Energy Proceedings
Vol 28, 2022
ISSN 2004-2965

Impact of Spatial Aggregation on Electricity Profiles

Alaa Krayem[1] *, Fredrik Wallin [1]

1 Future Energy Center, Department of Energy, Building and Environment, Mälardalen University, Västerås, Sweden

(alaa.krayem@mdu.se)

ABSTRACT

The increasing use of renewable energy and the spread of smart energy services require detailed studies of the system load. The data from the advanced metering infrastructure assist these requirements by providing detailed picture of electricity consumption patterns and profiles. However, high resolution electricity data can cause computational challenges and privacy concerns. As a result, the data are often spatially aggregated. This paper investigates the impact of data aggregation on the data understanding and the electricity load characteristics. The study looks at the similarity among different groups combinations within the same aggregation level, the variation in the load diversity and peaks occurrences, and on the hourly electricity variations between the individual customer and its aggregated group. The study concludes that the individual customers' behaviors are lost with the increasing levels of aggregation, and that the similarity among groups on the same aggregation level increases with the aggregation level.

Keywords: Electricity load, spatial aggregation, demand diversity, correlation, peak demand

1. INTRODUCTION

Due to the improvement of renewable energy systems and the increased use of distributed energy resources, the urban energy sector is evolving into a more complex system [1]. Furthermore, due to the expansion of Information and Communication Technologies (ICT) and the spread of smart energy services via smart metering infrastructures, the sector is enhancing its capacity for novel solutions [2]. Therefore, detailed studies of the system load are becoming increasingly important to increase energy system reliability and efficiency [3]. The massive amount of data measured by smart meters allow for more accurate and detailed load profiling, modeling, and forecasting. However, spatial aggregation is a common practice in energy studies, especially for appropriate sizing of the systems [4], [5], understanding the dynamics of the load at different levels of the distribution network [6], or overcoming computational challenges in the system simulation [7]. Another reason for data aggregation is the high sensitivity and potential privacy breaching. The aggregation practice in such cases can be based on different criteria. For instance, energy data of Los Angeles is available online after spatial aggregation based on the 15/15 rule. It requires that there be at least 15 non-residential customers within a certain location (or category) for any statistical summary of energy consumption to be provided online, and that no single customer within the group represents more than 15% of the total energy consumption of the group. A group including only residential customers must have a minimum of 100 customers [8]. These practices lead to masking many entries to ensure privacy but risk the usefulness of the data [9]. In the city of Västerås, Sweden, the authors had to comply to the GDPR regulations [10] while opening up urban energy data, under the project NRGYHUB[1]. Energy meters (electricity and district heating) were spatially aggregated in groups of at least 5 customers based on their geographical locations by the energy operator company in the city before even sharing them with researchers. In both research cases, the meters' addresses and boundaries make the aggregation difficult. In addition, in many cases, it depends on the person who performs the aggregation to decide to which group a meter belongs. This leads to different groups' sizes, but also to different possible combinations of groups.

The effects of spatial aggregation on energy systems have been explored. Elombo et al. [6] studied the variation of load profiles characteristics when interconnecting different sized groupings of customers (spatial aggregation) at different sampling resolutions (temporal aggregation). They inspected the After-diversity Maximum Demand ADMD which is defined as the simultaneous maximum demand within a group,

[1] www.nrgyhub.mdu.se (accessed: September 30, 2022)

divided by the number of customers per group. The ADMD was also evaluated for electricity and gas loads in district heating systems in UK [4] to assist the exploration of peak demand of individual and aggregated load profiles. Livingston et al. [11] applied K-means clustering to evaluate the similarity between the individual meters and their corresponding average building meter profile deduced from the aggregated group for non-residential buildings. They proposed a meter aggregation selection threshold to ensure tenant privacy. Sajjad et al. [12] carried out statistical tests in this regard. The results reveal the loss of individual customer's dynamics and its behavior influence on the aggregated load. In another study [13], the authors evaluate the aggregation impacts on the demand flexibility and proposed two probabilistic indicators to quantify the flexibility level of the aggregated profiles, which decreases with the increased level of aggregation. A statistical methodology framework was similarly conducted to investigate the variations of the peak demand factor in water networks and for different spatial aggregation levels [14].

In this paper, the impact of spatial aggregation at different aggregation levels on the characteristics of the electricity load in residential buildings is explored, using statistical metrics describing the time-series load data. In particular, the authors try to address the following questions:

1) How does the subjectivity of the person who performs the spatial aggregation affect the load characteristics?

2) How similar is a meter electricity profile to its group profile?

To answer these questions, cross correlation analysis between groups, diversity, peak distribution and trend difference between a meter and its group are analyzed.

The paper is structured as follows: Section 2 briefly describes the data and the analysis framework. Section 3 shows and discusses the results obtained. Section 4 concludes the discussion and suggest future applications.

2. MATERIAL AND METHODS

2.1 Description of input data

The data used in this study are individual electricity consumption profiles of 199 households, more precisely villas, in the city of Västerås. The data were obtained following a survey conducted in 2019, where participants gave consent to share their electricity data with Mälardalen University for research purpose. The data had different temporal span ranging from few months only to more than 3 years of hourly electricity consumption, due to the various customer needs to be upgraded to higher smart meter measurement frequency. The upgrade has historically been driven by enhanced regulation and/or customers need of data to use certain markets or contracts. Fig. 1 shows the meters data over a year. The meters with missing data were removed and 122 meters were left. A closer loop to a sample of the data is shown in Fig. 2. The difference in consumption between the meters and their daily patterns can be clearly observed. The electricity meters were divided into two categories: meters of customers connected (108 meters) or disconnected (14 meters) from the district heating (DH) system in the city, since this affects their electricity consumption based on their heating system.

To facilitate the analysis, one year of electricity recordings was considered for the analysis starting from September 2019 till August 2020.

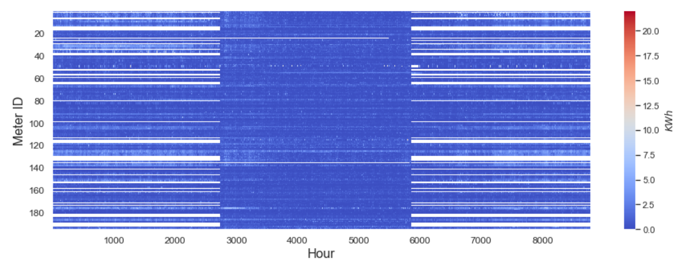


Fig. 1: Hourly electricity consumption of customers between September 2019 and August 2020. Some of the customers have shorter electricity records and they were excluded from this study.

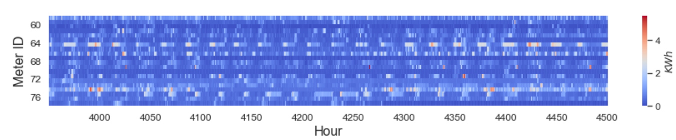


Fig. 2: A sample of the hourly electricity consumption of some of the meters depicting electricity patterns and variations among the customers.

2.2 Analysis framework

In the present study, the hourly electricity data is used to analyze the impact of aggregation level on the electricity profiles. The analysis is performed under progressive number of aggregation of meters. A Monte Carlo approach was adopted, which consists of multiple randomized selections fed as input to each metric computation. This is essential given the large possible combinations of meters at each aggregation level.

2.2.1 Similarity among groups

To evaluate impact of different possible grouping and the sensitivity of groups at different aggregation levels to each of their components (i.e individual meters), the similarity among the groups within the same aggregation level is computed. The aggregation level refers to the total number of meters included in the aggregation of the data. The aggregation levels range between 2 (i.e., two meters are aggregated) to 40. Correlation is used as a measure of two series similarity [15][16]. By considering only the meters of customers that are connected to DH, for each meters' grouping level, 50 different groups of meters were sampled for each meter (the same meter appears in all groups). After that, 50 combinations of two groups for each meter are randomly selected and the correlation coefficient R was estimated between their aggregated hourly electricity consumption. Then, the average correlation factor for each meter at each aggregation level was computed.

2.2.2 Electricity demand diversity

Individual households have different electricity demand patterns, and it is unlikely that all customers have their peak demands occurring simultaneously. This is known as electricity demand diversity [17]. As a result, the peak demand of the aggregated group will be less than the sum of the individual maximum demand of each household in the group. Diversity is defined in Eq. 1:

$$\text{Diversity factor} = \frac{\sum_{i=1}^{n} \text{Max}_i}{\text{Max}_{\text{group of n}}} \quad (1)$$

Where *i* is the i-th customer among *n* customers aggregated within a group.

The analysis of the electricity demand diversity is coupled with the hourly distribution of the peaks of individual customers, as well as groups of 5 and 40. The peaks are determined daily as the values above the 90^{th} quantile of the hourly profile of each meter or group of meters. For the two aggregation levels, 100 random combinations of customers were aggregated and used in this analysis.

2.2.3 Hourly trend Root Mean Square Deviation (RMSD)

To compare the trends in hourly electricity consumption between the individual customers and their respective groups at different aggregation levels, the Root Mean Square Deviation metric is calculated for each time step (one hour). For each aggregation level, 500 aggregations are executed. Then, for each aggregation, a random customer i is selected and its electricity profile is compared to the total of the group. The RMSD is defined in Eq. 2:

$$\text{RMSD}_k = \sqrt{\frac{\sum_{h=1}^{H-1}\left(\Delta E_{i,h} - \Delta E_{t,h}\right)^2}{H-1}} \quad (2)$$

Where k is the aggregation level, h the time range, i the individual customer, t the total of the group, ΔE the difference in electricity consumption for one hour time step and H the total number of hours.

3. RESULTS AND DICSUSSION

In this section, we present several metrics to demonstrate the impact of aggregation.
By computing the cross-correlation among groups at different aggregation levels, results in Fig. 3 show the lowest correlation at level 2 with a mean value of 0.2. The correlation factor increases with each level of aggregation and reaches 0.4 at the highest aggregation levels investigated in this study. The cross-correlation factor shows that as the number of meters per group increase, the similarity among the groups within the same level increases. In other words, the grouping becomes less dependent on which meter is being aggregated, and the groups become more correlated.

As expected, the diversity factor increases as function of the aggregation level (Fig. 4). By plotting the peaks distribution of the individual customers for weekdays, weekends, and holidays in Table 1, it is revealed that the maximum demands of individual customers occur at different times of the day, even at night, with most of the peaks occurring around 18:00 in all days, with an additional peak period around noon on weekends and holidays for customers connected to DH. By aggregating the DH customers, the peaks of the groups are more concentrated around 18 o'clock while the night peaks become less frequent (aggregation of level 5). For the aggregation of level 40, more the 76% of the groups' peaks occur around 18 o'clock, with an additional significant peak period in the morning on holidays. Peaks during nights and between the two peaks periods do not occur anymore.

The RMSD of the hourly variations between an individual customer and its aggregated group increases as function of the aggregation level as shown in Fig. 5. In other words, the hourly trends of the group are different from those of the individual customer, and the group as a unity depict different behaviors and electricity consumption patterns compared to its individual entities.

The peak distribution and RMSD values show that, for a customer, the larger the group it belongs to, the more diluted its individuality becomes. Profiling of the electricity data to classify customers is affected by the aggregation, especially that it becomes profiling of the groups, which have different sizes. In addition, any strategy that target the customers as part of their groups, such as demand response and flexibility margins, tariffs differentiation or energy modeling, should consider the groups' sizes as different impacts on the individual level are expected based on the group size the customer belongs to.

4. CONCLUSIONS

In this paper, the impact of spatial aggregation on the characteristics of the aggregated electricity profiles was studied. The analysis shows that with the increase of aggregation level, the particularities and impact of the individual customer's behaviors are lost. At low aggregation level, it is a matter of who while at higher levels it is a matter of how many. It was also shown that the results are affected by the multiple possible combinations of aggregation groups which can be subject to the person performing the spatial aggregation. The outcomes of this study can be useful for researchers and energy planners. It provides indication of the variation in electricity data characteristics caused by the aggregation. This highlights the importance of a trade-off between privacy and data usability, which can be achieved by developing a guidance framework. The latter would help properly performing energy studies, while not violating the GDPR regulations and not compromising the energy management strategies accuracies and the energy fairness among the customers.

ACKNOWLEDGMENT

We acknowledge support from the EU Regional Funds (Grant #20201324) through the project "World-class Energy Solutions" and the KK-foundation (Grant #20200073) through the project FLEXERGY.

REFERENCES

[1] M. H. Rehmani, M. Reisslein, A. Rachedi, M. Erol-Kantarci, and M. Radenkovic, "Integrating Renewable Energy Resources into the Smart Grid: Recent Developments in Information and Communication Technologies," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 7. pp. 2814–2825, 2018, doi: 10.1109/TII.2018.2819169.

[2] V. Marinakis *et al.*, "From big data to smart energy services: An application for intelligent energy management," *Futur. Gener. Comput. Syst.*, vol. 110, pp. 572–586, 2020, doi: 10.1016/j.future.2018.04.062.

[3] Y. Wang, Q. Chen, C. Kang, M. Zhang, K. Wang, and Y. Zhao, "Load profiling and its application to demand response: A review," *Tsinghua Science and Technology*, vol. 20, no. 2. pp. 117–129, 2015, doi: 10.1109/tst.2015.7085625.

[4] Z. Wang, J. Crawley, F. G. N. Li, and R. Lowe, "Sizing of district heating systems based on smart meter data: Quantifying the aggregated domestic energy demand and demand diversity in the UK," *Energy*, vol. 193, p. 116780, 2020, doi: 10.1016/j.energy.2019.116780.

[5] D. Griego, S. Schopfer, G. Henze, E. Fleisch, and V. Tiefenbeck, "Aggregation effects for microgrid communities at varying sizes and prosumer-consumer ratios," *Energy Procedia*, vol. 159, pp. 346–351, 2019, doi: 10.1016/j.egypro.2019.01.004.

[6] A. I. Elombo, T. Morstyn, D. Apostolopoulou, and M. D. McCulloch, "Residential load variability and diversity at different sampling time and aggregation scales," *2017 IEEE AFRICON Sci. Technol. Innov. Africa, AFRICON 2017*, pp. 1331–1336, 2017, doi: 10.1109/AFRCON.2017.8095675.

[7] Z. A. Khan, D. Jayaweera, and M. S. Alvarez-Alvarado, "A novel approach for load profiling in smart power grids using smart meter data," *Electr. Power Syst. Res.*, vol. 165, no. January, pp. 191–198, 2018, doi: 10.1016/j.epsr.2018.09.013.

[8] S. Pincetl, H. Gustafson, F. Federico, E. D. Fournier, R. Cudd, and E. Porse, *Energy Use in Cities*. 2020.

[9] B. L. Ruddell, D. Cheng, E. D. Fournier, S. Pincetl, C. Potter, and R. Rushforth, "Guidance on the usability-privacy tradeoff for utility customer data aggregation," *Util. Policy*, vol. 67, no. August, p. 101106, 2020, doi: 10.1016/j.jup.2020.101106.

[10] T. E. P. and the C. of the E. Union, "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation)," *Off. J. Eur. Union*, no. April, 2016, [Online]. Available: https://eur-lex.europa.eu/eli/reg/2016/679/oj.

[11] O. V. Livingston, T. C. Pulsipher, D. M. Anderson, A. Vlachokostas, and N. Wang, “An analysis of utility meter data aggregation and tenant privacy to support energy use disclosure in commercial buildings,” *Energy*, vol. 159, pp. 302–309, 2018, doi: 10.1016/j.energy.2018.06.133.

[12] I. A. Sajjad, G. Chicco, and R. Napoli, “Effect of aggregation level and sampling time on load variation profile - A statistical analysis,” *Proceedings of the Mediterranean Electrotechnical Conference - MELECON*. pp. 208–212, 2014, doi: 10.1109/MELCON.2014.6820533.

[13] I. A. Sajjad, G. Chicco, and R. Napoli, “Definitions of Demand Flexibility for Aggregate Residential Loads,” *IEEE Trans. Smart Grid*, vol. 7, no. 6, pp. 2633–2643, 2016, doi: 10.1109/TSG.2016.2522961.

[14] G. Del Giudice, C. Di Cristo, and R. Padulano, “Spatial aggregation effect on water demand peak factor,” *Water (Switzerland)*, vol. 12, no. 7, pp. 1–20, 2020, doi: 10.3390/w12072019.

[15] T. Hachino and V. Kadirkamanathan, “Time series forecasting using multiple Gaussian process prior model,” *Proc. 2007 IEEE Symp. Comput. Intell. Data Mining, CIDM 2007*, no. Cidm, pp. 604–609, 2007, doi: 10.1109/CIDM.2007.368931.

[16] A. Kianimajd *et al.*, “Comparison of different methods of measuring similarity in physiologic time series,” *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 11005–11010, 2017, doi: 10.1016/j.ifacol.2017.08.2479.

[17] G. Happle, J. A. Fonseca, and A. Schlueter, “Impacts of diversity in commercial building occupancy profiles on district energy demand and supply,” *Appl. Energy*, vol. 277, p. 115594, 2020, doi: 10.1016/j.apenergy.2020.115594.

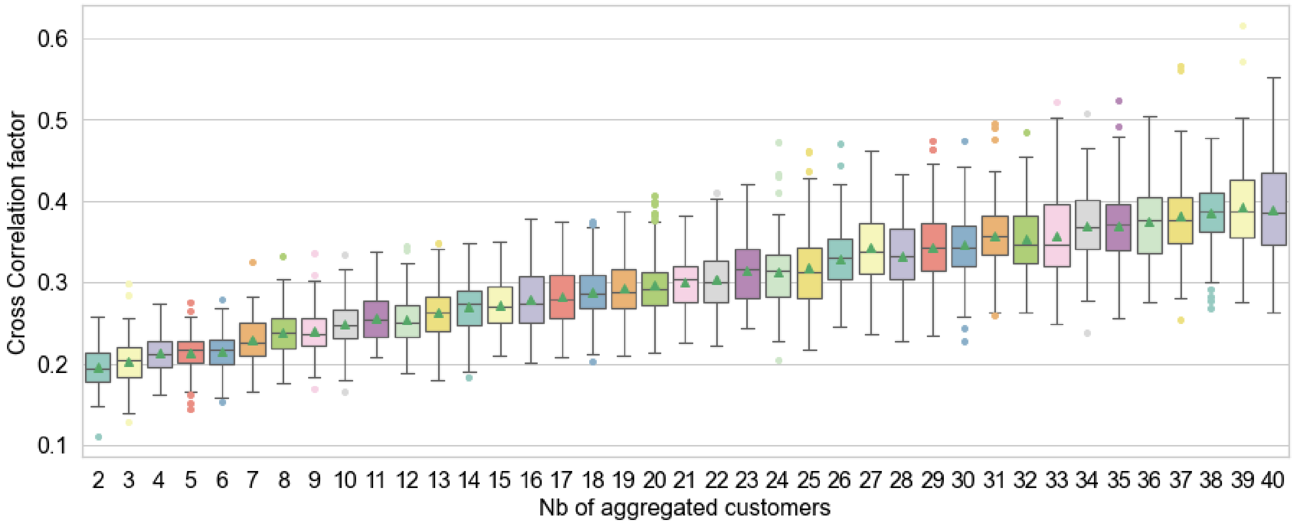


Fig. 3: Variations of the correlation coefficients (Pearson coefficient) as function of the aggregation level. The green shapes in each box plot represent the average correlation factor at each aggregation level.

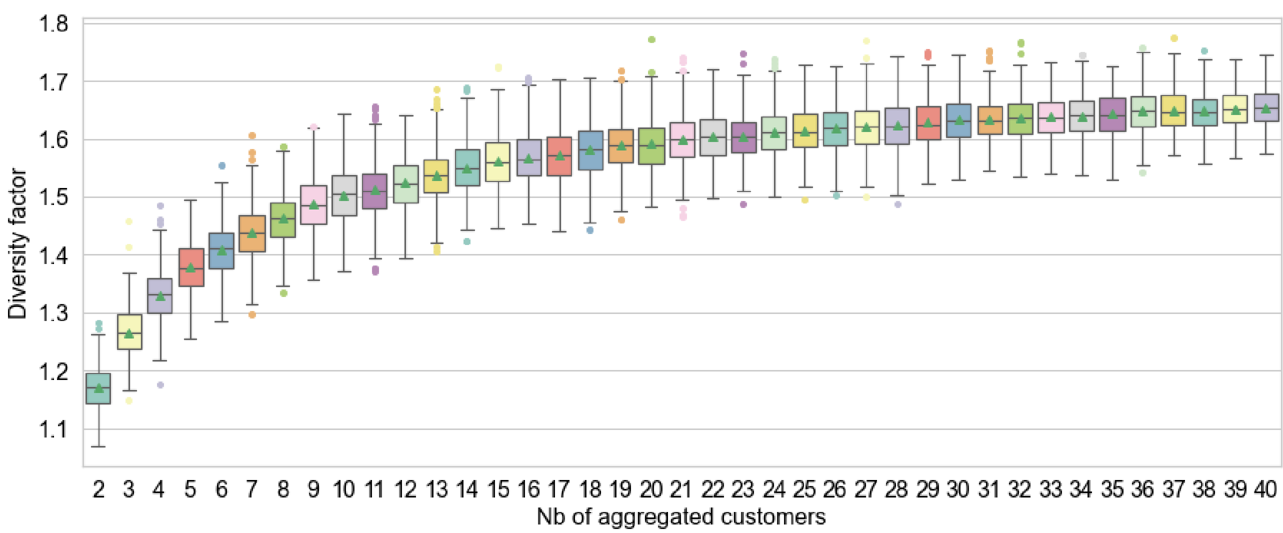
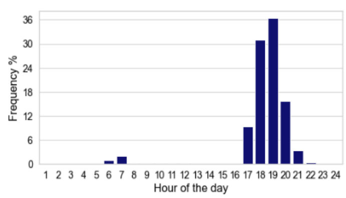
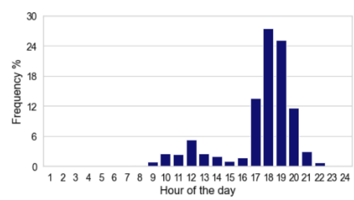


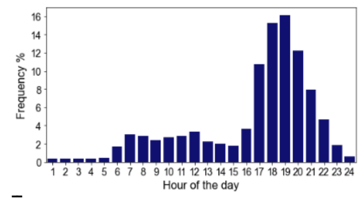
Fig. 4: Variations of the diversity factor of electricity loads as function of the aggregation level.

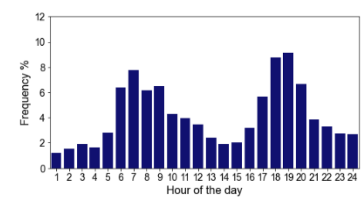
Total Customers	Customers groups of 40 (only with DH system)	Customers groups of 5 (only with DH system)	Customers not connected to DH	Customers connected to DH

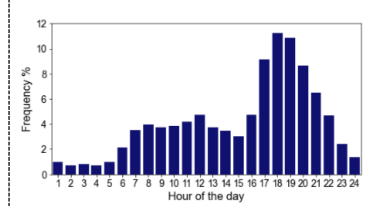
Weekdays



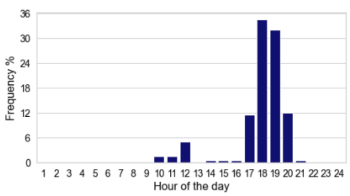


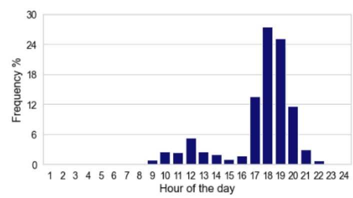


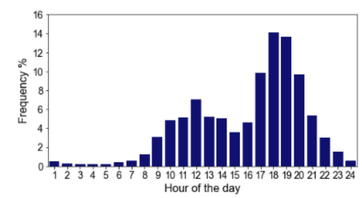


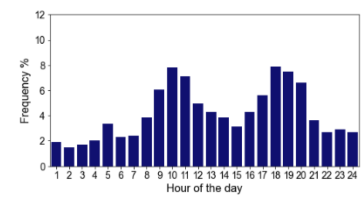


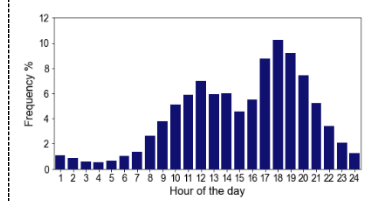
Weekends



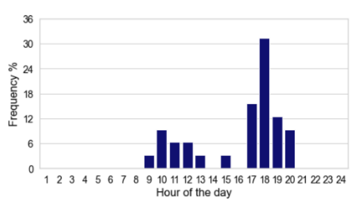


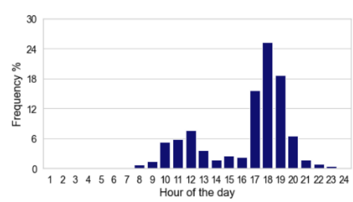


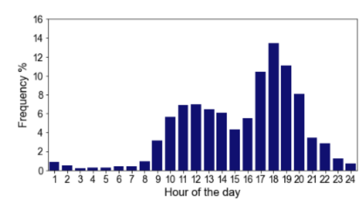


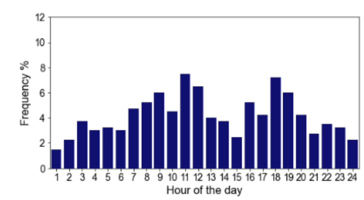


Holidays









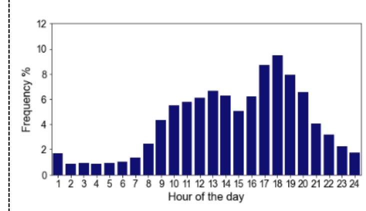


Table 1: Distribution of daily peaks of individual buildings connected (1st column) and not connected (2nd column) to district heating, and of groups of buildings (aggregation levels 5 and 40) (3rd and 4th columns), and of the total system of 108 customers in the final column. The peaks distributions are separated by weekdays, weekends and holidays.

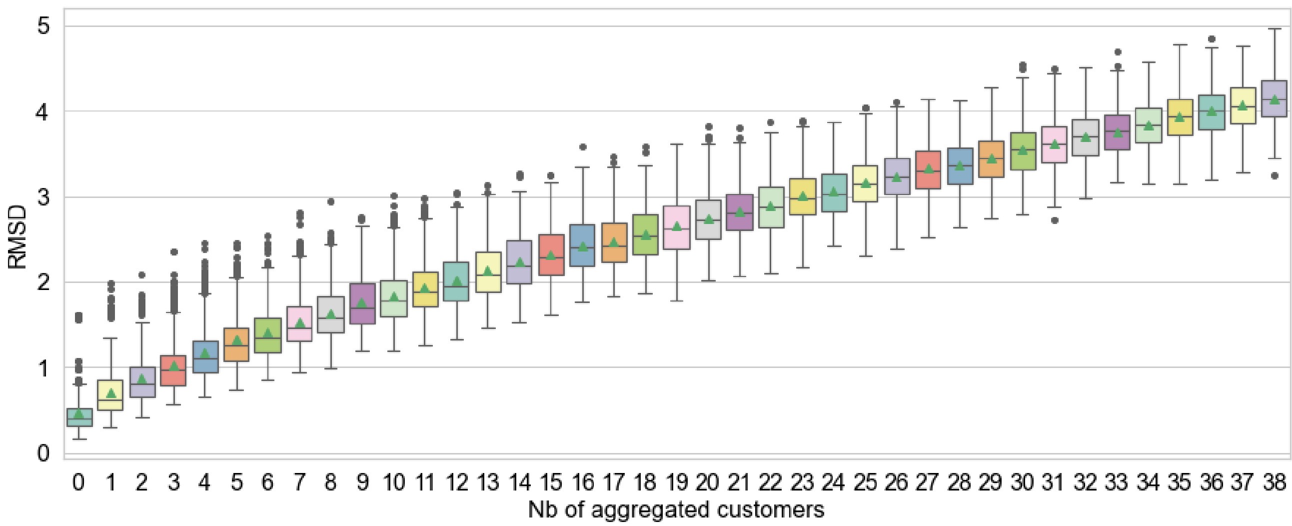


Fig. 5: Variation of RMSD as function of the aggregation level.