Building categorization revisited: A clustering-based approach to using smart meter data for building energy benchmarking

Article *in* Applied Energy · March 2020

DOI: 10.1016/j.apenergy.2020.114920

Building categorization revisited: a clustering-based approach to using smart meter data for building energy benchmarking

Sicheng Zhan[a], Zhaoru Liu[b], Adrian Chong[a,*], Da Yan[b,*]

[a]Department of Building, School of Design and Environment, National University of Singapore, 4 Architecture Drive, Singapore 117566, Singapore
[b]Department of Building Science, School of Architecture, Tsinghua University, Beijing 100084, China

Abstract

Current building energy benchmarking systems categorize buildings into peer groups by static characteristics such as climate zones and building types, which cannot account for the huge variation in building operations. Grouping buildings with diverse operations for benchmarking could result in misleading results. The smart meters provide an opportunity to feature the dynamic characteristics of building operations, but proper data mining techniques are needed to use the data for benchmarking. Accordingly, this paper proposes a framework that makes use of the time-series energy consumption data to categorize buildings by their operations and conduct energy benchmarking within each category.

The proposed framework is based on 3-step K-means clustering and consists of two main parts: 1) Operation quantification, and 2) Building categorization and benchmarking. The framework was tested on a dataset of 81 buildings in Singapore. Two baseline methods were also implemented for comparison. The results show that the proposed framework successfully categorized the buildings by their operational similarities and made a significant impact on the energy benchmarking results. Further, the superiority of operation-based energy benchmarking is manifested by investigating two typical buildings where the proposed framework disagreed with the baselines. It is necessary to integrate building operations in energy benchmarking so that the energy performance is evaluated more precisely and great energy saving potential can be achieved.

Keywords: Building energy benchmarking, Smart meter, Clustering, Building operation, Energy conservation,

1. Introduction

Buildings constitute around one-third of global greenhouse gas emissions and 30% of global energy consumption [1], making building energy conservation important in the campaign against global warming. Building energy benchmarking is referred to as the comparison of energy performance in buildings with similar characteristics [2]. It serves as an important component of most building rating systems and as a starting point to conduct energy audits

*Corresponding authors
Email addresses: adrian.chong@nus.edu.sg (Adrian Chong), yanda@tsinghua.edu.cn (Da Yan)

and analyze energy-saving scenarios. A typical statistical benchmarking approach is to compare the Energy Usage Intensity (EUI) of a building against the benchmark generated from its peer group.

1.1. Benchmarking systems

Most currently-used benchmarking systems define peer groups based on buildings' basic characteristics. For example, Energy Star, the most commonly-used program in the US, is based on a regression model built upon the CBECS (Commercial Buildings Energy Consumption Survey). It considers factors including location (climate zone), building type, floor area, etc. [3]. Green Mark, the benchmarking system applied in Singapore, generates EUI benchmarking quartiles considering building type and building size [4]. However, since there are many other factors affecting buildings' energy performance, the results of these single or low-dimensional benchmarking methods could be counter-intuitive. For example, an old school achieved a higher EnergyStar score as compared to a new school with many energy-efficient technologies because it deployed fewer facilities and therefore consumed less energy [5].

Towards a fairer comparison of building energy performance, studies were conducted to define the peer groups using more features. Using the data from CBECS, [6] showed that clustering buildings with more characteristics like height and age provided a more comprehensive benchmarking approach. With a database in South Korea, Park et al. demonstrated the effectiveness of using Decision Tree to account for factors like building use ratio for benchmarking [7]. One limitation of the databases used in these studies was the lack of detailed knowledge about the buildings. Consequently, these studies only showed the better prediction or categorization of EUI with the summarizing building features, instead of validating whether the proposed better evaluated the energy performance of a specific building. Besides, different operational patterns caused by various occupant behavior, as one of the major sources of uncertainty in building energy assessment [8], were not incorporated in the databases.

Considering that buildings are to provide a desirable environment for occupants, an ideal benchmarking system should be evaluating the buildings' energy performance when delivering similar amounts of services or following similar operational patterns [9]. Adopting improper peer groups for energy benchmarking may generate misleading results [10]. However, most buildings are mixed-use developments, and conventional approaches of categorizing them as building types (e.g. office or retail) could be spurious due to the inability to accurately describe the operational patterns [11]. Moreover, even buildings of the same use type could have a significant variation in operational patterns [12]. For example, a commercial building could be evaluated as not performing well with high EUI, not because of bad design or inefficient systems but just longer operating hours [13]. Additionally, a building could have a high base load and consume more energy because of a 24h-operated data center [14]. Thus, there is a need to account for the operational intensity and patterns in the energy benchmarking of buildings.

One approach to tackling this problem was to normalize the EUI with operating hours [15], which was adopted as an alternative of EUI in Singapore [16]. However, the diversified building functions and the various operational patterns were not reflected only in the length of operating hours but also features like base load, peak load, rising time, etc. Thus, while [17] showed the improved effectiveness of EUI normalized by operating hours, the variations in building operations were not fully characterized. Alternatively, another approach is to categorize buildings based on their operations and to conduct benchmarking within the peer groups. For example, if a building has a 24h-operated data center, it will be grouped with buildings with similarly high base loads. As the buildings in this peer group follow similar operational patterns, the bias caused by the heavier operations is addressed when benchmarking. Due to the complexity of building operations, more building operational data and proper data mining methods are required to realize operation-based categorization.

1.2. Operation quantification and building categorization

According to the smart metering surveys done in the US [18] and Europe [19], the deployment of smart meters was escalating during the past decades, making hourly total building electricity consumption easily available. Using data mining techniques, load profiles can be extracted from the smart meter data [20]. The load profiles represent building operations, which describe the variation of building usage intensity, including when (temporal trend) and by how much (load ratio) buildings consume energy. Since buildings are usually different from each other, and it is hard to obtain detailed information such as floor area composition and equipment deployment, unsupervised learning methods were recognized to be a promising approach of load profiling [21]. Many studies applied unsupervised algorithms such as K-means clustering [22], density-based clustering [23], hierarchical clustering [24], K-shape clustering [25], "follow the leader" clustering [26] and Self-Organizing Map [27] on single buildings. These methods captured the operational patterns of individual buildings by identifying representative daily profiles, where the variations in building operations are reflected in many differences, such as the number of profiles, the profile shapes, and the corresponding dates of each profile. Thus, aggregating the representative profiles by concatenating or averaging cannot precisely describe the operations and will lead to different lengths of profiles. On the other hand, the form of features needs to be identical for different buildings to quantify the in-between operational similarity. Therefore, profiles extracted from individual buildings are not readily useful for operation-based building categorization.

To quantitatively categorize the buildings based on their operations, all buildings need to be parameterized in the same way. Several studies applied extra steps to interpret the profiles from individual buildings. Li et al. identified nine typical profiles of all the 40 buildings through a 2-step clustering and accordingly identified abnormal consumption patterns. The study showed the advantage of applying an extra step of clustering by comparing it against single-step methods [28]. Tsekouras et al. featured buildings for clustering with one representative profile per building that was manually selected from the typical profiles identified in the previous steps [29]. This approach standardized the buildings by discarding less important profiles and was time-consuming due to its manual selection process. McLoughlin et al. first clustered the customers based on their daily profiles for each day in the half-year and then labeled the day for each customer with the corresponding cluster number. The customers' operation mode was characterized as the cluster number that took up the most days [30]. However, this approach disregarded the operational patterns on minor days and assumed that the overall operation of all customers as a whole was consistent on different days. Park et al. combined the daily profiles from hundreds of buildings and clustered them into three "fundamental" types. The buildings were then categorized into four groups based on the dominance by the three "fundamental" profiles [11]. However, a concern is the loss of detailed segregation of operating patterns if only three typical profile types were used for the categorization of buildings.

In summary, the studies simplified the buildings' operations to achieve standard featurization and subsequent building categorization, mainly by limiting the number of representative profiles. However, using only one or a few typical profiles might not be sufficient for the characterization of a building's operational patterns [31]. Different types of buildings usually have their basic operational patterns, and it is the operating characteristics at a higher resolution that enlarge the operational variations between buildings of the same use type. The information loss during simplification made the categorization results not very different from that based on building usage types. Hence, more effort is required to realize a finer operation-based building categorization.

1.3. Aim and objectives

This study aims to build an operation-based building categorization system and to use that for building energy benchmarking. It contributes to the literature by achieving the following three objectives:

- Design a framework to characterize the buildings' operations using the hourly electricity consumption data, and to categorize the buildings for energy benchmarking according to their operational similarity;
- Evaluate the effectiveness of this proposed method by testing it on a dataset with many buildings and comparing with baseline methods;
- Demonstrate the necessity of operation-based benchmarking by selecting typical buildings as case studies and conducting in-detail analyses on their energy performance.

Accordingly, this paper is structured as follows. Section 2 illustrates the workflow and mechanism of the proposed framework, as well as the methods used for evaluation and validation. Section 3 described the dataset of 81 buildings in Singapore used for testing. Section 4 displays the results of applying the framework on the dataset. Section 5 presents the comparison results of the baseline methods and the detailed analyses of two typical buildings. Section 6 discusses the benefit introduced by the proposed method and potential directions for further development. Finally, section 7 summarizes the contribution and significance of this paper.

2. Methodology

2.1. The benchmarking framework

As shown in figure 1, the proposed building benchmarking framework consists of two main parts. Part I Operation quantification identifies the typical load profiles of all buildings through a 2-step K-means and quantifies the buildings' operational characteristics as the proportion of days subject to each profile. Part II Building categorization and benchmarking categorizes the buildings based on their similarity in operation using K-means and accordingly benchmarks their energy performance as EUI ranking percentiles.

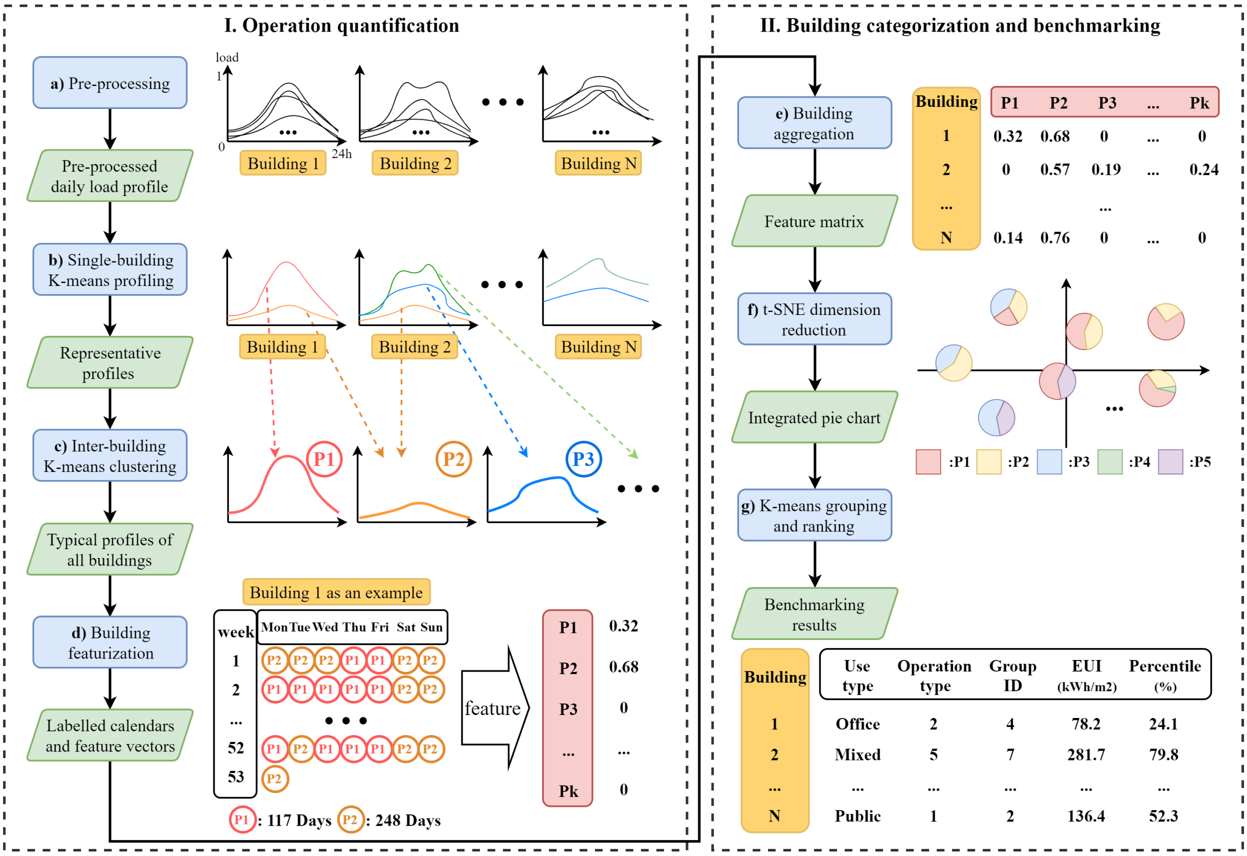

Figure 1: Schematic of the operation-based benchmarking framework.

2.1.1. *Operation quantification*

This part of the framework includes four steps: a) pre-processing, b) single-building K-means profiling, c) inter-building K-means clustering, and d) building featurization.

Pre-processing The raw accumulative electricity consumption data is first converted to hourly electricity consumption by incremental subtracting. Subsequently, the data is filtered, normalized (equation 1), and reshaped into daily load profiles for each day in the year. Considering that a load shift on a single hour can make a profile different, days with any missing value are filtered off.

$$\widehat{x}_{normalized} = \frac{x_{raw}}{x_{max}} \tag{1}$$

Maximum normalization [32, 33] is applied instead of [0, 1] normalization [34] or z-score standardization[35]. This maintains information about the base load, which is an important feature for the characterization of buildings' operational patterns. In equation 1, $\hat{x}_{\mathrm{normalized}}$ is the normalized hourly load ratio, x_{raw} is the raw hourly consumption and x_{max} is the maximum value in the year instead of the day [1].

[1]The impact of different normalization is further discussed in section 6.1

After pre-processing, daily electricity load profiles over one year are obtained for each building (figure 1.a). The following steps automatically quantify the operational characteristics for all the buildings so that 1) their variant operational patterns on different days are accounted for, and 2) they are featured in the same way to enable further categorization. K-means clustering is applied twice for both single-building and inter-building profile identification because it was found to be efficient and reliable in profiling [36].

Single-building K-means profiling K-means clustering is applied to cluster the daily load profiles. The number of clusters K is to be selected for every single building, so the tuning process is automated to improve the scalability. Tested among 2 and 7 using the Calinski Harabaz score [37], the K with the highest score is adopted. The centroids of each cluster are extracted as the representative profiles of the source buildings (figure 1.b). Compared with directly using the raw profiles to represent a building's operation, an extra step of clustering eliminates the variance caused by buildings' irregular stochastic behavior, which facilitates the following step.

Inter-building K-means clustering All the representative profiles are aggregated together to run inter-building clustering and identify the typical profiles of all buildings in the database (figure 1.c). With hundreds or thousands of profiles, the feature space will be crowded, and the distance between adjacent samples will be small, where the Clustering Validation Indices becomes less effective in evaluating the quality of clusters [38]. Since the number of clusters is only determined once for the whole dataset, it is not necessary to automate the procedure. Therefore, K is selected among 2 and 12 by applying the elbow method on the average within-cluster distance, which reflects the variance of operational patterns within the same cluster.

Building featurization Through typical profiles identification, all the representative profiles from individual buildings are assigned a cluster number. Subsequently, the cluster number is used to label the profiles and dates corresponding to the representative profiles for the source building. For the exemplary building 1 in figure 1.d, the two representative profiles are respectively clustered into the 1^{st} and 2^{nd} cluster during typical profile identification, so the dates subject to these two profiles are respectively labeled with 1 and 2. Finally, the percentage of labels in the year is used to feature the buildings and quantify their operation. In the example, assuming there are in total six typical profiles and the two profiles respectively take up 32% and 68% of the dates, building 1 is featured as a vector [0.32, 0.68, 0, 0, 0, 0].

2.1.2. Building categorization and benchmarking

This part of the framework includes two main steps: f) t-SNE (T-Distributed Stochastic Neighbor Embedding) dimension reduction, and g) K-means grouping and ranking.

t-SNE dimension reduction Aggregating vectors of all the buildings to form a feature matrix (figure 1.e), the operation dissimilarity between buildings is quantified as the Euclidean distance between the vectors. Through dimension reduction, the operation quantification results are visualized in an integrated pie chart (figure 1.f). Segments of the pie charts represent the proportion of typical profiles over one year, and the distance between pie charts represents the operational dissimilarity between buildings. t-SNE is applied to reduce the feature dimension from the number of typical profiles to 2 for visualization. Compared with traditional methods like Principal Component Analysis, t-SNE is better at preserving the information of pairwise distances [39].

K-means grouping and ranking From the pie chart, the number of pie clusters, which implies the number of typical operation type, is discovered. Accordingly, K-means clustering is applied so that buildings with different operations are distinguished from each other. Compared with manually defining the categories by observing the pie chart, the clustering more accurately decides the affiliation of buildings around the cluster borders. With the operation-based categorization results, buildings with the same usage type and operation type are grouped together for benchmarking (figure 1.g). The buildings are then ranked in their peer group by ascending EUI. The EUI is obtained by summing up to the annual total consumption and then dividing it by the total floor area. The buildings' EUI percentile evaluates their energy performance. For example, a building benchmarked as 70% indicates that it has higher EUI and worse energy performance than 70% buildings with the same usage type and similar operations.

2.2. Comparison and validation

2.2.1. Effectiveness of the categorization

We applied the proposed framework on an actual dataset and demonstrated the effectiveness by comparing against baseline methods from two perspectives: 1) the proposed framework effectively categorized the buildings based on their operation, and 2) the new categorization had an impact on the energy benchmarking results.

Two baseline methods were implemented for comparison. Baseline 1 grouped buildings only according to their usage types [4]. Baseline 2 categorized buildings according to their operations using the commonly used methods in the literature. This method differed from the proposed framework in three aspects: daily standardization was applied for pre-processing (used in [35], [40], and [41]), 1-step K-means clustering was used for typical profile identification (used in [22], [42], and [43]), and buildings were categorized according to the dominant profiles (used in [11], [29], and [30]). After defining the peer groups, both baselines conducted benchmarking by comparing EUI within the groups.

To demonstrate that the proposed framework achieved more accurate operation-based building categorization, the similarity of operations between buildings was visualized in two ways. First, the buildings were scattered in a two-dimensional plot using coordinates generated in step f of the proposed framework (left of figure 2). By coloring the points according to different categorization results, the operation similarity within each category was evaluated. Supplementarily, the pairwise Euclidean distances between different buildings' operational patterns were calculated as equation 2 to form a distance matrix and displayed as a heat map. As shown in the right of figure 2, blocks with lighter color corresponded to smaller distances and more similar operational patterns. By reordering buildings according to their categorization, the operational variations within each category were reflected in the color distribution.

$$D_{bldg1,bldg2} = \sqrt{\sum_{t=1}^{8760} \left(\widehat{x}_{bldg1,t} - \widehat{x}_{bldg2,t}\right)^2} \tag{2}$$

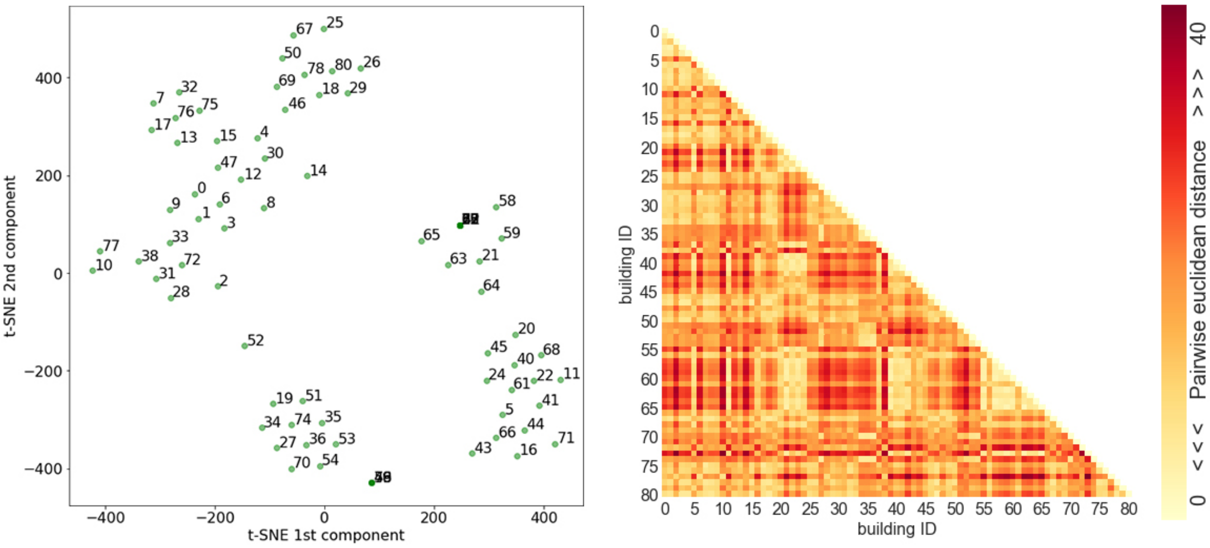

Figure 2: Left: operation-based scatter plot of the buildings using t-SNE; right: heat map of pairwise Euclidean distances between buildings.

Following the same ranking and labeling approach, each building was assigned three percentiles with respect to the three categorization methods. Considering that the population is non-Gaussian and paired, Wilcoxon signed-rank test [44] was applied between every two methods to see if the proposed categorization made a significant impact on the energy benchmarking.

2.2.2. Necessity of operation-based benchmarking

Further, we showed that the proposed benchmarking framework did help estimate buildings' energy performance by selecting case study buildings and conducting detailed analyses. The benchmarking results of the proposed and baseline methods were percentiles for each building. Lower percentile meant better energy performance. The difference between percentiles quantified how much difference the proposed framework made. The proposed framework giving lower percentiles meant the baselines underestimated the energy performance of the building. Vice versa, increased percentiles indicated that the buildings were supposed to consume less considering their operations, and the energy-saving potential was only revealed by the proposed. Two typical buildings with increased and decreased EUI percentiles were identified to show the superiority of the proposed. The two buildings were demonstrated respectively using the Measurement & Verification results and an energy model calibration approach.

3. Dataset description

The dataset consists of hourly total building electricity consumption for 105 buildings located on the campus of the National University of Singapore for the year 2018. After removing datasets with data quality issues, 81 buildings remained for further analysis. Of the buildings with data quality problems (23%), 14% were datasets with more than

10% missing data, 2% had frequent meter reset, and the remaining 7% had missing metadata. EUI was calculated using the respective total floor area of each building. Measurement & Verification (M&V) submission and data from the Building Management System were used to conduct detailed analyses on the selected case study buildings.

Among the 81 buildings, 71 are mixed-use, and the remaining 10 are residential. The frequency distribution of the buildings in terms of the number of valid days, gross floor area, and EUI (kWh/m^2) are shown in figure 3. The EUI quartiles are summarized in table 1. Next, the results of the proposed framework and baselines are presented.

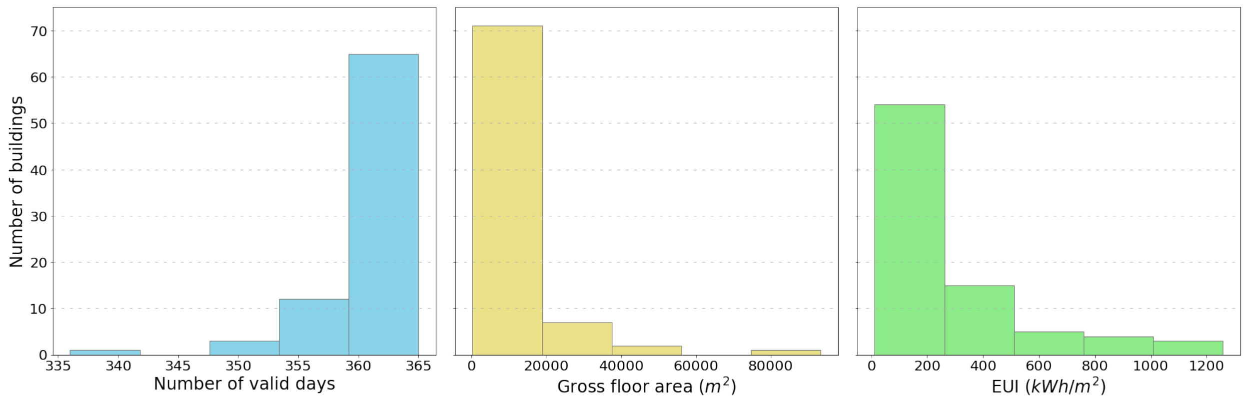

Figure 3: Summarized basic information of buildings in the dataset.

Table 1: EUI quartiles of residential and mixed-use buildings.

Group	**No. of buildings**	**Top quartile (1% - 25%)**	**2nd quartile (26% - 50%)**	**3rd quartile (51% - 75%)**	**Bottom quartile (76% - 100%)**
Mixed-use	71	≤ 99	99 - 172	172 - 409	> 409
Residential	10	≤ 56	56 - 68	68 - 108	> 108

4. Results of the proposed framework

4.1. Building operation quantification

Figure 4 shows the **single-building K-means profiling** results of six exemplary buildings (four mixed-used and two residential). Centroids of the k clusters are extracted as the representative profiles of each building, highlighted in bold lines, and labeled with C1 to Ck. Under most circumstances, K-means effectively separated the representative profiles. The operations of different mixed-used buildings differentiated from each other. For example, the first three buildings always consumed more energy during the working hours, while building #4 sometimes reached its peak in the evening. Building #2 strictly followed schedules as the others had larger variances during the operating hours. Meanwhile, residential buildings had relatively similar operational patterns as decreasing during the daytime and increasing in the evening

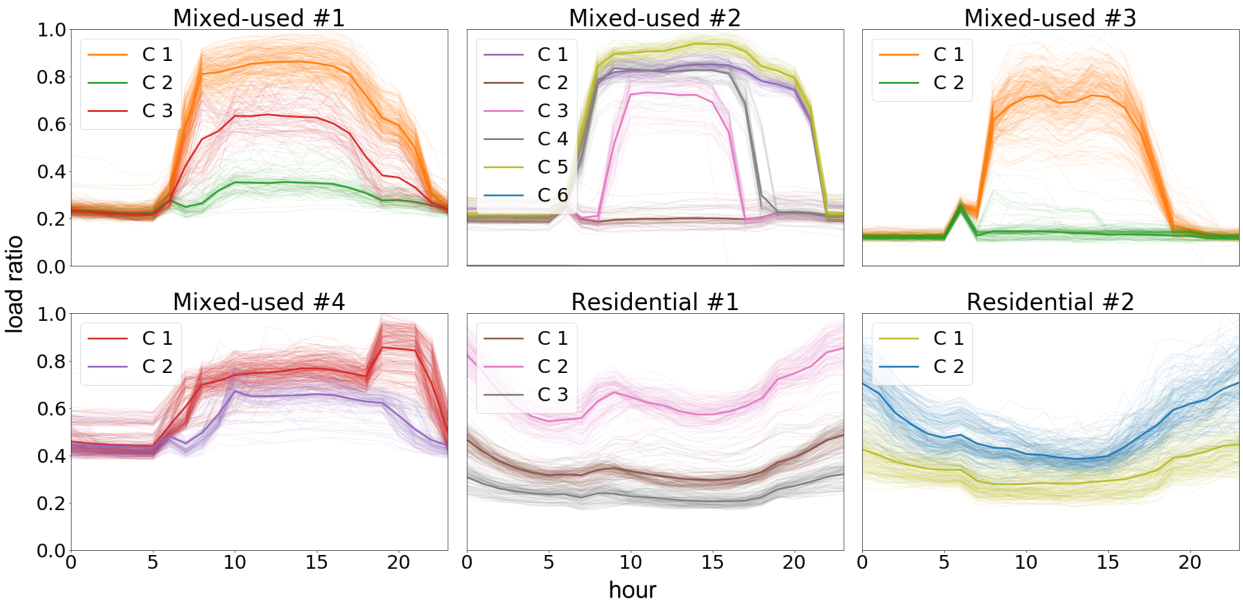

Figure 4: Examples of single-building K-means profiling results (legends are only effective within each subplot).

Out of the 81 buildings, 75 buildings were represented with two to four profiles, five of them required six representative profiles, and one building identified seven. In total, 221 representative profiles were extracted from all buildings, which were aggregated and plotted on the left of figure 5. Compared with the plot of 29257 raw profiles on the right of figure 5, the density of profiles was considerably reduced so that different operational patterns can be separated more easily.

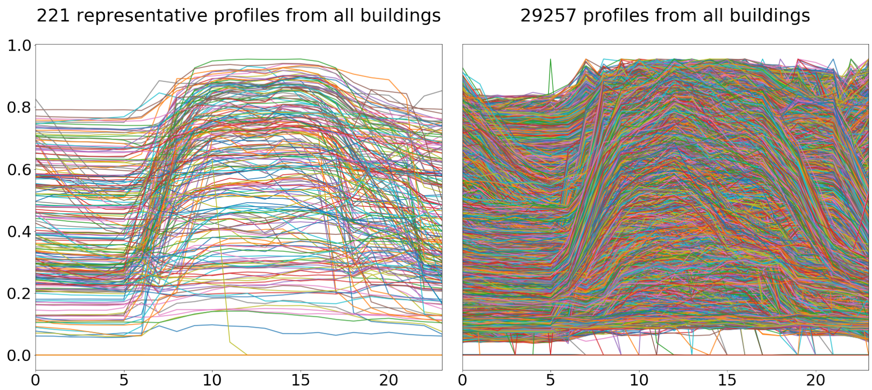

Figure 5: Plots of the aggregated representative profiles (left) and raw profiles (right).

For **Inter-building K-means clustering**, the trend of average within-cluster distance when varying the number of clusters K is plotted in figure 6. The average within-cluster distance quickly decreased as the number of cluster K increased from 2 to 6. When K continued to exceed 6, the algorithm started identifying similar clusters, and the decrease of average distance slowed down. Using the elbow method, K was selected as 6. Figure 7 displays the six typical profiles identified for all the buildings. Light lines are the representative profiles in each cluster, and the centroids (thick and bold lines) are the typical profiles identified.

The first three typical profiles were relatively flat, indicating the buildings were operated around a stable intensity over the day. Profiles from the residential buildings also fell into typical profile 2 and 3 because the increase of usage in the evenings was relatively small. Meanwhile, the last three typical profiles covered most working-in-daytime profiles, where the buildings were partially turned off before and after the working hours. Most working periods started at around 7 to 8 am and gradually ended after 6 pm. The difference between these three mainly laid in load ratios (peak load and base load). Typical profile 4 and 5 respectively had low and high load ratios, while typical profile 6 had a high peak load and a low base load. Note that the typical profiles represent general operational profiles of all buildings. In contrast, the small operational variations between profiles in the same cluster were neglected when categorizing buildings in the next step.

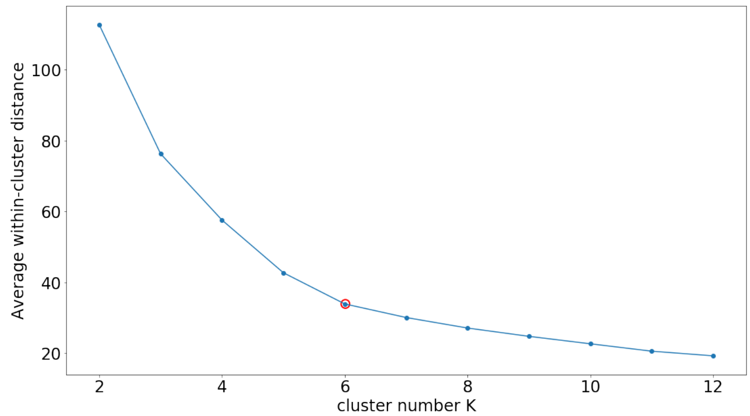

Figure 6: Trend of the average within-cluster distance as K varying between 2 and 12.

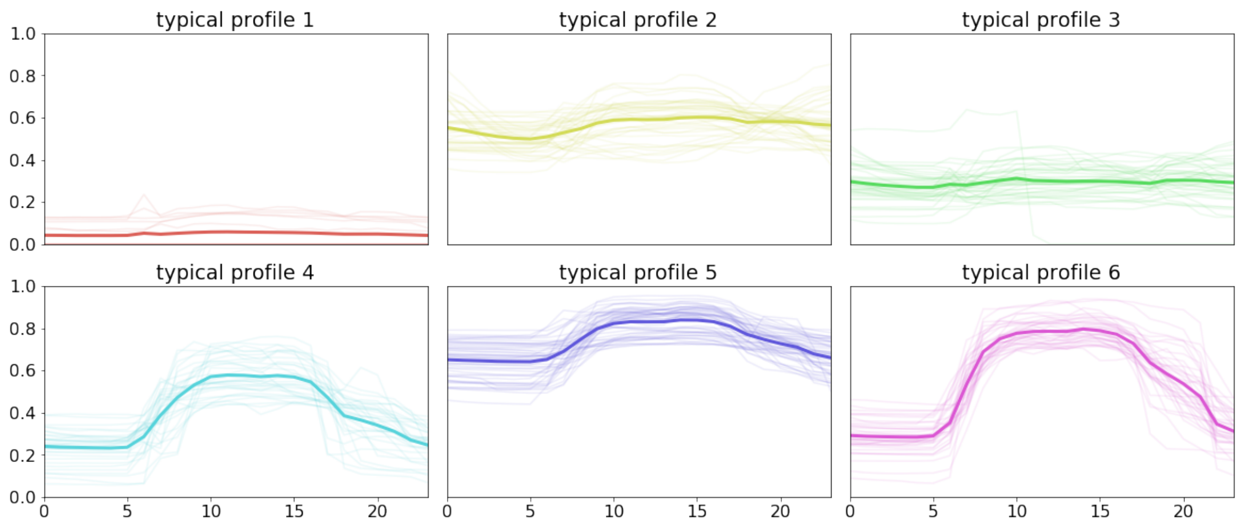

Figure 7: 6 clusters of representative profiles, centroids (thick lines) as the typical profiles.

Figure 8 displays three examples of **building featurization**. The calendar color maps visualize the results of labeling each day with the cluster number of the corresponding typical profile. Accordingly, the operations of each building were featured as six percentages of days that each typical profile took place in the year (titles of each calendar map in figure 8). For illustration, the labeled profiles of these buildings are plotted in figure 9.

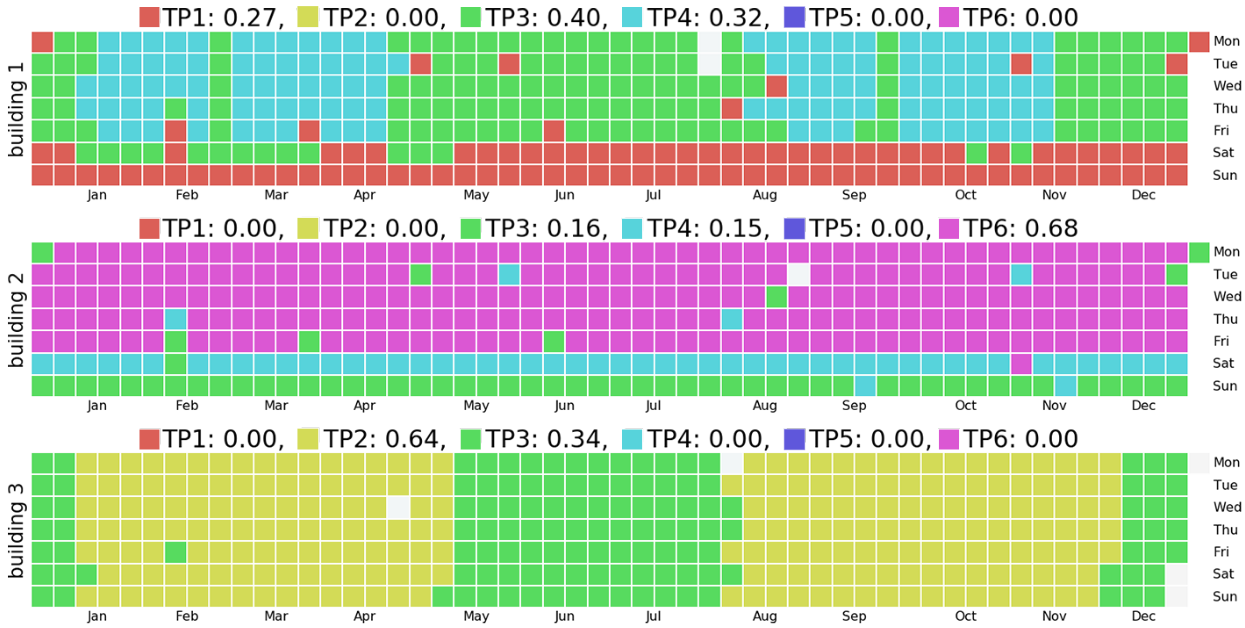

Figure 8: Examples of typical profile (TP) distribution in the year as colored calendar maps. Titles are the 6 percentages of days that feature the operation of each building.

Building 1 is a typical institutional mixed-use building, where the working days during the semesters (32%) followed typical profile 4 as a regular schedule and was partially on as typical profile 3 during the vacations (40%). On weekends and public holidays (27%), building 1 was almost off, holding a low base load as typical profile 1. Therefore, it was featured as [0.27, 0, 0.4, 0.32, 0, 0]. Building 2, also a mixed-use building, was operated more intensely. It followed typical profile 6 on almost all working days (68%), partially ran on Saturdays (15%), and rested on Sundays (16%). Both the base load during the off hours and the peak load during working hours were higher than building 1. The resulting features were [0, 0, 0.16, 0.15, 0, 0.68]. Buildings 3, as a dormitory, conducted a different operation from the mixed-use buildings. During semesters (64%), the load was higher beyond the working hours, subject to typical profile 3. Meanwhile, the flatter typical profile 2 was committed during vacations (34%), leading to the features of [0, 0.64, 0.34, 0, 0, 0]

4.2. Operation-based categorization and benchmarking

With the operation of all 81 buildings featured as 6-dimensional vectors, the integrated pie chart was generated by getting the buildings' coordinates through **t-SNE dimension reduction**. In addition to the buildings' operational pattern and in-between similarities, EUI is also combined and visualized as the size of pies (figure 10).

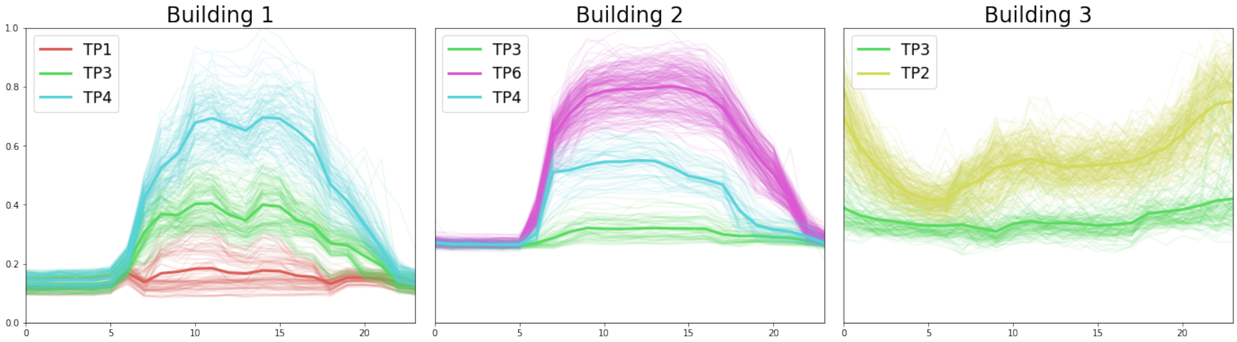

Figure 9: All profiles of the three exemplary buildings, labeled with typical profile (TP) numbers.

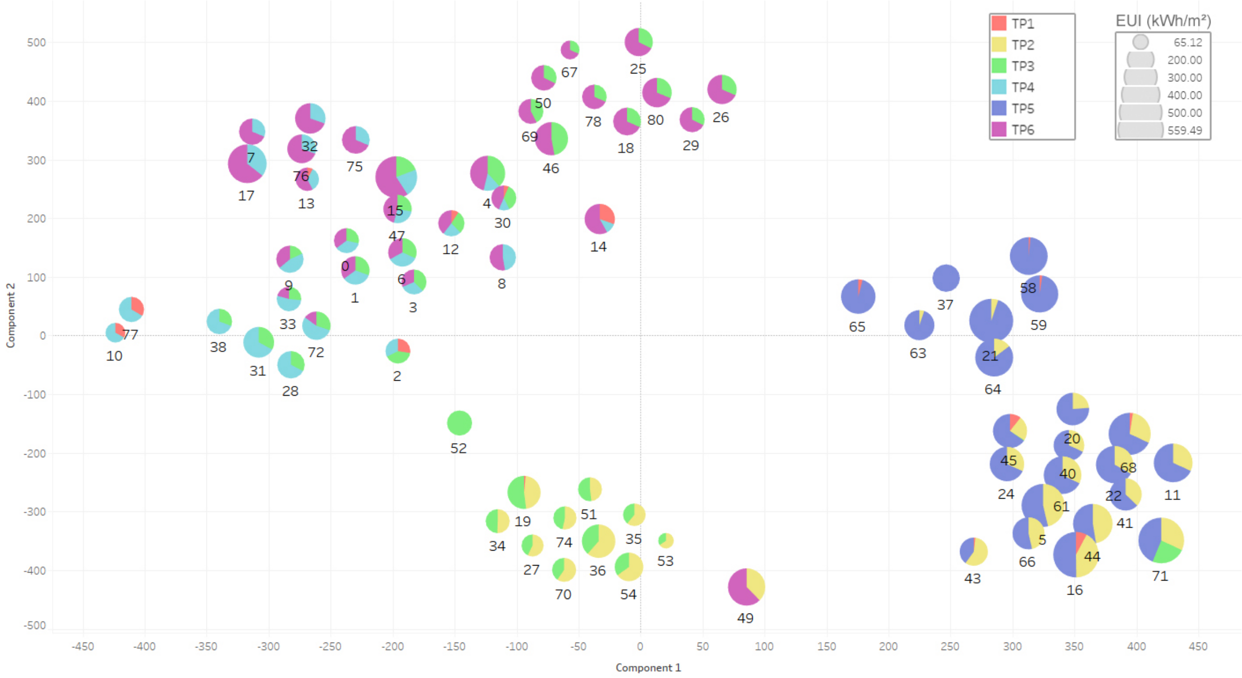

Figure 10: Integrated pie chart that visualize the buildings' typical profile ratios (pie segmentation), operational similarity (pie locations), and EUI (pie sizes).

Figure 10 shows approximately seven types of operation combinations among the 81 buildings, and most are dominated by 2 to 3 typical profiles. Accordingly, the number of clusters K was set as seven, and the **K-means grouping results** are shown in figure 11, where the building categories are highlighted by color. Both the scatter plot (close buildings clustered together) and the heat map (lighter color in the squares) shows that the buildings were well separated according to their operational characteristics. All residential buildings remained in the 6th category as typical profile 2 during semesters and 3 during vacations, while the mixed-use buildings were spread over the rest six categories.

Combining the operation-based categorization with the building use type, the framework defined seven **peer**

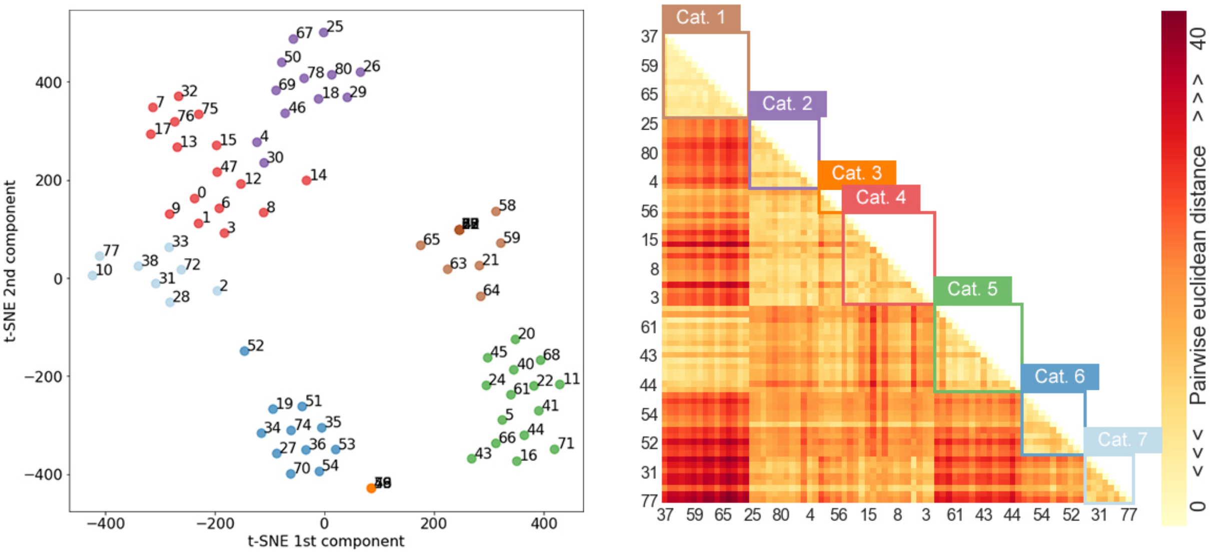

Figure 11: Operation-based scatter plot (left) and distance heat map(right) with the proposed building categorization.

groups for benchmarking. The detailed EUI benchmarks are summarized in table 2. Restricted by the number of buildings, one mixed-use building in the 6th operation category was excluded at this step. The 3rd and the 7th peer group could not generate valid benchmarks for the same reason. All the residential buildings had similar operations and relatively low EUI, while both the operation and EUI varied a lot for mixed-use buildings. Some peer groups such as the 1st and the 5th generally had higher EUI due to heavier operation (typical profile 5), whereas groups like the 2nd consumed less.

Table 2: Summary of the benchmarking peer groups and corresponding EUI quartiles.

Group No.	Building type	Operation category	No. of buildings	Top quartile (1% - 25%)	2nd quartile (26% - 50%)	3rd quartile (51% - 75%)	Bottom quartile (76% - 100%)
1	Mixed	1	15	≤ 360	360 - 526	526 - 927	> 927
2	Mixed	2	12	≤ 80	80 - 123	123 - 152	> 152
3	Mixed	3	4 [a]	≤ 112	112 - 369	369 - 389	> 389
4	Mixed	4	16	≤ 98	98 - 130	130 - 170	> 170
5	Mixed	5	15	≤ 226	226 - 399	399 - 650	> 650
6	Residential	6	10 [b]	≤ 56	56 - 68	68 - 108	> 108
7	Mixed	7	8 [a]	≤ 79	79 - 86	86 - 130	> 130

[a] Not enough buildings in the peer group to conduct effective benchmarking.

[b] The only mixed-use building in this operation category was excluded for benchmarking.

5. Evaluation and validation

5.1. Comparison with baselines

5.1.1. Results of the baselines

The benchmarking results of baseline 1, which categorized building by usage type for benchmarking, are summarized in table 1. As for baseline 2, the typical profiles identified are plotted in figure 12. Though the number of clusters was set as 3 for K-means empirically following the literature, it is worth noting that the Calinski Harabsz scores kept decreasing as K increase (left). The resulting typical profiles include the 1st valley-like profile representing the residential buildings and the other 2 representing mixed-used buildings with different working hours. The second daytime dominated profile (P2) is similar to P3 but shifted by one hour.

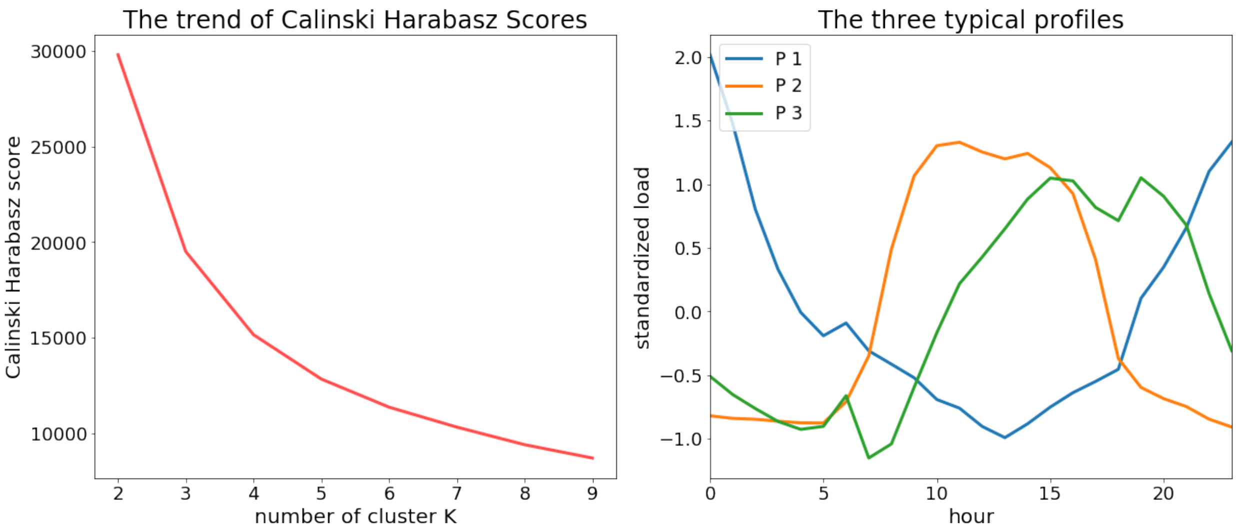

Figure 12: Typical profile identification results of baseline 2.

Using the typical profiles, the buildings were grouped disregarding the use type. As shown in figure 13, the first three groups are dominated by the corresponding profiles (> 50%), and the 4th group contains buildings with no dominant profile. The corresponding benchmarking results are summarized in table 3.

Table 3: Summary of baseline 2 benchmarking results.

Group No.	Operation group	No. of buildings	Top quartile (1% - 25%)	2nd quartile (26% - 50%)	3rd quartile (51% - 75%)	Bottom quartile (76% - 100%)
1	1	10	≤ 56	56 - 68	68 - 108	> 108
2	2	23	≤ 159	159 - 278	278 - 485	> 485
3	3	42	≤ 84	84 - 131	131 - 337	> 337
4	4	6 [a]	≤ 88	88 - 113	113 - 332	> 332

[a] Not enough buildings in the peer group to conduct effective benchmarking.

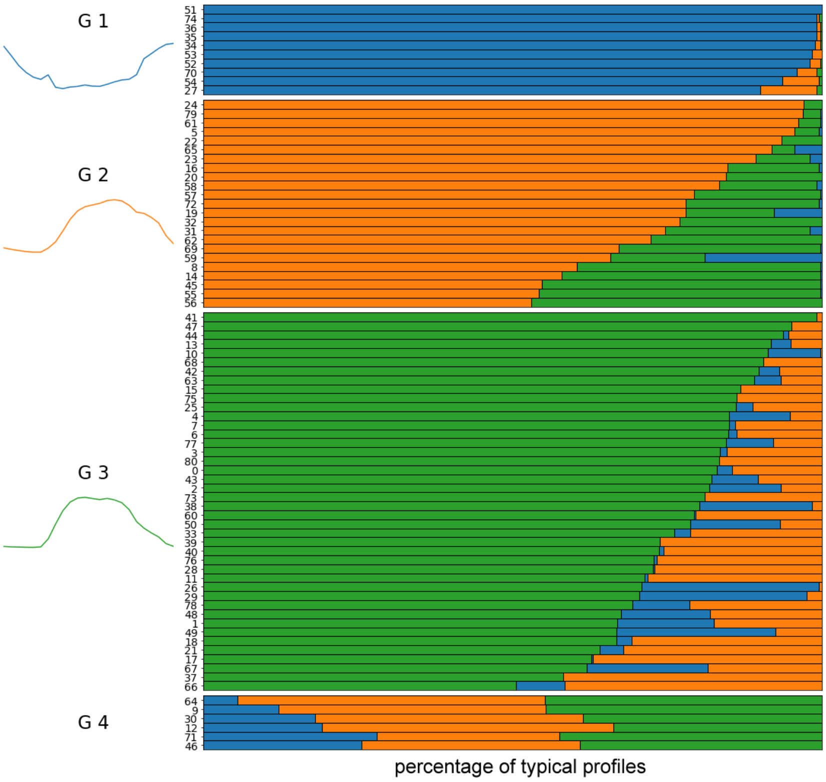

Figure 13: 4 groups of building categorized by baseline 2 (ratio of typical profiles as bar segmentation).

5.1.2. *Effective operation-based building categorization*

The building categorization results of the baseline methods are visualized respectively in figure 14 and figure 15. The residential buildings, where the operations were similar and different from other buildings, are distinguished by all three categorization methods (category 6 in the proposed, residential in baseline 1, and group 1 in baseline 2). Meanwhile, as shown in figure 14, the operational variations in mixed-use buildings are much more substantial. This indicates that baseline 1, which grouped buildings by use type, hardly accounted for the operational variations. According to figure 15, buildings in the three groups are almost randomly spread in the scatter plot, and the pairwise distances of each group are still disordered despite the small light yellow areas in group 2 and 3. Thus, while baseline 2 also categorized buildings based on their operational similarities, a finer resolution was achieved by the proposed method.

5.1.3. *Significant impact on benchmarking results*

Buildings in the too-small groups (3rd and 7th of the proposed, and 4th of baseline 2) were excluded when conducting the Wilcoxon test. The P-value between the proposed method and baseline 1 was 0.0265, and that between the proposed and baseline 2 was 0.0849. According to the 95% confidence level, this indicated the significant change made by the proposed method compared with baseline 1 and almost significant compared with baseline 2. By contrast,

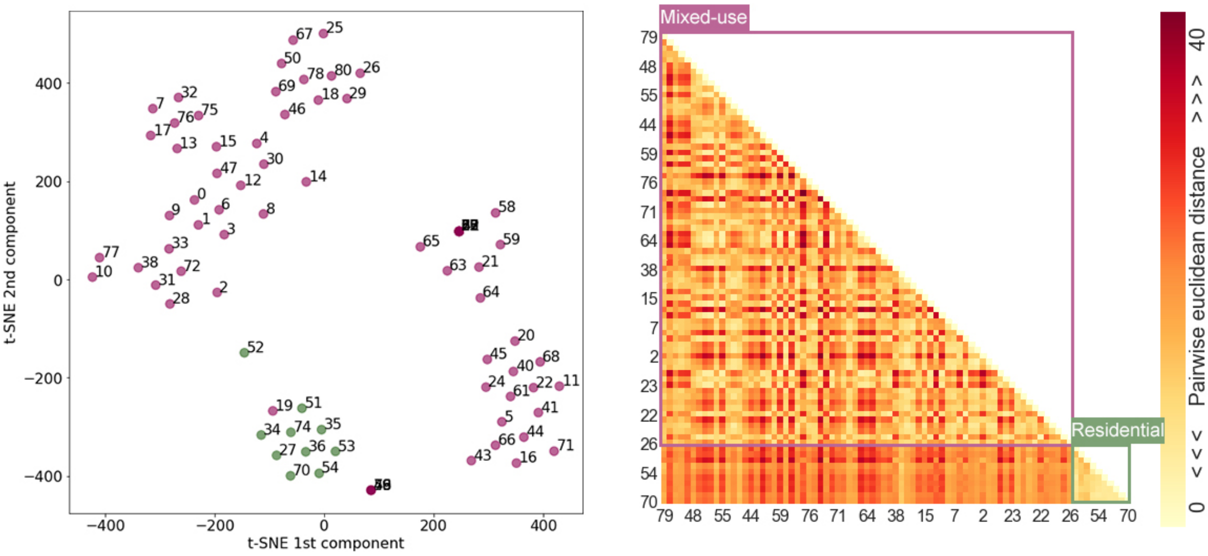

Figure 14: Operation-based scatter plot (left) and distance heat map(right) with building use type categorization.

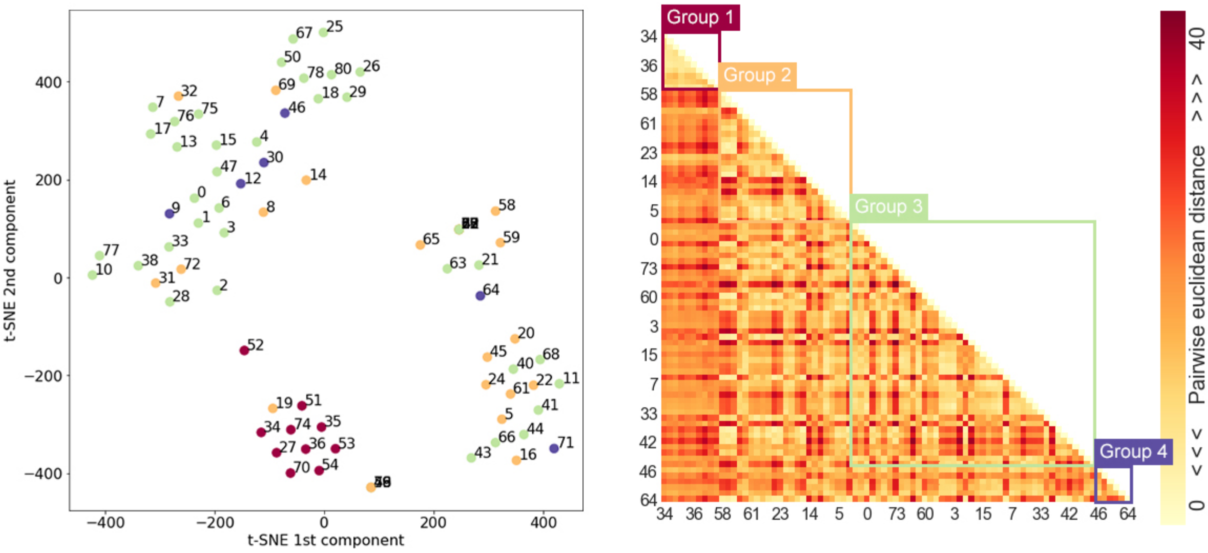

Figure 15: Operation-based scatter plot (left) and distance heat map(right) with baseline 2 groups.

baseline 2 got a P-value of 0.8149 against baseline 1, showing no significant difference. Therefore, an appropriate categorization is critical for building energy benchmarking.

5.2. *The two case studies*

5.2.1. *The first building*

The first identified building is a mixed-use building and held a EUI of 156 kWh/m^2, receiving the benchmarking percentile of 46.48% from baseline 1 and 57.14% from baseline 2. Considering that it was awarded a green building certification in 2015 after newly built, the energy performance according to the baseline benchmarking results was not as desired. In the proposed framework, the building was clustered into peer group 1 and became the least consuming building in the group with 0% percentile. Therefore, we hypothesized that the higher consumption was caused by the large amount of service delivered, instead of bad design or low efficiency.

With 40.13% of the area as galleries and labs that required 24-hour air conditioning, the base load of this building was around 50%. According to the one-week M&V done for the certification, the building achieved an overall lighting power density of 5.99 W/m^2 and an average plant room efficiency of 0.773 kW/RT, leading to 29.8% lighting energy saving and 13.7% chiller plant energy saving. Also, 17.7% of energy was saved for air-side fans through proper sizing. Therefore, the building consumed the lowest EUI in the 1st peer group with heavy operations, and it was performing very well despite the around median EUI among mixed-use buildings.

5.2.2. The second building

The second building is a mixed-use building with a relatively high EUI of 223 kWh/m^2, but was considered acceptable according to the baselines (57.74% by baseline 1 and 50% by baseline 2). On the contrary, the proposed method ranked the building at 83.33% as a bad performing building. Looking into the segmentation of energy consumption, we found that this building had normal electricity consumption (86 kWh/m^2) but very high cooling consumption (137 kWh/m^2). Accordingly, we tried to identify the cause of high cooling consumption through an energy audit and calibrated energy simulation. Applying the Morris method for sensitive analysis, the cooling consumption were found to be sensitive to several parameters and are summarized in table 4. The zone temperature and supply air temperature setpoints were set according to the BMS, and the window thermal properties are defined by looking up the observed material in the database. Besides, the infiltration rate and fan efficiency are found to be sensitive but hard to estimate. To provide confidence in the calibration, Bayesian calibration was carried out following the formulation proposed by Kennedy and O'Hagan [45], which explicitly models (1) parameter uncertainty, (2) model discrepancy, and (3) observation errors. For a fuller explanation of Bayesian calibration and its application, interested readers are referred to [46] and [47].

Table 4: Summary of EnergyPlus parameters.

Factor	EnergyPlus object	Source	Value
Zone temperature setpoint	ThermostatSetpoint:SingleCooing	BMS	23°C
Supply air temperature setpoint	SetpointManager:Scheduled	BMS	12°C
Windows SHGC	WindowMaterial:SimpleGlazingSystem	clear single layer	0.905
Windows U-Factor	WindowMaterial:SimpleGlazingSystem	clear single layer	5.894W/m^2-K
Infiltration rate	ZoneInfiltration:DesignFlowRate	to calibrate	N/A
Fan total efficiency	Fan:VariableVolume	to calibrate	N/A

Figure 16 shows the results of Bayesian calibration. According to the posterior distribution of the two parameters (a and b), the average infiltration rate is 1.13 air change per hour, and the average fan efficiency is 0.46. As shown in figure 16.c, the model well captured the measured data with small uncertainty after calibration. Thus, the energy performance of this building was not desirable. With all the parameters identified, the corresponding retrofit scenario was tested by modifying the EnergyPlus parameters. The zone temperature and supply air temperature setpoints were respectively set as 24 and 14°C, a low emissivity window with air gap was applied (U-Factor = 1.493W/m^2-K, SHGC = 0.373), the infiltration rate was reduced to 0.2, and the fan efficiency was improved to 0.6. The simulated monthly energy consumption after retrofitting is plotted in green as well in figure 16.c. The cooling consumption was remarkably reduced and 24% of energy saving was achieved. The analysis results accorded with the fact that a large portion of the building envelop is exposed to direct solar radiation. Therefore, better passive design should improve energy performance. This case study revealed that the proposed benchmarking successfully identified inefficient buildings that were overlooked by the baselines, where great energy-saving potential can be achieved.

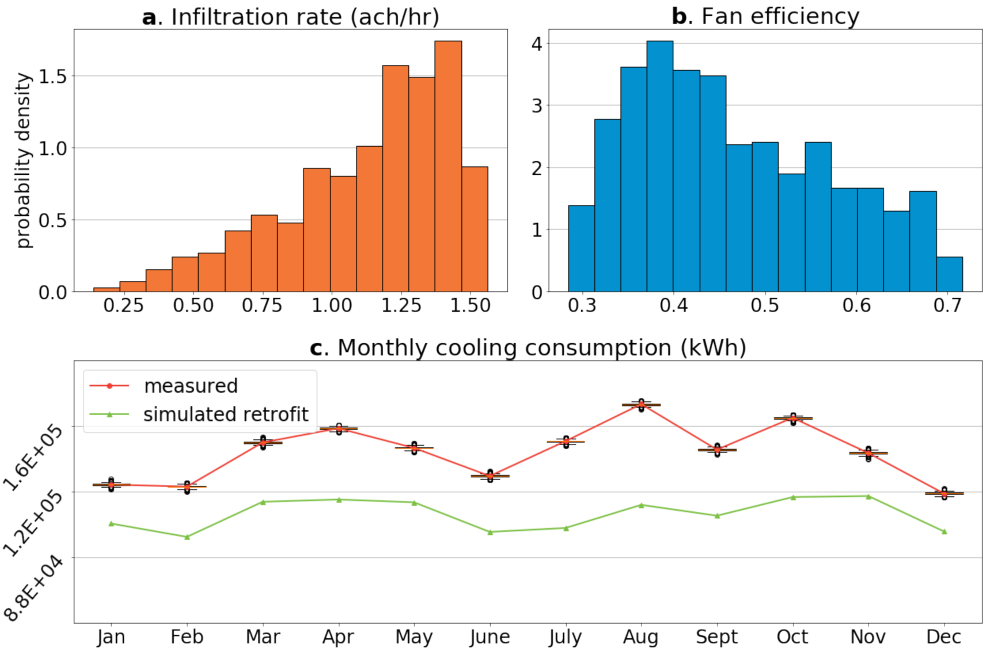

Figure 16: Results of applying Bayesian calibration on the EnergyPlus model (a. posterior of the infiltration rate; b. posterior of the fan efficiency); c. calibrated and proposed monthly cooling consumption.

6. Discussion

6.1. Effective operation-based categorization

As shown in the results, the buildings were effectively categorized for energy benchmarking based on the operational similarity. Similarly based on clustering methods, the proposed framework improved on the literature in two main aspects: annual maximum normalization is used for pre-processing, and K-means clustering is applied three times for three purposes. The methods are tailored according to the aim of this framework.

Normalization, as a step of pre-processing, is meant to extract the buildings' operation patterns by calculating the load ratio and preserving the trend of operation intensity. Since the aim is to quantify the buildings' operational similarity, the dissimilarity between profiles of different buildings and on different days should be preserved. The operational variations are composed of many temporal features, such as base load, peak load, rise time, duration time, and fall time. From this perspective, they should be maintained over pre-processing even though they may not yield to the statistically best clustering results. Figure 17 displays two examples of annual maximum normalization, compared with two typical methods used in the literature: daily standardization [11] and [0, 1] normalization [48].

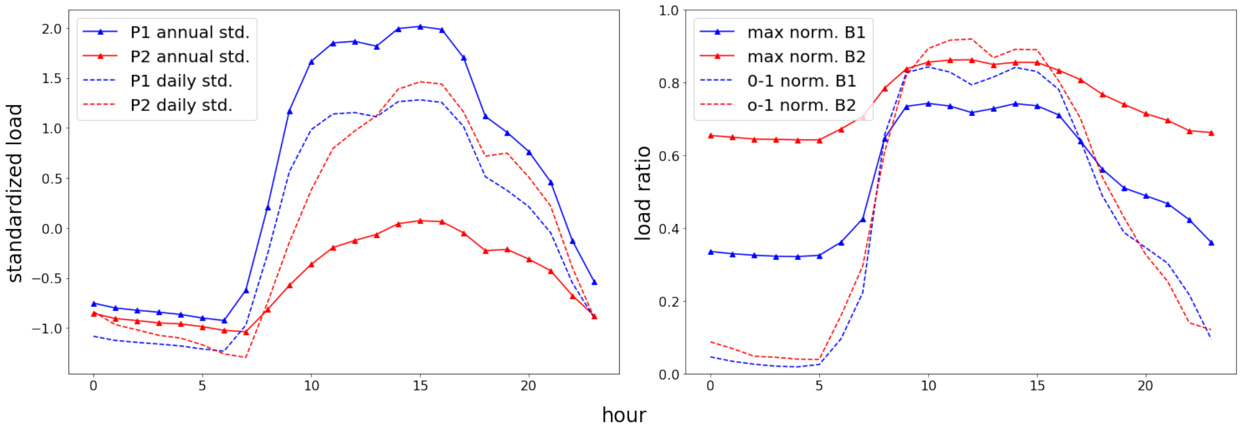

Figure 17: Comparison between different normalization methods using exemplary profiles.

As shown on the left, two profiles from the same building but on different days are standardized respectively with annual (solid) and daily (dashed) statistics (means and standard deviations). The peak load of these two profiles is originally different, which is reflected in the annually-standardized profiles. However, the profiles become very similar after daily standardization. This indicates the loss of different operational intensity on different days caused by daily standardization. On the right are two representative profiles from different buildings with annually maximum (solid) and [0, 1] (dashed) normalization. The results of maximum normalization show that building 1 maintains a much lower base load than building 2. In contrast, [0, 1] normalization stretches the profiles to fill the range between 0 and 1, and discards information about the different base load. To avoid such mistakes, normalization or pre-processing should be done towards appropriate information filtering instead of smaller variance or "cleaner" data. For buildings' operation quantification, maximum normalization should be applied on the annual scale.

As the annual maximum normalization keeps more information, the profiles aggregated together also become noisier and denser in the range between 0 and 1 (e.g., figure 5). Consequently, if K-means is directly applied, dozens of typical profiles would be identified, many of which would be similar and hinder further analysis. In this sense, the first step of single-building K-means clustering can also be considered as an extra step of pre-processing to account for the variations within individual buildings. The aggregated representative profiles are much more sparse, making it possible to apply the 2$^{\text{nd}}$ K-means for typical profile identification. The resulting typical profiles are distinct, and the characteristics are easily interpretable, enabling the labeling and featuring of each building. Lastly, the 3rd K-means is applied to group buildings using a dataset with the same number of dimensions as the typical profiles. While the number of features that characterizes the operational patterns is reduced from thousands to the number of typical profiles over the process of three-step clustering, the variations among different buildings during one year are represented. Thus, in the situation of larger datasets and more diversified building operations, the number of typical profiles is expected to increase, but the method will still be applicable.

6.2. More insightful energy benchmarking

According to the statistical test results and the in-detail case studies, it is necessary to conduct the energy benchmarking based on the building usage types and the operation-based categorizations. First, there is no point comparing the energy performance of different types of buildings as different types of service are delivered. For example, the residential buildings had fewer facilities and lower load density. Therefore, the EUI was much smaller than the mixed-use buildings. Meanwhile, within the same usage type, both the operational intensity and the trend of operational patterns have their impact on the buildings' energy behavior.

Being of the same type, buildings can have higher EUI simply because the operation is heavier or more service is delivered. For example, the bottom EUI quartile of the 4th peer group was lower than the top quartile of the 1st because the operation was much lighter. This was reflected in that the 4th peer group was dominated by typical profile 6, the base load of which was around half of typical profile 5 that dominated the 1st group. The first case study serves as another example, where the operation was intensive due to the galleries and labs that require 24-hour conditioning. Consequently, the energy performance was underestimated. These situations are not describable by just the length of operating hours, illustrating the need of conducting operation-based categorization for benchmarking.

Moreover, when the service is delivered or how the building is operated matters as well. For example, both the 1st and the 5th group had intensive operations. The light yellow intersection area in the right of figure 11 implies that their pairwise euclidean distances are small, and they had close accumulated load over the year. However, while buildings in the 1st group followed typical profile 5 almost every day, buildings in the 5th pursued typical profile 5 on most days, but typical profile 2 on minor days. As a result, all the EUI quartiles of the 5th group are smaller than the 1st. It has been shown that normalizing by operating hours cannot account for factors such as base load. This example further explains why normalizing by accumulated loads is not sufficient either to account for the variations in building operations.

The proposed framework is not meant to draw a perfectly correct conclusion on whether a building is performing good or not, but rather to give a better estimation of the performance and facilitate potential subsequent activities. It is also worth noting that the difference made by the proposed framework was more on buildings with around median EUI. The buildings with very high (low) EUI would be determined as bad (good) no matter how the peer groups were defined, while buildings in the middle were more likely to be misevaluated by the baseline methods. This, in turn, showed the greater capability of the proposed framework, as demonstrated in the 2nd case study, to exploit the energy-saving potential of those buildings that were overlooked.

6.3. Potential further investigation

6.3.1. Other possible applications

While the benchmarking framework was used to evaluate the buildings' energy performance during the operation phase in this study, the results can also be useful in the design stage. With the knowledge about a building's expected operation schedule, the building can be clustered into an appropriate peer group and get a more realistic benchmark as the goal for building system design. For example, the 1st case study building was operated intensely due to the galleries and labs. There are other types of space in modern buildings that require long operation hours, such as data centers. In such cases, the base load should be expected to be higher, and correspondingly the requirement on EUI can be relaxed.

Furthermore, the intermediate results of the framework can be further developed for different applications in building energy management and building simulation. The identified typical profiles can be used to catch abnormal operation patterns for a cluster of buildings by using some outlier detection techniques [49]. The operation-based building categorization can help with the building archetype identification for urban building energy modeling [50]. Besides, occupant behavior is a major source of uncertainty in building energy simulation, varies a lot among different buildings, and has been hard to quantify [8]. This study currently focused on the resulting energy consumption patterns, but the methods can potentially be used to infer the occupant behavior, given its significant impact on the energy consumption [51].

6.3.2. Testing on more comprehensive datasets

The dataset used in this study was obtained from the NUS campus. Giving the opportunity to conduct more in-depth analyses on typical buildings, it was also restricted by the number and the diversity of buildings. Considering that the energy performance was benchmarked as EUI ranking percentiles and that the median of a group of buildings would remain consistent as the group size increased [52], the benchmarking results were valid and representative. However, a few buildings were left not benchmarked because of the dataset size. For example, there were only four buildings in the 3rd peer group, which made this group not suitable for further benchmarking. Since the proposed framework is based on unsupervised clustering and is scalable for larger datasets, it would be interesting to test it on more buildings with more usage types. Besides, with the buildings located in Singapore, the climate is almost consistent over the year, and the annual trend of consumption is mainly caused by variations such as academic semesters or other operational behavior. If the method is applied in places with summers and winters, some decomposition techniques might be needed for pre-processing to remove the seasonal trend [53].

6.3.3. Integrating abnormal operation detection

This study achieved a fairer EUI-based energy benchmarking by removing the effect of different operations and did not account for the correctness of the operation patterns. On the other hand, buildings being not well operated is an essential factor that led to undesirable extra energy consumption [54]. For example, a building might have a high base load just because the HVAC system was wrongly kept on after people left. In other words, the proposed framework well benchmarked the energy performance of building systems and envelops, while there is another side of operational efficiency to be considered. Using the same kind of data, unreasonable or unexpected operation profiles can be identified based on knowledge [32] or some data mining techniques [55]. Therefore, a major part of future work lies in integrating these methods into the framework to benchmark a building's energy efficiency more comprehensively.

7. Conclusion

Aiming to a fairer building energy benchmarking, this study designed and implemented a framework that made use of the smart meter data to categorize buildings according to their operational similarities. Compared with existing methods in the literature, the proposed framework captured the building operational characteristics more precisely by adopting a more preservative normalization approach and three steps of clustering. The testing results on the 81 buildings showed that the proposed framework categorized the buildings better than both baseline methods.

The operation-based building categorization showed a significant impact (P=0.0265) on the benchmarking results. Detailed analysis of the relationship between the buildings' EUI and their operation patterns demonstrated the superiority of the new benchmarking framework. Further, the potential benefit of giving a more accurate estimation on buildings' energy performance was manifested by investigating two typical buildings.

Considering the integrated use type and diversified energy behavior of buildings, the application of this proposed method would be promising for both building design and operation. With the energy efficiency of building systems and envelops well evaluated, the framework can be further developed to account for the operation performance. Besides, the operation-based building categorization could serve as a starting point of applications such as occupant behavior quantification and urban building energy modeling.

References

[1] T. Abergel, B. Dean, J. Dulac, Towards a zero-emission, efficient, and resilient buildings and construction sector: Global status report 2017, UN Environment and International Energy Agency: Paris, France (2017).

[2] L. Pérez-Lombard, J. Ortiz, R. González, I. R. Maestre, A review of benchmarking, rating and labelling concepts within the framework of building energy certification schemes, Energy and Buildings 41 (3) (2009) 272–278. doi:10.1016/j.enbuild.2008.10.004.

[3] E. Star, Energy star® performance ratings technical methodology, Energy Star. from: http://www. energystar. gov/index. cfm (2011).

[4] B. Singapore, BCA Building Energy Benchmarking Report 2018 (2018).
URL https://www.bca.gov.sg/GreenMark/others/BCA_BEBR_Abridged_FA_2018.pdf

[5] A. Hinge, S. E. Partnerships, N. E. E. Partnerships, E. Smyth, R. Analytics, Back to school on energy benchmarking, Proceedings of the 2002 ACEEE Summer Study. Monterey, Calif (2002).

[6] X. Gao, A. Malkawi, A new methodology for building energy performance benchmarking: An approach based on intelligent clustering algorithm, Energy and Buildings 84 (2014) 607–616. doi:10.1016/j.enbuild.2014.08.030.

[7] H. S. Park, M. Lee, H. Kang, T. Hong, J. Jeong, Development of a new energy benchmark for improving the operational rating system of office buildings using various data-mining techniques, Applied energy 173 (2016) 225–237. doi:10.1016/j.apenergy.2016.04.035.

[8] W. Tian, Y. Heo, P. De Wilde, Z. Li, D. Yan, C. S. Park, X. Feng, G. Augenbroe, A review of uncertainty analysis in building energy assessment, Renewable and Sustainable Energy Reviews 93 (2018) 285–301. doi:10.1016/j.rser.2018.05.029.

[9] W. O'Brien, I. Gaetani, S. Carlucci, P.-J. Hoes, J. L. Hensen, On occupant-centric building performance metrics, Building and Environment 122 (2017) 373–385. doi:10.1016/j.buildenv.2017.06.028.

[10] S. Wang, C. Yan, F. Xiao, Quantitative energy performance assessment methods for existing buildings, Energy and buildings 55 (2012) 873–888. doi:10.1016/j.enbuild.2012.08.037.

[11] J. Y. Park, X. Yang, C. Miller, P. Arjunan, Z. Nagy, Apples or oranges? identification of fundamental load shape profiles for benchmarking buildings using a large and diverse dataset, Applied energy 236 (2019) 1280–1295. doi:10.1016/j.apenergy.2018.12.025.

[12] F. McLoughlin, A. Duffy, M. Conlon, Evaluation of time series techniques to characterise domestic electricity demand, Energy 50 (2013) 120–130. doi:10.1016/j.energy.2012.11.048.

[13] T. Hong, M. A. Piette, Y. Chen, S. H. Lee, S. C. Taylor-Lange, R. Zhang, K. Sun, P. Price, Commercial building energy saver: an energy retrofit analysis toolkit, Applied Energy 159 (2015) 298–309. doi:10.1016/j.apenergy.2015.09.002.

[14] J. Ni, X. Bai, A review of air conditioning energy performance in data centers, Renewable and sustainable energy reviews 67 (2017) 625–640. doi:10.1016/j.rser.2016.09.050.

[15] W. Chung, Y. Hui, Y. M. Lam, Benchmarking the energy efficiency of commercial buildings, Applied energy 83 (1) (2006) 1–14. doi:10.1016/j.apenergy.2004.11.003.

[16] B. Singapore, Green Mark Criteria for Non-residential Existing Building (2019 (accessed December 9, 2019)).
URL https://www.bca.gov.sg/green_mark/cdl/GreenMarkExistingBuildingsCriteria.pdf

[17] W. Chung, Y. V. Hui, A study of energy efficiency of private office buildings in hong kong, Energy and Buildings 41 (6) (2009) 696–701. doi:10.1016/j.enbuild.2009.02.001.

[18] E. I. A. (EIA), Advanced Metering Count by Technology Type (2019 (accessed October 20, 2019)).
URL https://www.eia.gov/electricity/annual/html/epa_10_10.html

[19] R. Hierzinger, M. Albu, H. Van Elburg, A. J. Scott, A. Łazicki, L. Penttinen, F. Puente, H. Sæle, European smart metering landscape report 2012, SmartRegions Deliverable 2 (2012).

[20] Y. Wei, X. Zhang, Y. Shi, L. Xia, S. Pan, J. Wu, M. Han, X. Zhao, A review of data-driven approaches for prediction and classification of building energy consumption, Renewable and Sustainable Energy Reviews 82 (2018) 1027–1047. doi:10.1016/j.rser.2017.09.108.

[21] C. Miller, Z. Nagy, A. Schlueter, A review of unsupervised statistical learning and visual analytics techniques applied to performance analysis of non-residential buildings, Renewable and Sustainable Energy Reviews 81 (2018) 1365–1377. doi:10.1016/j.rser.2017.05.124.

[22] A. Lavin, D. Klabjan, Clustering time-series energy data from smart meters, Energy efficiency 8 (4) (2015) 681–689. doi:10.1007/s12053-014-9316-0.

[23] S. Jalori, P. T Agami Reddy PhD, A new clustering method to identify outliers and diurnal schedules from building energy interval data, ASHRAE Transactions 121 (2015) 33.

[24] C. Fan, F. Xiao, C. Yan, A framework for knowledge discovery in massive building automation data and its application in building diagnostics, Automation in Construction 50 (2015) 81–90. doi:10.1016/j.autcon.2014.12.006.

[25] J. Yang, C. Ning, C. Deb, F. Zhang, D. Cheong, S. E. Lee, C. Sekhar, K. W. Tham, k-shape clustering algorithm for building energy usage patterns analysis and forecasting model accuracy improvement, Energy and Buildings 146 (2017) 27–37. doi:10.1016/j.enbuild.2017.03.071.

[26] M. S. Piscitelli, S. Brandi, A. Capozzoli, Recognition and classification of typical load profiles in buildings with non-intrusive learning approach, Applied Energy 255 (2019) 113727. doi:10.1016/j.apenergy.2019.113727.

[27] S. V. Verdu, M. O. Garcia, C. Senabre, A. G. Marin, F. J. G. Franco, Classification, filtering, and identification of electrical customer load patterns through the use of self-organizing maps, IEEE Transactions on Power Systems 21 (4) (2006) 1672–1682. doi:10.1109/TPWRS.2006.881133.

[28] K. Li, Z. Ma, D. Robinson, J. Ma, Identification of typical building daily electricity usage profiles using gaussian mixture model-based

clustering and hierarchical clustering, Applied energy 231 (2018) 331–342. doi:10.1016/j.apenergy.2018.09.050.

[29] G. J. Tsekouras, N. D. Hatziargyriou, E. N. Dialynas, Two-stage pattern recognition of load curves for classification of electricity customers, IEEE Transactions on Power Systems 22 (3) (2007) 1120–1128. doi:10.1109/TPWRS.2007.901287.

[30] F. McLoughlin, A. Duffy, M. Conlon, A clustering approach to domestic electricity load profile characterisation using smart metering data, Applied energy 141 (2015) 190–199.

[31] G. Chicco, R. Napoli, F. Piglione, Comparisons among clustering techniques for electricity customer classification, IEEE Transactions on Power Systems 21 (2) (2006) 933–940. doi:10.1109/TPWRS.2006.873122.

[32] X. Luo, T. Hong, Y. Chen, M. A. Piette, Electric load shape benchmarking for small-and medium-sized commercial buildings, Applied energy 204 (2017) 715–725. doi:10.1016/j.apenergy.2017.07.108.

[33] F. Iglesias, W. Kastner, Analysis of similarity measures in times series clustering for the discovery of building energy patterns, Energies 6 (2) (2013) 579–597. doi:10.3390/en6020579.

[34] I. P. Panapakidis, T. A. Papadopoulos, G. C. Christoforidis, G. K. Papagiannis, Pattern recognition algorithms for electricity load curve analysis of buildings, Energy and Buildings 73 (2014) 137–145. doi:10.1016/j.enbuild.2014.01.002.

[35] Z. Ma, R. Yan, N. Nord, A variation focused cluster analysis strategy to identify typical daily heating load profiles of higher education buildings, Energy 134 (2017) 90–102. doi:10.1016/j.energy.2017.05.191.

[36] D. Hsu, Comparison of integrated clustering methods for accurate and stable prediction of building energy consumption data, Applied energy 160 (2015) 153–163. doi:10.1016/j.apenergy.2015.08.126.

[37] T. Caliński, J. Harabasz, A dendrite method for cluster analysis, Communications in Statistics-theory and Methods 3 (1) (1974) 1–27. doi:10.1080/03610927408827101.

[38] S. Zhan, A. Chong, A robust unsupervised framework for high-resolution building energy consumption profiling, Proceedings of the International Building Performance Simulation Association 16 (2019).

[39] L. v. d. Maaten, G. Hinton, Visualizing data using t-sne, Journal of machine learning research 9 (Nov) (2008) 2579–2605.

[40] A. Schaefer, E. Ghisi, Method for obtaining reference buildings, Energy and Buildings 128 (2016) 660–672. doi:10.1016/j.enbuild.2016.07.001.

[41] C. Fan, F. Xiao, S. Wang, Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques, Applied Energy 127 (2014) 1–10. doi:10.1016/j.apenergy.2014.04.016.

[42] P. Gianniou, X. Liu, A. Heller, P. S. Nielsen, C. Rode, Clustering-based analysis for residential district heating data, Energy conversion and management 165 (2018) 840–850. doi:10.1016/j.enconman.2018.03.015.

[43] P. Xue, Z. Zhou, X. Fang, X. Chen, L. Liu, Y. Liu, J. Liu, Fault detection and operation optimization in district heating substations based on data mining techniques, Applied energy 205 (2017) 926–940. doi:10.1016/j.apenergy.2017.08.035.

[44] F. Wilcoxon, Individual comparisons by ranking methods, in: Breakthroughs in statistics, Springer, 1992, pp. 196–202. doi:10.1007/978-1-4612-4380-9_16.

[45] M. C. Kennedy, A. O'Hagan, Bayesian calibration of computer models, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63 (3) (2001) 425–464.

[46] A. Chong, K. Menberg, Guidelines for the bayesian calibration of building energy models, Energy and Buildings 174 (2018) 527–547. doi:10.1016/j.enbuild.2018.06.028.

[47] A. Chong, K. P. Lam, M. Pozzi, J. Yang, Bayesian calibration of building energy models with large datasets, Energy and Buildings 154 (2017) 343 – 355. doi:https://doi.org/10.1016/j.enbuild.2017.08.069.
URL http://www.sciencedirect.com/science/article/pii/S0378778817318741

[48] V. Figueiredo, F. Rodrigues, Z. Vale, J. B. Gouveia, An electric energy consumer characterization framework based on data mining techniques, IEEE Transactions on power systems 20 (2) (2005) 596–602. doi:10.1109/TPWRS.2005.846234.

[49] C. Fan, F. Xiao, Z. Li, J. Wang, Unsupervised data analytics in mining big building operational data for energy efficiency enhancement: A review, Energy and Buildings 159 (2018) 296–308. doi:10.1016/j.enbuild.2017.11.008.

[50] T. Hong, Y. Chen, X. Luo, N. Luo, S. H. Lee, Ten questions on urban building energy modeling, Building and Environment (2019) 106508doi:10.1016/j.buildenv.2019.106508.

[51] D. Yan, W. O'Brien, T. Hong, X. Feng, H. B. Gunay, F. Tahmasebi, A. Mahdavi, Occupant behavior modeling for building performance simulation: Current state and future challenges, Energy and Buildings 107 (2015) 264–278. doi:10.1016/j.enbuild.2015.08.032.

[52] S.-M. Hong, G. Paterson, E. Burman, P. Steadman, D. Mumovic, A comparative study of benchmarking approaches for non-domestic buildings: Part 1–top-down approach, International Journal of Sustainable Built Environment 2 (2) (2013) 119–130. doi:10.1016/j.ijsbe.2014.04.001.

[53] P. Jonsson, L. Eklundh, Seasonality extraction by function fitting to time-series of satellite sensor data, IEEE transactions on Geoscience and Remote Sensing 40 (8) (2002) 1824–1832. doi:10.1109/TGRS.2002.802519.

[54] P. De Wilde, The gap between predicted and measured energy performance of buildings: A framework for investigation, Automation in Construction 41 (2014) 40–49. doi:10.1016/j.autcon.2014.02.009.

[55] A. Capozzoli, M. S. Piscitelli, S. Brandi, D. Grassi, G. Chicco, Automated load pattern learning and anomaly detection for enhancing energy management in smart buildings, Energy 157 (2018) 336–352. doi:10.1016/j.energy.2018.05.127.