Apples or Oranges? Identification of fundamental load shape profiles for benchmarking buildings using a large and diverse dataset

June Young Park\textsuperscript{1}, Xiya Yang\textsuperscript{1}, Clayton Miller\textsuperscript{2}, Pandarasamy Arjunan\textsuperscript{3}, Zoltan Nagy\textsuperscript{1,*}

\textsuperscript{1}Intelligent Environments Laboratory, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin, Austin, TX, USA
\textsuperscript{2}Building and Urban Data Science (BUDS) Lab, Department of Building, School of Design and Environment (SDE), National University of Singapore (NUS), Singapore
\textsuperscript{3}Berkeley Education Alliance for Research in Singapore (BEARS), Singapore

Abstract

Buildings are responsible for 30—40\% of the anthropogenic greenhouse gas emissions and energy consumption worldwide. Thus, reducing the overall energy use and associated emissions in buildings is crucial for meeting sustainability goals for the future. In recent years, smart energy meters have been deployed to enable monitoring of energy use data with hourly or sub-hourly temporal resolution. The concurrent rise of information technologies and data analytics enabled the development of novel applications such as customer segmentation, load profiling, demand response, energy forecasting and anomaly detection. In this paper, we address load profiling and benchmarking, i.e., determining peer groups for buildings. Traditionally, static characteristics, e.g., primary space use (PSU) together with the annual energy-use-intensity, EUI, have been used to compare the performance of buildings. Data-driven benchmarking approaches have begun to also consider the shape of the load profiles as a means for comparison. In this work, we identify three fundamental load shape profiles that characterize the temporal energy use in any building. We obtain this result by collecting a dataset of unprecedented variety in size (3,829 buildings) and primary use (75 programs), and applying a rigorous clustering analysis followed by entropy calculation for each building. The existence of fundamental load shape profiles challenges the man-made, artificial classification of buildings. We demonstrate in a benchmarking application that the resulting data-driven groups are more homogeneous, and therefore more suitable for comparisons.
between buildings. Our findings have potential implications for building and urban energy simulations, portfolio management, demand response and renewable energy integration in buildings.

**Keywords:** Building Energy, Load Profile, Energy Benchmarking, Unsupervised Learning, Visual Analytics, Data Analytics

1. Introduction

The building sector represents the largest portion of energy consumption and greenhouse gas emission worldwide. In the United States alone, residential and commercial buildings account for 40% of the energy consumption and 38% of CO$_2$ emissions [1]. With increasing per capita energy usage and rapid urbanization, the energy demand in the building sector continues to increase at unprecedented level [2]. Thus, reducing the overall energy usage and associated emissions in buildings is crucial for meeting sustainability goals. As a result, there are tremendous research and entrepreneurial activities by both public and private stakeholders to optimize energy usage in buildings [3, 4, 5, 6].

To this end, smart meters have been deployed around the world during the last decade. For example, 70 million smart electricity meters were installed in the US by 2016 [7]. The availability of smart meter data enables both utilities and consumers to have a better understanding of how energy is spent in buildings. In general, the rise of information technologies fused with energy system has resulted in energy-cyber-physical systems, or e-CPSs, enabling the development of several fine-grained energy management applications, such as consumer segmentation and load profiling [8, 9, 10, 11], demand-response [12, 13], energy forecasting [14, 15], and anomaly detection [16, 17, 18].

1.1. In Search of Alternative Means of Classifying Buildings

In this paper, we address the topic of data-driven load profiling and benchmarking. Traditionally, buildings are classified into man-made categories, e.g., residential, commercial, and various sub-categories, such as education, office, and retail, based on their Primary Space Usage (PSU). Primary space use, also known as primary space activity, is a concept that is used extensively within benchmarking systems and energy consumption surveys, including the Commercial Buildings Energy Consumption Survey (CBECS) in the United States [19]. These PSU classifications are the key component when defining the Building Type Definition, a label that is in
most aspects of performance analysis, including benchmarking. The problem with these classifications is that they are inflexible to the reality of modern buildings: entities that are considered a whole building do not often wholly fit into these categories due to an increasing diversity of uses and loads in buildings. These buildings are often referred to as mixed use colloquially, but are often still officially given a rigid building type label. The CBECS data collection protocol instructs that buildings used for more than one of the activities described are assigned to the activity occupying the most floorspace. This type of fuzzy classification creates a situation in which a number of buildings are placed in peer groups that may under or over-estimate their relative energy performance.

Using the concept of PSU and building type, several building energy benchmarking and labeling methods have been proposed in the literature [20, 21, 22, 23]. The objective of these benchmarking methods is to derive groups of similar buildings, which can highlight whether or not a specific building in this group is performing better or worse than its peer group. One of the widely used benchmarking methods is EUI, or energy-usage-intensity, which is simply the annual energy consumption divided by the square footage of the building.

Since the EUI oversimplifies the energetic performance of a building, and capitalizing on the aforementioned deployment of advanced metering infrastructures, data-driven building energy benchmarking methods have been proposed. In contrast with using the static attributes of the building, e.g., the floor area, and coarse-grained energy usage data (monthly or annual bills), data-driven methods use fine-grained energy usage data, typically sampled at every hour, and attempt to capture unique load characteristics independent of the artificial man-made attribute.

1.2. Research Contribution

In this paper, we hypothesize that fundamental load shape profiles exist that characterize the energy use of a building. Fundamental profiles are independent of a building’s man-made, artificial label, and, if exist, would allow to label buildings by their temporal energetic behavior. As a consequence, natural, data-driven peer groups buildings can be formed, with similar energetic behavior rather than with similar artificial label, resulting in much more meaningful comparisons. We discover these profiles using both (1) a diverse dataset composed of an unprecedented variety in size (3,829 buildings), primary use (75 building programs), and location (Fig. I), and (2) a thorough clustering analysis.
We add to the existing literature by identifying fundamental load shape profiles of a building in two steps. First, we demonstrate that the daily load profiles of almost all buildings (≈94%) in our dataset can be clustered into three representative groups. Second, analyzing the entropy of the formed clusters for each building, we show that almost all buildings exhibit a consistent energy use pattern, i.e., one of the three load profiles is dominant: it occurs for more than 50% of the days. Thus, we identify three fundamental load shape profiles that can be qualitatively characterized by having either a morning, a mid-day or an evening peak of energy use, respectively. We then show that regrouping the buildings according to these profiles, e.g., for benchmarking, results in much more homogeneous groups.

The paper is organized as follows. The next section presents an overview of the related literature. In Section 3, we present our dataset and methodology. Section 4 details the results, while Section 5 discusses the implications and possible applications of the discovered load shape profiles. Section 6 concludes the paper.

2. Literature Review

2.1. Load profiling

Existing data-driven load profiling approaches are broadly divided into direct and indirect clustering methods [12, 24]. While direct clustering approaches, as the name implies, directly use the raw meter data to the clustering algorithms, indirect clustering approaches use the features extracted from the meter data.

The most commonly used clustering algorithms for load profiling are k-means, fuzzy k-means, weighted fuzzy average k-means, follow-the-leader, hierarchical clustering, and Self Organizing Maps (SOM). Their advantages and disadvantages with respect to different similarity measures and validity metrics have been studied in [25, 26, 27, 28, 29]. These traditional clustering methods have also been extended for modeling some specific attributes of high-volume time-series energy data. For example, in [30], authors proposed a dynamic clustering method, by extending the traditional k-means, for capturing time-dependent seasonal trends.

Whereas, in indirect clustering, suitable features are extracted from the raw smart meter data before using the clustering algorithm. Smart meter data are inherently time-series with high dimensionality. Hence dimensionality reduction methods are applied before clustering. The most common dimensionality reduction method is Principle Component Analysis (PCA),
which is explored in [31]. Other similar methods are using Support Vector Clustering (SVC) [32], K-Medoids [33], and Neural Networks [34], and C-Vine copula mixture model [35].

While a plethora of studies exist, they are limited in their generalization due to either small sample sizes, specific to a particular building use type, limited geographic variation, focus on algorithms, or case study character.

2.2. Benchmarking studies

Several benchmarking methods have been proposed in the literature with varied complexities. The EUI is one of the widely used methods as it is simple, and easy to compute and interpret [36]. However, EUI makes a strong assumption that energy usage and gross floor area scale linearly, which is not the case with many buildings. Further, it fails to normalize other important factors (e.g., age, occupancy, electrical systems, etc.), thus making it unreliable when comparing heterogeneous building use types [20].

EPA’s Energy Star [37] is another popular rank-based benchmarking system. It can normalize energy usage for a variety of factors and it finds the average consumption for a group of input buildings using national survey data. The Energy Star scores are based on residuals from ordinary least square (OLS) regression models, but that includes statistical noise, measurement errors, any unknown factors, and it is sensitive to outliers in the data. This rating system uses the CBECS survey as a data source to create the peer groups for submitted buildings.

Several existing studies utilized advanced machine learning based approaches, such as Artificial Neural Networks (ANN) [38, 39, 40, 41], cluster analysis [42, 43], decision trees [44], data envelopment analysis [45], and stochastic frontier analysis [46, 47], for developing benchmarking models. While these systems address specific issues with existing benchmarking systems, such as generalizable, interpretable, robustness, etc., they have been monotonously validated for specific building use types (hotels [42], schools [40], office [44], government [45], residential [41, 46], commercial [39]), geography and climate zones, thus limiting their wide applicability across the world. Further, most of these studies used a limited set of building characteristics (floor space, age, occupancy, number of floors, etc.) for benchmarking.

In contrast, in this paper, we identify three fundamental load shape profiles from smart meter readings, as a baseline for grouping similar buildings for benchmarking.
3. Methodology

3.1. Data sources and organization

Multiple sources are used to collect building energy data for this study (see Tab. 1). Each dataset contains hourly energy consumption with labeled building informations, i.e., location, program, and gross area. In total, we collected hourly data from 3,829 buildings with 2,365,563 daily profiles of energy consumption. Fig. 1 illustrates the various global locations from where the data were collected. Our dataset contains both residential and non-residential buildings, and each individual dataset has different data collection period. Notably, all data sources are publicly available for academic purpose.

We joined the individual datasets into a single hierarchical data format (HDF5) file to serve as our database [54]. The HDF5 data format is particularly useful for our study, because our dataset contains large amount of building energy data with hierarchical information. Fig. 2 shows the structure of the database. It contains the unique identifier of the buildings, and temporal and a metadata folders for each building. In the temporal folder, hourly energy consumptions are stored for each year based on their availability. The first meta folder stores categorical meta data, e.g., industry, sub-industry, primary space usage (PSU), and climate zone. Industry and sub-industry is high level category of buildings, i.e., residential, educational, governmental, and others. More precisely, we detailed the program of each building by PSU types which is defined in similar studies [55, 56, 57]. The PSU indicated for each building was either collected from the facilities management department of the source institutions, scraped from web-based resources that accompanied the raw temporal data, or through a best guess estimate from the research team based on discussions and analysis. The PSU categories for these buildings mostly mirror those used for the CBECs
<table>
<thead>
<tr>
<th>Dataset</th>
<th>Location</th>
<th>No. of buildings</th>
<th>Type</th>
<th>Date Range</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anonymous Building Data Genome (BDG)</td>
<td>Various</td>
<td>342</td>
<td>Non-Residential</td>
<td>2010-01-01—2015-12-31</td>
<td>48</td>
</tr>
<tr>
<td>Arizona State University (BDG)</td>
<td>Tempe, AZ, USA</td>
<td>174</td>
<td>Non-Residential</td>
<td>2015-01-01—2015-12-31</td>
<td>56</td>
</tr>
<tr>
<td>BuildSMART DC</td>
<td>Washington DC, USA</td>
<td>499</td>
<td>Non-Residential</td>
<td>2016-01-01—2016-12-31</td>
<td>40</td>
</tr>
<tr>
<td>Cardiff Council/Carbon Culture (BDG)</td>
<td>Cardiff, UK</td>
<td>161</td>
<td>Non-Residential</td>
<td>2015-11-30—2016-12-01</td>
<td>45</td>
</tr>
<tr>
<td>CER Smart Meter Data</td>
<td>Ireland</td>
<td>1,781</td>
<td>Residential</td>
<td>2009-07-14—2010-12-31</td>
<td>50</td>
</tr>
<tr>
<td>EnerNOC Green Button Data (BDG)</td>
<td>Various</td>
<td>348</td>
<td>Non-Residential</td>
<td>2012-01-31—2014-12-31</td>
<td>56</td>
</tr>
<tr>
<td>MIT</td>
<td>Cambridge, MA, USA</td>
<td>87</td>
<td>Non-Residential</td>
<td>2014-01-01—2016-12-31</td>
<td>51</td>
</tr>
<tr>
<td>Pecan Street Inc.</td>
<td>Austin, TX, USA</td>
<td>113</td>
<td>Residential</td>
<td>2012-03-19—2017-09-16</td>
<td>57</td>
</tr>
<tr>
<td>UK Government Buildings/Carbon Culture (BDG)</td>
<td>UK</td>
<td>34</td>
<td>Non-Residential</td>
<td>2014-12-01—2015-11-30</td>
<td>45</td>
</tr>
<tr>
<td>University College London/Carbon Culture (BDG)</td>
<td>London, UK</td>
<td>53</td>
<td>Non-Residential</td>
<td>2014-12-01—2015-11-30</td>
<td>52</td>
</tr>
<tr>
<td>University of California - Berkeley (BDG)</td>
<td>Berkeley, CA, USA</td>
<td>29</td>
<td>Non-Residential</td>
<td>2012-01-01—2016-12-01</td>
<td>48</td>
</tr>
<tr>
<td>University of Greenwich/Carbon Culture (BDG)</td>
<td>Greenwich, UK</td>
<td>46</td>
<td>Non-Residential</td>
<td>2014-12-01—2015-11-30</td>
<td>48</td>
</tr>
<tr>
<td>University of Texas at Austin</td>
<td>Austin, TX, USA</td>
<td>111</td>
<td>Non-Residential</td>
<td>2009-01-15—2017-08-20</td>
<td>53</td>
</tr>
<tr>
<td>University of Southampton (BDG)</td>
<td>Southampton, UK</td>
<td>51</td>
<td>Non-Residential</td>
<td>2014-12-01—2015-11-30</td>
<td>48</td>
</tr>
</tbody>
</table>

Table 1: Summary of data sources
survey \[19\]. In addition, climate condition is labeled based on building location and International Energy Conservation Code (IECC) climate zone map \[58\]. Secondly, the numerical meta data folder contains gross area (m\(^2\)) and energy use intensity (EUI) (kWh/m\(^2\)/year).

3.2. Discovering fundamental load shape profiles

To investigate fundamental building energy consumption patterns, we developed a load profile based clustering framework, shown in Fig. 3. It consists of three steps: (1) Preprocessing to eliminate incomplete load profiles and apply Z-normalization. (2) Clustering using unsupervised learning techniques, i.e., K-means, Bisecting K-means, and Gaussian Mixture Models. (3) All clustered profiles are then re-assembled on a building level, and we calculate the cluster distribution, i.e., the frequency of each cluster, for each building. We detail each process in the following.

1. Preprocessing

We extract daily profiles of energy consumption from our dataset. Let \( t \in [1, 24] \) be the hour of day, and \( L_d(t) \) the hourly energy consumption of a building on day \( d \) in kWh. The daily profile is expressed as 24 data points, i.e., \( L_d(1), \ldots, L_d(24) \). The number of daily profiles varies for each building due to the different data collection periods of buildings.

We first remove daily profiles that do not have complete 24 data points. Then, we normalize daily profiles for further analysis using Z-normalization as \[59, 60\]

\[
Z_d(t) = \frac{L_d(t) - \mu}{\sigma}
\]

where \( \mu \) and \( \sigma \) are the mean and standard deviation of \( L(t) \). Z-normalization allows us to capture the shape of the profile rather than
1. Preprocessing

2. Clustering profiles

3.1 Clustering distributions and fundamental profiles

<table>
<thead>
<tr>
<th></th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.6</td>
<td>0.3</td>
<td>bldg1</td>
</tr>
<tr>
<td>0.2</td>
<td>0.2</td>
<td>0.6</td>
<td>bldg2</td>
</tr>
<tr>
<td>0.7</td>
<td>0.1</td>
<td>0.2</td>
<td>bldg3</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

3.2 Grouping buildings

Figure 3: Overview of data analytics framework
Algorithm 1: K-Means clustering

Determine the number of clusters ($k$)
Initialize $k$ number of centroid randomly

repeat
  for every data point do
    for every centroid do
      calculate the distance between the data point and the centroid
      assign the point to the cluster with the lowest distance away
    end
  end
  for every cluster do
    calculate the cluster mean
    assign the cluster to the mean
  end
until no data point has changed cluster assignment

The magnitude as the resulting mean for all profiles will be close to 0, while the standard deviation will be close to 1. We now elaborate how we cluster these profiles.

2. Profile Clustering The objective of clustering is to group the given data points, load profiles in our case, into a certain number $k$ of clusters that show similarity. We use three clustering algorithms in our study, but other unsupervised learning algorithms can be also used.

The first algorithm that we investigate is K-means clustering (see Alg. 1). Due to its simplicity, this algorithm has been widely applied in various domains [61], and has been shown to be the most popular approach for smart meter and portfolio analysis [62], which are potential applications of our study.

The drawback of K-means is its randomness in the initialization of the $k$ initial centroids, which sometimes results in local minimum rather than a global one [63]. To mitigate this potential issue, we also apply the Bisecting K-means [64] algorithm to our dataset. The main difference compared to basic K-means is that Bisecting K-means starts to cluster dataset with $k = 2$ (see Alg. 2), calculates the sum of squared error (SSE) of each cluster, divides one of the clusters into two new ones, and proceeds iteratively until a number of $k$ clusters have been determined.

Both basic K-means and Bisecting K-means are deterministic in nature, i.e., they use the mean as centroid of clusters and assign the clus-
Algorithm 2: Bisecting K-Means clustering

Determine the number of clusters ($k$)
Start with basic K-means clustering ($k=2$)
repeat
  for every cluster do
    measure the SSE of the clusters
    select the cluster with higher SSE
  end
  for selected cluster do
    K-means clustering ($k=2$)
  end
until the number of clusters reached $k$

Algorithm 3: Gaussian Mixture Model based clustering

Determine the number of clusters ($k$)
Obtain $k$ centroids using basic K-Means
Initialize weights, means and variances based on the $k$ centroids obtained
repeat
  for every data point do
    calculate the responsibility of the data point for each mixture component using the updated weights, means and variances
  end
  compute the estimates for weights, means and variances that maximize the expected complete data log likelihood given the calculated responsibilities
until the expected likelihood converged

The use of a mixture model allows for the estimation of probability or uncertainty on the cluster assignment [65].

In this regard, Gaussian mixture model (GMM) can estimate a mixture of multi-dimensional Gaussian probability distributions of each cluster (see Alg. 3). Compared to K-means, GMM is more flexible in terms of cluster covariance. GMM is based on a two step expectation-maximization approach: 1) Expectation: for each data point, find weights encoding the probability of membership in each cluster, and 2) Maximization: for each cluster, update its location, normalization, and shape based on all data points, making use of the weights [65].

We calculate three metrics to evaluate the clustering performance of each algorithm. The first one, Cohesion, measures the similarity of profiles within a cluster by evaluating the sum of squared distances
from each data point to the respective centroid:

$$\text{Cohesion} = \sum_{i=1}^{k} \sum_{x \in C_i} ||x - c_i||^2$$  \hspace{1cm} (2)$$

where $k$ is the number of clusters, $C_i$ is cluster $i$, $x$ is a point in cluster $C_i$ and $c_i$ is the centroid of cluster $C_i$.

Conversely, Separation measures how well dissimilar profiles are grouped into separate clusters by evaluating the sum of squared distances from each centroid to the overall centroid adjusted by the number of data points in the respective clusters:

$$\text{Separation} = \sum_{i=1}^{k} |C_i|||c_i - c||^2$$  \hspace{1cm} (3)$$

where $|C_i|$ is the number of points in each cluster and $c$ is the overall centroid of the data.

Third, we use the Calinski-Harabasz (CH) Score, which offers a trade-off between separation and cohesion by using both the average between- and within-cluster sum of squares as

$$\text{CH Score} = \frac{\sum_{i=1}^{k} |C_i|||c_i - c||^2/(k - 1)}{\sum_{i=1}^{k} \sum_{x \in C_i} ||x - c_i||^2/(n - k)}$$  \hspace{1cm} (4)$$

where $n$ is the number of data points.

We described our selection of clustering algorithms and the evaluation metrics of clustering performance. Next, we explain how this clustering result is interpreted with respect to fundamental load shapes.

3. Cluster distribution and fundamental load shapes

Each of the found $k$ clusters represent a distinct energy consumption pattern in our dataset. If we aggregate these clustered daily profiles at the building level, then each building has proportions of cluster assignment. Dominant clusters, i.e., those that occur often can be identified using the entropy computation as

$$E(j) = -\sum_{i=1}^{k} p_j(C_i) \log_2 p_j(C_i)$$  \hspace{1cm} (5)$$

where $p_j(C_i)$ is the proportion of cluster type $i$ in building $j$. The entropy quantifies how distinct the load shape profiles of the building
are. If the building has only one cluster, then $E(j) = 0$. Larger values
for $E(j)$ indicate that various consumption patterns are occurring with
similar distribution, i.e., no dominant profiles are present.

The existence of dominant clusters, i.e., buildings with low entropy,
indicates that the building consumed energy in a relatively consistent
pattern. Reversely, if a building has evenly distributed $k$ clusters, i.e.,
no dominant cluster, then this building behaved with various energy
consumption patterns. If the same dominant profile is present in a
large amount of buildings, it is considered a fundamental load shape.

Finally, we group the buildings by cluster assignment, i.e., the building
belongs into group $G_k$ if its dominant cluster is $C_k$:

$$G_k = \{ \text{Bldg}(j) \mid \exists! k \text{ such that } p_j(C_k) > 0.5 \}.$$  \hspace{1cm} (6)

If there is no dominant cluster, then these buildings are classified as,

$$G_0 = \{ \text{Bldg}(j) \mid p_j(C_k) < 0.5 \ \forall k \}.$$  \hspace{1cm} (7)

To summarize, the groups $G_k$ have been derived only via clustering
and are hence representative of the energy consumption pattern. The
dominant profiles of each group are considered the fundamental load
shape profiles of the buildings.

3.3. Application: Data-driven load profile based benchmarking

As a case study of possible applications, we apply the derived fundamen-
tal load profiles to data-driven benchmarking. As stated earlier, the main
difference to the conventional approach is that the objective of the proposed
benchmarking is to group a large amount of buildings into the groups of
buildings with similar load shape profiles.

Once we group the buildings based on their load shape profiles, we fur-
ther investigate the meta data distribution of each group. In addition, we
evaluate the results of the two benchmarking strategies by (1) EUI, which is
widely used for comparing building performance between buildings and (2)
energy consumption pattern, which is the main topic of this paper.

3.4. Computing facility

Our dataset contains 2,365,563 daily profiles from 3,829 buildings. The
proposed framework is computationally demanding, especially the cluster-
ing and performance metric computations. Thus, we employed the Maverick
high performance computing system from the Texas Advanced Computing
Center (TACC) [68]. The computation time was approximately 16 hours to
perform data preprocessing, three clustering analyses with performance met-
ic calculations, regrouping of buildings, and the data-driven benchmarking
study. We used Python for pre- and post-processing. The clustering algo-
rithms themselves have been implemented using the scikit-learn library [69].
All our code is organized in Jupyter notebooks and released on Github [70].

4. Results

4.1. Data exploration

Fig. 4 shows the summary of the data by industry type, PSU, and gross
area. There are 1,910 residential and 1,919 non-residential buildings in our
dataset. Residential buildings are mainly single family houses, while non-
residential buildings are from education, government and other industries.
The major constitution of non-residential buildings are from education in-
dustry (1,038 buildings). In terms of geographical locations, residential
buildings are predominately from climate zone 4, 10% being located in cli-
mate zone 2. Similarly, most of the non-residential buildings are located in
climate zone 4. Approximately, 20% of buildings are from climate 2 and
5, respectively, and only a few buildings are located in climate zones 1 and
3. Regarding building size, most of non-residential buildings are larger than
residential buildings in our dataset. For residential buildings, the majority of
the buildings (64%) are between 100 and 200 m\(^2\). On the other hand, most
of non-residential buildings are larger than 3,000 m\(^2\), and we also have very
large facilities (> 10,000m\(^2\)), i.e., auditorium, stadium, and gymnasium in
education industry.

4.2. Clustering

Fig. 5 shows the clustering result: Each column represents an algorithm,
i.e., K-means, Bisecting K-means, and GMM, and each row indicates a dif-
ferent \(k = 2 \ldots 10\), resulting in 27 individual sub-figures. In each sub-figure,
the colored lines represent the cluster centroid, i.e., the average of the daily
profiles in each cluster. The line thickness is scaled according to the number
of profiles for the respective cluster, i.e., a thicker line indicates that that
cluster contains more daily profiles.

We can evaluate the clustering results qualitatively first. As an example,
the first sub-figure is the result of K-means clustering for \(k = 2\): all the daily
profiles are clustered into the two representative load profiles. The blue load
profile has the peak around noon, while the orange one consumes less energy
around noon but has two shallow peak points during morning and evening
time. Also, there are more load profiles clustered to the orange load profile compared to the blue one.

Further, in Fig. 5, we also observe that both K-means and Bisecting K-means result in rather similar load profiles. Increasing the cluster numbers until \( k=4 \), both methods subdivide the orange cluster into morning peak and evening peak precisely (orange, green, red). By increasing from \( k=5 \) to \( k=10 \), K-means details more evening peak clusters, while Bisecting K-means generates various morning peak clusters. This is because Bisecting K-means selects the cluster of higher SSE and again clusters (\( k=2 \)) on the data points of said cluster. For example, the purple profile in Bisecting K-means (\( k=5 \)) emerged due to this reason and this recursive approach generates different outcomes afterward. On the other hand, the GMM based clusters differ. This method clearly clusters noon peak profiles after \( k=5 \) case, but most of profiles are compounded on each other, which indicates that GMM may not be a suitable clustering method to find distinct profiles.

The clustering performance metrics, i.e, cohesion, separation and CH score are shown in Fig. 6 as a function of \( k \). In each case, K-means has the lowest cohesion, and both the largest separation and largest CH score, indicating that K-means clustering provides the best results in Fig. 5. In addition, increasing the number of clusters generally leads to lower cohesion and larger separation. Both K-means and Bisecting K-means show decreasing CH score with increasing \( k \), while GMM shows low CH score with little variation. Although there is no optimal procedure to find the optimal \( k \)
Figure 5: Clustering result for both residential and non-residential buildings (K-means clustering result with k=3 is highlighted and used for further analysis)
for clustering analysis in general, based on our metrics we conclude that in Fig. 6, the best balance between cohesion and separation is achieved with $k = 3$, which is also supported by the highest CH score.

4.3. Dominant clusters and fundamental load shape profiles

Fig. 7 visualizes the dominant clusters for our dataset for $k = 3$. Each horizontal line represents one building, shown with cluster assignment and color-coded meta data information for reference. The cluster assignment column shows the proportions of the three clusters: The buildings in $G_1$ exhibit a dominant cluster whose profile peaks at noon. $G_2$ and $G_3$ exhibit predominantly morning and evening profiles, respectively. Finally, buildings in $G_0$ have proportions of the three clusters each less than 50%, i.e., no dominant cluster. The last five columns visualize meta data information, aggregated in Fig. 8.

Using the proportions of clustering assignments, we calculate the entropy for each building. This entropy value quantifies the consistency of the load shape profiles of a building. Fig. 8a) visualizes the distributions of entropy of the buildings for each group. $G_0$ shows the highest entropy compared to the other groups. This indicates that it is comparatively more difficult to identify a fundamental load profile of the buildings in $G_0$. On the other hand, groups $G_1$—$G_3$ have lower average entropies, with $G_2$ having the lowest. Since this indicates that the buildings have consistent energy use patterns, we conclude that the identified dominant load shape profiles for $k = 3$ are indeed fundamental, in that they are characteristic of the energy
Figure 7: Grouping result by dominant clusters (the horizontal lengths of skyblue, blue, and navy represent the proportions of noon, morning, and evening peak load profiles of each building; legends for meta data are in Fig. 8)
Figure 8: Aggregated meta data information on each group
use of buildings. In addition, $G_0$ contains only 194 (~6%) buildings in total and mixed with residential and non-residential buildings with various PSUs. This relatively small number of buildings in $G_0$ indicates that most of buildings indeed exhibit at least one fundamental load shape profile. We show these profiles again in Fig. 9 for reference.

$G_1$ and $G_2$ have 1,005 and 478 buildings, respectively, and 90% of these buildings are non-residential buildings (Fig. 8b)). The distributions of PSU and gross area on group 1 & 2 were similar which suggests that building program and building size are not the primary factors to define the dominant load profile (Fig. 8c) & Fig. 8e)). Although our data collection has slightly skewed climate zones, $G_2$ has buildings from climate zone 2, 4, and 5 evenly, while $G_2$ mainly contains the buildings from climate zone 4 (Fig. 8d)). Regarding EUI, Fig. 8f) shows similar EUI distributions for both groups, confirming that EUI based benchmarking is an inappropriate approach for differentiating different energy consumption patterns of buildings.

About one half of the buildings are categorized as $G_3$ (1,645 buildings), and they are primarily residential buildings (single family houses) with a few dormitory buildings (Fig. 8b) & c)). This indicates that the majority of residential buildings are characterized as an evening peak energy load profile, which is a typical daily occupant behavior pattern in residential buildings. In terms of climate condition, climate zone 4 takes primary portion on this group due to the fact that majority of residential buildings are located in climate zone 4.

In conclusion, our results show that about 94% of the buildings have been assigned a dominant cluster, i.e., a cluster that is representative for the daily energy consumption pattern of the building for more than 50% of
the days. The centroid of the dominant cluster, therefore, can be interpreted as a fundamental load shape profile (Fig. 9). Given that our clustering result suggested an optimal value for the number of clusters as $k=3$, it follows that there exist three fundamental load shape profiles that appropriately capture the temporal energy use of buildings, regardless of other artificial, man-made labels.

4.4. Load profile based versus PSU based benchmarking

We compare our data-driven benchmarking result with the conventional approach, i.e., PSU based grouping. We adopt two different perspectives in terms of building energy, i.e., energy consumption pattern and EUI, to analyze the benchmarking results.

First, Fig. 10 visualizes the distribution of the resulting normalized load profiles for each benchmarking approaches. The top nine sub-figures show the distributions of load profiles of common PSUs in our dataset, and bottom four sub-figures are our data-driven benchmarking results for the same buildings. In general, the bottom sub-figures have clear load shape profiles with smaller interquartile ranges (IQR) compared to the profiles of PSU based grouping results. However, three PSUs (single family house, dormitory, and library) show relatively small IQR, which suggests that these PSUs have their unique load shape profile. Notably, the shape of single family house is similar to the shape of $G_3$, because the main constitution of $G_3$ is the single family house type (Fig. 4b)). Also, the buildings in $G_1$ is mainly collective of office, college laboratory, community center, library, and primary/secondary classrooms PSU types. Fig. 8c) details distribution of PSUs. The buildings of $G_0$ show fairly constant energy consumption pattern because they have evenly distributed fundamental load profiles, which average out each other throughout the day.

To evaluate how each benchmarking strategy grouped the buildings in terms of load profile, we show the clustering performance metrics in Tab. 2. The result indicates smaller cohesion, larger separation and larger CH score for data-driven benchmarking, meaning that the fundamental load shape profile approach is superior in both grouping and separating buildings with similar and dissimilar load profiles, respectively. This confirms that the proposed benchmarking method is particularly suitable to discover the peers with similar energy consumption pattern. In addition, this reduction from nine PSUs to four groups ($G_0$–$G_3$) suggests that we would only need four groups to investigate building performance comprehensively.

Second, we also investigated the resulting EUI distributions of the two benchmarking methods. Fig. 11 shows that while both benchmarking meth-
Figure 10: Profile distributions of PSU based (top) and data-driven benchmarking (bottom) (basic and dashed lines indicate mean and quartiles (25% & 75%) of normalized load profiles, respectively; All sub-figures are at the same scale)
### Table 2: The performance metrics indicate that clustering based on fundamental profiles results in better groups (low cohesion, large separation and large CH score)

<table>
<thead>
<tr>
<th>Benchmark type</th>
<th>Cohesion</th>
<th>Separation</th>
<th>CH score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary space use</td>
<td>$4.73 \times 10^7$</td>
<td>$3.85 \times 10^6$</td>
<td>$1.20 \times 10^4$</td>
</tr>
<tr>
<td>Fundamental load shape</td>
<td>$4.28 \times 10^7$</td>
<td>$8.30 \times 10^6$</td>
<td>$1.52 \times 10^5$</td>
</tr>
</tbody>
</table>

ods result in outliers, our data-driven benchmarking groups buildings with similar EUI values, i.e., smaller variations in EUI in each group compared to PSU benchmarking. This suggests that we can potentially utilize the proposed method for benchmarking not only load shape profiles but also the EUI of buildings.

### 5. Discussion

Several areas of building performance research in the last three decades have relied upon the approximation of typical daily, weekly and seasonal patterns of energy use. The rapid increase in the availability and quality of raw measured data from the built environment enables the wider use of such patterns for various performance applications. In this paper, we questioned whether the extraction of daily performance patterns can impact the way buildings are labeled for the purpose of benchmarking. The result indicates that 94% of buildings can be grouped by three dominant load profiles. In addition, since buildings share these dominant load profiles, we consider
them fundamental, and conclude that three fundamental profiles can be used as load shape characteristics of buildings.

Our proposed method clearly differentiates between residential and non-residential buildings by their energy consumption patterns. Most of residential buildings are grouped into $G_3$, which is an evening peak profile, whereas non-residential buildings are divided into the two groups, $G_1$ and $G_2$, with noon and morning peak respectively. This is because the energy consumption of residential buildings is largely determined by occupant behavior, i.e., the occupant’s building system and appliance usage [71]. Although occupant behavior is also an important factor to understand energy consumption in non-residential buildings, most of education or governmental facilities have predefined schedules to operate buildings. One may, therefore, rightfully question whether it would make sense to separate residential from non-residential buildings to perform our analysis. We opted to keep them together to reinforce the fundamental nature of our results. However, we have performed the same clustering analysis for the two separated datasets and show it in the Appendix. In both cases, $k=2$ emerges as a good value for the number of clusters, suggesting that two load shape profile exist in each case, which is not so different from $k=3$ in the combined case. Further research is necessary to investigate the differences and opportunities in separating the two major use types.

Although our dataset is one of the most diverse that has been analyzed so far in literature, it may still be biased. For example, most of the residential buildings are from the CER Smart Meter Data [50], from the same location (Cork/Ireland), in one particular climate zone, and might result in similar occupant behavior. With other residential datasets, we can further investigate occupant behavior patterns. In addition, non-residential buildings are also mainly from educational buildings, i.e., university campuses, which might have similar predefined schedules by the facility management. Since the datasets are public, and our proposed clustering framework is open source [70], we invite researchers to add to our dataset and reanalyze the clustering results to improve the robustness of our approach.

Various parameters can affect the results of the clustering and subsequent benchmarking analysis. First, the temporal resolution of the smart meter data is important. In [72], authors studied the impact of using different temporal resolution meter data (2 minutes to 2 hours) and concluded that 4-60 minutes resolution data is ideal for robust load profiling. We used 60 minute interval in our analysis. In addition, the day type of each daily profile could be considered separately as it is likely that, for example, load profiles of weekdays and weekends are different. Lastly, we used a domi-
nance threshold of 50%. By increasing this number, each group would have a smaller entropy value. However, there would be more buildings assigned to \( G_0 \). The threshold value can be varied based on the purpose of benchmarking.

Conventional building benchmarking systems seek to establish how much better or worse a building performs as compared to its peers. A prominent example is the EnergyStar building rating system in the United States \[37\]. EnergyStar utilizes data collected from the Commercial Building Energy Consumption Survey (CBECS) to create a distribution of performance for typical building typologies. A building is benchmarked by comparing annual consumption normalized by area and schedule. These self-reported schedules are often intended or \textit{best guess} on the part of the operations staff. The opportunity arises for the use of clustered daily profiles to automate the process of establish the \textit{use intensity} of a building beyond self-reported schedules. Future benchmarking systems will likely require submission of hourly or sub-hourly performance data that can be used to automatically establish the use intensity of a building.

5.1. Other Potential Applications

In addition to the application to benchmarking, the clustering of behavior from collected empirical data will be useful for building simulation input analysis, portfolio management, demand response and renewable energy planning and allocation.

5.1.1. Simulation Input

In the same way as benchmarking, daily use patterns are used in the predictive simulation of buildings using tools such as EnergyPlus. These day-type patterns are utilized to establish the status quo of full or partial operation of lighting, heating, ventilation, and air-conditioning systems and to approximate the flow of occupants in and out of the various parts of the building. For example, day-typing is a procedure established by the ASHRAE Research Project 1093 in the late 1990’s to extract standardized load schedules in the form of diversity factors for use in building performance simulation \[73\]. This research has been used extensively since its release as it creates a set of defaults that building professionals often use in the first passes of the simulation process. Novice simulation users often use these defaults without even understanding their impact. These diversity schedules have more influence on typical simulation results than the data set used to create them can justify. Only 46 building were used to develop the various non-residential diversity factor schedule from this project. The results of
this paper illustrate the creation of diversity factors from a much larger set of buildings and this could form the foundation for simulation default schedules.

5.1.2. Portfolio Management

Facilities management of a collection of buildings requires the alignment of operations policies across potentially hundreds of buildings. These policies dictate how buildings should respond to use requirements from the functional needs of the building. University campuses are a good example of an organization that often own and operate numerous buildings and seek to manage energy consumption and keep occupants satisfied. Standard operating schedules are often used to create consistency in these types of organizations, however these policies are not often verified in a data-driven way. Automated fault detection and diagnostics systems are often used to detect these schedule mismatches, however these systems have limited market penetration and are overly sophisticated. The ability to compare the extracted daily and weekly patterns of buildings in a portfolio empowers the automated comparison to standard operating schedules.

Fig. 12 shows such a data-driven portfolio analysis for a random selection of 100 buildings in our dataset grouped according to $G_0$—$G_3$, each dot representing a building colored according to its PSU. From such a graph, building managers can understand their building performances comprehensively, i.e., the fundamental load shape profile and the EUI distribution.

5.1.3. Renewable Energy Integration and Demand Response

Demand response and renewable energy integration are similar challenges in that they rely on the characterization of patterns of use in the time domain. In demand response applications, building owners need to understand the peak regions of energy use across numerous buildings and develop strategies to offset those collective maximums. The ability to characterize the load profiles of buildings in an automated and way facilitates this analysis.

6. Conclusion

In this paper, we investigated the existence of fundamental building load shape profiles using unsupervised machine learning methods, and applied them to a data-driven benchmarking study. With K-means clustering, three
fundamental profiles, i.e., morning, noon, and evening peak energy consumption pattern, are discovered. Calculating the distribution of each clustering assignment, we grouped the buildings with respect to their dominant profiles. We found that 94% of the buildings are assigned to one of the three fundamental profile shapes. This novel grouping result is further compared with a conventional building usage type based benchmarking and has evidenced its potential applications for shaping a sustainable built environment.

Appendix

The appendix shows the clustering results separated for residential (Fig. 13) and non-residential (Fig. 14) buildings, as well as the cluster quality metrics (Fig. 15). These results show that K-means clustering method with $k=2$ provides the best grouping in each case.

References


22. L. Pérez-Lombard, J. Ortiz, R. González, I. R. Maestre, A review of benchmarking, rating and labelling concepts within the framework of


Figure 13: Clustering result for residential buildings
Figure 14: Clustering result for non-residential buildings
Figure 15: Clustering performance metrics from residential buildings (top) and non-residential buildings (bottom)