METHODOLOGY FOR EXTRACTING DYNAMIC STANDARD LOAD PROFILES FROM SMART METER DATA

Krischan KEITSCH[*1], Hendrik KONDZIELLA[1], Thomas BRUCKNER[1,2]

1 Fraunhofer-Zentrum für Internationales Management und Wissensökonomie, Neumarkt 9-19, 04109 Leipzig, {krischan.keitsch | hendrik.kondziella | thomas.bruckner} @moez.fraunhofer.de

2 Universität Leipzig, Institut für Infrastruktur und Ressourcenmanagement (IIRM), Grimmaische Straße 12, 04109 Leipzig, bruckner@wifa.uni-leipzig.de

Kurzfassung: The German Federal Ministry for Economic Affairs and Energy published a white book for a new power market design to support the transition of the national energy system ("Energiewende"). One planned measure is to increase the balance area loyalty. This underlines the necessity for accurate standard load profiles to increase the forecasting quality of the electrical demand of customers in balance areas.

With the increasing roll-out of smart meters, fine-grained logging data of electricity consumption in households and enterprises becomes available. In a step by step analysis, smart meter profiles are analyzed. The goal is to identify clusters of similar demand profiles and then to develop optimized dynamic standard load profiles for accurate forecasts. An increased forecasting quality is achieved by the usage of dynamic functions. The dynamic functions are used to adjust the standard load profiles for different seasons. The parameters of the dynamic functions are identified by using a least square approach and an Evolution Strategy. The application of optimized dynamic functions increases the forecasting accuracy measured with the mean average percentage error (MAPE) and the normalized rooted mean square error (NRMSE) by 21% and 28% respectively in this case study compared to the sole application of non-optimized standard load profiles.

Keywords: Standard Load Profile, Clustering, Smart Meter

1 Introduction

The liberalized German power market is organized in several balancing areas. Every balance responsible party (BRP) is obligated to aim for balanced portfolios by matching demand and production. Discrepancies between demand and production are handled by the grid operators. The resulting costs of the reserve energy is accounted afterwards. The electrical demand of customers in balancing areas is therefore commonly anticipated by standard load profiles (SLPs) [1].

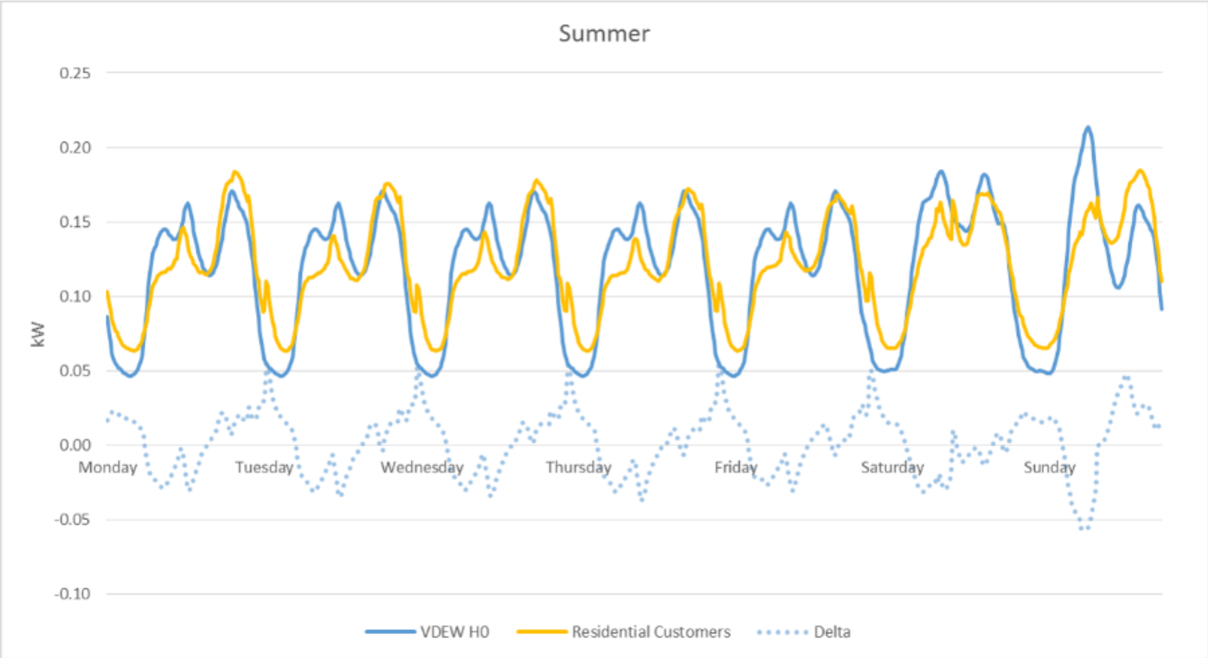

Figure 1: Comparison of the electrical demand of residential customers from a project described in [2] and the commonly used standard load profile "H0".

The German Federal Ministry for Economic Affairs and Energy published a white book for a new power market design to support the transition of the national energy system ("Energiewende"). One planned measure is to increase the balance area loyalty („Bilanzkreistreue") [3] to keep the demand and production of electricity leveled and to decrease the demand of ancillary services and reserve energy. In addition, the proposed law to digitize the "Energiewende" in Germany will make the installation of smart meters mandatory for customers with a yearly consumption of more than 6 000 kWh [4].

This surplus of detailed electrical load data may increase the accuracy of demand forecasts for customers above 6 000 kWh/a. However, the electrical demand of customers without smart meters will still be estimated with standard load profiles. It should be noted that the commonly used SLPs for the German power market are based on a small data set of 332 household profiles from 1981 for instance [1]. Figure 1 is an example of the changed electrical demand pattern of residential customers from a recent project [2] and the need for new SLPs. In [5] the authors underline the need for new SLPs. They emphasize the risk for grid operators and power traders as decentralized generation from renewable sources increases and new tariffs may influence the electrical demand. The authors propose additional SLPs for households with, for instance, solar power systems [6]. The increased roll-out of smart meters allows to generate new standard load profiles for residential households and small to medium sized enterprises (SME). Precise forecasts reduce the demand of expensive energy balancing and ancillary services [7].

2 Methodology

This paper suggests a standardized methodology to develop new SLPs based on data analysis of smart meter electrical demand profiles from residential and small to medium sized enterprises (SME).

First, the smart meter data is preprocessed to identify type days and to reduce the data size. Similar groups of profiles in a given data set are identified by incorporating k-means clustering. In order to evaluate the methodology, synthetic loads (one year mid-term forecasts) are generated. Commonly used evaluation metrics and the available meta-data are used to interpret the results.

The second part of the paper focuses on the optimization of dynamic standard load profiles for higher forecasting accuracy and reduction of the influence by seasonal changes. An increased forecasting quality is achieved by using dynamic functions. The dynamic functions are used to adjust standard load profiles for different seasons. The parameters of the dynamic functions are identified using a least square approach and an Evolution Strategy (ES). This methodology is applied to a publicly available smart meter data set [8] and for two different customer groups for the purpose of demonstration only.

2.1 Data analysis

In order to demonstrate the process, a data set from the "CER Smart Metering Project"[1] is used. The regulator for the electricity and natural gas sectors in Ireland is the Commission for Energy Regulation (CER). The CER initiated the project to gather information and experience on smart metering technologies. A data set containing more than six thousand smart meter profiles from residential (4225 profiles) and small to medium enterprises (SME, 485 profiles) as well as profiles with no classification (labeled as "others") has been made available for scientific research. The CER data set was recorded from mid July 2009 until the end of 2010 with a resolution of 0.5 h values. The data set is anonymous and contains no information to where the smart meters are localized.

In order to identify standard load profiles, a format for the SLP is defined. The standard load profiles consist of three seasons (winter, transition and summer). The transition season represents spring and fall.

Table 1: Configuration of the season settings for this exemplary case study.

Season	Sub- Season	Begin	End	Duration [d]
Winter	Winter 1	January 1st	February 28th	136
	Winter 2	October 16th	December 31st	
Transition	Spring	March 1st	April 15th	91
Summer	Fall	September 1st	October 15th	138
	Summer	April 16th	August 31st	

In this case study, the week days from Monday to Sunday are represented by five type days: Monday, Tuesday+Wednesday+Thursday, Friday, Saturday and Sunday. A type day represents the average load profile of all days of its kind and season. This configuration of the weekday is chosen because the load profiles of the weekdays differ. Mondays and Friday show a unique profile, whereas Tuesday, Wednesday and Thursday have a more similar profile. The length of the seasons and the type days differ from the approach of the commonly used VDEW load profiles [1] for the German power market.

[1] http://www.ucd.ie/issda/data/commissionforenergyregulationcer/

Another necessity is a calendar to assign the type days and holidays. National holidays may be treated as Sundays, days between a holiday and a weekend (bridging days) may be represented by the type day Saturday.

2.2 Clustering

One common approach to group data is to use algorithms such as k-means. K-means is a well-established standard and an algorithm of high performance. K-means does not determine the amount of clusters by itself. The amount of cluster k has to be defined by the user first. As k-means randomly chooses initial centers, a perfect clustering (finding the global optimum) is not guaranteed [9]. Therefore, an iterative approach (Monte Carlo search) is chosen to repeat the clustering n times for better results. The clustering algorithm identifies a group of profiles with similar characteristics. The result is the mean of a cluster. Each cluster represents a different customer portfolio.[2]

The silhouette coefficient (sc) and silhouette plot may be used to determine the quality of the clustering attempts. The silhouette coefficient is an indicator for the quality of a cluster procedure. The sc is defined as:

$$sc = \frac{1}{n}\sum_{i=1}^{n} s_i$$

Where s_i is the silhouette of the item i in a cluster and n is the number of the silhouette. The silhouette s_i is defined as:

$$s_i = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

With $a(i)$ and $b(i)$ being the average dissimilarities of i items to all other objects in cluster A and B [10].

However, the sc does not answer the question of how good a cluster is in order to represent a certain customer portfolio in terms of load forecasting.

Therefore, the normalized root mean square error (NRMSE) and the mean absolute percentage error (MAPE) are used to determine the quality of the synthetic load (generated from the standard load profiles) compared to the original CER load.

The NRMSE is defined as:

$$NRMSE = \sqrt{\frac{\sum_{t=1}^{n}(\bar{y}_t - y_t)^2}{\sum_{t=1}^{n} y_t^2}}$$

[2] Please note that customers may be grouped into different portfolios by analyzing available meta-data. A representative load profile may then be assigned to each customer portfolio. If reliable meta-data is not available or the quality and accuracy does not match the requirement, a clustering technique may be used. In this case study the available meta-data is used to verify the clustering results.

Where y_t is the load profile from the CER customers and $\overline{y_t}$ is the synthetic load at time t. The MAPE is defined as:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} |(\overline{y_t} - y_t) \cdot (y_t)^{-1}|$$

2.3 Dynamic functions

In order to increase the forecasting quality and to reduce the gap during seasonal change, modifications have to be applied to the standard load profiles. In this part of the paper, the necessary steps are described. A cosine function and a polynomial function of grade 4 and 6 are used. The cosine function guarantees a continuous course. A polynomial function of grade 6 is additionally chosen as it offers a higher degree of flexibility to adjust the shape of the curve than the grade 4 polynomial.

The Cosine function is defined as:

$$x = x_0(\cos(\frac{\pi}{t_{end}} \cdot (2t + A)) \cdot B^{-1} + C)$$

Where x is the adjusted value, x_0 is the original value of the load profile and t is an index ranging from 1 (January 1st) to 17520 (December 31st) in 0.5h steps. A is a parameter to adjust the phase shift, B is responsible for the amplitude and C for the height level.

The polynomial function of grade 4 and 6 are defined as:

$$x = x_0(A't^4 + B't^3 + C't^2 + D't + E')$$

$$x = x_0(A''t^6 + B''t^5 + C''t^4 + D''t^3 + E''t^2 + F''t + G')$$

Where the parameter A' to E' respectively A'' to G'' are the polynomial coefficients.

Before a dynamic function can be applied, load profiles with a strong seasonality need to be normalized. The normalization is done for each season separately. In this example (refer to Figure 3) the type days of the Winter season are scaled down, while the Summer type days are scaled up. The overall characteristic of the type days needs to be preserved as well as the overall size of the integral. The normalization requires the following steps:

1. Determine helper values: Evaluating the sum of the load profile (lp) and determine the maximum load for each season ("Winter (Wi)", "Transition (Tr)" and "Summer (Su)"):

$$I_{lp} = \sum_{t=1}^{t_{end}} lp_t$$

$$m_{season} = max(lp_{t,season})$$

2. Normalization of the sub-load profiles for each season towards a peak of 1 kW:

$$lp_{t,season}^{1kW,norm} = \frac{lp_{t,season}}{m_{season}}$$

3. Integration of seasons as helper values and profile normalizers:

$$I_{lp}^{1kW,norm} = \sum_{t=1}^{t_{end}^{season}} lp_{t,season}^{1kW,norm}$$

$$lp_t^{norm} = \frac{lp_{t,season}^{1kW,norm}}{I_{lp}^{1kW,norm}} \cdot I_{lp}$$

Where I_{lp} is the sum of the standard load profile and m_{season} represents the maximum value for each season of the load profile.

After steps 1 to 3 are performed, $\sum_{t=1}^{t_{end}} lp_t = \sum_{t=1}^{t_{end}} lp_t^{norm}$. The original profile and the normalized profile share the same size of their integrals and the shape of the type days of the three seasons as shown in Figure 3.

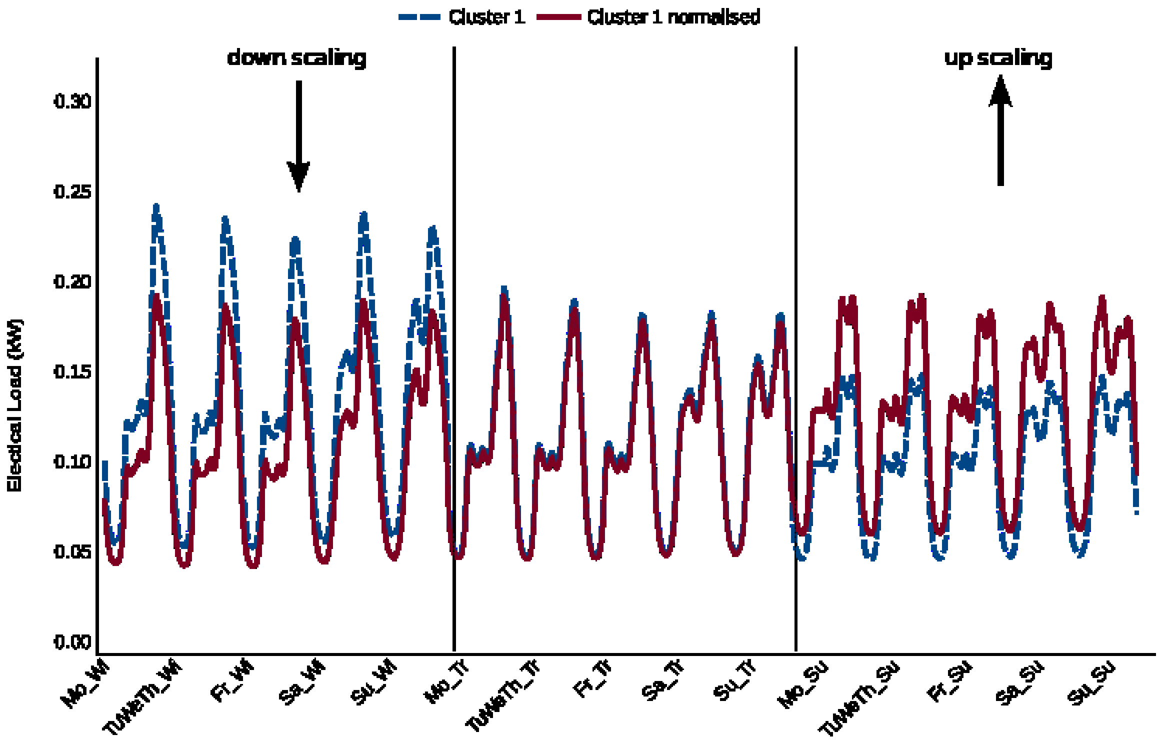

Figure 2: Visualization of the normalization procedure for the seasons Winter (Wi), Transition (Tr) and Summer (Su).

As a standard method for curve fitting, a least square approach is used to determine the parameters of polynomial functions of grade 4 and 6. The load profile of each cluster needs to undergo some adjustments to apply the least square method. First, the load of each cluster is normalized in such a way that all values range from 0 to 1. In the second step, the mean of the load profile is determined. Third, an adjustment of the load profile is performed to raise the mean of the load profile to 1. The curve fitting is then applied to retain parameter for polynomial functions of grade 4 and 6.

Besides the standard curve fitting method least square, the application of an Evolution Strategy (ES) for better curve fitting results is discussed in this paper as a second approach.

As a member of the Evolutionary Algorithms family, ES have proven to perform well on nonlinear optimization targets. An ES mimics some aspects of the natural evolution and was introduced by [11].

The NRMSE is used as a utility function to calculate the overall quality of the forecasting time series. The following equation represents the optimization problem for the two types of dynamic functions:

$$NRMSE_k^{cos} = \sqrt{\sum_{t=1}^{n} \left((lp_k^{norm} \cdot \tilde{a}) - y_{t,k}\right)^2 \cdot (\sum_{t=1}^{n} y_{t,k}{}^2)^{-1}}$$

$$NRMSE_k^{polyX} = \sqrt{\sum_{t=1}^{n} \left((lp_k^{norm} \cdot \tilde{b}) - y_{t,k}\right)^2 \cdot (\sum_{t=1}^{n} y_{t,k}{}^2)^{-1}}$$

With $\tilde{a} = \cos(\frac{\pi}{t_{end}} \cdot (2t + A)) \cdot B^{-1} + C$ and $\tilde{b} = \sum_{n=1}^{m} a_{n,k} x^n$. Each utility function is separately minimized: $NRMSE_k \xrightarrow{yields} min$. The quality of all clusters k is determined by $NRMSE = \sum_{k=1}^{p} NRMSE_k \cdot k^{-1}$.

3 Results and Conclusion

The clustering process identified a homogeneous group of residential and SME customers each, matching the available meta-data. During post-processing the identified standard load profiles where optimized by applying a dynamic function.

Using an ES to find appropriate parameters for dynamic functions leads to better results than the least square approach. Improvements of up to 21% measured by the mean average percentage error (MAPE) or 28% using the normalized rooted mean square error (NRMSE) are possible using a dynamic function with a polynomial function of grade 4 (compared to using no dynamic functions). Accuracy is improved especially during the peak time range.

The described methodology may be utilized to identify new and reliable SLPs also on a regional level to increase balance area loyalty during the “Energiewende”. The proposed clustering may help to automatically identify different types of residential household demand patterns.

3.1 Clustering Results

As this paper focuses on the methodology to optimize standard load profiles, the following section will only discuss results for $k = 2$ exemplarily. The methodology is applied to just two clusters for simplicity reasons.

Figure 4 represents the silhouette plot for the two clusters. The bottom part of the plot represents cluster 1, the smaller plot on top cluster 2. Figure 2 compares the clustering results for $k = 2$ with the available meta-data. The residential profile from cluster 1 represents 5950 smart meters and 80% of the electrical load. Cluster 2 consists of mainly SME profiles, representing 20% of the load. The silhouette plot in Figure 3 indicates a rather homogeneous cluster of residential profiles.

By conclusion, choosing two clusters divides the profiles into a residential and a SME group.

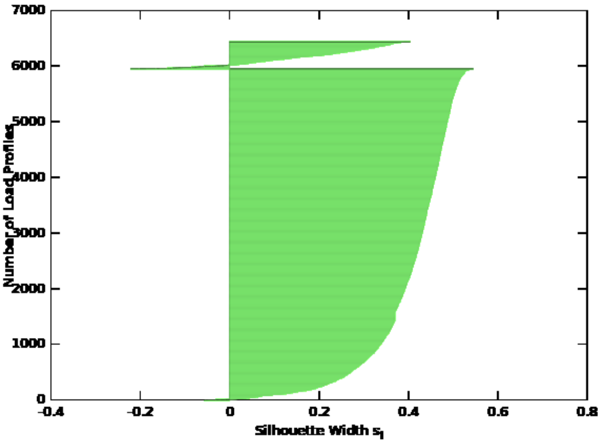

Figure 3: Silhouette plot of the two clusters. The plot for the SME is shown on top.

Table 2: Comparison of the clustering results with the available meta-data of the CER data set.

	Cluster 1	Cluster 2
Members (from total 6435)	5950	485
Percentage of total consumption	80.24%	19.76%
Residential profiles	4190	35
SME profiles	166	319
Other profiles	1594	131

3.2 Dynamic Functions

During post-processing, the standard load profile for $k = 2$ is normalized and three functions (one cosine and two polynomials of grade 4 and 6) for dynamic adaptation are applied. The parameters of the dynamic function are determined by the least square approach. In an additional step, the parameters of the dynamic functions are optimized with an evolutionary algorithm. The results of the parameter optimization are shown in Table 3. Please note that each dynamic function is optimized separately. It should also be noted that applying an ES does not guarantee to find the global optimal solution. The use of multiple populations increases the search area and decreases the possibility to "get trapped" in a local optimum.

Table 3: Parameters of the dynamic functions for each cluster.

	cos		poly4		poly6	
	C1	C2	C1	C2	C1	C2
A, A', A''	3.79	4.24	$-1.36 \cdot 10^{-16}$	$-1.62 \cdot 10^{-16}$	$-1.84 \cdot 10^{-24}$	$-2.43 \cdot 10^{-24}$
B, B', B''	-519.63	-3785.56	$4.81 \cdot 10^{-12}$	$5.91 \cdot 10^{-12}$	$9.44 \cdot 10^{-20}$	$1.14 \cdot 10^{-19}$
C, C', C''	0.99	0.95	$-4.68 \cdot 10^{-8}$	$-6.55 \cdot 10^{-8}$	$-1.75 \cdot 10^{-15}$	$-2.07 \cdot 10^{-15}$
D', D''			$7.29 \cdot 10^{-5}$	$1.8 \cdot 10^{-5}$	$1.76 \cdot 10^{-11}$	$1.93 \cdot 10^{-11}$
E', E''			1.24	1.1	$-9.23 \cdot 10^{-8}$	$-9.74 \cdot 10^{-8}$
F'					$1.59 \cdot 10^{-4}$	$1.59 \cdot 10^{-4}$
G''					1.1	1.1

Using an ES to find appropriate parameters for dynamic functions leads to better results. The optimized results are shown in Table 4. Improvements of up to 21% (MAPE, base) or 28% (NRMSE, base) are possible using a dynamic function with a polynomial function of grade 4 (compared to using no dynamic functions) if an ES is used. Accuracy is improved especially during the peak (8:00h – 20:00h) time range. The usage of the least square approach leads to a decrease in quality compared to the case where no dynamic function is used at all.

Table 4: Comparison of quality achieved by applying optimized dynamic functions.

Year 2010	NRMSE			MAPE		
Dyn.Func.	base	peak	off peak	base	peak	off peak
none	0.10708	0.09964	0.12406	8.40%	7.26%	9.53%
least square 4	0.15260	0.14370	0.17327	13.98%	12.85%	15.12%
least square 6	0.15247	0.14355	0.17317	13.96%	12.85%	15.12%
cos	0.07965	0.06874	0.10234	6.76%	5.49%	8.04%
poly4	0.07639	0.06486	0.09995	6.57%	5.17%	7.97%
poly6	0.09299	0.08160	0.11714	7.47%	6.23%	8.72%

Figures 5 and 6 show a comparison between a synthetic load profile and the total original CER load profile without application of a dynamic function (Figure 5) and with post-processed SLPs with an optimized dynamic function of grade 4 (Figure 6). The synthetic load profile uses a poly4 dynamic function and is assembled from two SLPs for residential and SME customers. Comparing these two Figures shows that the gap during the seasonal shift is reduced.

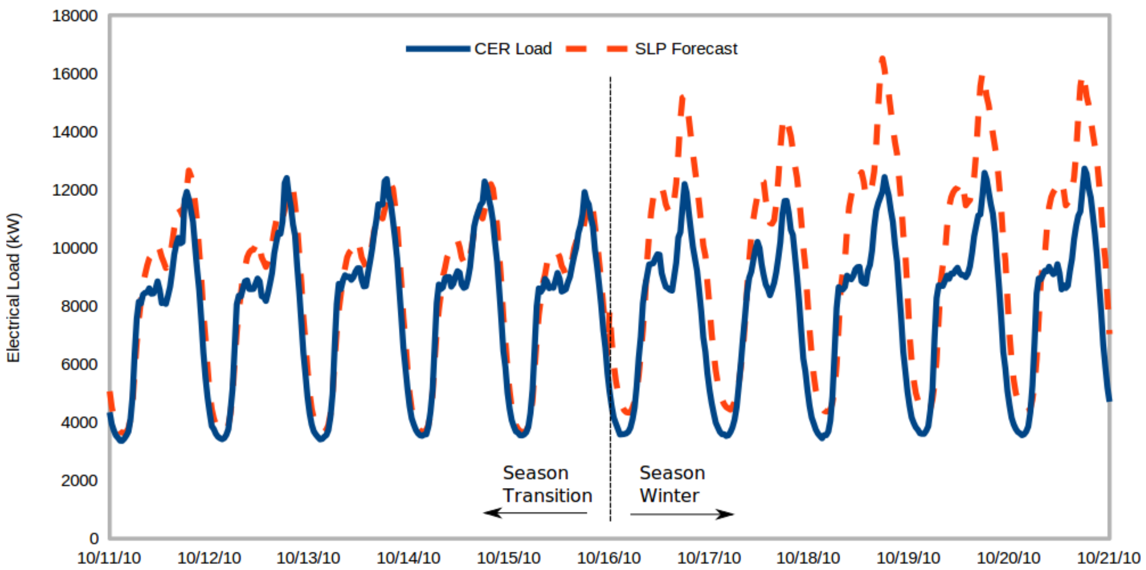

Figure 4: Comparison of the CER load and the synthetic load profile without a dynamic function at ~~a~~ seasonal change.

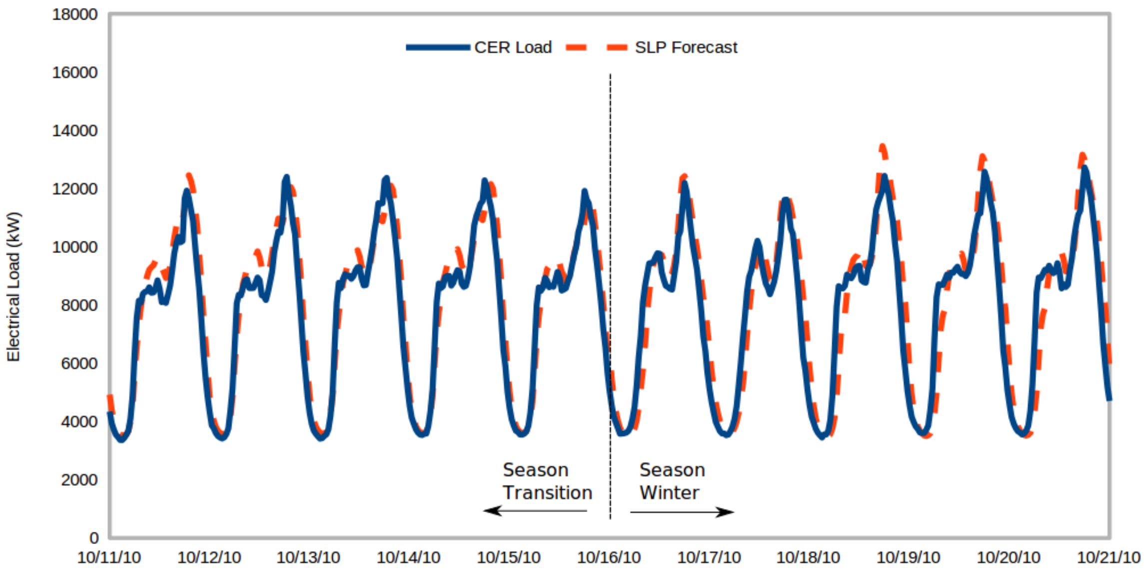

Figure 5: Comparison of the CER load and a synthetic load profile generated from a dynamic standard load profile at ~~a~~ seasonal change.

3.3 Concluding Remarks

This paper described the extraction of SLPs from a smart meter data set provided by the Irish CER. The data set is iteratively clustered using a k-means algorithm. Synthetic load profiles are then evaluated with commonly used metrics such as the MAPE and the NRMSE for different time frames (base, peak and off-peak). Due to strong seasonality of the mainly residential smart meter data, a dynamic function is required to improve the quality of synthetic load profiles.

Three different dynamic functions are evaluated. A multi population ES is used to optimize the parameters of the dynamic functions for each cluster separately. In the analyzed case for $k = 2$ a polynomial function of grade 4 outperformed a cosine and a polynomial function of grade 6. The application of a dynamic function improved the overall quality of the synthetic load profile.

The described methodology may be applied to identify new and reliable SLPs also on a regional level to increase the balance area loyalty during the “Energiewende”. The clustering as suggested may help to automatically identify different type of residential household demand patterns.

However, this approach should be considered as proof of concept. Many obstacles remain and should be kept in mind by the reader of this paper:

In this analysis, one example with two clusters ($k = 2$) was chosen for the purpose of demonstrating the process. Without further information on the households or the SME, a clear classification of the identified profiles remains difficult. Analyzing a load profile data set with additional meta-data could help to find recommendations for k and to clearly assign dynamic load profiles to customer segments.

It should also be noted that the small data basis of 1.5 years may lead to an overfitting of the dynamic functions.

In this analysis the influence of temperature, global radiation, wind chill effects and other influence factors are ignored. This could be considered as a general problem of SLP load forecasting. For reliable forecasts, this additional input should especially be used for short term forecasts. For mid to long term forecasts statistical temperature profiles as well as extremes should be used in order to increase performance and fulfill risk management requirements.

Acknowledgement

We would like to thank the Commission for Energy Regulation (CER) and the Irish Social Science Data Archive (ISSDA) for making the smart meter data sets of residential and SME customers available for research. Furthermore I thank the Fraunhofer Center for International Management and Knowledge Economy for their support and Mr. U. Groth for proof reading.

Bibliography

1. Schieferdecker, B. and Funfgeld, C. and Meier, H. and Adam, T., "Repräsentative VDEW-Lastprofile", VDEW-Materialien M-28/99, Frankfurt (1999).
2. Scheller, F.; Keitsch, K.; Kondziella, H.; Reichelt, D. G.; Dienst, S.; Kühne, S.; Bruckner, T.: Evaluation von Geschäftsmodellen im liberalisierten Energiemarkt. BWK 67 Nr. 11 (2015).
3. BMWi, Bundesministerium für Wirtschaft und Energie, "Ein Strommarkt für die Energiewende (Weißbuch)" (2014).
4. BMWi, Bundesministerium für Wirtschaft und Energie, "Referentenentwurf Gesetz DIGITALISIERUNG DER ENERGIEWENDE", http://www.bmwi.de/BMWi/Redaktion/-PDF/P-R/referentenentwurf-entwurf-gesetz-digitalisierung-energiewende (2015).
5. Gerblinger, A. and Finkel, M. and Witzmann, R, "Neuen Standardlastprofile für den Eigenverbrauch von elektrischer Energie bei Haushaltskunden", Energiewirtschaftliche Tagesfragen (2014).
6. Gerblinger, A. and Finkel, M. and Witzmann, R., "Entwicklung und Evaluation von neuen Standardlastprofilen für Haushaltskunden" (2014).
7. Bundesnetzagentur, Bundeskartelamt, "Monitoringbericht 2014" (2014).
8. CER, “Electricity Smart Metering Customer Behaviour Trials (CBT) Findings Report” (2011).
9. MacQueen, J., Some methods for classification and analysis of multivariate observations, 281-297 (1967).
10. Rousseeuw, P. J., Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics 20, 53-65 (1987).
11. Rechenberg, I., Evolutionsstrategie'94. Frommann-Holzboog Verlag, Stuttgart (1994).