*Federal association of the energy and water economy*). These profiles, having been created during the 1990s, have not been adjusted to the technological advances and are thus deemed as an increasingly bad model to forecast the energy demand [28]. Due to the growing competition and legal obligations like the liberalization of the energy market and the policies towards green energy usage, there is an increasingly high interest and necessity to convert the information available thanks to modern technologies into valuable knowledge [15]. In this paper we focus on applying *fuzzy clustering* [5] on *smart metering time series* to generate load profiles as a means to provide such knowledge. We also assess our approach in the way electricity companies in Germany implement load profiles; we do think however that assessments in other regions might be comparable.

2 Related Work

Along with the increasing amount of data electricity companies and regulators can measure, process and analyze during the past decades, research towards understanding and making use of this kind of data gained significant attention [10, 19, 24]. These fields of research include, but are not limited to, outlier detection [29], marketing and tariff optimizations [9, 21]. [2, 22] present methods to predict long-term changes of the aggregated energy load. Publications that rely on clustering often choose *K-Means* to analyze the data [4, 11, 16, 23, 25] or approaches based on it [20]. Newer publications also increasingly use *fuzzy clustering* [19, 27, 29] and include newer technologies like *Smart Metering* [3, 12].

In this paper, we continue our previous work [6] and present an approach based on *fuzzy clustering* [5] for electricity companies to generate load profiles using smart metering time series. We focus on not only identifying *customer groups* to aggregate customers with similar consumption behavior, but also to dynamically find *day types*, which help to model long- and short-term periodic changes in consumer behavior. We also assess our approach using two real-world smart metering datasets and grade the quality of the energy forecast in a way closely related to how electricity companies would evaluate the generated load profile in a production environment.

3 Structure of Load Profiles

When electricity companies forecast the energy demand and negotiate corresponding capacities with producers, they typically assign one *load profile* and a *year consumption forecast (YCF)* to each customer. Load profiles work by segmenting all calendar days into groups of so called *day types* and by segmenting all customers into *customer groups*. The idea behind the usage of day types is that customers tend to have a finite set of genuinely different daily routines. Examples for this type of behavior are that office employees typically consume less

energy at home on working days and more on weekends, while seasonal events like funfairs orientate themselves by local holidays. Load profiles allow for different day types to accommodate for these trends. However, each customer group gets assigned exactly one consumption pattern on a given day type, which is why this model expects a customer to behave the same for all calendar days belonging to the same day type. This is a rather strong simplification as customers rarely behave exactly the same, even on days with the same daily routine. Thus, load profiles expect a consumption pattern assigned to a set of customers to be representative and to even out deviations between individual consumer behaviors and their associated consumption pattern. This characteristic makes load profiles bad at forecasting the energy demand of a single user, but useful at the use case relevant for electricity companies, which is predicting the aggregated total demand. Thus, a good day type segmentation groups calendar days in a way such that the total demand over the course of a day on calendar days belonging to the same day type are as similar as possible while aiming to make the total demand on calendar days belonging to different day types as dissimilar as possible. This task description closely matches the goal of generic clustering algorithms which is why we opted to rely on *fuzzy clustering* for our approach.

Another important aspect of load profiles is the aforementioned *year consumption forecast (YCF)* assigned to each customer. This *YCF* is needed because the consumption patterns used by load profiles are represented by one normalized time series. The normalization is done to ensure that the consumption patterns describe only the shape of the consumer behavior, rather than the total amount used. When used in practice, the model will first determine what day type the calendar day for which a prognosis shall be generated belongs to. Depending on the regularity and complexity of day type patterns, this can be done via rule sets or using the help of an analyst. The appropriate consumption pattern of each load profile is then chosen depending on the day type. The aforementioned *YCF* of each customer is then used to scale the applicable consumption pattern to match the estimated demand. Specifically, scaling is done so that the estimated total energy demand from the beginning to the end of a given year equals the *YCF*, which in turn requires the model to be able to forecast the day type segmentation for an entire year. Including the *YCF* as part of the model accommodates for customers with similar daily routines but different total consumption. In the next section, we describe an approach to generate such load profiles.

4 Generating Load Profiles

Due to the way load profiles are structured when used in practice, our approach for building them consists of three stages:

1. Determine the optimal number of day types as well as their segmentation onto the individual calendar days. In addition, compose rules to classify future calendar days.

2. For each day type, determine the optimal number of consumption patterns and their characteristics.
3. Compile load profiles by combining former results and assign a profile to each customer.

4.1 Day Type Segmentation

The most important criteria for the day type segmentation is finding calendar days where the total energy consumption is sufficiently dissimilar. In addition, the quality of the process of building the partitions must satisfy the properties of being independent of the number of measurements available and focussing on the *shape* of the consumption time series rather than on the *amount* of energy consumed. Requiring these properties is based on multiple reasons. First and foremost, independence from the number of customers is desirable because it enables electricity companies with both small and big customer bases to use the approach. It also helps to diminish the impact of missing values introduced by temporary malfunctions in the way measurements are gathered, transmitted and processed. As we demonstrate in Sect. 5.1, management of missing values in the available time series is a non-negligible task in real-world datasets. In addition, focusing on the *shape* of the consumption time series is required because the daily routine that will be predicted by the load profiles is scaled using the *year consumption forecast (YCF)* as mentioned in Sect. 3. Thus, we want customers whose consumption behaviors differ almost only by a scalar to be assigned to the same load profile. The *YCF* of a given customer for the current year itself is usually known in advance; common ways electricity companies calculate the *YCF* include setting it equal to the total energy consumed in the previous year by said customer, or using a moving average of the total consumption over the last periods. Because of this, even though the *YCF* is used in combination with the load profiles when forecasting the energy demand, it is not part of the process of building the load profiles themselves. As a result, for the purpose of this paper, we assume the *YCF* to be known when evaluating our approach.

To construct the day type segmentation, we first build a new time series $X = \{x_1, x_2, \dots, x_T\}$ for each point in time $t_j, 1 \leq j \leq T$ using the smart metering time series $S_i, 1 \leq i \leq N$ as follows:

$$x_j := \frac{1}{N_j} \sum_{i=1}^N \frac{s_{i,j}}{YCF_{i,j}} \quad (1)$$

Here, N stands for the number of distinct customers and N_j represents the total number of measurements available for t_j . The term $YCF_{i,j}$ is a time series specific scalar we use to normalize each customer; it represents the aforementioned *year consumption forecast* for the customer associated with S_i for the year that t_j belongs to. This enables us to solely concentrate on the *shape* of the consumption time series during clustering. The time series X can be vividly

described as an average time series of all normalized smart metering time series. Using X we subsequently construct dataset D with elements d_l as follows:

$$D := \left\{ d_l := (x_j, \dots, x_{j+m}) \mid \begin{array}{l} \forall a \text{ with } 1 \leq j \leq a \leq j + m \leq T : \\ t_a \text{ belongs to the } l\text{-th calendar day} \end{array} \right\} \quad (2)$$

The term $(m + 1)$ describes the number of measurements per day. Since smart metering time series are typically measured at fixed points in time, e.g. every 15, 30 or 60 min, each d_l corresponds to a 96-, 48- or 24-tuple, respectively. Afterwards we apply clustering on the dataset to retrieve a good day type segmentation. In principle, an arbitrary clustering algorithm can be used for this task. For our purposes, we have opted to use *Fuzzy-C-Means* [5] as the clustering algorithm and repeat the clustering process with different values for the number of clusters c ; the optimal value for c is then determined using a variety of *Cluster Validity Indices* [7]. The reason we chose Fuzzy-C-Means is its tendency to build spherical clusters [7] as this better conforms to the way load profiles are expected to even out derivations between the individual consumption time series belonging to the same customer group by using a representative consumption pattern as outlined in Sect. 3. The optimal clustering yielded by this procedure is the desired day type segmentation. By knowing which d_l got assigned to the same cluster and which calendar days they represent, it is possible to determine which calendar days belong to the same day type. For categorizing future calendar days we rely on the expertise of an analyst to review the day type segmentation and derive rulesets based upon observed regularities.

4.2 Identifying Typical Consumption Patterns

To determine the optimal number and characteristics of customer groups and their corresponding consumption patterns, we look at the smart metering data available for each day type separately. For this purpose, let $K_n, 1 \leq n \leq L$ be the sets of day types built in Sect. 4.1 where each K_n contains its matching t_j . We then construct the disjoint sets $P_n, 1 \leq n \leq L$ with elements $p_{e,n}$ as follows:

$$P_n := \left\{ (y_{i,j}, \dots, y_{i,j+m}) \mid \begin{array}{l} \forall a, b \text{ with } 1 \leq j \leq a, b \leq j + m \leq T : \\ t_a, t_b \in K_n \text{ and } y_{i,a}, y_{i,b} \text{ belong to} \\ \text{the same calendar day} \end{array} \right\} \quad (3)$$

with $y_{i,j} := \frac{s_{i,j}}{YCF_{i,j}}$

Each P_n is, similar to D , a dataset where smart metering measurements have been aggregated to $(m + 1)$ -tuples. P_n however does only contain data belonging to the day type K_n and contains the individual normalized measurements $y_{i,j}$ rather than the average of the normalized measurements. Each P_n is then individually segmented using clustering. Contrary to identifying the day type segmentation however, this time we are restricted to centroid-based clustering algorithms. The reason for this is that the cluster prototypes $C_{q,n}, 1 \leq q \leq c_{n,optimal}$

for a given P_n that have been deemed optimal by the algorithm directly correspond to the desired typical consumption patterns for the day type K_n . For our experiments, similar to the day type segmentation, we have opted to use *Fuzzy-C-Means*, try different values for the number of clusters c and evaluate each segmentation using *Cluster Validity Indices*.

4.3 Compiling Load Profiles

Load profiles as outlined in Sect. 3 can be represented as a set of L -tuples where the n -th entry contains the consumption pattern to use for calendar days assigned to K_n . In order for a set of load profiles to be usable however, we also require a way to assign each customer exactly one load profile; ideally the one that best suits him. Thus we propose to individually build the optimal load profile for a given customer based on the available consumption patterns and assign the constructed load profile to the customer in the process. For this purpose, let each customer be represented by its smart metering time series S_i . We then propose that for the day type K_n a given customer gets assigned to the consumption pattern $C_{q,n}$ if the highest membership degrees of S_i -based elements of P_n most commonly point to $C_{q,n}$.

Algorithm 1. Compiling load profiles

Input: $S_i, P_n, K_n, C_{q,n}, U_n$

Output: set of all load profiles G , set of profile assignments Z

```

1:  $G \leftarrow \emptyset$ 
2:  $Z \leftarrow \emptyset$ 
3: for  $i = 1$  to  $N$  do
4:   for  $n = 1$  to  $L$  do
5:      $H[n] \leftarrow C_{q,n}$  with  $q := \arg \max_{q'} \left\{ \left. \begin{array}{l} p_{e,n} \\ \wedge \nexists q'' : u_{q'',e,n} > u_{q',e,n} \end{array} \right| \exists j : (y_{i,j}, \dots, y_{i,j+m}) = p_{e,n} \right\}$ 
6:   end for
7:    $G \leftarrow G \cup H$ 
8:    $Z \leftarrow Z \cup (S_i, H)$ 
9: end for
10: return  $G, Z$ 

```

This procedure is illustrated in Algorithm 1. Here, the profile assignments Z are described by a set of 2-tuples where the first entry contains the customer S_i and the second entry the load profile H constructed for him. H itself is a L -dimensional array with $H[n]$ containing the consumption pattern for the day type K_n . The $u_{q,e,n} \in U_n$ used in Algorithm 1 correspond to the final membership degree of the dataset-tuple $p_{e,n}$ towards the consumption pattern $C_{q,n}$ we determined in Sect. 4.2 via *Fuzzy-C-Means*.

5 Evaluation

5.1 Description of Datasets

To evaluate the performance of our approach we have used two real world smart metering datasets which are both visualized in Fig. 1.

The first one, which we will call the *BTU-Dataset*, contains a total of 7668 distinct customers with a resolution of 1 measurement every hour over the course of 26 months. Because this dataset is provided in cooperation with a German electricity company who had a complete rollout of smart meters, we are able to test our approach under realistic conditions. The dataset is maintained and made available by the *BTU EVU Beratung GmbH* [1].

The second dataset, which we will refer to as the *CER-Dataset*, consists of 6445 distinct Irish customers with 1 measurement every 30 min over the course of 18 months. It is provided by the *Irish CER (Commission for Energy Regulation)* and accessed via the *Irish Social Science Data Archive (ISSDA)* [8].

Since both datasets contain real world data they are also subject to temporary technical failures, e.g. by any of the smart metering devices installed in the homes of the consumers or by network transmission errors. In either of these cases, a *missing value* is introduced into the respective dataset.

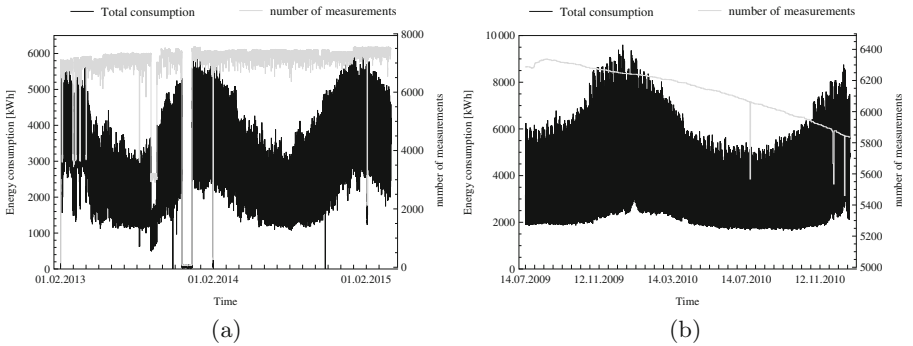


Fig. 1. Overview of (a) the *BTU-Dataset* and (b) the *CER-Dataset*. The black colored graphs show the sum of the energy consumption of all customers (applied on the primary axis); the grey colored graphs show the number of non-missing values from distinct customers available for a given point in time (applied on the secondary axis).

5.2 Results

To derive the optimal segmentation of the day types and consumption patterns for both datasets, we have preprocessed each dataset according to Eqs. 2 and 3. As for the *year consumption forecast* required to normalize each time series, we used the total energy consumed per customer per year:

$$YCF_{i,j} := \sum_{j' \in Z_j} s_{i,j'} \quad \text{with} \quad Z_j := \{j' \mid t_{j'} \text{ belongs to the same year as } t_j\} \quad (4)$$

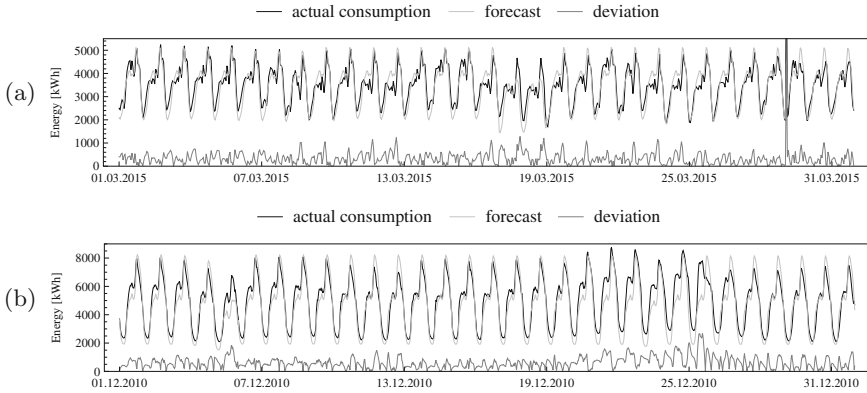


Fig. 2. Comparison of the actual consumption (black graph) and the consumption predicted using the load profiles based on 2 day types and 2 consumption patterns per day type (light gray graph) for (a) the *BTU-Dataset* and (b) the *CER-Dataset*. The dark gray graph shows the absolute difference of the actual consumption and the forecast. The apparent outlier on march 29 is due to a switch from *standard time* to *daylight saving time*.

To handle the *missing values* present in both datasets, we have incorporated the *Partial Distance Strategy* [14] adaptation of *Fuzzy-C-Means*, which has shown a solid performance in experimental evaluations [17], using different values for c ($2 \leq c \leq 25$). Since *Fuzzy-C-Means* uses random coordinates as the starting configuration of the cluster prototypes, we have opted to improve the statistical significance of our results by independently repeating the clustering process for every value of c a total of 100 times each. We have then used the *Cluster-Validity-Indices Partition Coefficient* [7], *Normalized Partition Coefficient* [17], *Compactness & Separation by Xie and Beni* [7], *Compactness & Separation by Bouguessa, Wang and Sun* [7], *Fuzzy Hypervolume* [13] and *Partition Density* [13] to evaluate each clustering and choose the best day type segmentation as the basis for the following consumption pattern analysis, as well as choose the best consumption pattern segmentation as the basis for the following load profile compilation.

Some of these day type segmentations are visualized in Fig. 4 and are further discussed in Sect. 5.3. In order to test the accuracy of the profiles, we have excluded the last month of smart metering data from both the *BTU-Dataset* and the *CER-Dataset* while building the load profiles; we have then used the formerly excluded month to compare the actual consumption in that month with the one predicted by the profiles. Some of our results are exemplary shown in Fig. 2. If an electricity company were to use these load profiles, they would plan their buy-in of energy according to the *forecast*-graph. Because the forecast is known and necessary capacities can be planned for in advance, they are relatively cheap from a business standpoint. Deviations from the actual total consumption however, both by overestimating and underestimating the actual demand, require extremely short-term adjustments in

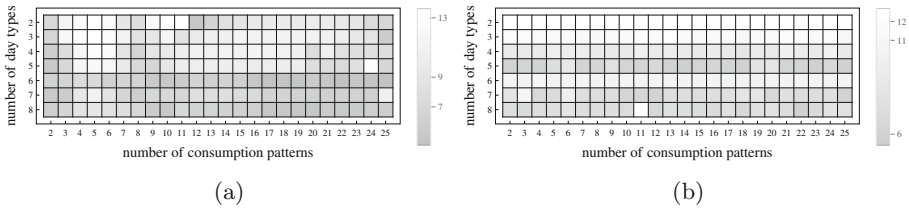


Fig. 3. Ratio of the deviations and the actual consumption in percent yielded by the load profiles generated using different values for the number of day types and the number of consumption patterns. The graphs visualize the results for (a) the *BTU-Dataset* and (b) the *CER-Dataset*.

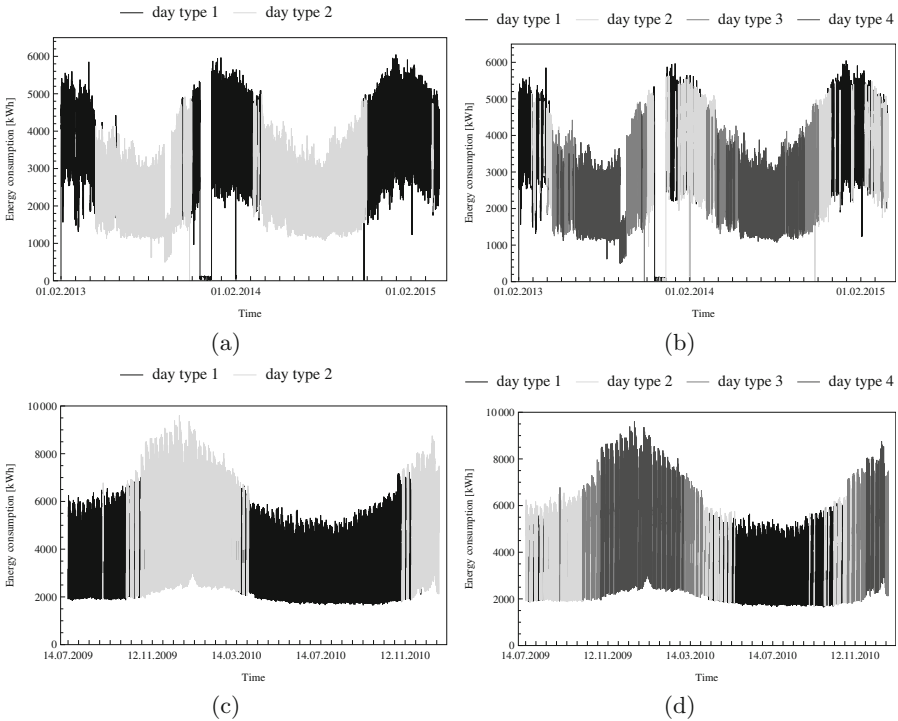


Fig. 4. Overview of the segmentations of day types for the (a) (b) *BTU-Dataset* and (c) (d) *CER-Dataset* yielded by using different values for the numbers of day types. The graphs have been colored depending on which cluster the total consumption time series has been assigned to on a given day.

the amount of energy circulating in the energy grid. Because of the ad hoc nature, their limited availability and the importance of these adjustments in terms of preventing electricity outages, the cost of these reserves are generally much higher than long-term agreements with producers. Thus, electricity companies typically aim

to keep deviations to a minimum and assess load profiles by the amount of energy they are required to trade using the aforementioned short-term reserves to meet the actual demand. This performance can be made comparable between electricity companies of different sizes by looking at the ratio of the deviations and the actual consumption [18]. The results of our approach are shown in Fig. 3. These results pose a significant improvement over the standard load profiles that are in use by most electricity companies today. Electricity companies using standard load profiles typically achieve ratios of roughly 14% [18]. In the next section, we will present and discuss ideas to further improve the accuracy of load profiles generated by our approach.

5.3 Improvements

The idea behind the segmentation of the total energy consumption is to identify periods of time in which consumption patterns genuinely differ from one another. Since we are only interested in accurately forecasting the total energy demand rather than focussing on predicting the consumption of individual customers, it is reasonable to use the same consumption patterns for each customer on days where the total consumption is not expected to be significantly dissimilar. Figure 4 shows a subset of the segmentation we got for the day types for the *BTU-Dataset* and the *CER-Dataset*. One striking property of the clustering using four or less day types is that it roughly resembles the seasonal segmentation manually chosen in other publications, e.g. in [26, 27]. Increasing the number of day types further however, it becomes progressively clear that, for both datasets, the identification of day types has approximately resulted in segmenting the data according to a threshold filter. This fact by itself does not necessarily mean the segmentation is flawed; however, while the load profiles based on these day type segmentations have resulted in a significant performance improvement compared to the standard load profiles as pointed out in Sect. 5.2, the repeating sine-shaped pattern visible in the graphs has prevented our approach to compare the daily consumption tuples based on their shape.

This yearly periodic pattern with its peak near the end of december is something we also see in a similar fashion in many other (non smart metering based) total consumption time series. Because of this, we propose to apply a *high-pass filter* on the dataset to create the load profiles based only on the daily consumption behavior of the individual customers. We propose this high-pass filter to be used during both the day type segmentation as well as the identification of typical consumption patterns. The yearly periodic pattern is then reapplied onto the time series when the load profiles have been later used to forecast the energy demand. A candidate to fulfill these requirements is the *Fourier transformation*, where the lowest-frequency terms from the total consumption time series can be used to describe and filter the yearly periodic patterns from the dataset.

Another optimization we propose is to change the function for computing the dissimilarity between tuples. For our experimental results presented in Sect. 5.2 we have used the *partial distance*, an adaptation of the *euclidean distance* for

missing values originally introduced in [14]. However, to make the process of generating load profiles more sensitive to minimize these costly deviations between the total energy consumption and its forecast, we propose to adapt the *manhattan distance* to handle missing values. Using this distance function and the high-pass filter derived by using the *Fourier transformation* more closely complies to our original vision to compare consumer behavior based on their shape and optimize for low deviations between forecast and the actual total consumption.

6 Conclusion and Future Work

In this paper we have introduced a clustering method for generating load profiles using smart metering time series. In order to tailor our approach to the specific needs of electricity companies we have incorporated the use of consumption patterns and day types the same way they are treated by the industry. Furthermore, our method does predetermine neither the number nor the shape of the consumption patterns or day types. Our findings show that using the presented approach results in a significant improvement regarding the deviations between the forecasted total demand and the actual energy consumption compared to the standard load profiles typically in use. This helps electricity companies to better plan the buy-in of energy ahead of time, which lowers costs and improves the security of the energy supply. In addition, we have presented and discussed possible enhancements for our method which we plan to further investigate in future publications.

As of now, our approach requires the load profiles to be generated from scratch each time new smart metering data is available. While this is not a major concern for typical use cases since load profiles are usually changed at most once per year, there is an interest to reduce the required computation time, e.g. by looking into ways incrementally add new smart metering data as it becomes available. This might be one potential area for future research.

References

1. BTU EVU Beratung GmbH. <http://www.btu-evu.de>
2. Andersen, F., Larsen, H., Boomsma, T.: Long-term forecasting of hourly electricity load: identification of consumption profiles and segmentation of customers. *Energ. Convers. Manag.* **68**, 244–252 (2013)
3. Beckel, C., Sadamori, L., et al.: Revealing household characteristics from smart meter data. *Energy* **78**, 397–410 (2014)
4. Benítez, I., Quijano, A., et al.: Dynamic clustering segmentation applied to load profiles of energy consumption from Spanish customers. *Int. J. Electr. Power Energ. Syst.* **55**, 437–448 (2014)
5. Bezdek, J.C.: *Pattern Recognition with Fuzzy Objective Function Algorithms*. Springer, New York (1981)
6. Bock, C.: Clustering-Ansatz zur Erstellung von Lastprofilen zur Vorhersage des Stromverbrauchs. In: *Proceedings of the 28th GI-Workshop Grundlagen von Datenbanken (GvDB 2016)*, pp. 21–26 (2016)

7. Bouguessa, M., Wang, S., Sun, H.: An objective approach to cluster validation. *Pattern Recogn. Lett.* **27**(13), 1419–1430 (2006)
8. CER – The Commission for Energy Regulation. Accessed via the Irish social science data archive. <http://www.ucd.ie/issda>
9. Chicco, G., Ilie, I.S.: Support vector clustering of electrical load pattern data. *IEEE Trans. Power Syst.* **24**(3), 1619–1628 (2009)
10. Diamantoulakis, P.D., Kapinas, V.M., Karagiannidis, G.K.: Big data analytics for dynamic energy management in smart grids (2015). CoRR, abs/1504.02424
11. Figueiredo, V., Rodrigues, F., et al.: An electric energy consumer characterization framework based on data mining techniques. *IEEE Trans. Power Syst.* **20**(2), 596–602 (2005)
12. Fusco, F., Wurst, M., Yoon, J.: Mining residential household information from low-resolution smart meter data. In: 2012 21st International Conference on Pattern Recognition (ICPR), pp. 3545–3548, November 2012
13. Gath, I., Geva, A.: Unsupervised optimal fuzzy clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* **11**, 773–780 (1989)
14. Hathaway, R.J., Bezdek, J.C.: Fuzzy c-means clustering of incomplete data. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **31–35**, 735–744 (2001). IEEE
15. Hayn, M., et al.: Electricity load profiles in Europe: the importance of household segmentation. *Energ. Res. Soc. Sci.* **3**, 30–45 (2014)
16. Hernández, L., Baladrón, C., et al.: Classification and clustering of electricity demand patterns in industrial parks. *Energies* **5**(12), 5215 (2012)
17. Himmelspach, L.: Fuzzy clustering of incomplete data. Ph.D. thesis, Heinrich-Heine-University, Institute of Computer Science (2016)
18. Kolo, A., Kretschmann, C.: Hebung finanzieller Potentiale in der Strombilanzierung. et – *Energiewirtschaftliche Tagesfragen* **7**, 43–45 (2015)
19. Yang, S.L., Shen, C., et al.: A review of electric load classification in smart grid environment. *Renew. Sustain. Energ. Rev.* **24**, 103–110 (2013)
20. López, J.J., Aguado, J.A., et al.: Hopfield-k-means clustering algorithm: a proposal for the segmentation of electricity customers. *Electr. Power Syst. Res.* **81**(2), 716–724 (2011)
21. Mahmoudi-Kohan, N., Moghaddam, M.P., Sheikh-El-Eslami, M.: An annual framework for clustering-based pricing for an electricity retailer. *Electr. Power Syst. Res.* **80**(9), 1042–1048 (2010)
22. Misiti, M., Misiti, Y., et al.: Optimized clusters for disaggregated electricity load forecasting. *Revstat* **8**(2), 105–124 (2010)
23. Rodrigues, F., Duarte, J., et al.: Proceedings of Machine Learning and Data Mining in Pattern Recognition: Third International Conference, MLDM 2003, Leipzig, Germany, 5–7 July 2003, pp. 73–85. Springer (2003)
24. Rudin, C., Waltz, D., et al.: Machine learning for the New York city power grid. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(2), 328–345 (2012)
25. Räsänen, T., Voukantsis, D., et al.: Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data. *Appl. Energ.* **87**(11), 3538–3545 (2010)
26. Schäfer, H., et al.: Analysing the segmentation of energy consumers using mixed fuzzy clustering. In: *Fuzzy Systems (FUZZ-IEEE)*, pp. 1–7, August 2015
27. Viegas, J.P.L., Vieira, S.M., Sousa, J.M.C.: Fuzzy clustering and prediction of electricity demand based on household characteristics. In: 2015 Conference on International Fuzzy Systems Association and European Society Fuzzy Logic and Technology (IFSA-EUSFLAT-15)

28. von Roon, D.-I.S., et al.: Statusbericht zum Standardlastprofilverfahren Gas, 11 2014. <https://www.ffegmbh.de/kompetenzen/system-markt-analysen/508-status-bericht-standardlastprofile-gas>
29. Zhang, X., Sun, C.: Dynamic intelligent cleaning model of dirty electric load data. *Energ. Conver. Manag.* **49**(4), 564–569 (2008)