Author Accepted Manuscript version of the paper by Kasper Emil Thorvaldsen, Venkatachalam Laksmanan and Hanne Sæle
in 2023 International Conference on Smart Energy Systems and Technologies - SEST (2023)
http://dx.doi.org/10.1109/SEST57387.2023.10257550

Analysis of Accuracy of Flexibility Baseline Prediction Methods for Office Building at Different Measuring Points

Kasper Emil Thorvaldsen
Dept. of Energy Systems
SINTEF Energy Research
Trondheim, Norway
kasper.thorvaldsen@sintef.no

Venkatachalam Lakshmanan
Dept. of Energy Systems
SINTEF Energy Research
Trondheim, Norway
venkat@sintef.no

Hanne Sæle
Dept. of Energy Systems
SINTEF Energy Research
Trondheim, Norway
Hanne.Saele@sintef.no

***Abstract*—Customer baseline load (CBL) prediction plays an important role in calculating the volume and value of the flexibility provided by end-users. In this paper, two different CBL methods are applied to investigate their prediction accuracy for a given load with high resolution metered data. One of the CBL methods makes use of historical data, named CBL_{XofY}, while the other makes use of the load pattern before/after the CBL-prediction, denoted as $CBL_{B/A}$. A real office with high resolution load data is used to investigate CBL prediction accuracy at multiple measuring points, for the different CBL methods. The results show that $CBL_{B/A}$ has a high level of accuracy at the office-building level, due to an internal 200-kW threshold for import that made the load profile flat during the midday. As the load increases throughout the morning, both methods undershoot the accuracy, where $CBL_{B/A}$ undershoots by 13%, while CBL_{XofY} undershoots by 4-5%. In the electric vehicle (EV) parking lot, there is a noticeable offset for both CBL-methods, as the lot is the internal throttling mechanism for maintaining the 200-kW threshold at the building level. This analysis has captured the importance of measuring points for calculating CBL when an internal demand response is available within the building, which can cause noise and inaccuracy.**

***Index Terms*—Demand profile, Distribution networks, Electric vehicles, Flexibility baseline.**

I. Introduction

Power system operation is becoming challenging due to an increasing share of variable renewable resources [1]. The non-dispatchable renewable resources require flexibility from other resources, namely demand and storage. In Norway, where over 90% of the generation is dispatchable hydroelectric, flexibility is important for maximising the network's capacity utilisation and avoiding new high capital investments. For example, short-term demand peaks and mismatch between distributed photovoltaic (PV) generation and local demand causes network congestion and voltage problems in transmission and distribution networks, which could be avoided with flexibility at demand side [2]. Historically, both implicit and explicit flexibility have been used, but explicit flexibility has mainly focused on larger customers. Explicit flexibility enables the network operator or a third party to activate flexibility directly by accessing the customers' flexibility resource [3]. However, to pay the customer accordingly, a load baseline is needed [4]. Customer baseline load (CBL) is a means of calculating the expected load for a consumer when they have performed a flexible action. The deviation between the CBL and measured energy usage is equal to the amount of flexibility delivered.

Multiple CBL estimation methods have been described in literature, with their own merits and demerits [5]–[17]. As the demand pattern is not same for all types of customers, these methods must be quantitatively evaluated. Some of the CBL estimations are simple and transparent, while others are mathematically complex. When a CBL estimation method is used, the level of the customer's technical expertise matters in order to avoid any dispute while financial settlements are made between the parties providing flexibility, the flexibility user, and the aggregator transacting the flexibility between them.

In this paper, different CBL methods provided in the literature [5], [6] are quantitatively evaluated for commercial consumer segment in Norway. Additionally, some of the different methods discussed in the literature are shortly described and are evaluated using high resolution smart grid data recorded from an office building in Norway with a parking lot for EVs. Methods using measurements before/after activation of flexibility, and through use of historical data, are carried out in this analysis. The analysis investigates the performance of these CBL-methods over a given workday in October 2022. Due to access to detailed consumption data, both in the EV parking lot and at the office building level, accuracy at different measurement levels were investigated. The analysis was also applied to each hour during a workday, to capture the accuracy in prediction based on time of day.

Our main contributions are the following:

- Two different CBL methods are applied to investigate the accuracy of their prediction for a given load.
- Real-life office load data at different measuring points with high resolution is used to investigate the accuracy of CBL prediction at 5-minute resolution.
- The accuracy of CBL prediction is analyzed at different measurement levels, capturing the influence from active in-house demand response.

The research leading to this publication received funding from CINELDI - Centre for Intelligent Electricity Distribution, an 8-year Research Centre under the FME-scheme (Centre for Environment-friendly Energy Research, 257626/E20). The authors gratefully acknowledge the financial support from the Research Council of Norway and the CINELDI partners. Additionally, thanks to Lede AS and the partners in the Strømfleks project (https://lede.no/stromfleks/) for the load data.

The structure of the paper is as follows: First, a literature study on different CBL methods will be presented in Section II. The chosen CBL methods are described in Section III, with the case study in Section IV. The results and general discussion surrounding the results will be provided in Section V, followed by the conclusion and future work in Section VI.

II. Literature study

The International Energy Agency (IEA) states that "Power system flexibility is the ability of a power system to reliably and cost-effectively manage the variability and uncertainty of demand and supply across all relevant timescales" [18]. The deviation from unaltered demand or generation profile is the actual flexibility provided. Once the flexibility is activated, the unaltered profile is not available for calculation of flexibility. Fig. 1 shows an example of flexibility activation, consequent demand adjusted, and resulting flexibility provided. Different methods for calculating flexibility in the absence of unaltered profile are proposed in literature [5]–[17].

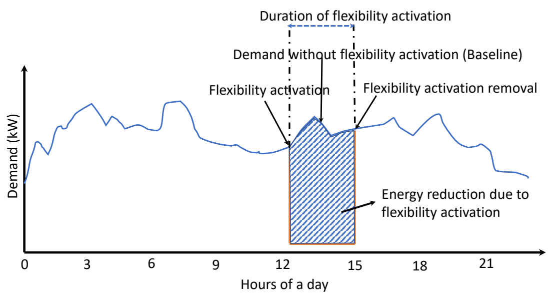

Fig. 1. Baseline demand and flexibility activation.

A. Window before

The 'window before' method considers the last measured power before flexibility activation as reference for calculating the baseline [5], [8]. The baseline profile is assumed to be a straight line with a value corresponding to the last consumption before flexibility activation. This method is very simple and transparent. The main disadvantage is its lack of immunity to manipulation. The flexibility resource owner can artificially alter their consumption if the hour of activation is known well before the activation.

B. Window before and after

The 'window before and after' method considers the measurements before and after the flexibility activation. Like the 'window before' method, the baseline power profile is assumed to be a straight line with a value that is the average of the last measurement before flexibility activation and the measured value after flexibility activation is removed [7]. It is also a simple and transparent method, but with the same disadvantage of manipulation of measurements before and after. In addition, the flexibility resources like electric water heaters and space heater, which have a rebound effect, will affect the 'after' measurement, while not a forced manipulation.

C. Historical or averaging (also called XofY)

In the historical or averaging method, an average value from the past is taken as the baseline. It is also called XofY, which means an 'X' number of days from the past 'Y' number of days are considered [5]. As flexibility activation hides the true consumption and potential, the 'X' number of days taken for averaging are the days without flexibility activation in the last 'Y' number of days [9]. As loads can vary depending on the days being weekdays or -ends, filtering the 'Y' days to only include similar days can be performed. In the case discussed in this paper, the assumption is that there has been no flexibility activation in the past. Therefore, only the recent 10 days are the reference.

D. Calculated

The baseline can be calculated in two simple ways. One transparent way is a linear interpolation between the two measurements before flexibility activation and after flexibility activation removal [10]. Another way is to use a regression technique with measured data before and after the flexibility period [11]. For a common customer, regression could be complex method and not a transparent method.

E. Machine learning

Machine learning (ML) follows a similar approach to regression by using data set for baseline calculation. Reference [6] describes an ML-based approach for a residential demand baseline calculation, which uses a neural networks-based approach. A clustering-based approach for baseline calculation is detailed in [12]. ML could use data from multiple days, find the similarity in the data set, and produce baseline data for the flexibility activation period. Unlike regression, ML does not necessarily need to fit the data into one mathematical function. For the common customer, ML is a black box and the method is not transparent. Multiple methods could be combined to calculate the baseline. For example, reference [13] uses multiple regression methods, which could be combined to find the coefficients of the equation that represents the baseline, or a deep learning and quantile regression on a pool of data where there has been no flexibility activation to extract the coefficients to calculate a baseline.

F. Control group

The control group uses other customers' demand during flexibility activation as a reference for baseline. In a population of consumers, a segment for which flexibility is not activated forms a reference group. The group for which the flexibility is activated is called control group [14]. The control group may be selected randomly or be well defined and non-experimental [15]. The delivered flexibility is the difference between the reference and control group. For this study, it is assumed the control group and reference group is the same or has a negligible difference during the period of flexibility activation. In reality, the method is not simple nor transparent, as the data of the reference group is not available to everyone.

G. Prognosis

The prognosis provides the expected behaviour of the flexible consumer, which is taken as the baseline for reference. There could be many forecasting methods that could be used for prognosis. The main difference between the prognosis and the other method is that the baseline is estimated before flexibility activation [16], [17]. It is like the balance responsible parties (BRPs) commitment to transmission system operator (TSO). The BRPs are obligated to provide their hourly generation plan to the TSO, so that any deviations can be calculated as imbalance. Other methods estimate baseline during settlement phase and prognosis is useful for bidding phase.

III. Methodology

As described in Section II, multiple CBL methodologies can be applied that account for behaviour around the activation window, make use of historical data to predict the load pattern, and utilise machine learning or control groups. The methods vary in both their complexity and need for data.

This work investigates two different CBL methodologies based on the existing methods described in Section II. First, the ”window before and after” method, which looks at the measurement data before and after activation is used. This analysis only considers the average data before/after activation, named $CBL_{B/A}$. This method only relies on measurement point before and after the activation period.

The second method is the ”historical or averaging” method, denoted as CBL_{XofY} in this work. By using historical data points, the method creates an expectation of the load pattern, finding the historical average load. A key strength of this method is that it limits manipulation of data points by limiting influence from extreme load data, unless the flexibility provider frequently manipulates load. One weakness is that the data used can be sensitive to seasonal, daily and hourly variations, meaning that the reference data should be chosen carefully. Therefore, the data points collected can be filtered to exclude extreme data points. The filtering process used in this work is based on the interquartile range technique, using percentiles to define outliers [19]. The percentile range should be balanced with the data used, to avoid extreme data points being included. The analysis performed in this work uses measurements without flexibility activation to evaluate the accuracy of the CBL prediction methods.

The two CBL methodologies were chosen due their simplicity and transparency, and ability to predict CBL without extensive amounts of metered data. These considerations are the basic requirement by the regulatory authority to avoid disputes in settlement between the parties involved in the flexibility value chain [5].

IV. Case Study

The case study for this analysis is a modern office building located in south-eastern Norway, part of the ”Strømfleks” project [20]. The office building has multiple flexible loads available, and detailed monitoring of their electricity consumption during operation, including cooling/heating, PV-system, and an EV parking lot with 50 charging points. During summer 2022, the parking lot was subject to several manual throttling cases, where the charging level was throttled down to 20 kW. Therefore, this case study assumes the EV parking lot to be the source of flexibility activation, which would prompt the need for calculating the CBL.

Several metering points for measuring the electricity consumption are placed within the office building. For this case study, two of these metering points are of interest: the whole office building, and the EV parking lot. The data is given at 5-minute resolution, providing detailed characteristics of the load profiles and their variation within an hour.

The office building has incorporated some internal demand responses using the EV parking lot. The following internal demand responses are known:

- The EV parking lot has a dynamic throttling system throughout the day, where it is throttled down to 40 kWh/h during the busiest morning hours (6-9 AM).
- The office building has a 200-kW threshold measured at building-level during winter months, when the EV parking lot is actively throttled to stay below this threshold.
- PV production at the building level increases the throttling level for the EV parking lot.

Internal demand response originating from the EV parking lot can interrupt the accuracy of the CBL-calculation. Since the 200-kW threshold measured at building-level is maintained by actively throttling the EV parking lot, this external influence on charging capacity makes it valuable to investigate for comparing the CBL accuracy at both measurement points.

A. Case runs

To test the effectiveness of the two CBL-methodologies, the office building is analysed for Tuesday October 4th 2022. During this period, heating demand increases and the load pattern more frequently experiences a 200-kW cap throughout the day. The analysis is conducted for each hour from 7 AM to 5 PM, to capture the time-dependent variations. For all cases, 5-minute resolutions are used within the CBL analysis. Each CBL activation lasts for 1 hour. For CBL_{XofY}, 5-minute resolutions are gathered for the last 10 weekdays, and the filtering process assume percentile range of 25th-75th. $CBL_{B/A}$ assume hourly average (60 minutes). Additionally, two measurement points within the office building will be subject for the CBL analysis; the building-level and the EV parking lot. Overall, this will provide a detailed analysis of how the temporal variance of each CBL behaves, at both measurement levels.

V. Results & Discussions

The different CBL methodologies estimate what the load would be if a flexible action is not performed. Comparing the accuracy of the CBL methodologies is done by analysing the variance of CBL and the actual load. This is done by simulating a flexible action for different time steps, enabling us to compare to the actual baseline load. Since we analyse different CBL methodologies, as well as different measurement points of CBL calculation, it is relevant to compare the accuracy of both the methodologies and measurement points. The CBL accuracy of both methods, measured at the building level, are presented in Section V-A. The accuracy of CBL calculation for the EV parking lot will be shown in Section V-B, followed by a discussion and comparison of accuracy at the given measurement levels in Section V-C.

Author Accepted Manuscript version of the paper by Kasper Emil Thorvaldsen, Venkatachalam Laksmanan and Hanne Sæle
in 2023 International Conference on Smart Energy Systems and Technologies - SEST (2023)
http://dx.doi.org/10.1109/SEST57387.2023.10257550

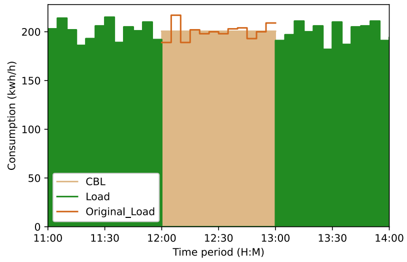

Fig. 2. Load profile measured at building-level for 11-2 PM, with CBL-calculation $CBL_{B/A}$ between 12-1 PM.

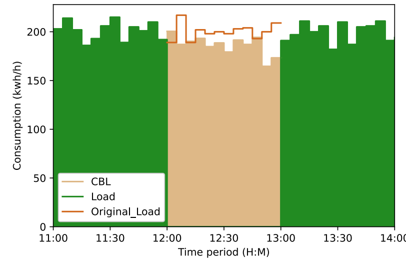

Fig. 3. Load profile measured at building-level for 11-2 PM, with CBL-calculation CBL_{XofY} between 12-1 PM.

A. CBL accuracy at building-level

The upper part of table I showcases the temporal performance of both the $CBL_{B/A}$ and CBL_{XofY} methods at the building level from 7 AM to 5 PM. In general, the load for the office building increases during the morning, reaching the 200-kW threshold from 9 AM. This load is maintained until 4 PM, when it decreases.

During the morning, the accuracy of each method varies, and undershoots in most cases. For $CBL_{B/A}$, it undershoots the most at 7 AM, with a 13% offset compared to the actual load during this hour. CBL_{XofY} also undershoots during this period, but with a more stable quantity and less offset, at about 5-7 kWh (about 4-5% offset).

During the midday, the overall load is flat at around 200 kW, which makes the $CBL_{B/A}$ very accurate. With this flat load profile, the average of previous and next hour provides high accuracy for load prediction. For the CBL_{XofY}, there is a noticeable trend that it undershoots the expected load during this period. Here, some variation for the different number of used data points is showcased, where the case with outliers performs worse than when the outliers are filtered out. Within the last ten workdays, there has been some behavior that has assisted in undershooting the expected CBL volume, which the filtering process has managed to remove, providing better accuracy. This demonstrates that there is some value in filtering the input data, since there are occurrences that would create noise or unwanted trends in the load pattern.

Figs. 2 and 3 showcase the accuracy of both CBL methodologies for 12-1 PM. Due to the flat consumption profile before and after this period, the $CBL_{B/A}$ is very accurate during this period on an hourly average, but is not able to capture the variations within the hour. In Fig. 3, the CBL_{XofY} noticeably undershoots the estimated electricity consumption, with a total offset by 14 kWh (7%). For the last two intervals, the estimation is off by up to 40 kWh/h, indicating that the historical data points are struggling to accurately predict the load pattern at this hour.

From 4 PM and onward, there is a noticeable inaccuracy from both CBL methodologies. For $CBL_{B/A}$, the decrease in load at the end of the day has a detrimental effect on its accuracy. The same trend is seen for the historical data points in CBL_{XofY}, but the load decrease starts much earlier, from around 1-2 PM. The inaccuracy of the CBL prediction increases until the last hour of the analysis, which suggests that the high load during these hours are not normally captured by the historical data points.

The historical offset for CBL_{XofY} at the building level suggests that the electricity consumption has increased more than expected. The offset during the midday indicates that the recent historical occurrences did not experience 200-kW levels. This load increase is most likely due to increasing heating load in response to colder temperatures. The increase in load at the end of the day implies that there might be a rebound effect in the system or load increase, which might origin from the EV parking lot, as it throttles more actively to keep the 200-kW level.

B. CBL accuracy for the EV parking lot

The lower part of table I presents the temporal performance of CBL measured for the EV parking lot. The load pattern for the EV parking lot increases throughout the day as more EVs arrive, with load between 70-90 kWh/h during midday. The variation in load indicates the EV charging capacity is continuously throttled to keep the overall building-level load at 200 kW. Throughout the entire day, the performance of the $CBL_{B/A}$ varies to a larger degree than it does at the building level. During the morning, it does not manage to capture the 40 kW limit between 7-9 AM, but both under- and overshoots the estimate. The accuracy increases at midday, with some offset at certain hours. Since the EV load fluctuates during the day, this method is not able to accurately portray the changes in load. The accuracy of the historical data in CBL_{XofY} also fluctuates at certain times of the day. During the morning, the 40 kW limit is captured well, with accurate predictions of the limit regardless of filtering. During midday, the trend for the historical data points overshoot by about 10 kWh, predicting the load to be around 90 kWh from 11 AM to 2 PM when filtering data points.

In the evening, the historical data points dramatically undershoot the CBL. The CBL_{XofY} estimates load of 48 kWh at 3 PM, and 14-15 kWh at 4 PM, but the actual load is 74 kWh and 35 kWh, respectively. This change in volumetric accuracy indicates that there has been an active internal demand response for EV charging during midday, leading to a rebound effect of EV load at the end of the workday. This is not captured in the historical data, causing a significant error during both periods. This mismatch indicates that the throttling is more actively used now than before, resulting in changes at the end of the workday.

Author Accepted Manuscript version of the paper by Kasper Emil Thorvaldsen, Venkatachalam Laksmanan and Hanne Sæle
in 2023 International Conference on Smart Energy Systems and Technologies - SEST (2023)
http://dx.doi.org/10.1109/SEST57387.2023.10257550

TABLE I
OVERVIEW OVER CBL-PERFORMANCE AT BUILDING-LEVEL AND EV PARKING LOT [IN KWH], FOR BEFORE/AFTER, AND XOFY WITH 10 DATA POINTS.

Case\Time [h]	7	8	9	10	11	12	13	14	15	16	17
$P^{Building}_{Load}$	167	175	199	199	201	200	200	200	194	137	98
$CBL^{Building}_{B/A}$	144.96	183.17	187.00	200.17	199.79	200.54	200.29	196.67	168.79	145.75	109.25
$CBL^{Building}_{x=y=10}$	159.63	170.7	194.66	185.36	188.48	180.35	174.93	161.66	141.62	102.85	93.51
$CBL^{Building}_{x=7,y=10}$	161.82	169.02	194.54	184.94	195.08	186.17	178.32	161.66	141.07	101.31	92.18
P^{EV}_{Load}	38.48	42.28	76.44	71.36	80.49	91.65	83.9	74.29	74.31	34.68	7.93
$CBL^{EV}_{B/A}$	28.06	57.46	56.82	78.47	81.50	82.20	82.97	79.10	54.48	41.12	19.53
$CBL^{EV}_{x=y=10}$	38.57	41.19	71.81	71.67	84.56	83.84	84.67	70.95	48.66	15.75	4.97
$CBL^{EV}_{x=7,y=10}$	40.55	41.90	72.80	74.26	90.0	89.77	93.15	73.43	48.66	14.43	3.80

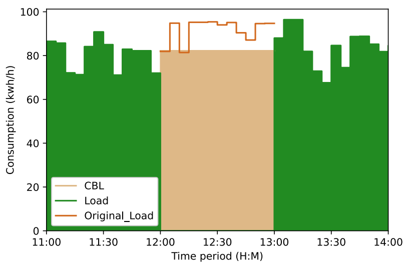

Fig. 4. Load profile measured for the EV parking lot for 11-2 PM, with CBL calculation $CBL_{B/A}$ between 12-1 PM.

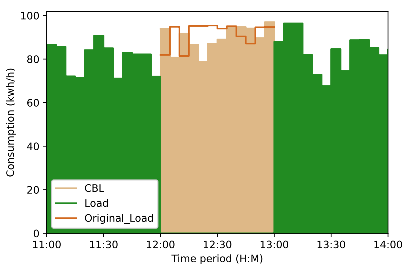

Fig. 5. Load profile measured for EV parking lot for 11-2 PM, with CBL calculation CBL_{XofY} between 12-1 PM.

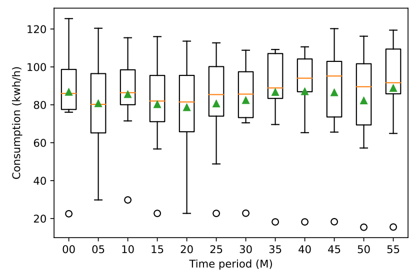

Fig. 6. Boxplot of historical data points at the EV parking lot used in CBL_{XofY} for each 5-minute resolution between 12-1 PM.

Figs. 4 and 5 showcase the accuracy of the CBL calculation from 12-1 PM for the EV parking lot. Fig. 4 illustrates that the load fluctuates at midday, which causes the $CBL_{B/A}$ to have some mismatch during this hour. Looking at each 5-minute resolution, the load shifts between 70-100 kWh/h frequently during the 3 hours included, making it difficult to accurately predict the CBL using only trends before and after activation. Similar patterns are seen using historical data in Fig. 5, as the 5-minute resolution is off most of the time between 12-1 PM. The variation indicates that the load pattern is not accurately predictable, and that load changes during the hour do not necessarily follow any specific trend.

The variation in load data used as basis for the CBL_{XofY} at the EV-level is showcased in Fig. 6, without filtering. For each 5-minute resolution, the data points for the last 10 workdays before October 4th show a large variation in consumption. There is at least one day where the consumption was extremely low, at around 20 kWh/h for the period, which is flagged as an outlier for almost all cases. The percentiles limit range between 25-75 does not filter every single outlier during this period in Fig. 6. Tuning the filtering process could have improved the accuracy by ignoring the extreme data points for all time steps. The 20 kWh/h outliers explains why filtering leads to higher average load for this hour. Most of the data points operate between 60-120 kWh/h as indicated by the whiskers, which is a significant variation for these 10 workdays. The data points on the upper part of the whisker, indicate that the load has been much higher previously, but other factors make it decrease in amplitude. Since the whiskers also have a large span, this also indicates that the load fluctuates to some degree for all historical events, and the mean and median values just capture an "expected trend".

C. Comparison of accuracy at the measurement levels

The performance of the two CBL methodologies change depending on the measurement level. For the $CBL_{B/A}$ in Table I, it is apparent that the building has the best performance due to a flat consumption profile during the midday. The EV parking lot has higher variation in load, making this CBL method inaccurate from hour to hour. There are some inaccuracies for the morning and evening hours due to the increasing load in the building and EV parking lot. The 200-kW cap at the building level works as a good guidance for $CBL_{B/A}$, causing some predictability in the load pattern.

For the CBL_{XofY}, both measurement levels experience inaccuracies throughout the day. During the morning, the EV parking lot accurately captures the load, due to its 40-kW cap. During the midday, this method undershoots the expected load at the building level. For the EV parking lot, this trend is the opposite; the historical data predicts a higher EV load during this period, causing the CBL method to overshoot the expected load. The offset at both levels imply a change in load for the building; the 200-kW cap initiates with increasing loads in other sectors of the building, causing a throttling of the EV parking lot. This causes a rebound effect in the evening, where the load at both measurement levels is higher than expected.

The inaccuracies from the historical data points are a result of noise and change in load behavior for the building overall. The building level has more noise affecting historical prediction due to increasing load from other sources. The inaccuracy could have been larger if not for the 200-kW cap. If CBL_{XofY} is used to predict the load in winter in the future, this offset can be expected to decrease, and the 200-kW cap would be reflected within the prediction shortly. Thus, the offset is generally the result of a seasonal transition in the load within the building, which has not stabilised for the historical pattern. Other noise, such as EV load and PV production, would also cause inaccuracies, but the 200-kW cap dominates more during the midday. For the EV parking lot, noise from the internal demand response is the main factor behind the inaccuracies. Since the demand response is an external factor, it is difficult to accurately account for this within the predictions, especially when a seasonal transition can lead to a higher external influence on the charging capacity. During the morning, the 40-kW cap is accurately portrayed, but the midday and the evening can experience more variation, depending on the stability of the external noise. Since a decrease in the midday load can lead to an increase in the evening period, there are longer periods at the EV parking lot where it is difficult to accurately predict the load.

VI. Conclusion

CBL is highly relevant for flexibility settlement between the parties involved in flexibility procurement and activation. This paper has quantitatively evaluated two different methods for calculating CBL, namely $CBL_{B/A}$ and CBL_{XofY}, using smart meter data from an office building. The data showcases the delicacy and difficulty of predicting the customer load in order to produce a baseline consumption pattern for flexibility purposes. The accuracy of calculating the CBL was investigated using two methods: one surrounding load patterns around activation, and another using historical data. The results indicate that internal demand response actions influence the accuracy both positively and negatively. At the building level, the $CBL_{B/A}$ method performed the best during midday. With the historical data, the building level load undershot most of the time, as the measurement history did not capture the demand cap. For the EV parking lot, both methods experienced variation in accuracy. The use of historical data points overshot the prediction due to an expected higher load during the midday. During the morning, when the demand started to increase in the office building, the $CBL_{B/A}$ method undershot the actual level by 13%. In comparison, the CBL_{XofY} method demonstrated a better accuracy, undershooting at 4-5%. $CBL_{B/A}$ is suitable for aggregated level, with low load variations. $CBL_{B/A}$ and CBL_{XofY} are not suitable for random, fast varying and rebounding loads like EV parking lots when actively throttled. This paper only evaluated two CBL methods for one type of consumer. The flexibility market value chain features many stakeholders, such as the flexibility provider, aggregator, and end-user. It is not necessary to use the same methodology for all stakeholders as long as the method is accurate enough. Therefore, the authors propose that the other CBL methodologies be evaluated for different consumer types as well as other stakeholders in the future.

References

[1] A. Monterrat, C. Carrejo, S. Hilliard, and F. Devaux, "Integration Cost of Variable Renewable Resources to Power Systems - A Techno-economic Assessment in European Countries," *10th IEEE International Conference on Renewable Energy Research and Applications, ICRERA 2021*, pp. 210–215, 2021.

[2] H. Chang and A. Moser, "Benefits of a combined flexibility utilisation between TSO and DSO for congestion management," *CIRED - Open Access Proceedings Journal*, vol. 2020, no. 1, pp. 758–760, 2020.

[3] H. Liao and J. V. Milanović, "Flexibility Exchange Strategy to Facilitate Congestion and Voltage Profile Management in Power Networks," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 4786–4794, 9 2018.

[4] F. G. Venegas, M. Petit, and Y. Perez, "Can DERs fully participate in emerging local flexibility tenders?" *International Conference on the European Energy Market, EEM*, vol. 2019-September, 9 2019.

[5] C. Ziras, C. Heinrich, and H. W. Bindner, "Why baselines are not suited for local flexibility markets," *Renewable and Sustainable Energy Reviews*, vol. 135, 1 2021.

[6] J. Jazaeri, T. Alpcan, R. Gordon, M. Brandao, T. Hoban, and C. Seeling, "Baseline methodologies for small scale residential demand response," *IEEE PES Innovative Smart Grid Technologies Conference Europe*, pp. 747–752, 12 2016.

[7] O. Valentini, N. Andreadou, P. Bertoldi, A. Lucas, I. Saviuc, and E. Kotsakis, "Demand Response Impact Evaluation: A Review of Methods for Estimating the Customer Baseline Load," 7 2022.

[8] N. Rossetto, "POLICY BRIEF Measuring the Intangible: An Overview of the Methodologies for Calculating Customer Baseline Load in PJM," Tech. Rep. [Online]. Available: https://www.ferc.gov/industries/electric/indus-act/demand-response/dem-res-adv-metering.asp.

[9] T. K. Wijaya, M. Vasirani, and K. Aberer, "When Bias Matters: An Economic Assessment of Demand Response Baselines for Residential Customers," *IEEE Transactions on Smart Grid*, vol. 5, no. 4, pp. 1755–1763, 2014.

[10] L. Hatton, P. Charpentier, and E. Matzner-Løber, "Statistical Estimation of the Residential Baseline," *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 1752–1759, 2016.

[11] X. Zhou, N. Yu, W. Yao, and R. Johnson, "Forecast load impact from demand response resources," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*, 2016, pp. 1–5.

[12] M. Sun, Y. Wang, F. Teng, Y. Ye, G. Strbac, and C. Kang, "Clustering-Based Residential Baseline Estimation: A Probabilistic Perspective," *IEEE Transactions on Smart Grid*, vol. 10, no. 6, pp. 6014–6028, 2019.

[13] K. Li, F. Wang, Z. Mi, M. Fotuhi-Firuzabad, N. Duić, and T. Wang, "Capacity and output power estimation approach of individual behind-the-meter distributed photovoltaic system for demand response baseline estimation," *Applied Energy*, vol. 253, p. 113595, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261919312693

[14] S. Mohajeryami, M. Doostan, and P. Schwarz, "The impact of Customer Baseline Load (CBL) calculation methods on Peak Time Rebate program offered to residential customers," *Electric Power Systems Research*, vol. 137, pp. 59–65, 8 2016.

[15] K. Li, B. Wang, Z. Wang, F. Wang, Z. Mi, and Z. Zhen, "A Baseline Load Estimation Approach for Residential Customer based on Load Pattern Clustering," *Energy Procedia*, vol. 142, pp. 2042–2049, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1876610217361581

[16] USEF Foundation, "USEF: THE FRAMEWORK EXPLAINED," USEF Foundation, Tech. Rep., 11 2015.

[17] Hans de Heer (USEF), Marten van der Laan (USEF), and et al., "USEF: WORKSTREAM ON AGGREGATOR IMPLEMENTATION MODELS," Tech. Rep., 9 2017. [Online]. Available: https://www.usef.energy/app/uploads/2017/09/Recommended-practices-for-DR-market-design-2.pdf

[18] International Energy Agency, "Status of Power System Transformation 2019: Power system flexibility," *OECD Publishing*, 2019.

[19] "Finding outliers in dataset using python — by Renu Khandelwal — DataDrivenInvestor." [Online]. Available: https://medium.datadriveninvestor.com/finding-outliers-in-dataset-using-python-efc3fce6ce32

[20] "StrømFleks - Lede." [Online]. Available: https://lede.no/stromfleks/