

# Lawrence Berkeley National Laboratory

## LBL Publications

### Title

Characterizing patterns and variability of building electric load profiles in time and frequency domains

### Permalink

<https://escholarship.org/uc/item/3jr7k7wf>

### Authors

Li, Han

Wang, Zhe

Hong, Tianzhen

et al.

### Publication Date

2021-06-01

### DOI

10.1016/j.apenergy.2021.116721

Peer reviewed



Building Technologies & Urban Systems Division  
Energy Technologies Area  
Lawrence Berkeley National Laboratory

# Characterizing Patterns and Variability of Building Electric Load Profiles in Time and Frequency Domains

Han Li<sup>1</sup>, Zhe Wang<sup>1</sup>, Tianzhen Hong<sup>1</sup>, Andrew Parker<sup>2</sup>, Monica Neukomm<sup>3</sup>

<sup>1</sup>Building Technology and Urban Systems Division, Lawrence Berkeley National Laboratory

<sup>2</sup>National Renewable Energy Laboratory

<sup>3</sup>Building Technologies Office, Department of Energy,

Energy Technologies Area  
June 2021

DOI: 10.1016/j.apenergy.2021.116721



This work was supported by the Assistant Secretary for Energy Efficiency and Renewable Energy,  
Building Technologies Office, of the U.S. Department of Energy  
under Contract No. DE-AC02-05CH11231.

Disclaimer:

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

# Characterizing Patterns and Variability of Building Electric Load Profiles in Time and Frequency Domains

Han Li<sup>1</sup>, Zhe Wang<sup>1</sup>, Tianzhen Hong<sup>1\*</sup>, Andrew Parker<sup>2</sup>, Monica Neukomm<sup>3</sup>

<sup>1</sup>Building Technology and Urban Systems Division, Lawrence Berkeley National Laboratory  
One Cyclotron Road, Berkeley, CA 94720, USA

<sup>2</sup>National Renewable Energy Laboratory, 15013 Denver W Pkwy, Golden, CO 80401, USA

<sup>3</sup>Building Technologies Office, Department of Energy, 950 L'Enfant Plaza SW, Washington, DC 20024, USA

\* Corresponding author: [thong@lbl.gov](mailto:thong@lbl.gov), (+1) 510-486-7082

## ABSTRACT

The rapid development of advanced metering infrastructure provides a new data source—building electrical load profiles with high temporal resolution. Electric load profile characterization can generate useful information to enhance building energy modeling and provide metrics to represent patterns and variability of load profiles. Such characterizations can be used to identify changes to building electricity demand due to operations or faulty equipment and controls. In this study, we proposed a two-path approach to analyze high temporal resolution building electrical load profiles: (1) time-domain analysis and (2) frequency-domain analysis. The commonly adopted time-domain analysis can extract and quantify the distribution of key parameters characterizing load shape such as peak-base load ratio and morning rise time, while a frequency-domain analysis can identify major periodic fluctuations and quantify load variability. We implemented and evaluated both paths using whole-year 15-minute interval smart meter data of 188 commercial office building in Northern California. The results from these two paths are consistent with each other and complementary to represent full dynamics of load profiles. The time- and frequency-domain analyses can be used to enhance building energy modeling by: (1) providing more realistic assumptions about building operation schedules, and (2) validating the simulated electric load profiles using the developed variability metrics against the real building load data.

## Keywords:

Building electrical load profile, smart meter, time-domain analysis, frequency-domain analysis, building energy modeling, load variability

## 1. Introduction

### 1.1 Background

Buildings are a significant energy consumer and carbon emitter. Building Energy Modeling (BEM) plays a significant role in design and operation of energy efficient buildings through right-sizing heating, ventilation, and air conditioning (HVAC); energy retrofit analysis; model predictive controls; and district energy system planning. Two fundamental questions about BEM remain unresolved: (1) how model accuracy can be improved, (2) and how it can be evaluated. With the rapid development of advanced metering infrastructures (AMI), smart meters have been deployed in an increasing number of buildings. The smart meter data with high temporal resolution provide us with new information [1], such as occupant energy behavior and building energy performance [2], in a non-intrusive way [3] that could be valuable to enhance BEM.

Building operation schedules, including HVAC operation, occupancy, lighting, and plug load schedules are key input variables and also major uncertainty sources for building energy analysis [4]. The current widely adopted

approach in BEM is to use over-simplified homogeneous static schedules, ignoring the diversity of space use and time [5]. Bianchi et al. (2020) extracted distributions of occupancy-related operations from smart meter data and applied the extracted schedules to simulate energy consumption of building stocks in Los Angeles. This approach proved to be capable of enhancing BEM accuracy [6].

As the building energy behavior varies significantly due to various factors such as building type, vintage, and local weather, the characteristics of the whole building population might not be obvious. A common practice is to divide the whole building stock into several clusters, and then analyze the characteristics of each cluster with similar patterns. Clustering is usually a first step for smart meter data analysis focused on mining valuable information [7]. Clustering can be done with the raw temporal smart meter data [8] or with features extracted from the raw data. Principal component analysis (PCA) and auto-encoder have been used to reduce the dimensionality and extract key features for load clustering in studies by Koivisto et al. [9] and Varga et al. [10]. One shortcoming of the features extracted from PCA and auto-encoder is their physical implications are difficult to explain. To deal with this challenge, Luo et al. [11] and Haben et al. [12] extracted features with clear implications (such as peak-base load ratio, operating duration, and key load change point) for load clustering.

In addition to the commonly used time-domain analysis, frequency-domain analysis, although not as common, provides another angle to quantify the building load profiles. While the time-domain features show how a load profile changes over time, frequency-domain features reveal the power spectrum of the load profile over a range of frequencies. The strengths of frequency-domain features include (1) extracting periodic patterns with the frequency and amplitudes, and (2) dimension reduction of the data after the time- to frequency-domain transformation. The applications of frequency-domain analysis for the smart meter data include several categories. The first category of application involves load-profile clustering. Zhong and Tan (2015) [13] and Kazaki and Papadopoulos (2018) [14] used Fast Fourier Transformation (FFT) to extract characteristics attributes in frequency domain, and then used the extracted attributes for clustering. The second category of applications involves system and operation pattern identification. Bier et al. used frequency-domain features and an artificial neural network (ANN) to detect appliance types and operation events in residential buildings [15]. Chalmers et al. proposed a method to disaggregate smart meter data and classify hospital patients' daily routine behaviors using frequency-domain features [16]. The third category of application involves load monitoring and abnormal detection. Wrinch et al. analyzed smart meter data in the frequency domain with a weekly traveling time window to detect anomalies such as inappropriately configured thermostat setback [17]. Frequency-domain features are also utilized to identify disturbances that could affect the quality of electricity that is delivered to consumers [18], [19]. Despite those efforts, there is a lack of research and application of frequency-domain analysis of smart meter data to improve BEM.

## 1.2 Objectives

Smart meter data analytics could be categorized into three types: (1) descriptive (what do the data look like), (2) predictive (what is going to happen), and (3) prescriptive (what decisions can be made) [20]. This study focuses on the descriptive analysis of smart meter data, which is usually the first step and prerequisite for further analysis.

Time-domain and frequency-domain analyses have different focuses and strengths. Time-domain analysis is more capable of capturing the general trend of daily profiles such as when the load starts to change. The frequency-domain analysis is more capable of capturing the load's periodic variabilities. Though previous studies applied either a time-domain or frequency-domain approach to analyze smart meter data, these two approaches have rarely been conducted and validated by one single dataset. Also, there is a lack of application using insights from both domains to enhance building energy modeling at both single building and building stock scales.

In this study, we proposed a new load-profiling approach that combines the analyses from both the time and frequency domains. We then demonstrated the approach with a dataset of actual smart meter data from office buildings. The goal of this smart meter data study was to (1) infer assumptions for more accurate building energy modeling, and (2) define and evaluate metrics to quantify the load profile variabilities. The major contribution of this study is that it analyzed the building load profiles from both the time and frequency domains. We implemented and compared these two different approaches and identified how they could be used for various purposes.

The remainder of this paper is organized as follows. Section 2 introduces the analytical framework, including the data pre-processing (Section 2.1), the metrics and workflows of time (Section 2.2), and the frequency domain analysis (Section 2.3). Section 3 presents the analytical results, using real AMI data (Section 3.1) as an example to illustrate what information can be extracted from the time (Section 3.2) and frequency (Section 3.3) domain analysis. Section 4 first compares the two analytical approaches (Section 4.1), then discusses how the results could be used to enhance BEM (Section 4.2) and summarizes the study’s contribution and limitations. Section 5 draws conclusions and proposes future research.

## 2. Method

In this study, we tried to characterize building electric load profiles from both the time domain and frequency domain. Figure 1 shows the overall workflow.

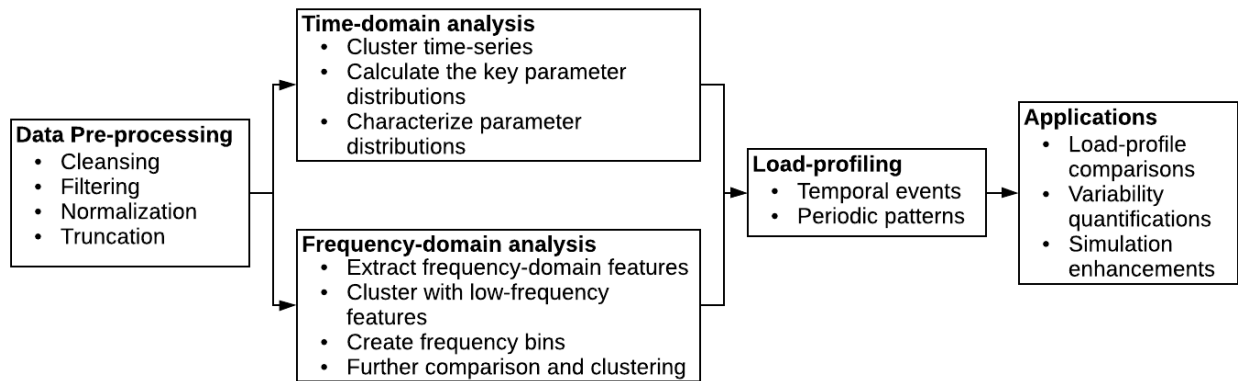


Figure 1. The enhanced load-profiling approach integrating the time- and frequency-domain analysis

### 2.1 Load profile pre-processing

Data pre-processing is usually the first step in real-world data analytics. In general, the goal of pre-processing includes: (1) detecting and correcting unrealistic records, such as missing values and outliers via data cleansing; and (2) down-selecting and reformatting raw data to fit them for specific purposes via data editing and reduction. The research objective of this study was to characterize electric load profiles with a focus on short-term (i.e., hourly to daily) variations. We performed the following activities before further analysis:

- **Data Cleansing:** Electric meter data usually contain corrupted data due to sensor, data storage, or other hardware and software failures. With a preliminary quality check, we found that occasional extreme values were the most common issue with our electric load profiles. In this case, values that are more than three standard deviations away from the mean were considered outliers and were replaced with the mean value.
- **Data Filtering:** The quality and availability of information vary from building to building in real electric meter datasets. For instance, load profiles may have different durations. Some buildings might have building type, location, and floor area information available, while others do not. It is necessary to filter the raw data so all the instances in the subset have the same level of information. In this study, we extracted an entire year’s data for each building, and only kept the date, time, and consumption information. Building type was implicitly kept because all buildings in the dataset were offices.
- **Normalization:** Depending on the type and size of the building, electric load profiles can vary significantly in terms of their absolute values. Therefore, normalization was needed to allow comparisons among load profiles with different magnitudes. Building floor area and the load-profile maximum value are two popular denominators for normalization. In this study, since the building floor area information was missing, we used the peak value of each load profile to normalize them.
- **Truncation:** Different information is associated with load profiles that have different durations. For instance, an annual load profile might reveal seasonal patterns of energy consumption, while a weekly load profile might reveal day-to-day energy consumption variations. In this study, we were interested in

the short-term variations on a daily basis. It was not reasonable to have the whole year’s load profiles as the time- and frequency-domain analysis inputs, as there would be too many timestamp features for further clustering analysis. Therefore, we divided the load profiles into daily chunks. As the smart meter data for each building covers a whole year, this division results in 365 daily profiles for each building. We did not differentiate seasons or consider other factors such as working vs. non-working days.

The details about the data used in this study are described in Section 3.

## 2.2 Characterize building load profiles in the time domain

### 2.2.1 Key parameters

The building electricity loads demonstrated a clear periodic behavior. For office buildings, building load rises in the morning (morning ramp up), peaks around noon, starts to decrease in the afternoon (afternoon setback), and returns to the base load at night. To characterize the building load shape, we learned from the previous study conducted by Price [21] using nine key parameters: two loads (base load, peak load); four times (morning rise, high-load start, high-load finish, afternoon fall finish); and three time intervals (rise time, high-load duration, fall time), as shown in Figure 2 and defined in Table 1.

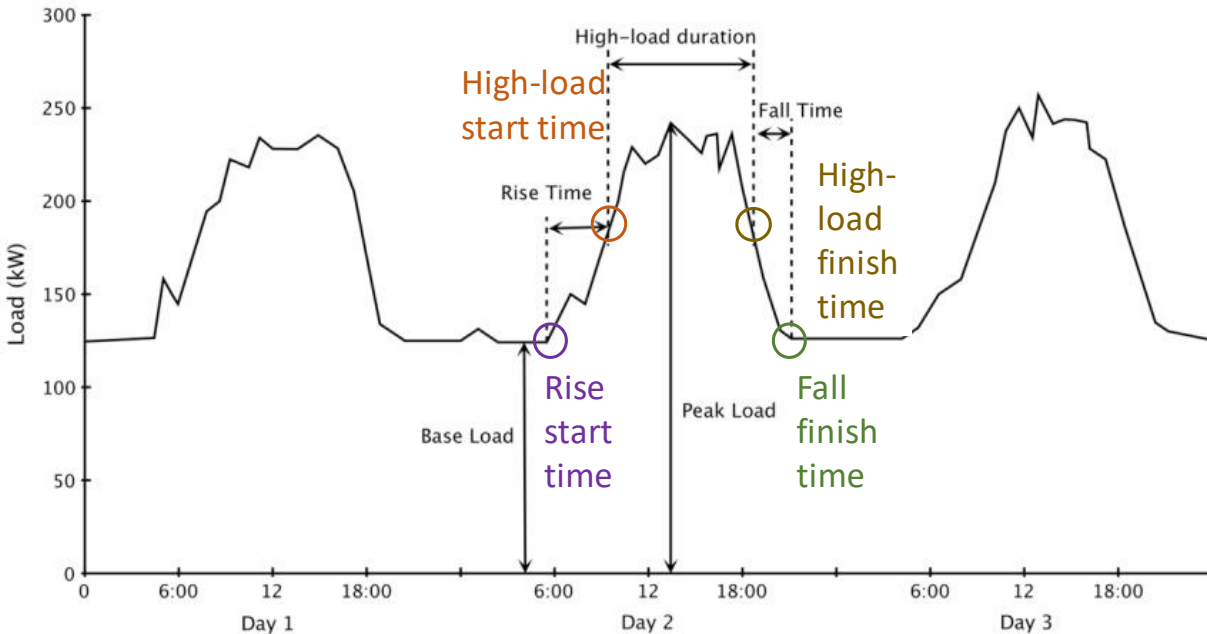


Figure 2: Key parameters to characterize building load shape from the time domain

Table 1: Key parameters to characterize the building load shape from the time domain

Parameters	Definition
Peak load	97.5 percentile of daily load
Base load	2.5 percentile of daily load
Rise start time	The latest time in the morning when the load is less than: $\text{base load} + 0.05 * (\text{peak load} - \text{base load})$

High load start time	The earliest time during each day when the load is more than halfway to the 97.5 percentile load for the day
High load finish time	The latest time during each day when the load is more than halfway to the 97.5 percentile load for the day
Fall finish time	The earliest time in the afternoon when the load is less than: $\text{base load} + 0.05 * (\text{peak load} - \text{base load})$
Rise time	The duration it takes for the load to increase halfway to the peak load; the time interval between rise start time and high load start time
High load duration	The duration the load stays above the halfway mark; the time interval between the high load start time and high load finish time
Fall time	The duration it takes the load to fall from the halfway point back to the base load; the time interval between the high load finish time and fall finish time

As shown in Table 1, we used the 97.5 and 2.5 percentile of daily load to define the peak and base load, because it is possible that in some buildings, the very highest 15-minute data point is substantially higher than any other data point, which could be an outlier or due to some extreme events. Excluding 2.5% extreme values could result in both more stable and more relevant values, rather than the absolute maximum and minimum values [21].

Based on the base load and peak load, we defined the four time points that are critical to describe the building load curve. Rise start time characterizes when the building load starts to increase in the morning, which might be due to the operator turning on building services such as air conditioning. High load start time describes when the building enters full-load operation. High load finish time and fall finish time define when the building load starts to decrease and return to the base load, respectively. Built upon the four time points, we defined three time intervals: rise time, high load duration, and fall time, as illustrated in Figure 2 and Table 1.

### 2.2.2 Workflow

To characterize the building electric load profiles, we proposed the workflow shown in Figure 3.

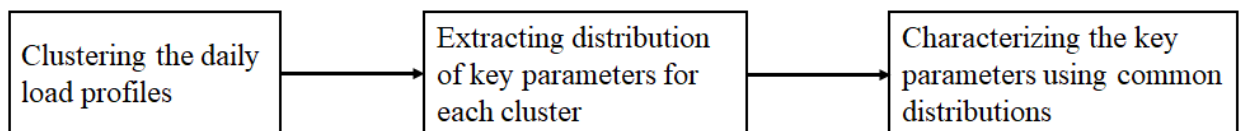


Figure 3: Workflow to characterize building load profiles from the time domain

The first step was to divide the building load profiles into different clusters. Building loads have significantly different profile patterns, for instance between working day and non-working day. Mixing different patterns together might deliver meaningless analytical results. Therefore, the first step of our workflow was to cluster the daily load profiles based on the similarity of the load shape and characterize the load curves for each cluster.

The second step was to extract the distribution of each key parameter for each cluster. As the building load curve can be characterized well by the key parameters introduced above, we use the extracted key parameters to capture the general trends of daily load profiles.

The third step was to use simple distributions (such as normal distribution or uniform distribution) to approximate the true distribution of each key parameter. The motivation for this step was that it can be



challenging to mathematically describe the original distribution. We chose to use simple distributions to approximate the original distributions after considering two factors. First, simple distributions are easier to incorporate into building simulation tools. Second, simple distributions such as the normal distribution are more generalizable and widely observed in many fields. Based on the Large Number Theorem, a normal distribution is capable of representing an event if the event is a sum of independent, identically distributed variables, which is true of some key parameters of load profiles. For instance, the rise start time of a load profile is a result of multiple independent, identically distributed events (e.g., turning on one of the many individual electricity appliances in a building). Once the original distribution is accurately approximated by simple distributions, the result can be easily documented and used by other researchers. For instance, we could use the distribution of rise start time and fall finish time to enhance the assumption of the HVAC operation schedule.

## 2.3 Characterize building load profiles in the frequency domain

### 2.3.1 Frequency-domain feature extraction

Frequency-domain characterization refers to the analysis of a signal's periodic patterns (i.e., cycling and amplitudes) with respect to frequency. The first step was to extract frequency-domain features from the original time-series load profiles. Theoretically, any time-series signal can be converted to a frequency spectrum with a mathematical transform operation. Among different types of operations, the Fourier Transform is the most commonly used method. Suppose the original time-series signal is  $f(t)$ , the Fourier Transform of it can be defined by Equation (1):

$$F(\xi) = \int_{-\infty}^{\infty} f(t)e^{-j2\pi\xi t} dt \quad \text{Equation (1)}$$

Where  $\xi$  is the frequency. The frequency-domain spectrum contains two important features—frequency and magnitude—which describe how the original signal's energy is distributed across different frequencies. Figure 4 illustrates the conversion of a signal from a time domain to a frequency domain.

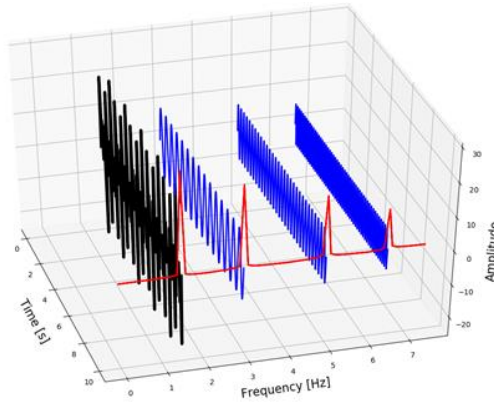


Figure 4. Time domain to frequency domain transformation

For discrete time-series data like the electric load profile, we applied Discrete Fourier Transform (DFT), as defined by Equation (2):

$$\xi_k = \sum_{k=0}^{N-1} f_k e^{-\frac{j2\pi kn}{N}} \quad \text{Equation (2)}$$

Where  $\xi_k$  is the  $k$ th frequency,  $N$  is the total number of observations, and  $n$  is the  $n$ th observation in the time domain. It should be noted that DFT has an inherent limitation when dealing with finite-duration signals. Because DFT assumes the input signal is periodic, when the start and end of the signal are not equal, there will be an abrupt transition which causes unexpected oscillations in the frequency domain. This phenomenon is called *spectral leakage* [22]. Applying a window function to the original signal is a solution to this problem. A window function is a bell-shaped function that equals one at the middle and equals zero outside of the chosen time-series chunk. By multiplying the window function to the original signal, the start and end values of the signal both become zero, which avoid the abrupt transition in the DFT process. Various window functions are available [23], [24]. The Hanning window was used this study, and it could be expressed as Equation (3):

$$w(n) = 0.5 \left( 1 - \cos \cos \left( 2\pi \frac{n}{N} \right) \right) \quad \text{Equation (3)}$$

The DFT and associated feature extraction process is implemented in a Python environment with the SciPy library.

### 2.3.2 Workflow

Figure 5 shows the overall workflow of the frequency-domain load profile characterization.

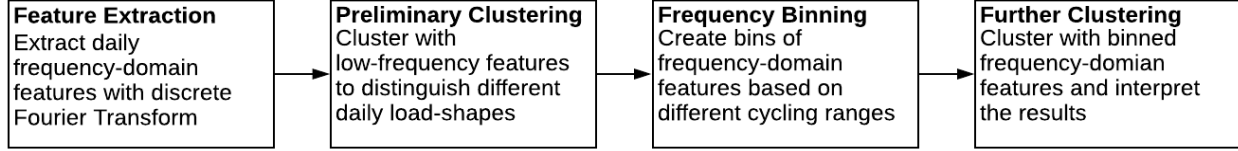


Figure 5. Workflow of frequency-domain load profile characterization

The first step was to extract the frequency-domain features from the time-series data. Since we were interested in the load-profile variations at the daily basis, we divided all the load profiles into daily chunks and then extracted the frequency spectrums for each of them. Then a preliminary clustering with the low-frequency features was performed, to distinguish different daily curves at a high level. Next, the frequency-domain features were grouped into bins by the cycle range (e.g., 15-min to 30-min, 1 h to 2 h, etc.). The frequency features were binned because even though the DFT from daily load profiles yields plenty of unique frequencies, many of them are very close to one other. Binning the features not only groups frequency components with the similar periodic cycles together, but also reduces the dimension of the data, which can benefit the subsequent clustering analysis. Finally, the binned frequency-domain features were clustered. The clustering results can be used to quantify the variabilities of the load profiles and compare different groups of load profiles. It should be noted that the frequency components in the same bin may be caused by different physical events. Inferring the cause behind the frequency components require additional efforts in load disaggregation, which was beyond the scope of this analysis.

## 3. Results

This section will describe how we used the workflow and metrics described in the previous section to quantify the building load profiles and applied this methodology to real smart meter data. We briefly introduce the data source and then present the result of the time-domain and frequency-domain analysis.

### 3.1 AMI data

The smart meter data we used to test our method contains 286 office buildings in Northern California. Each building has 15-minute interval whole-building electricity use data from 2015. The data was provided by a utility company and the measurement accuracy was not specified. However, based on a laboratory test of 156 utility-grade smart meters [25], all met the  $\pm 0.2\%$  accuracy standard established by the meter manufacturer, which also satisfied the CPUC (California Public Utilities Commission) accuracy requirement of  $\pm 2.0\%$ . In total, we had more than 100,000 daily load profiles. However, there were some missing values and non-typical load profiles for those daily load profiles (identified in Section 3.3). After removing the entries with missing values, we had more than 68,000 daily profiles from 188 buildings.

### 3.2 Characterize building load in the time domain

#### *Step 1: Clustering*

We used k-means to cluster the building daily load profile in the time domain. Silhouette score [26] and Within-cluster Sum of Squared Distances (WSSD) were used to select the optimal number of clusters. The number of clusters leading to a higher Silhouette score and a lower WSSD is preferred. We identified three clusters, as shown in Figure 6.

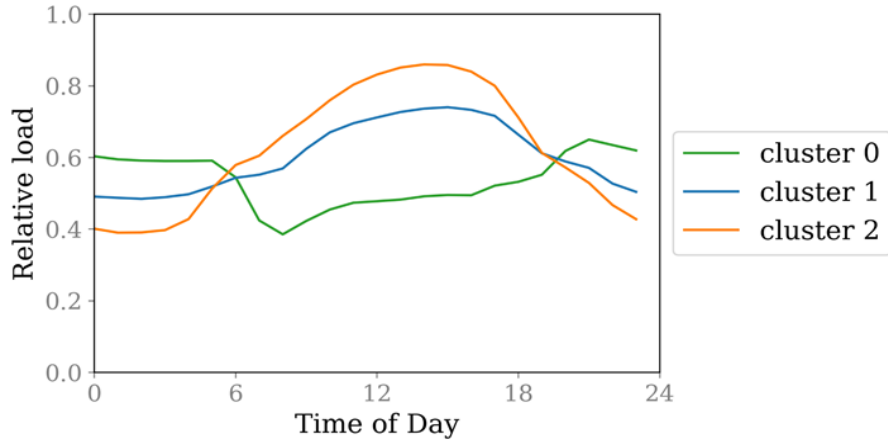
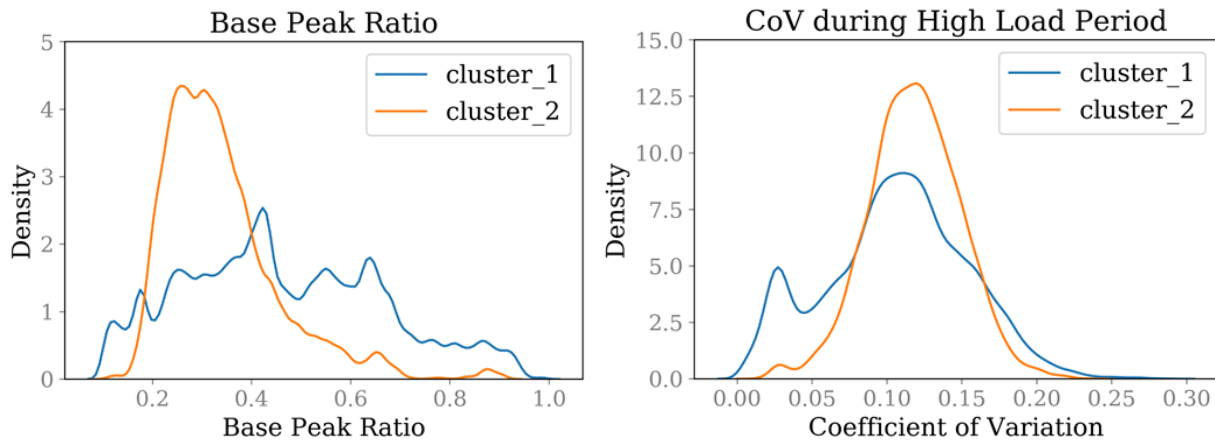


Figure 6: Clustering of electricity load profiles in the time domain

Cluster 0 has high loads during the night time, which might be a result of nighttime exterior lighting consumption and therefore not analyzed later in this study. Cluster 1 and Cluster 2 showed a typical working pattern for office buildings. Among the dataset studied in this paper, 70% of the working days belonged to Cluster 1 and 30% belonged to Cluster 2. The building load rose in the morning and decreased in the late afternoon. As the key statistics introduced in the previous section are proposed for typical working days of office buildings, we only calculated those parameters for Cluster 1 and Cluster 2.

**Step 2: Key parameters**

We extracted the distribution of key parameters for Cluster 1 and Cluster 2, as shown in Figure 7. We were especially interested in six key parameters that can inform more accurate building energy modeling and model validation. The base peak ratio could be used to determine the base load of non-occupied hours. Here, we used a base-to-peak ratio rather than a peak-to-base ratio because the base-to-peak ratio is guaranteed to be in the narrow range of 0 to 1 and therefore easier to depict mathematically. The load coefficient of variation (CoV, defined as the ratio of the load standard deviation to the load mean value) during the high load period quantified the load variability of the target building. The morning rise start, high load start, high load finish, and afternoon fall finish can be used to enhance the assumption of building operation schedules, which are important inputs for building energy modeling.



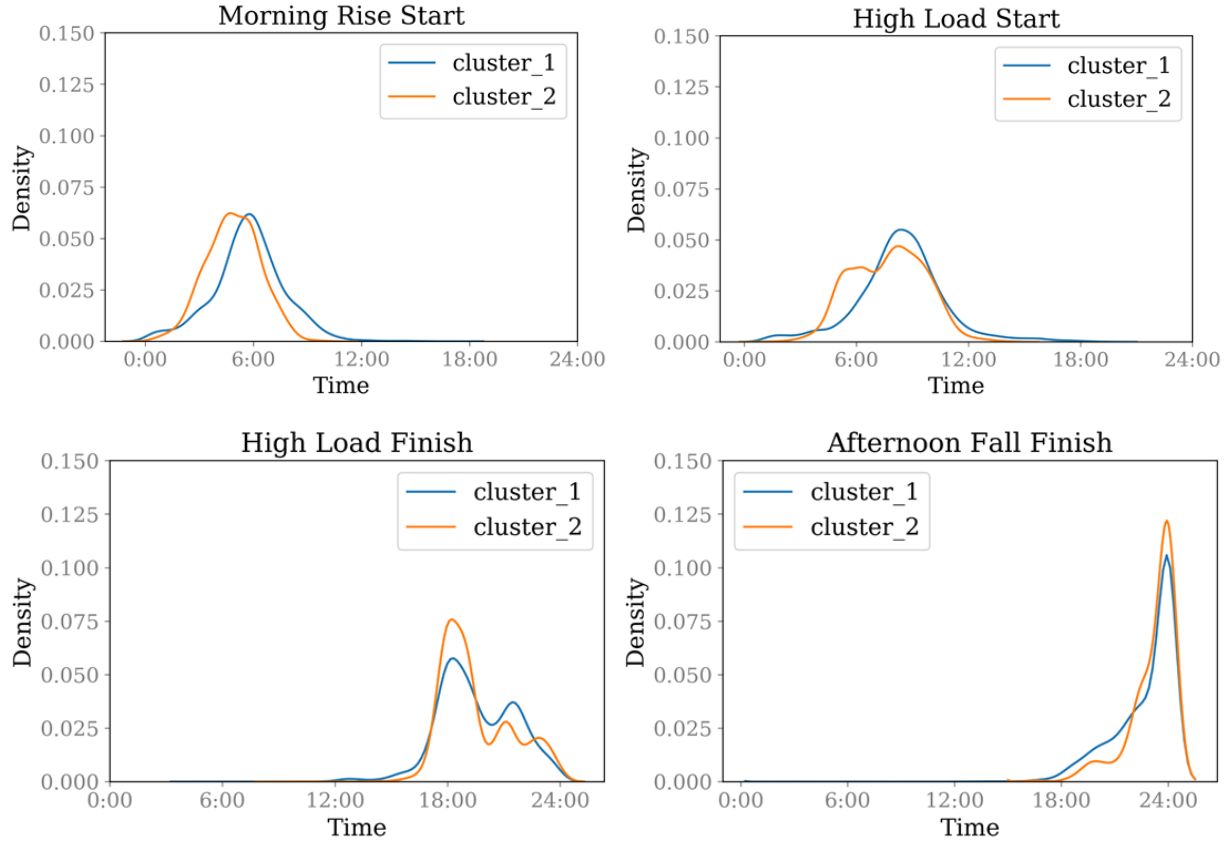
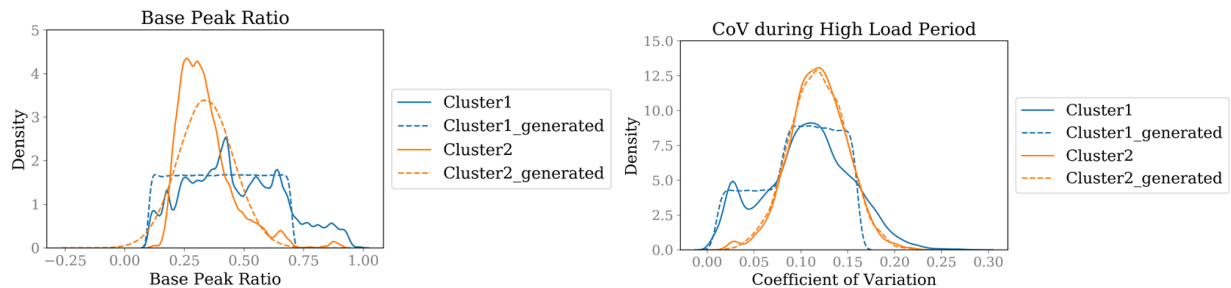


Figure 7: Exact key parameters for each cluster

### Step 3: Quantify the distribution

The final step of the time-domain analysis is to approximate the distribution of key parameters using simple distributions that are commonly used. Quantifying the distribution can help researchers share this knowledge and facilitate its use in other projects.

In this study, we used two simple distributions—normal distribution and uniform distribution—as shown in Figure 8 and Table 2. In Figure 8, the original distributions are shown in solid lines; while the approximated distributions generated from simple distributions are shown in dotted lines. Table 2 shows the mathematical forms of the approximated distributions. Since we limited ourselves to using normal and uniform distributions only, we could not completely ensure that we approximated the original distributions accurately. As shown in Figure 8, the general trends and shapes of each distribution were well captured. The results of Table 2 can be used by other researchers to enhance their building simulation assumptions (such as equipment schedule) or to validate their simulation results.



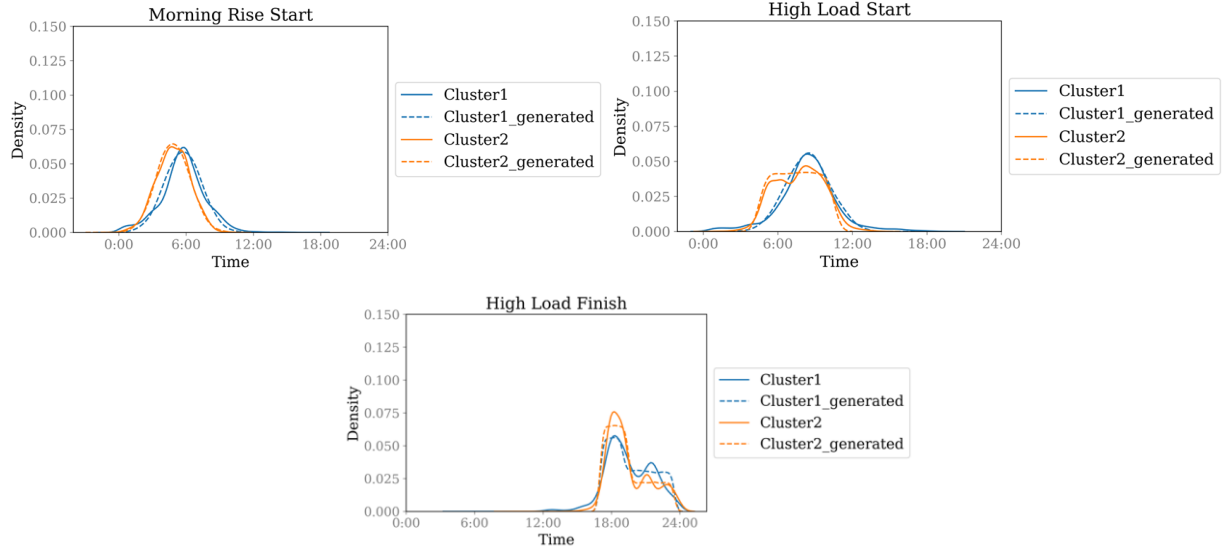


Figure 8: Quantifying the distribution

Table 2: Quantifying the distribution

Parameter	Cluster	Probability
Base peak ratio	Cluster 1	$p(x) = \begin{cases} 1.67, & 0.1 \leq x \leq 0.7 \\ 0, & \text{else} \end{cases}$
	Cluster 2	$p(x) = \frac{1}{0.12\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-0.34}{0.12}\right)^2}$
CoV during high load period	Cluster 1	$p(x) = \begin{cases} 4.28, & 0.01 \leq x \leq 0.08 \\ 8.75, & 0.08 < x \leq 0.16 \\ 0, & \text{else} \end{cases}$
	Cluster 2	$p(x) = \frac{1}{0.03\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-0.12}{0.03}\right)^2}$
Morning rise start time	Cluster 1	$p(t) = \frac{1}{1.6\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-5.75}{1.6}\right)^2}$
	Cluster 2	$p(t) = \frac{1}{1.5\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-4.83}{1.5}\right)^2}$
High load start time	Cluster 1	$p(t) = \frac{1}{1.7\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-8.33}{1.7}\right)^2}$
	Cluster 2	$p(t) = \begin{cases} 0.17, & 4.5 \leq t \leq 10.5 \\ 0, & \text{else} \end{cases}$

High load finish time	Cluster 1	$p(t) = \begin{cases} 0.23, & 17 \leq t \leq 19 \\ 0.12, & 19 < t \leq 23.5 \\ 0, & \text{else} \end{cases}$
	Cluster 2	$p(t) = \begin{cases} 0.26, & 17 \leq t \leq 19.5 \\ 0.09, & 19.5 < t \leq 23.5 \\ 0, & \text{else} \end{cases}$

### 3.3 Frequency-domain analysis results

#### Step 1: Frequency-domain feature extraction

In this step, frequency features were extracted from the 104,390 daily load profiles (286 office buildings x 365 days). A total of 48 frequency features can be extracted from each daily load profile (15-minute interval, 96 timestamps per day) with DFT. Figure 9 shows two examples of the frequency spectrum extracted from the time-series load profiles. The subplots in the top row indicate two daily load profiles, and the subplots in the second row indicate the extracted frequency spectrums. The Example 1 load profile has a smooth daily curve; its consumption started to increase at about 03:00 and reduced to the base load at about 21:00. This daily pattern is reflected by the frequency spectrum, where the low frequency (2.3e-5 Hz, which corresponds to a 12-hour cycle) has a large amplitude. In contrast, in addition to the daily cycle, the Example 2 load profile has many short-term spikes during the 10:00 to 18:00 period. Those variations were captured by the frequency spectrum, too. The frequency-domain features are powerful indicators of the frequencies of short-term cycles and their corresponding magnitudes, which can help us characterize and distinguish load profiles.

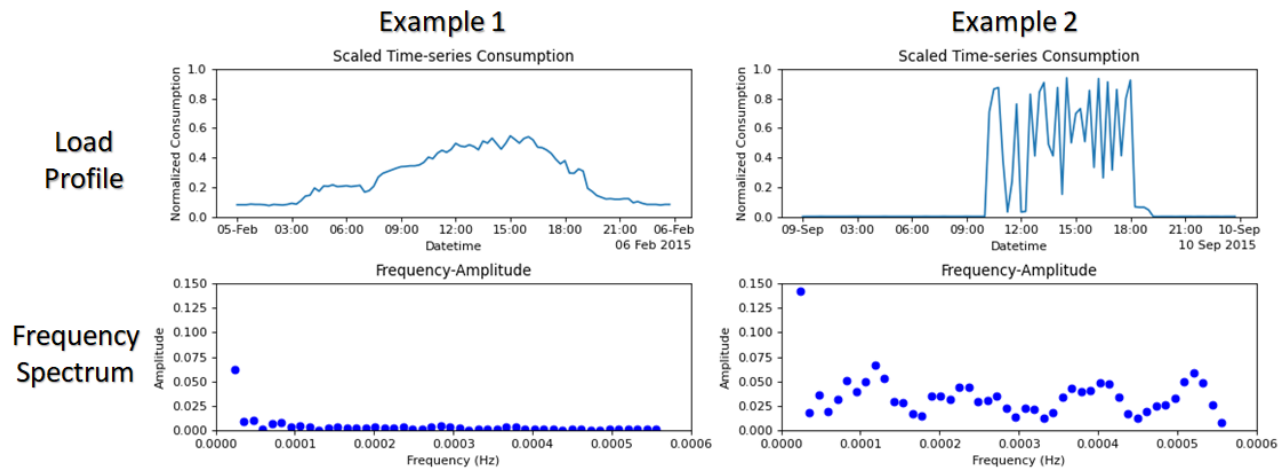


Figure 9. Example frequency spectrums extracted from daily load profiles

#### Step 2: Preliminary clustering

As shown in Section 2.2.1, low-frequency features are good indicators of how the daily load curves look. Through visual inspection of the data, we found some atypical load profiles with high load during nighttime and low load during daytime. To automate this step, three low frequency components (which correspond to 24 h, 12 h, and 8 h cycles) and their amplitudes were extracted from all daily load profiles and used as input parameters for the K-means clustering to identify those atypical load profiles. Similar to the time-domain feature clustering, Silhouette score WSSD were used as the metrics to select the optimal number of clusters. Figure 10 below shows all the load profiles in each of the six clusters identified. Each subplot shows all the daily load profiles in the corresponding cluster, with time of day on the horizontal axis and normalized consumption on the vertical axis.

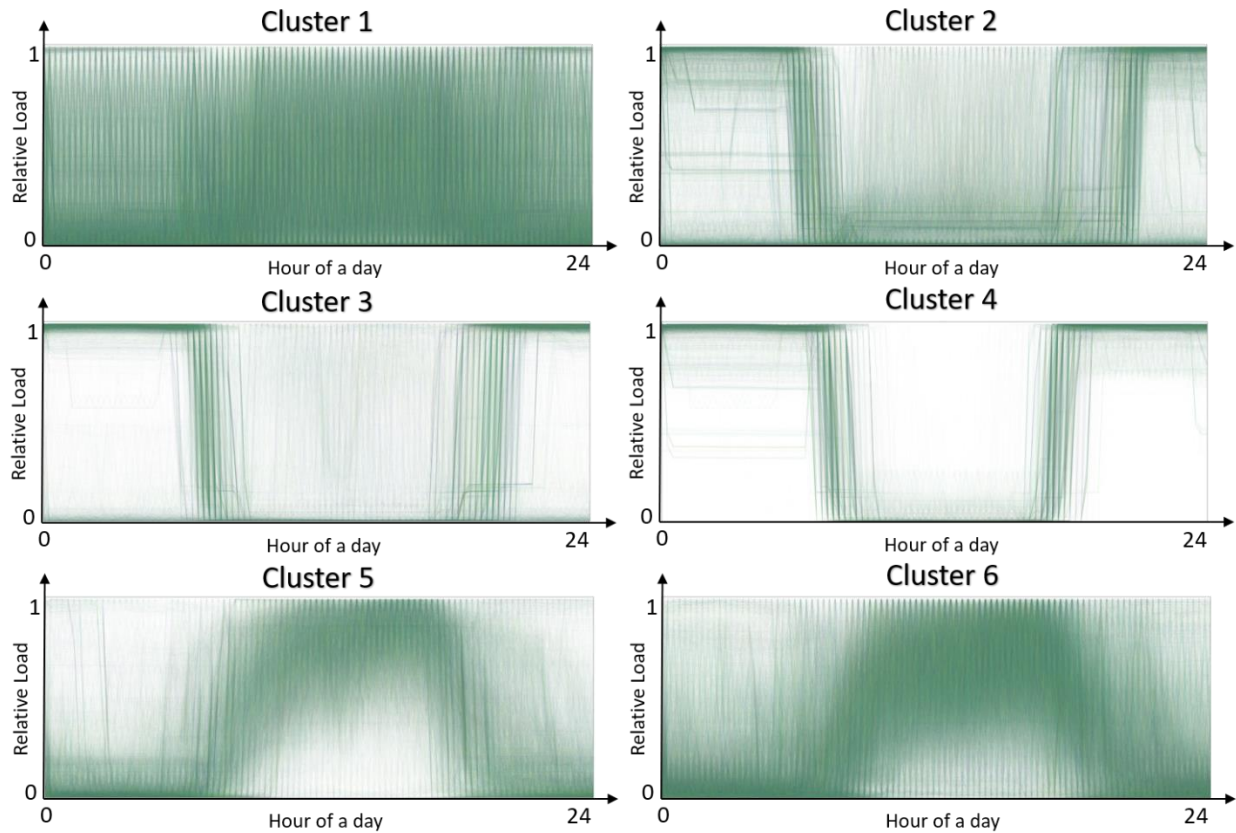


Figure 10. Preliminary clustering results with three low-frequency features

The visualization shows three distinct patterns. First, the load profiles in Cluster 1 appear to occupy the whole subplot because of normalization. Those load profiles have low variability, which corresponds to days with low energy consumption. About 57% of the days in Cluster 1 are weekends and holidays, and 43% of the days are weekdays.

Second, load profiles in clusters 2, 3, and 4 show a pattern of high consumption at night and very low consumption during the day. Annual heat maps were plotted to further investigate why this pattern occurred. Figure 11 below shows two examples of annual load profiles in those three clusters.

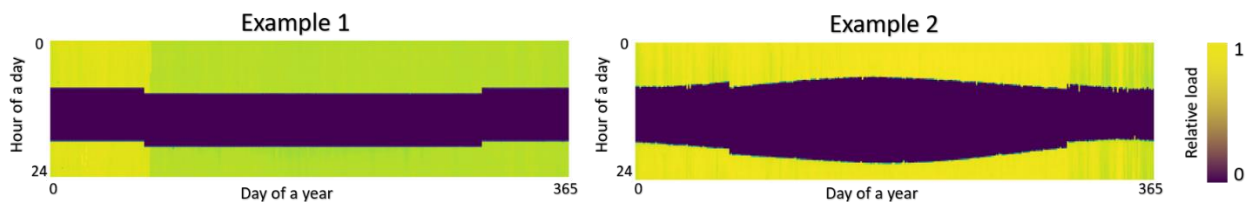


Figure 11. Heat maps of two example load profiles in clusters 2, 3, and 4

In both examples, the dark area corresponds to the low daytime energy consumption and the bright areas correspond to high nighttime energy consumption. There are also one-hour shifts on the daylight-saving start and end dates. Therefore, it can be inferred that those load profiles are exterior lighting or related consumption. In addition, Example 2 shows a change of nighttime high-load duration with seasonal variations, which is probably caused by daylighting controls which automatically control lights based on the changing solar sunset/sunrise time over the course of the year. This explains the different nighttime high-load durations in those three clusters.

Third, the load profiles in clusters 5 and 6 have a high daytime consumption and low nighttime consumption

pattern, and load profiles in Cluster 6 appear to be more variable. In Step 3, we further explore load profiles from clusters 1, 5, and 6.

**Step 3: Frequency spectrum binning**

As discussed in Section 2.2.2, binning the frequency features could group features that are different but have similar physical meaning behind them. It also helps to reduce the dimension of the dataset, which reduces the computation time needed for further analysis. Table 3 below shows the binning of the frequency features. Based on the cycle range, the frequencies were divided into nine bins. Bin 1 to Bin 4 correspond to cycles between 30 minutes to 90 minutes, which are local spikes typically caused by frequent equipment on and off [27]. Bin 5 to Bin 7 correspond to cycles between 1.5 hours to 4 hours, which are short-term cycles such as morning and afternoon high loads. Bin 8 and 9 correspond to cycles between 4 hours to 12 hours, which are determined by the basic daily consumption.

Table 3. Frequency bins

<b>Bins</b>	<b>Frequency Range (Hz)</b>	<b>Cycle Range (hour)</b>	<b>Load Shape</b>	<b>Physical Interpretation</b>
Bin 1	0.00037 to 0.00056	0.5 to 0.75	Local spikes	Equipment on and off, short-term plug-load consumption
Bin 2	0.00028 to 0.00037	0.75 to 1	Local spikes	
Bin 3	0.00022 to 0.00028	1 to 1.25	Local spikes	
Bin 4	0.00019 to 0.00022	1.25 to 1.5	Local spikes	
Bin 5	0.00016 to 0.00019	1.5 to 1.75	Short-term cycling	Loads with relatively constant schedules such as morning and afternoon high loads
Bin 6	0.00014 to 0.00016	1.75 to 2	Short-term cycling	
Bin 7	0.000069 to 0.00014	2 to 4	Short-term cycling	
Bin 8	0.000035 to 0.000069	4 to 8	Daily load curve	Base load such as lighting, non-occupant related electric loads
Bin 9	0.000023 to 0.000035	8 to 12	Daily load curve	

Figure 12 shows two load profiles in a week and their daily binned frequency feature distribution boxplots. The time-series plot shows that Example A has high short-term variabilities, which is captured by Bin 1 (0.5 hour to 0.75 hour) in the boxplot. In comparison, both the mean value and the interquartile range of Bin 1 in Example B are lower than those of Example A, which is because the load profile in Example B has lower working hour demands and lower short-term fluctuations. The examples indicate that the binned frequency features are good metrics for quantifying load-profile variabilities, as they not only indicate the magnitude of the variations, but also pinpoint the frequency and possible reasons for those variations.



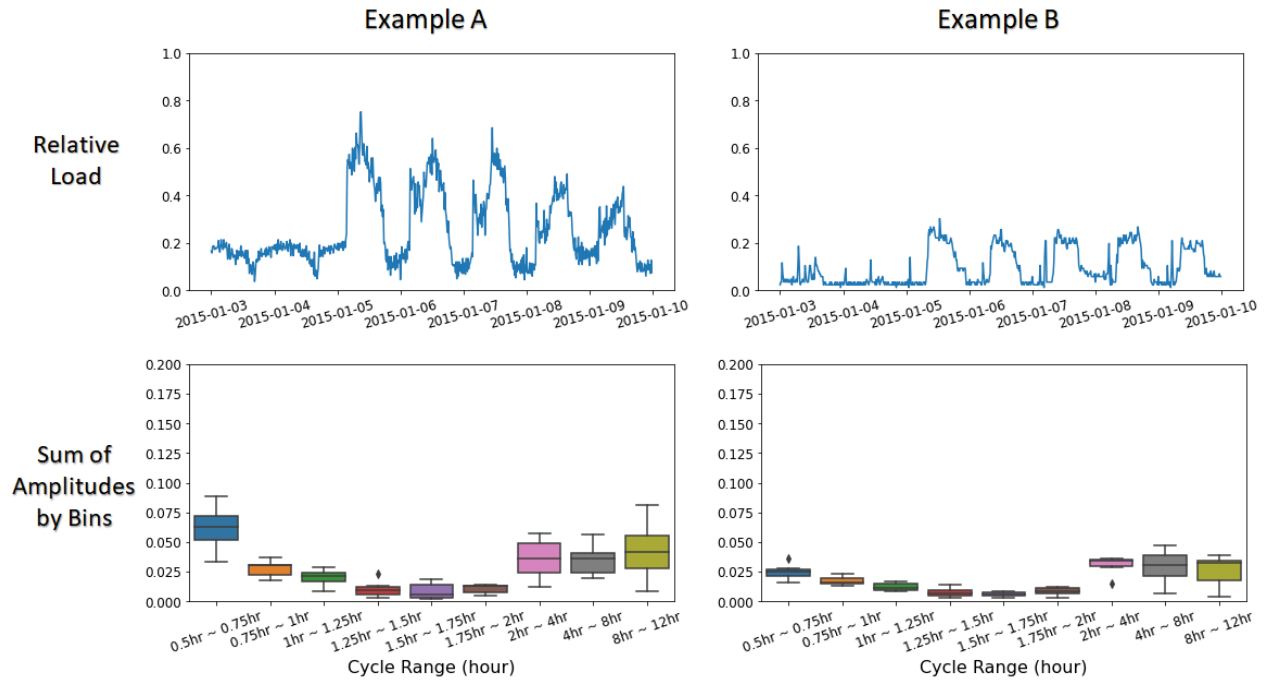


Figure 12. Example load profiles and the binned daily frequency features

**Step 4: Clustering with binned features**

The binned frequency features allow a group of buildings' load-profile patterns to be characterized by clustering them. In the last step, we conducted a second round of K-means clustering with the "typical" daily load profiles in clusters 1, 5, and 6 from Step 2. An optimal number of five clusters were determined from the Silhouette score and WSSD. Figure 13 shows the centers and distributions of the five clusters (0-4).

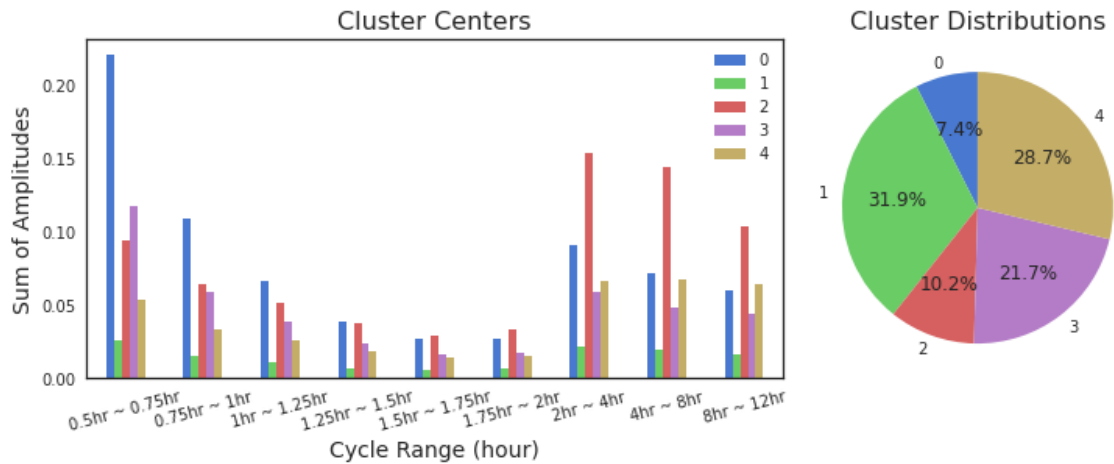
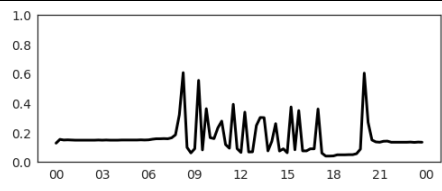
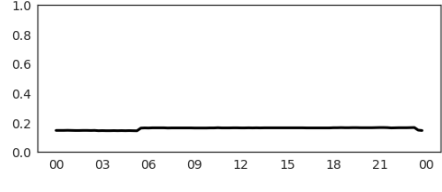
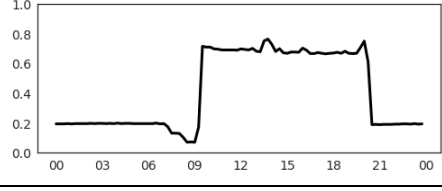
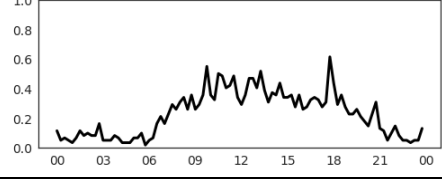
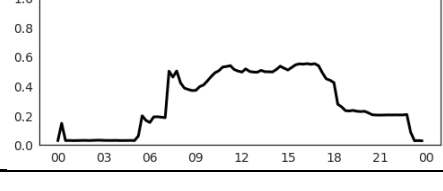


Figure 13. Cluster centers and distribution with the binned frequency features

The typical daily load profiles from each cluster and the interpretations are summarized in Table 4 below.

Table 4. Typical daily load profiles from each cluster

Cluster	Typical Load Profile	Interpretation
Cluster 0		Load profiles with large short-term spikes and relatively low daily base loads. Those are loads that switch on and off frequently, like equipment cycling.
Cluster 1		Load profiles with constantly low consumption during the entire day. These are usually non-working days like weekends and holidays.
Cluster 2		Load profiles with a high daily base load with very small short-term variations. Those are usually buildings with a dominant fixed-schedule electric load.
Cluster 3		Load profiles with normal daily base load and relatively large short-term spikes. Those are usually buildings with some scheduled loads and stochastic (e.g., occupant-related) electric loads.
Cluster 4		Load profiles with a normal daily base load and relatively small short-term spikes. Those are usually buildings with mostly scheduled loads and little stochastic electric loads.

## 4. Discussion

### 4.1 Applications

#### *Time-domain Analysis*

Figure 14 summarizes the outcome of analyzing the building load shape from the time domain. Starting from the raw data, which is difficult to extract useful information from, we proposed a workflow to extract useful information that could be shared with other researchers or used for other purposes.

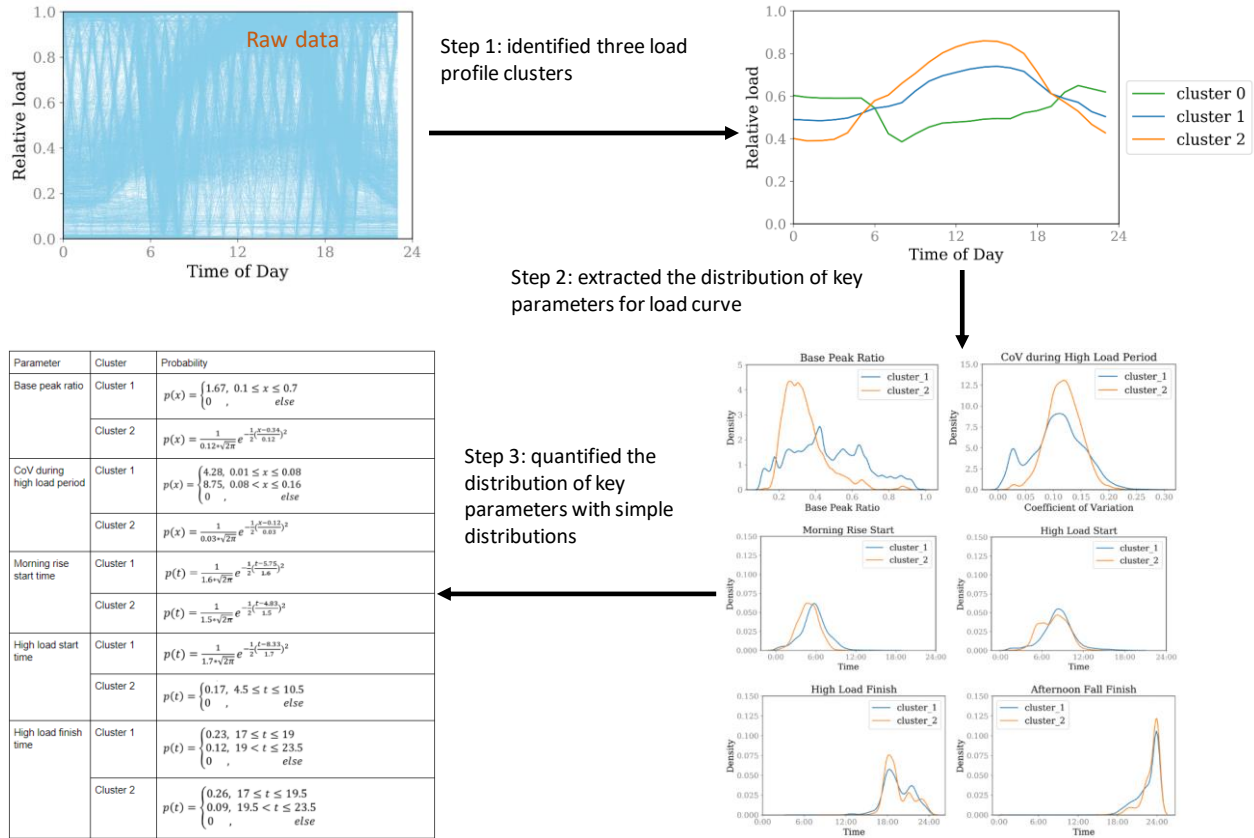


Figure 14: Time-domain analysis workflow and outcomes

Two potential applications were identified based on the results of the time-domain analysis. First, we could enhance the assumptions of building energy modeling. Conventional building energy modeling usually assumes fixed schedules for HVAC operations, which does not reflect the variabilities in real building operations. The distributions of the temporal parameters we extracted from the time-domain analysis could serve as the basis of more realistic HVAC operation or occupancy schedule modeling assumptions. Taking the schedule of an office building on a working day as an example, we could do a random sampling (represented by rolling a dice in Figure 15) to decide whether this building belongs to Cluster 1 or Cluster 2. Then we could sample again from the distribution of morning rise start time of the corresponding cluster. This time could serve as the HVAC operation start time for energy modeling. The whole process is shown in Figure 15.

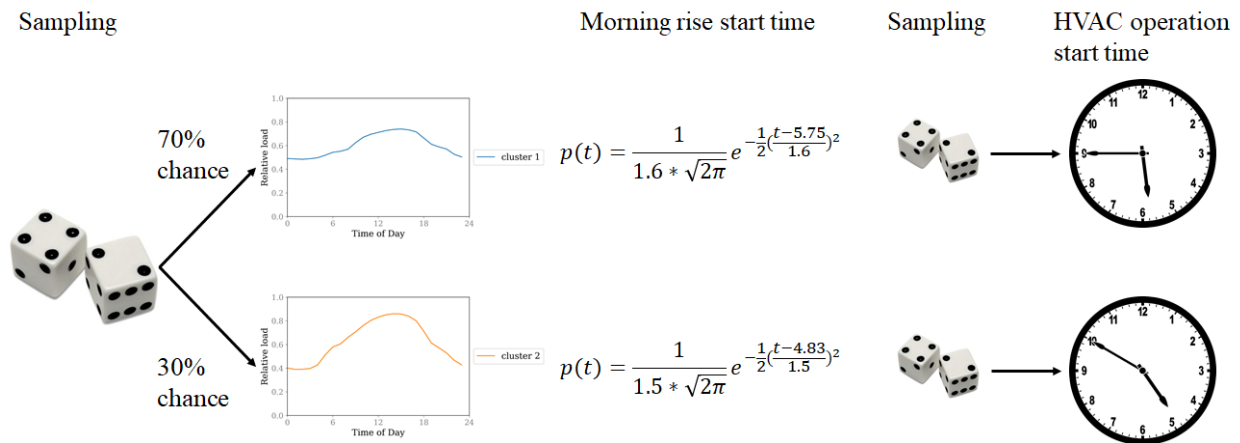


Figure 15: Time-domain analysis application 1: enhance building energy modeling

The second application is that we could use the result to validate building energy simulation models. Due to the randomness of building operation, it is virtually impossible that the simulated building loads would be exactly the same as the actual loads for each time step. By extracting key building load parameters and quantifying the distribution of those key parameters from real building data, as we did in this study, we could compare the distribution of key parameters of modeled building load to the distribution we exacted from real building operation data. The building energy modeling would be validated if those key parameters matched the real building data.

### ***Frequency-domain Analysis***

Frequency-domain analysis has been widely applied in signal processing. However, its applications in the building field are still limited. Previous studies have focused on hourly interval data. This study proposed a workflow to extract the frequency-domain features, reorganize them into bins, and cluster the load profiles. The workflow has been tested on real commercial building electric load profiles. There are two potential applications of the frequency-domain analysis. First, it could be used to describe periodic electricity consumption patterns of a building. As discussed in Section 3.3, a daily load profile's fluctuation can be divided into local spikes, short-term cycles, and daily base curves. With the binned frequency features, it's easy to tell what dominant periodic patterns a building has. For real buildings, it can be used to identify suboptimal operation patterns like frequent equipment cycling. For simulation calibrations, it can tell us which model assumptions should be improved. For instance, if the simulated load profile has lower high-frequency spikes, we might need to tweak the model to add more stochastic dynamics.

Second, the frequency-domain features could be used as a new metric to compare load profiles at both single building or building stock levels. At the single building level, we could use the binned frequency feature distributions to check if two buildings have similar amplitude values at different cycling ranges. At the building stock level, we could first cluster the load profiles with the frequency features and then check if two building stocks have similar cluster components and cluster centers. With this new metric, we can check if two real building stocks have similar load profile components. We can also validate to determine if the simulated building stock load profile variabilities are similar to those of a real building stock.

These potential applications are to be tested in future research. Authors are using these time- and frequency-domain analysis approaches in a project to verify the simulated load profiles from physics-based EnergyPlus modeling against the actual load profiles for commercial buildings at regional scale.

### **4.2 Comparisons and complements**

As discussed above, both time-domain and frequency-domain feature analysis can provide unique insights into building electric load profiles. Time-domain analysis is good at identifying key temporal events, such as the high-load start and end timestamp, and the high-load durations. Frequency-domain analysis is good at identifying the amplitudes and durations periodic patterns. However, either one is limited if it is used alone. For instance, assume we have two daily load profiles, as shown in Figure 16 below. Time-domain analysis might consider them to be the same type, while frequency-domain could distinguish between them by the durations and amplitudes of the periodic patterns. On the other hand, time-domain analysis could easily identify the start and end time of the high-load, while a frequency-domain analysis could not.

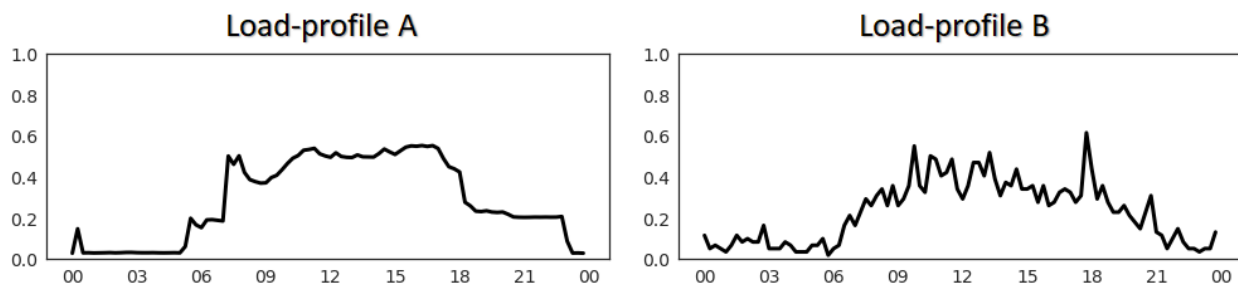


Figure 16: Two example daily load profiles

Table 5 describes the comparisons of the time-domain and frequency-domain analysis in terms of data requirements, load-profile characterization, and potential applications.

Table 5. Comparison of time-domain and frequency-domain load profiling

Category	Feature	Time Domain	Frequency Domain
<b>Data Requirements</b>	Duration	Load profiles need to be truncated into daily chunks to identify key timestamps.	Load profile truncation is not mandatory for detecting periodic patterns.
	Temporal resolution	Lower resolution (e.g., one hour) is sufficient for quantifying a basic daily curve, while higher resolution (e.g., 15-min) is needed to pinpoint the key events.	High resolution (e.g., 15-min or more granular) is needed to identify high-frequency load patterns.
<b>Load Profile Characterization</b>	Characteristics identification	Key parameters such as high-load start and end timestamp, duration, and high-load variations.	Periodic load patterns, including the durations, amplitudes, and phases.
	Variability quantification	Via the high-load variations.	Via original and binned frequencies and amplitudes.
<b>Potential Applications</b>	Enhance building energy simulations	Improve model assumptions with inferred operation schedules (e.g., start and end time of the HVAC system).	Identify whether the stochastic dynamics assumptions (equipment on/off, occupant-related system operations) of the load profile are reasonable.
	Load-profile comparison	Compare the six key parameters.	Compare the frequency and amplitudes distributions.

Another interesting finding is that the results of the time- and frequency-domain analyses were consistent with each other. For instance, Figure 6 presents the clustering results of the time-domain features, while Figure 10 presents the clustering results of frequency-domain features. The Cluster 0 identified from time-domain clustering corresponds to the clusters 2, 3, and 4 identified from the frequency domain. The clusters 1 and 2 identified from time-domain clustering could be mapped to the clusters 5 and 6 identified from the frequency-domain domain. This mapping relationship confirms that these two different approaches are consistent with each other.

The strengths of time-domain and frequency-domain can be combined for some applications. For instance, to validate an energy simulation model, the time-domain analysis can ensure the general trend of the simulated load profile matches the real building smart meter data, while the frequency-domain analysis can be used to check the periodic variation patterns.

#### 4.3 Limitation and Contributions

A limitation of this study is we do not have information about the measurement uncertainty of the smart meter dataset provided by the local utility company. The impact of measurement uncertainty on the time-domain analysis is not significant as the time-domain analysis aims to capture the general trends of daily load profile and ignores small variations. As for the frequency-domain analysis, measurement errors, depending on their

occurrence patterns, may increase the high-frequency variation components but have limited impacts on the low-frequency variation components.

A novel contribution of this study is enhanced building electric load profiling with the strengths of both domains. This approach allows users to extract the key parameters and quantify the variabilities from a more comprehensive perspective. The source code of the enhanced load profiling method will be available open-sourced at GitHub for users to adopt.

## 5. Conclusion

The rapid development of advanced metering infrastructure provides a vast amount of building load data at high temporal resolutions. Analyzing building load profile data could generate useful information for building energy modeling. This study proposed a two-path (i.e., time-domain and frequency-domain) method to characterize building electric load profiles at the daily duration basis. The time-domain path focuses on extracting the key temporal parameters and their distributions, while the frequency-domain path focuses on extracting the key periodic patterns and their distributions.

The method was tested with a dataset composed of 15-minute interval load profiles from 188 commercial office buildings for an entire year. We demonstrated the workflow of each path and their applications. The time-domain analytics clusters the load profiles, extracts key parameter distributions, and then quantifies the distributions of those key parameters. The frequency-domain path could distinguish the basic daily load curves, extract amplitudes for different frequencies, and quantify periodic fluctuations. The information extracted from both paths together form a set of metrics for enhanced load profile characterization. The findings from the study are helpful for (1) quantifying the variabilities of a load profile, (2) comparing two groups of load profiles in terms of temporal and frequency patterns, and (3) improving building energy modeling accuracy by providing more realistic assumptions about building operation schedules.

Future work can further evaluate the time- and frequency-domain characterization method using electric load profile data from other building types and climate zones, and adopt the method to benchmark simulated load profiles against actual smart meter data.

## Acknowledgment

This research was supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, Office of Building Technologies of the United States Department of Energy, under Contract No. DE-AC02-05CH11231.

## References

- [1] Jixuan Zheng, D. W. Gao, and Li Lin, "Smart Meters in Smart Grid: An Overview," in *2013 IEEE Green Technologies Conference (GreenTech)*, Denver, CO, Apr. 2013, pp. 57–64, doi: 10.1109/GreenTech.2013.17.
- [2] I. P. Panapakidis, T. A. Papadopoulos, G. C. Christoforidis, and G. K. Papagiannis, "Pattern recognition algorithms for electricity load curve analysis of buildings," *Energy Build.*, vol. 73, pp. 137–145, Apr. 2014, doi: 10.1016/j.enbuild.2014.01.002.
- [3] D. Benyoucef, P. Klein, and T. Bier, "Smart Meter with non-intrusive load monitoring for use in Smart Homes," in *2010 IEEE International Energy Conference*, Dec. 2010, pp. 96–101, doi: 10.1109/ENERGYCON.2010.5771810.
- [4] Z. Wang, T. Hong, and R. Jia, "Buildings.Occupants: a Modelica package for modelling occupant behaviour in buildings," *J. Build. Perform. Simul.*, vol. 12, no. 4, pp. 433–444, Jul. 2019, doi: 10.1080/19401493.2018.1543352.
- [5] S. D'Oca, T. Hong, and J. Langevin, "The human dimensions of energy use in buildings: A review," *Renew. Sustain. Energy Rev.*, vol. 81, pp. 731–742, Jan. 2018, doi: 10.1016/j.rser.2017.08.019.
- [6] C. Bianchi, L. Zhang, D. Goldwasser, A. Parker, and H. Horsey, "Modeling occupancy-driven building loads for large and diversified building stocks through the use of parametric schedules," *Appl. Energy*, vol. 276, p. 115470, Oct. 2020, doi: 10.1016/j.apenergy.2020.115470.
- [7] K. Zhou, S. Yang, and C. Shen, "A review of electric load classification in smart grid environment," *Renew.*

- Sustain. Energy Rev.*, vol. 24, pp. 103–110, Aug. 2013, doi: 10.1016/j.rser.2013.03.023.
- [8] R. Granell, C. J. Axon, and D. C. H. Wallom, “Impacts of Raw Data Temporal Resolution Using Selected Clustering Methods on Residential Electricity Load Profiles,” *IEEE Trans. Power Syst.*, vol. 30, no. 6, pp. 3217–3224, Nov. 2015, doi: 10.1109/TPWRS.2014.2377213.
- [9] M. Koivisto, P. Heine, I. Mellin, and M. Lehtonen, “Clustering of Connection Points and Load Modeling in Distribution Systems,” *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1255–1265, May 2013, doi: 10.1109/TPWRS.2012.2223240.
- [10] E. D. Varga, S. F. Beretka, C. Noce, and G. Sapienza, “Robust Real-Time Load Profile Encoding and Classification Framework for Efficient Power Systems Operation,” *IEEE Trans. Power Syst.*, vol. 30, no. 4, pp. 1897–1904, Jul. 2015, doi: 10.1109/TPWRS.2014.2354552.
- [11] X. Luo, T. Hong, Y. Chen, and M. A. Piette, “Electric load shape benchmarking for small- and medium-sized commercial buildings,” *Appl. Energy*, vol. 204, pp. 715–725, Oct. 2017, doi: 10.1016/j.apenergy.2017.07.108.
- [12] S. Haben, C. Singleton, and P. Grindrod, “Analysis and Clustering of Residential Customers Energy Behavioral Demand Using Smart Meter Data,” *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 136–144, Jan. 2016, doi: 10.1109/TSG.2015.2409786.
- [13] S. Zhong and K.-S. Tam, “Hierarchical Classification of Load Profiles Based on Their Characteristic Attributes in Frequency Domain,” *IEEE Trans. Power Syst.*, vol. 30, no. 5, pp. 2434–2441, Sep. 2015, doi: 10.1109/TPWRS.2014.2362492.
- [14] A. G. Kazaki and T. A. Papadopoulos, “Cluster analysis of university campus smart meter data,” in *2018 53rd International Universities Power Engineering Conference (UPEC)*, Sep. 2018, pp. 1–6, doi: 10.1109/UPEC.2018.8541941.
- [15] T. Bier, D. O. Abdeslam, J. Merckle, and D. Benyoucef, “Smart meter systems detection & classification using artificial neural networks,” in *IECON 2012 - 38th Annual Conference on IEEE Industrial Electronics Society*, Montreal, QC, Canada, Oct. 2012, pp. 3324–3329, doi: 10.1109/IECON.2012.6389365.
- [16] C. Chalmers, P. Fergus, C. A. Curbelo Montanez, S. Sikdar, F. Ball, and B. Kendall, “Detecting Activities of Daily Living and Routine Behaviours in Dementia Patients Living Alone Using Smart Meter Load Disaggregation,” *IEEE Trans. Emerg. Top. Comput.*, pp. 1–1, 2020, doi: 10.1109/TETC.2020.2993177.
- [17] M. Wrinch, T. H. M. EL-Fouly, and S. Wong, “Anomaly detection of building systems using energy demand frequency domain analysis,” in *2012 IEEE Power and Energy Society General Meeting*, Jul. 2012, pp. 1–6, doi: 10.1109/PESGM.2012.6344790.
- [18] F. A. S. Borges, R. A. S. Fernandes, I. N. da Silva, and C. B. S. Silva, “Feature Extraction and Power Quality Disturbances Classification Using Smart Meters Signals,” *IEEE Trans. Ind. Inform.*, 2016, doi: 10.1109/TII.2015.2486379.
- [19] I. PARVEZ, M. AGHILI, A. I. SARWAT, S. RAHMAN, and F. ALAM, “Online power quality disturbance detection by support vector machine in smart meter,” *J. Mod. Power Syst. Clean Energy*, vol. 7, no. 5, pp. 1328–1339, Sep. 2019, doi: 10.1007/s40565-018-0488-z.
- [20] Y. Wang, Q. Chen, T. Hong, and C. Kang, “Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges,” *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 3125–3148, May 2019, doi: 10.1109/TSG.2018.2818167.
- [21] P. Price, “Methods for Analyzing Electric Load Shape and its Variability,” Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States), LBNL-3713E, May 2010. doi: 10.2172/985909.
- [22] D. Lyon, “The Discrete Fourier Transform, Part 4: Spectral Leakage,” *J. Object Technol.*, vol. 8, no. 7, p. 23, 2009, doi: 10.5381/jot.2009.8.7.c2.
- [23] A. V. Oppenheim, R. W. Schaffer, and with J. R. Buck, *Discrete-time Signal Processing. Vol. 1*. Upper Saddle River [N.J.]: Prentice Hall: imprint of Pearson, 2001.
- [24] S. Rapuano and F. J. Harris, “An introduction to FFT and time domain windows,” *IEEE Instrum. Meas. Mag.*, vol. 10, no. 6, pp. 32–44, Dec. 2007, doi: 10.1109/MIM.2007.4428580.
- [25] Structure Consulting Group. PG&E Advanced Metering Assessment Report. September 2010. <http://www.pge.com/includes/docs/pdfs/myhome/customerservice/meter/smartmeter/StructureReport.pdf>
- [26] P. J. Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis,” *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, Nov. 1987, doi: 10.1016/0377-0427(87)90125-7.
- [27] J. L. Mathieu, P. N. Price, S. Kiliccote, and M. A. Piette, “Quantifying Changes in Building Electricity Use,

With Application to Demand Response,” *IEEE Trans. Smart Grid*, vol. 2, no. 3, pp. 507–518, Sep. 2011, doi: 10.1109/TSG.2011.2145010.