Analyzing the Energy Usage of a Community and the Benefits of Energy Storage

JOHN WAMBURU, IBM Research Africa, Kenya
STEPHEN LEE, University of Pittsburgh, USA
SRINIVASAN IYENGAR, Microsoft Research, India
DAVID IRWIN and PRASHANT SHENOY, University of Massachusetts Amherst, USA

13

Understanding the energy usage of a community is crucial for policymaking, energy planning, and achieving sustainable development. The advent of the smart grid has made it feasible to gather fine-grain energy usage data at large-scales, providing us with new opportunities to understand demand patterns at different spatial and temporal scales. In this paper, we conduct a large-scale empirical study of energy usage of 14,849 residential and commercial energy consumers from a small city in the United States. We conduct a wide ranging analysis of energy usage at multiple granularities—citywide, transformer-level, and individual home levels. In doing so, we demonstrate how city-wide smart meter datasets can answer a variety of questions on energy consumption, such as the impact of weather on energy usage. For example, we show that extreme weather events significantly increase energy usage, e.g., by 36% and 11.5% on hot summer and cold winter days, respectively. As another example, we show 19.2% of transformers in the grid get overloaded during peak load periods. Finally, we evaluate the impact of incorporating varying amounts of energy storage within the distribution grid and the impact such deployments will have on the peak demand patterns seen by the grid as well as the ability to reduce overloads seen by distribution transformers during peak periods.

CCS Concepts: • **General and reference → Empirical studies**; • **Information systems → Clustering;**

Additional Key Words and Phrases: Electricity, grid, cluster analysis, energy storage

ACM Reference format:
John Wamburu, Stephen Lee, Srinivasan Iyengar, David Irwin, and Prashant Shenoy. 2024. Analyzing the Energy Usage of a Community and the Benefits of Energy Storage. *ACM J. Comput. Sustain. Soc.* 2, 2, Article 13 (May 2024), 23 pages.
https://doi.org/10.1145/3637209

1 INTRODUCTION

The building sector contributes an estimated 40% of the energy and 70% of the electricity consumed in the United States each year [2]. As a result, there is a significant interest in understanding and optimizing building energy usage. Recently, a number of inexpensive consumer/utility-grade

Authors' addresses: J. Wamburu, IBM Research Africa, P.O. Box 35475, Nairobi, 00200, Kenya; e-mail: john.wamburu@ibm.com; S. Lee, University of Pittsburgh, 210 S. Bouquet Street, Pittsburgh, PA, 15260, USA; e-mail: stephen.lee@pitt.edu; S. Iyengar, Microsoft Research, "Vigyan", No. 9, Lavelle Road, Bengaluru, Karnataka, 560 001, India; e-mail: sriyengar@microsoft.com; D. Irwin, University of Massachusetts, Amherst, 151 Holdsworth Way, Amherst, MA, 01003-9284, USA; e-mail: irwin@ecs.umass.edu; P. Shenoy, University of Massachusetts Amherst, 140 Governors Drive, Amherst, MA, 01003-9264, USA; e-mail: shenoy@cs.umass.edu.

2834-5533/2024/05-ART13 $15.00
https://doi.org/10.1145/3637209

advanced smart meters have come on the market, which monitors building energy usage, e.g., electric or gas, at high resolutions in the order of minutes to seconds. Electric and gas utility companies have deployed an estimated 58.5 million smart meters in the U.S., of which 88% have been deployed in the residential sector [2]. By tracking energy usage at a fine granularity, data from smart meters can reveal numerous insights into when and how a building and its occupants consume energy and how its usage changes over time. Understanding these insights is important for energy planning and management, as well as evaluating potential for energy-efficiency improvements and optimizations.

Indeed, there have been several studies that have analyzed electricity data from smart meters across several homes [30, 31, 54]. Public datasets, including Dataport database [1], the ECO dataset [9], and the UMass Smart* [7] are also available for research use. However, due to the difficulty in instrumenting buildings and collecting data, most of these studies and datasets span a few tens to a few hundred homes within a region. Due to the relatively small size of these datasets, they provide limited insights into energy usage across a larger population or across a contiguous geographical region under a single administrative domain, e.g., a city, town, or county.

Another key consideration for utility companies is to understand the peak load profiles within the distribution grid and analyze the impact of deploying energy storage on the peak loads seen by distribution grid transformers. Traditionally electric utilities have deployed edge transformers based on their expected load at each transformer. Due to fluctuations in end user demand and the proliferation of intermittent renewable energy sources into the grid, edge distribution transformers tend to see more fluctuations and may even become overloaded. To mitigate such problems, the emergence of the smart electric grid has resulted in new technologies for more flexible demand-side load management and load mitigation in the grid. In particular, grid-level energy storage is emerging as a key technology for supporting future smart grids, since it can smooth out fluctuations from intermittent renewable energy sources, such as solar and wind, as well as enable grid optimizations, such as shaving peak loads and serving as backup power to reduce outage durations [36, 40, 41].

In this paper, we study the electricity demand experienced across a small city in the New England region of the United States. A novel aspect of our work is that we analyze energy demand at three different resolutions—city-wide, transformer-level and home-level—to understand temporal and seasonal variations in the demand and how these differ at different granularity. We also analyze peak demand seen at different scales and analyze the impact of deploying energy storage at large scales to mitigate local peaks and reduce transformer overloads. In conducting our longitudinal empirical analysis, this paper makes the following contributions:

Citywide Aggregate Demand Analysis. We first examine the aggregate grid demand across an entire city and quantify how it correlates with changes in weather, the seasons, and time-of-day. While these general trends are well-known, we quantify their specific level of correlation on our city-scale dataset. In addition, we also quantify the periodicity of aggregate demand, and how demand deviates from expected values during extreme weather events, e.g., hot and cold days.

Home-level Demand Analysis. We next analyze the demand seen by all individual homes in the entire city. Specifically, we cluster homes into different groups based on the characteristics of their load profile, which is based on the daily routines and patterns. We then analyze the loads seen by individual homes over time and variability in energy usage based on the weather.

Transformer-level Demand Analysis. Finally, we analyze the demand seen by all distribution-edge transformers in the city and quantify their different load profiles. Surprisingly, we observe that most transformers are not over provisioned in the network and all transformers are already

[1] https://dataport.pecanstreet.org/

designed to gracefully handle temporary overloads. Moreover, we find that 19.2% of transformers are heavily overloaded, having a utilization of over 100%. We also decompose transformers into different groups based on the characteristics of their load profile, which is the aggregate of several homes connected to each transformer.

Impact of Energy Storage. We next analyze the impact of large-scale deployment of energy storage, either by individual home owners or at individual transformers by the utility. We specifically analyze how such energy storage deployments will change the peak loads seen at lower-levels of the distribution grids and can mitigate overloads seen by transformers during peak load periods. We examine the effect of two deployment strategies—the use of energy storage at the home and transformer level—to reduce transform overloads, extend their lifetime, and improve grid reliability. Our results show that local deployments of even a small amount of energy storage can dramatically reduce peak loads and the risk of failures in transformers.

2 BACKGROUND

In this section, we present background on the distribution grid and edge transformers, smart meters and grid-based energy storage.

2.1 Distribution Electric Grid

The architecture of the electric grid has three key components: generation, transmission, and distribution. The distribution grid is primarily responsible for supplying electricity to end consumers, which include industrial, commercial, and residential customers. While electricity is transmitted at high voltages through transmission lines, the distribution grid network uses a series of transformers to progressively step down the voltage and supply end-consumers with electricity at voltages of 110V (North America) or 230V (Europe and Asia). The distribution grid comprises sub-stations, feeders, and transformers that are responsible for supplying electricity to end consumers and can be viewed as a hierarchical network [6].

Such distribution edge transformers come in a range of capacities, varying from small 5-10 **kilo-Volt-Ampere (kVA)** pole-top transformers to larger 500, 1000 and 1500 kVA transformers. Note that transformer capacity is rated in kVA, which is the unit used for apparent power, i.e., the product of the **root mean square (rms)** of voltage and current in an AC power system. Small transformers may serve a small number of homes (e.g., two to four homes), while the larger ones serve apartment complexes or office buildings.

Electric utilities size edge transformers based on their expected load. However, typical capacity planning for transformers in the grid works differently from capacity planning in server farms and data centers, which is a well-studied problem [13, 34]. In particular, server capacities are computed based on their expected peak load, such that a server cannot service a peak load that exceeds its capacity, since they have a fixed computing capacity (based on their clock speed and bus bandwidth) that they cannot exceed. Transformer capacity is also sized based on its expected peak load, but a transformer is an analog device that *can supply electricity that exceeds its rated capacity.* Ultimately, the more power a transformer services, the more heat it generates. However, transformers have built-in safety mechanisms, specifically mineral oil, that can absorb some amount of excess heat generated from being over capacity. Thus, unlike servers, transformers are sized to operate over a wide range of their rated capacity, e.g., up to 1.25 of their rated capacity [27]. Even so, overloaded periods are undesirable as they reduce transformer efficiency (since the excess heat represents lost energy), and over time they can cause the insulating oil to evaporate. Once there is not enough oil to absorb the excess heat, it can melt the transformers coils and cause it to fail.

Moreover, overloaded transformers have a higher probability of experiencing outage. While we do not analyze outage information in this study as it is not available, prior studies have shown that overloading can be a good proxy for measuring grid reliability, and that preventing transformer overloading can reduce outages [39]. Thus, limiting the time periods (and magnitude) when the transformer load exceeds its rated capacity is important in reducing the negative impact on transformer lifetimes. For our analysis, we consider a transformer with a peak load between 0.9 to 1.25 its rated capacity to be highly utilized, a peak load between 1.25 to 1.5 to be overloaded and peak load exceeding 1.5 to be critically overloaded. Transformers with loads less than 0.9 are considered to have low to moderate utilization. We use the threshold of 1.25 of a transformer's rated capacity to indicate overload because the rules for sizing overcurrent protection in transformers are that it should not exceed 125% of the primary current [1, 3, 52].

2.2 Smart Electric Meters

Utility companies have started deploying smart meter infrastructure in place of traditional electricity meters. Smart meters are electric meters capable of two-way communication, allow fine-grained metering of energy usage and can send send energy reports in real time – compared to traditional meters that are read manually. Some smart meters also have wireless reporting capabilities that enable outage notifications. In this work, we use smart meters to analyze electricity usage at fine granularity. The electricity meters used in our dataset monitor and report usage at a resolution of 5 minutes.

2.3 Grid-based Energy Storage

Grid-level energy storage, in the form of batteries, has emerged as a promising approach for various grid optimizations. Battery-based grid energy storage can be deployed at various points in the grid's hierarchy—generation, transmission, or the distribution part of the grid network. Prior studies have shown that battery-based storage is especially appealing to handle the intermittency exhibited by renewable energy sources, such as solar and wind, by using storage to smooth out the fluctuations [8, 29]. Similarly, battery-based storage has been used for peak load shaving [40, 41].

Although the cost of battery-based energy storage remains high [32], prices are dropping more rapidly than expected even a few years ago, and commercial products and deployments are beginning to ramp up. At the same time, centralized grid-level large-scale battery storage systems capable of supplying a whole neighborhood have been shown to be a feasible alternative [55]. In this work, we consider the deployment of energy storage batteries at the grid level alongside distribution edge transformers to mitigate overloads caused by peak end user loads and enhance transfer lifetime—a use of batteries that has not seen much attention in the distribution network. Utilities are especially interested in using such application in the future as prices continue to fall.

3 PROBLEM AND RESEARCH METHODOLOGY

In this section, we present the problem and key research questions we address in the paper, and then describe the datasets and experimental methodology that we use to answer those questions.

3.1 Problem

The primary goal of our paper is to study the energy usage across an entire city at three spatial resolutions: at city-scale, at the level of distribution transformers, and at individual homes, and understand the distribution of peak energy demand experienced by distribution transformers, using real-world energy data. An additional goal is to understand whether emerging technologies, such as battery-based grid storage, can alleviate the overloads. Specifically, we seek to answer the following research questions.

(1) What is the aggregated demand profile of all consumers across an entire city? What is the distribution of energy usage across the customer base? How is it impacted by seasons, time

of day and the day of the week? What is the impact of extreme weather events such as a heat wave or a cold wave on the peak usage? What is the distribution of peak energy usage at city scale?

(2) What are demand profiles of individual homes and what clusters of usage patterns emerge from these demand profiles? What are the peak demand profiles of individual homes?
(3) How do the load profiles change when multiple homes are aggregated at each transformer? What are the daily and seasonal variations in this load? What kind of peak loads are seen at individual transformers?
(4) What is the impact of deploying storage at the home and transformer levels? What do these results reveal about the relative size and feasibility of energy storage as a mitigation strategy for transformer overloading?

3.2 Description of Datasets

The answers to these questions will vary from region to region, and clearly depend on the current state of the distribution grid in terms of its load over time, transformer capacities, and the resulting slack. In this paper, we use a city in the New England region of the United States and attempt to answer these questions for this city by conducting a city-wide data analysis. While the reliability of electric grids varies across regions around the world [21, 42], the distribution grid studied in this work is typical in many areas [28], and our high level insights are applicable in such regions.

Distribution Grid Dataset. Our dataset consists of electricity usage (load) data recorded by 14, 849 smart meters that serve residential (13, 458) and commercial (1, 391) users in the city. These 14, 849 meters are served by 1,270 distribution edge transformers. The dataset also includes a mapping of each meter to its edge transformer, and also includes a detailed specification of each transformer, including its rated capacity. The load data is recorded at a five minute granularity and spans from 2015 to 2019. The electric meters used to collect the data are advanced electronic meters which have the capability to be read remotely. All meter data is gathered using an automated meter reading system that allows meter readings to be sent directly to a central system via a secure communication network. Some meter readings contain spurious reads whose values are much higher than expected. In such cases, we limit the readings to the 99.9^{th} percentile. Further, since there has been little seasonal variation during the whole duration of data availability, for this analysis, we use data from 2018 and 2019.

Since these edge transformers are low voltage transformers that are directly connected to end-customers, the load on each transformer can be computed by summing the load recorded by each meter connected to that transformer. Doing so yields highly detailed load information experienced by each distribution edge transformer over the two year period of the study. The availability of detailed load information for all 1,270 edge transformers in a city is a distinguishing feature of our study. Most prior work has considered simulation based studies to analyze transformer load data. In contrast, we study probability distribution of loads as well as the time of day/seasonal impacts that other studies did not consider using real world electric usage data.

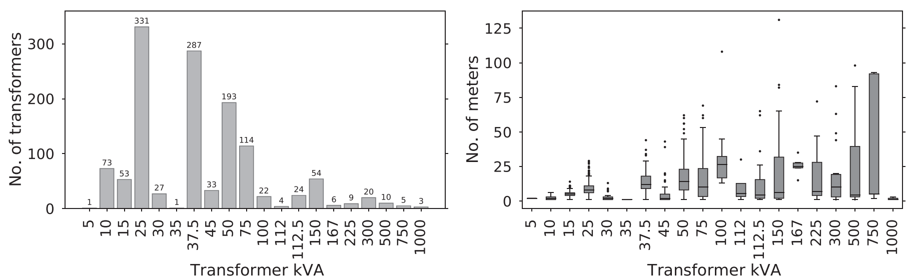
Table 1 summarizes the key characteristics of our dataset discussed above. Figure 1(a) then depicts the diversity of transformer capacities in the distribution grids and the distribution of transformers across varying sizes. Note that, since the rated capacity is in apparent power as kVA, in our later analysis, we use the average power factor to convert it into **kilowatts (kW)** to make our results more intuitive. We use the equation below for the conversion.

$$kW = kVA \cdot PF \tag{1}$$

Here, $0 \leq PF \leq 1$ is the power factor. For our analysis, we use power factor of 0.9 and 0.95 for summer and winter, respectively, which represents the average power factor in these seasons.

Table 1. Grid Dataset

No. of residential meters	13, 458
No. of commercial meters	1, 391
Total no. of meters	14, 849
No. of transformers	1, 270
Size of transformers	5 kVA to 1, 000 kVA
Electric meter granularity	5 minutes
Duration	2015 to 2019



(a) Distribution of transformer size

(b) Number of smart meters by transformer size

Fig. 1. (1(a)) Distribution of transformers by capacity, and (1(b)) Distribution of smart meters connected to transformers of varying capacity.

Figure 1(a) shows that transformer capacities can vary from 5 kVA all the way to 1500 kVA. Most of the deployed transformers are "small" and have a rated capacity of less than 150 kVA—a few transformers are large with a capacity of 500 kVA to 1500 kVA. Generally, the small transformers serve a small number of residential customers (e.g., two to four homes). The larger transformers serve apartment complexes, office buildings, other light commercial customers.

Figure 1(b) shows the distribution of meters connected to transformers of various sizes. We observe that the number of connected residential meters increases with the increase in transformer capacity. In contrast, fewer meters are connected to transformers that provide electricity to commercial buildings as they tend to consume higher energy. The median number of meters connected to these transformers ranges from 2 to 28.

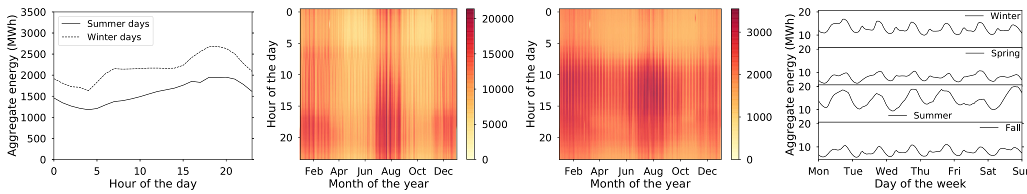
Weather Data. In addition to electricity data, we also supplement our dataset with weather data for the city. We gathered weather data at one-hour granularity for the city from the Dark Sky API[2]. Since temperature fluctuations are infrequent within the hour, this granularity for weather data is sufficient for our analyses.

In the sections that follow, we analyze this city-scale dataset to answer questions about aggregate usage: the distribution of energy usage across customers and the relationship of energy usage and weather. Later, we discuss the different types of customer segments that we observe in our dataset. Further, we study the impact of electricity usage on transformer overloading. Finally, we study the potential of storage at the home and transformer level to mitigate such overloading.

4 CITY-SCALE ENERGY ANALYSIS

In this section, we examine the energy usage of individual customers *en masse* and the aggregate usage across all customers. We specifically examine the impact of time of the day, the day of the

[2] https://darksky.net/dev



(a) Demand by time of day (b) Residential meters (c) Commercial meters (d) Demand by day of week

Fig. 2. (2(a)) Aggregate power consumption by time of day for all meters, and (2(b)) aggregate electricity consumption for all residential buildings across the year, and (2(c)) aggregate electricity consumption for all commercial buildings, and (2(d)), aggregate demand at the grid level for a representative week in each season.

week, seasons, and the weather on energy usage. We also perform customer segmentation analysis on the daily load profile of homes across the entire customer base.

Figure 2(a) depicts the electricity demand which varies between 1180 MWh and 2680 MWh—approximately a factor of 2 difference between off-peak and peak hours on average. The summer electricity demand shows increased usage during the day and a peak around 7-8PM, presumably due to higher usage of cooling equipment. The winter demand shows two peak periods— a morning peak around 8AM and an evening peak around 7–8PM. The winter evening peak demand is 37.4% higher than the summer peak. This is because northeastern cities sometimes have long and harsh winters and the temperature may go as low as -25°F, triggering use of electric heaters for longer durations.

Figures 2(b) and 2(c) are heat maps showing aggregate usage of electricity in residential and commercial buildings for each hour (shown on the Y-axis) of each day over the course of the year 2019 (shown on the X-axis), where darker colors indicate higher energy usage. In both figures, the aggregate electricity usage reveals the same morning and evening peaks seen in Figure 2(a). The figure also reveals high energy usage in peak summer and winter months, which indicate the use of **air conditioners (ACs)** and electric heaters, respectively.

Figure 2(d) demonstrates the aggregate electricity demand over the course of an entire week for four seasons: Winter, Summer, Fall, and Spring. The figure shows the time of day effects for each day—the overall pattern repeats across seasons, although the magnitude of the peaks and the average usage is higher in warmer seasons than in colder ones.

In summary: *(1) Energy usage shows time of day effects with morning & evening peaks as well as seasonal effects. (2) Electricity demand is higher in summer and winter due to the use of electric ACs and electric heaters, respectively.*

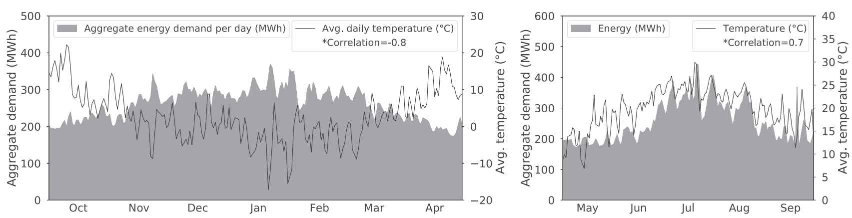
4.1 Impact of Weather on City-scale Energy Usage

Next, we study the impact of outside temperature on the energy usage—both during winter and summer. Figures 3(a) and 3(b) plot the aggregate daily electricity demand over the course of a day along with the average daily temperature for winter and summer, respectively. Figure 3(a) shows a strong negative correlation of -0.8 between temperature and electricity usage in the winter—as the temperature falls, the electricity usage rises, due to increased heating costs from heating water, space heaters and homes with electric heaters.

Figure 3(b) shows a moderate positive correlation of 0.7 between temperature and electric usage in summer. Due to generally milder summers in the North England region, many homes do not have AC or run them infrequently, leading to a moderate (rather than high) correlation to outside

Table 2. Criteria for Extreme Weather Events

Event	Criteria
Hot days	Daily average temperature above 90°F
Cold days	Daily average temperature below 12°F
Snow days	Days of snowfall
Summer days	All Days in Jun, Jul, Aug, Sep
Winter days	All Days in Dec, Jan, Feb, Mar



(a) Winter (b) Summer

Fig. 3. Variations in daily aggregated electricity consumption at the grid level for the entire year.

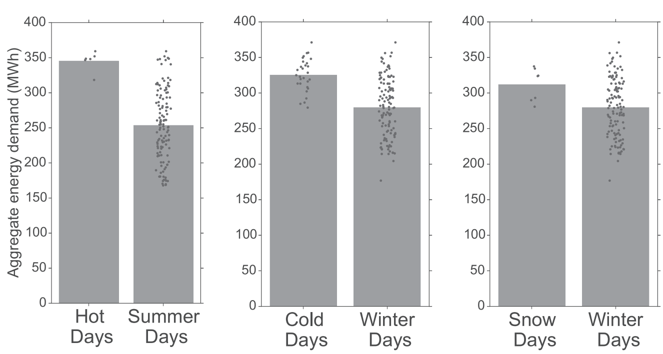
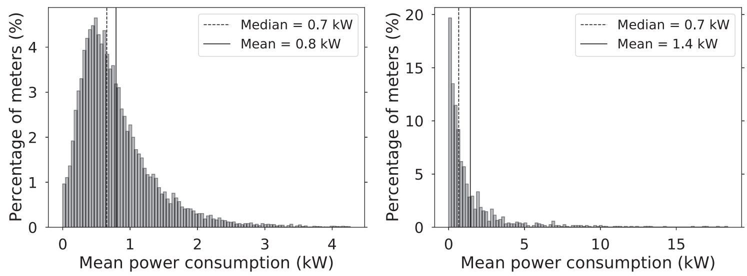


Fig. 4. Electricity consumption for extreme weather events.

temperature. The figure also shows that winter demand varies from 173 MWh to 369 MWh, while summer demand varies from 170 MWh to 440 MWh. In addition, the average absolute day-to-day change in the aggregate electricity demand is 5 MWh and 20.2 MWh in winter and summer, respectively.

Figure 4 compares average daily summer and winter usage to “extreme" weather days (e.g., hot summer day to an average summer day). Table 2 defines the criteria for “extreme" weather events. The scatter plots show the days selected using the criteria suggested in the table, whereas the bar chart show their average energy demand. Figure 4 shows demand can be 36% higher on a hot summer day than that on an average summer day—due to increased AC usage. The figure also shows 16.4% and 11.5% higher demand for cold winter days and during snowfall events over average winter days, respectively.

In summary: *(1) There is a strong correlation of electricity usage with temperature in winter, while only moderate correlation exists in the summer. (2) The demand on cold winter and hot summer days can be 11.5% and 36% higher than the average days in those seasons, respectively.*



(a) Residential meters

(b) Commercial meters

Fig. 5. (5(a)) Distribution of average power consumption per home, (5(b)) Distribution of average power consumption by commercial meters.

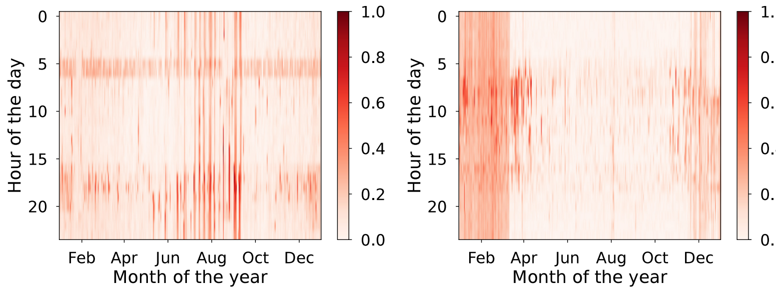
5 ENERGY USAGE OF INDIVIDUAL CONSUMERS

In this section, we analyze temporal and seasonal energy usage by individual homes and perform cluster analysis to identify the different types of home load profiles in the city. By analyzing data at an individual home level, privacy concerns such as occupancy detection can be introduced. Therefore, for this analysis, we analyze the aggregate energy consumption across all buildings in the city collectively. Since all meter data is anonymized, our analysis minimizes privacy risks that may come with load disaggregation.

5.1 Temporal and Seasonal Analysis

We first analyze the distribution of electricity usage across residential and commercial customers. Figure 5(a) depicts the histogram of average electricity usage for residential customers. The figure shows a mean power consumption of 0.8 kW and the distribution shows a long tail where the 99th percentile of the consumption is 4.6 times the mean. The mean consumption is lower than the average usage of 1.24 kW reported for a typical US household [2]. Figure 5(b) depicts the histogram of average electricity usage for commercial customers, and differs from residential consumption in a number of ways. First, the figure shows a mean power consumption of 1.4 kW, which is 1.75 times the residential mean consumption. This is because commercial customers typically have higher energy usage than residential ones driven by higher occupancy and energy for production needs. The figure also reveals that the highest consuming commercial buildings draw up to 20 kW, which is 14.3× the average consumption. Figure 2(a) shows the aggregate energy demand of all customers for each hour of the day for the summer and winter seasons. In this study, the months with most days having an average daily temperature greater than 60°F are categorized as summer, whereas the rest are categorized as winter. We chose this threshold because the resulting summer and winter months coincide with summer and winter seasons in the North Eastern region of the United States, and included September in summer months because we observed a high number of days with an average temperature of 60°F or more. Thus, unless stated otherwise, winter days are defined as the days from Jan-Apr and Oct-Dec 2019, whereas summer days are defined as the period from May-Sep 2019.

Figures 6(a) and 6(b) are heat maps showing electric usage from two individual homes. Here, the energy consumption for each home has been normalized using Min-Max scaling. In Figure 6(a), the electricity usage pattern reveals clear peaks during morning and evening hours over the entire year with somewhat higher usage during summer evenings. Figure 6(b), on the other hand, reveals higher usage in winter months than summer months. Figure 6(b) also reveals higher usage during the morning (around 7AM) throughout the year—presumably due to the need for hot water for



(a) Home A: Electric meter data

(b) Home B: Electric meter data

Fig. 6. (6(a)) Normalized energy consumption for an individual home with alternate heating energy source, e.g., gas (Home A), and (6(b)) normalized energy consumption for another home using electric heating (Home B).

showers. Prior work that has analyzed daily energy trends also revealed morning and evening peaks [37]. However, these analyses lack seasonal patterns in the data as it is available across a shorter period of time.

Figure 6(b) depicts a home in which electricity is used to provide heat during the winter. The figure shows higher electricity demand in winter for electric heating and also shows higher morning and evening peaks. It also reveals higher usage for a few days in August—presumably due to higher cooling demand. Prior work that has analyzed seasonal energy consumption of residential customers also found higher energy usage during summer and winter months [30]. However, due to the coarse granularity of the data used, our analysis reveals these patterns at a much more granular scale.

In summary: *(1) Energy usage at individual homes shows time of day effects with morning & evening peaks as well as seasonal effects. (2) Electricity demand is higher in summer and gas demand is higher in winter due to the use of electric ACs and gas heaters, respectively.*

5.2 Load Profile Analysis

Having examined the temporal and weather influence on energy use, we next study how different customers use energy on a day-to-day basis in their homes. For this analysis, we only cluster data from residential homes. Our hypothesis is that the energy usage within a home is largely determined by the daily routines and activities within a household, and depending on the characteristics of residents and their routines, different groups of customers will exhibit similar types of usage patterns. For example, homes, where everyone works during the day from 9AM-5PM, will have a different profile than a home with a retired person.

To validate this hypothesis, we perform customer segmentation analysis on the daily load profile of homes across the entire customer base. Since we are primarily interested in the pattern rather than the magnitude of the energy usage, we begin by normalizing the average daily load profile for each home between 0 and 1. We then use *k-means* clustering on these load profiles. *k-means* is a widely used clustering technique that takes a set of instances (individual homes) and their features (average energy consumption for each hour of the day) along with the desired number of clusters, k, as input. It uses an iterative approach to partition the data set into k groups such that the intra-cluster distance is small and inter-cluster distance is high. For this experiment, we used the sum of the squared distances between different load profiles. Typically, there exists statistical techniques to converge on the number of clusters such as ***Akaike Information Criterion*** **(AIC)** or ***Bayesian Information Criterion*** **(BIC)**. However, these are generic model selection criteria

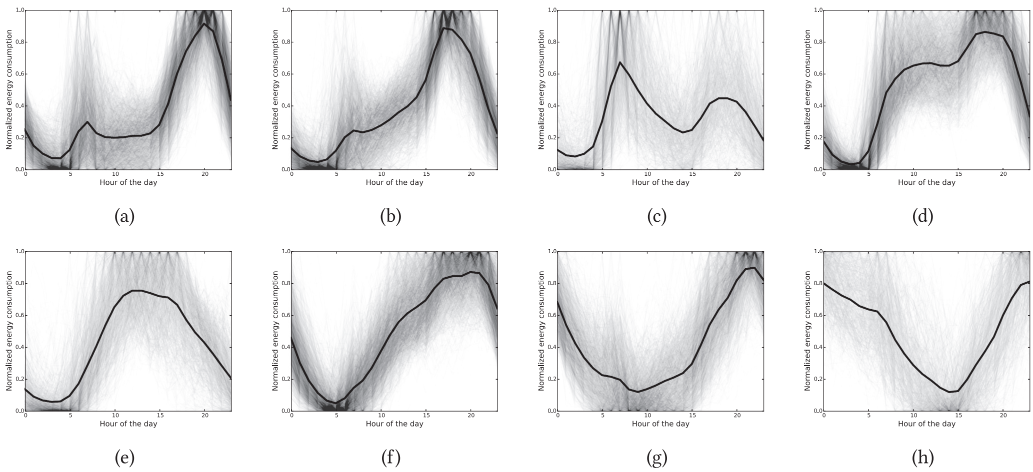


Fig. 7. Load profile clusters of buildings observed across all residential homes in the dataset. Clusters (a) to (d) show a bi-modal distribution with two peak periods (morning and evening). Clusters (e) to (h) show a unimodal distribution with the peak occurring at different times of the day.

and may not necessarily work well for all domains. Thus, we decided to employ visual model selection. By running *k-means* for different values of k, we found $k = 8$ to be the ideal choice for our dataset which did not result in any outlier clusters.

Figure 7 shows the eight clusters (customer segments) that resulted from our analysis. The lightly shaded lines are the profiles of each home present in that cluster and the bold line represents the centroid of the load pattern within each cluster. Broadly, there are four clusters that are bimodal with two peak usage periods of varying degrees and four clusters that are unimodal with a single peak usage period over the course of the day.

Table 3 summarizes the key characteristics of the customer segments within each cluster that include cluster type, peaks observed, the number of homes and their proportion in the dataset. As shown, around 7,216 homes (53.6% of total) exhibit bimodal usage (clusters *a,b,c,d*), 1,240 homes (9.2% of total) exhibit unimodal daytime peak usage (cluster *e*), 2,291 homes (17.0% of total) exhibit unimodal evening peak usage (cluster *f*), while 2,711 homes (20.1% of total) exhibit "nocturnal" usage (clusters *g,h*).

Figures 7(a)–7(d) depict the four bimodal clusters. Figures 7(a) and 7(b) are homes with a small morning peak and a more prominent evening peak. These homes usually correspond to homes with working/school routine. Figure 7(c) is the opposite with a greater morning peak and a less prominent evening peak. Figure 7(d) depicts households with large morning and evening peaks. The nature of these peaks reflect appliance usage with homes at different times of day. For example, a taller morning peak reveals greater appliance use in the morning (e.g., use of laundry machines), while those with taller evening peaks reveal homes where more of these activities are performed in the evenings. Figure 7(d) depicts a more uniform distribution of activities in the morning and evenings. These represent homes that are occupied during the day.

Figures 7(e)–7(h) depict four clusters with unimodal usage characterized by a single peak. Figure 7(e) depicts households where energy usage peaks in mid-day—presumably due to occupancy during daytime hours. Figure 7(f) depicts homes where peak usage occurring during evenings, with different peaks reflecting when daily chores are performed. Figures 7(g) and 7(h) represents nocturnal homes where the off-peak period occurs in the late morning or mid-afternoon

Table 3. Summary of Residential Load Clusters

Cluster	Type	Peaks	No. of homes	Proportion
(a)	Bi-modal	7AM and 9PM	2,131	15.8%
(b)	Bi-modal	7AM and 5PM	2,057	15.3%
(c)	Bi-modal	7AM and 6PM	1,097	8.2%
(d)	Bi-modal	12PM and 6PM	1,931	14.3%
(e)	Uni-modal	12PM	1,240	9.2%
(f)	Uni-modal	9PM	2,291	17.0%
(g)	Uni-modal	11PM	1,694	12.6%
(h)	Uni-modal	12AM	1,017	7.6%

and peak usage occurs during night hours. Presumably, these homes represent occupants who come home late at night.

Prior work that has studied segmentation based on energy usage also revealed similar unimodal and bimodal peaking profiles, with households that exhibit an evening peak accounting for the largest group of customers [31]. However, as with most other prior studies, the data used in this prior work is less granular (> 10×) than the data used in this work, and this enables segmentation at a much finer scale.

In summary: *Our customer segmentation reveals how the energy profiles correspond to their daily routines, with 53.6% of homes exhibiting bimodal energy usage, whereas, 26.2% and 20.1% of homes exhibit unimodal daytime and nocturnal energy usage, respectively.*

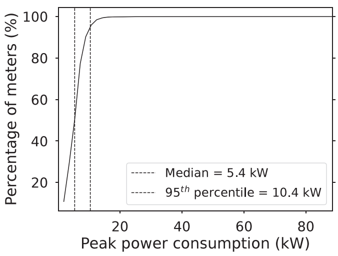
5.3 Peak Analysis

Next, we analyze the peak power consumption recorded by each meter. We define peak power as the maximum power recorded by a smart meter at any time during the year. The recorded power in the data contains a few instances of unusually high values, which we attribute to spurious meter readings. Therefore, to compute the peak, we take the the $99.9th$ percentile reading across all data to eliminate these spurious reads which may affect peak analysis.

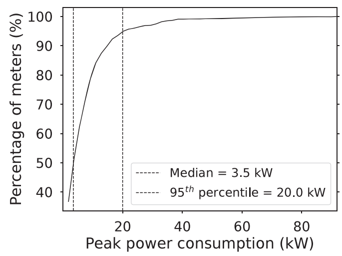
Figures 8(a) and 8(b) depict the cumulative distribution function of peak power drawn by residential and commercial meters during the year, respectively. Figure 8(a) shows that the median peak power across residential meters is 5.4 kW. In comparison, this is approximately 7.7× the average usage depicted in Figure 5(a). The figure also shows that some meters draw up to ≈88 kW at some point in time during the year, while the 95th percentile is 10.4 kW. On the other hand, Figure 8(b) shows that the median peak power across commercial meters is 3.5 kW, which is 5× the average usage depicted in Figure 5(b). The figure also shows that the 95th percentile is 20.0 kW, which is 2× the highest peak observed in 95% of all residential meters.

To examine the extremity of peak usage from the average usage, we compute the peak-to-average ratio for each home during the year. Figure 8(c) plots the distribution of this ratio. The median peak-to-average ratio is 6.9, indicating that the peak usage is approximately 7× the average usage. Similar to the peak usage, this distribution also depicts a long tail, with some homes experiencing as high as 40× the average power consumption. As we will see later in Section 6.2, the peak-to-average ratio experiences a smoothing effect at the transformer level, i.e., when multiple meters are combined into one transformer, their respective peaks occur at different times, and this leads to a lower aggregate peak at the transformer level.

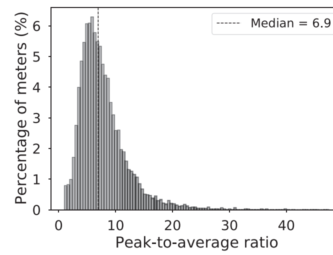
In summary: *Commercial meters experience higher peaks than residential meters. The peak-to-average ratio across all meters is 6.9, indicating that peak usage is approximately 7× the average usage.*



(a)



(b)



(c)

Fig. 8. (8(a)) Cumulative distribution function of absolute peak power usage by residential meters, and (8(b)) cumulative distribution function of absolute peak power usage by commercial meters, and (8(c)) distribution of power peak-to-average ratio over the year for all meters.

6 TRANSFORMER-SCALE ENERGY ANALYSIS

In this section, we analyze the probability distribution of loads, seasonal variation of loads and the peak loads experienced by edge distribution transformers.

6.1 Transformer Load Profiles

We begin with an analysis of the monthly and daily loads seen by the 1,270 edge transformers across the city. Figure 9 depicts the monthly load experienced by a representative transformer over the year 2019. The figure illustrates the seasonal variation in the load, and is characterized by two peak demand periods—winter and summer. The winter peak occurs due to increased use of electric heaters during the winter, while the summer load coincides with the increased use of air conditioning on hot summer days. Although the winter peak is slightly higher than the summer one, the summer peak has a greater impact on transformer efficiency and lifetimes. Prior studies have shown that a high ambient temperature can have an adverse impact on transformer lifetimes [19, 51], as a high ambient temperature contributes to the effect of overloading by further heating up (and evaporating) the insulation oil, which protects transformers from overheating. On the other hand, spring and fall seasons see lower peak loads, have cooler temperatures and transformers have more slack, and as a result, they are less vulnerable during these periods.

Next, we analyze the daily load profile of edge transformers to identify the most common types of transformers based on their load profile. For this analysis, we clustered the average daily profile of all transformers using *k-Means* clustering. Since the transformers are of different sizes, we normalize the daily load profile of each transformer to a range between 0 and 1 (e.g., using MinMaxScaler in scikit-learn), and then perform clustering. Figure 10 depicts the five clusters that emerge when using k-means with k = 5. After running k-means with multiple values of k, we selected k = 5 since 5 was the highest value of k that yielded clusters that were qualitatively different, and also did not yield an outlier cluster with few transformers. The bold line depicts the centroid of the clusters, while the lightly shaded gray lines show the energy profiles of all the transformers in the cluster. A summary of the clusters with the type, peak period, number of transformers and their respective proportion is shown in Table 4.

The five clusters reveal interesting patterns. For example, Figures 10(a) and 10(b) depict transformers that exhibit daytime peaks, while Figures 10(c), 10(d) and 10(e) depict transformers that exhibit evening peaks. The captions depict the number and percentage of transformers in each cluster. We hypothesize that the transformers exhibiting daytime peaks, in Figures 10(a) and 10(b), serve office buildings that have a 9 am to 5 pm workday, or businesses, such as retail stores, that have 9 am to 9 pm work hours. These transformers have a low load during the late evening and nighttime hours.

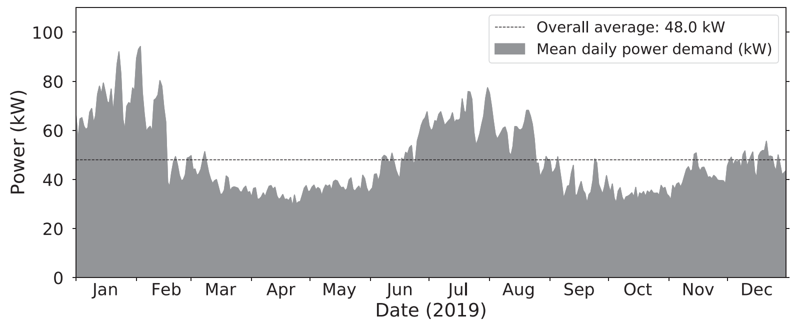


Fig. 9. Graph illustrating the load on a sample transformer over a year with demand peaking during winter and summer seasons.

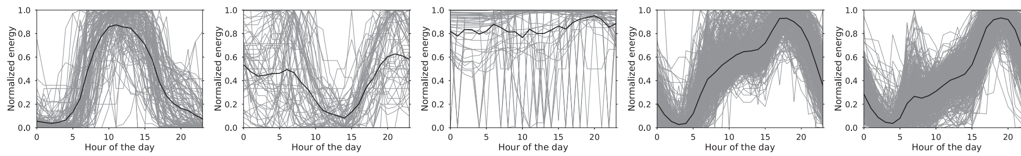
Table 4. Summary of Transformer Clusters

Cluster	Type	Peaks	No. of transformers	Proportion
(a)	Uni-modal	10AM-3PM	106	8.3%
(b)	Uni-modal	8PM	62	4.9%
(c)	Uni-modal	7PM	42	3.3%
(d)	Uni-modal	7PM	503	39.6%
(e)	Bi-modal	8AM and 8PM	557	43.9%

The clusters shown in Figures 10(c), 10(d), and 10(e) all exhibit evening peaks and also exhibit a nontrivial amount of nighttime usage—we hypothesize that these are large residential customers with different daily routines. The cluster in Figure 10(c) shows transformers that see a low load during the day—these are likely users who are away from home (i.e., working) during the day and at home in the evening and night. Figures 10(d) and 10(e) show residential customers with evening peaks, but also a non-trivial amount of daytime and nighttime usage. These are likely to be families where someone is at home during the day, where the increased evening activities result in an evening peak—these two clusters 10(d) and 10(e) also account for a large fraction of the transformers, 39.6%, and 43.9%, respectively.

Comparing transformer clusters (Figure 10) with home clusters (depicted in Figure 7) makes some interesting observations. First, the number of transformer clusters is lower than the number of home clusters depicted in Figure 7. This fewer number of distinct clusters is brought about by energy loads from multiple homes being aggregated at each transformer. This leads to a smoothing effect on the load observed at the transformer level, as well as merging some clusters into others. Second, predominant patterns in usage data at the individual building level, e.g., unimodal and bimodal daily usage, persist even at the transformer level. These patterns also contain the highest number of transformers in the dataset. This is due to the localization of transformer installations, where multiple homes that have the same characteristics, e.g., residential neighborhood, are connected to the same distribution transformer. Lastly, similar to the home profile analysis, a high number of transformers also exhibit high daytime usage, with only a small number of transformers exhibiting higher usage at night time (4.9%).

In summary: *The aggregation of workloads from several homes into one transformer load leads to smoothing and results in fewer transformer clusters than home clusters. However, basic daily usage patterns remain, with 43.9% of all transformers exhibiting bimodal usage, while 56.1% exhibit unimodal patterns.*



(a) 106 (8.3%) (b) 62 (4.9%) (c) 42 (3.3%) (d) 503 (39.6%) (e) 557 (43.9%)

Fig. 10. Demand profile clusters across transformers. The number and percentage of transformers in each cluster is listed in the caption. The clusters are qualitatively different, with some exhibiting daytime peaks and others exhibiting evening peaks. Note some of the differences in these clusters from the individual home clusters shown in Figure 7.

6.2 Transformer Peak Analysis

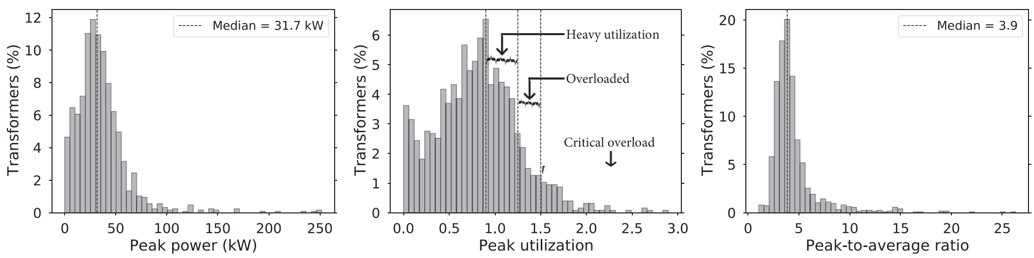
Next, we analyze the peak load experienced by transformers throughout the year. The availability of high granularity energy usage data at the transformer level (5 minutes) enables us to study peak distribution at a fine temporal scale, and this is a distinguishing feature of this work. We define the peak as the maximum power drawn at any point during the year, i.e., we take the max of the 5-minute granularity power profile of each transformer. Before summing up smart meter data to create a transformer profile, we take the 99.9^{th} percentile of meter readings to eliminate spurious reads which may affect the peak analysis of transformer loads. Using Equation (1), we compute the rated capacity of transformers in kW and then compute the utilization by normalizing the load observed at the transformer with its rated capacity.

We then group the transformers into four categories, explained in Section 2.1 and depicted in Table 5: low-to-moderate, highly utilized, overloaded, and critically overloaded. As noted in Section 2.1, we use the threshold of 125% of a transformer's rated capacity to indicate overload because the rules for sizing overcurrent protection in transformers are that it should not exceed 125% of the primary current [1, 3, 52]. Figure 11(a) depicts the peak load distribution of the transformer across the whole city, while Table 5 shows the number and percentage of transformers that fall in each category. Since transformers have a typical lifetime of 20-30 years, one would expect careful sizing, such that the peak load is well below the rated capacity. However, as shown in Table 5 and Figure 11(b), only 60.6% of the transformers service a peak load of less than 90% utilization over the course of the year. Around 26% of the transformers are heavily utilized and service a peak load of up to 125% of their rated capacity. Note that this implies that the transformer operated at or above its rated capacity for at least part of the time over the year. Around 7.2% of the transformers are overloaded and see a peak load that exceeds 125% utilization, while an additional 6.4% of the transformers are critically overloaded with the peak load exceeding 150%. As explained earlier, it is not "abnormal" for a transformer to exceed 100% utilization for short periods, since they have mineral oils to insulate them from overheating, although sustained overloads for long periods are dangerous. Therefore, we next analyze the duration of the overloads experienced by the edge transformers.

We consider only the transformers that are in the overloaded and the critically overloaded groups and compute the number hours over the year for which they service a load exceeding 125% of their rated capacity, and also compute the maximum "session duration" over which the transformer is continuously overloaded. Figure 12(a) plots the total number of hours for which transformers are overloaded or critically overloaded during the year. The figure shows that the overload distribution is long-tailed—the majority are overloaded for 162 hours over a year, while a few see overloads of as many as 1,000 - 3,000 hours. Figure 12(b) analyzes each continuous period that experiences an overload, and plots the longest continuous duration for which a transformer

Table 5. Summary of Transformer Groups by Peak Utilization

Group	Utilization	No. of transformers	Proportion
Low to moderate	<90%	769	60.6%
Heavily utilized	≥90% to <125%	329	25.9%
Overloaded	≥125% to <150%	91	7.2%
Critically overloaded	≥150%	81	6.4%



(a) (b) (c)

Fig. 11. (11(a)) Distribution of absolute power demand over a year, and (11(b)) distribution of transformer overloading over a year, based on the peak utilization, and (11(c)) distribution of transformer peak-to-average ratio over the year.

was overloaded. The figure, plotted on a log scale, shows the median duration of overload was 45 minutes, while some transformers see a sustained overload of 143 hours.

Next, we analyze the extremity of peak loads from regular demand. To do so, we compute the peak-to-average ratio of a transformer's peak load from the average load during the year. Figure 11(c) shows the distribution of the peak-to-average across all transformers in the dataset. The median peak-to-average ratio is 3.9, indicating that peak loads are approximately 4 times the average. Compared to individual homes, the median peak-to-average ratio is lower than that of individual homes (the median peak-to-average ratio for homes is 6.9 as shown in Figure 8(c)). This indicates some kind of statistical smoothing which is brought about by peaks occurring at different times at the home level. Because peaks at the home level do not directly coincide, a smoothing effect is experienced at the aggregate peak causing transformer peaks to be lower than that of individual customers.

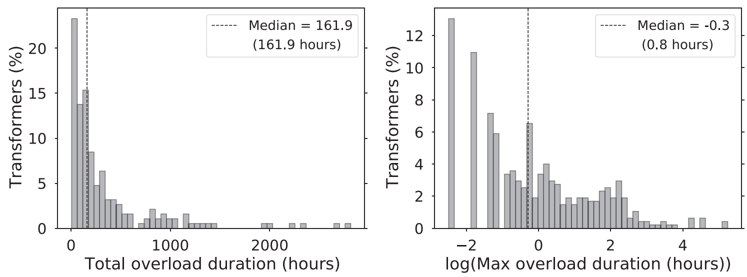
In summary: *Our analysis shows that roughly two-thirds of the transformers experience low-to-moderate peak loads. Conversely, around 26% of the transformers are heavily utilized, while around 6% are already overloaded or critically overloaded. Finally, one surprising aspect of our analysis is our finding that shows roughly 30% of the transformers routinely operate over capacity at least for a portion of time each year, with some experiencing long sustained overloads of many days.*

7 IMPACT OF ENERGY STORAGE ON DISTRIBUTION GRID LOADS

Having examined the load profiles of individual homes and edge transformers, we next evaluate the impact of incorporating energy storage in the grid. Specifically, we explore how deploying energy storage at the home and grid level can be used for peak load reduction and consequently, reducing the number of overloaded and critically overloaded transformers in the grid.

7.1 Energy Storage at the Home Level

For this analysis, we consider different storage penetration levels across homes and understand the impact of storage on the peak demand seen by homes and the transformers they are connected



(a) (b)

Fig. 12. (12(a)) Distribution of the total number of hours during which transformers were overloaded during the year, and (12(b)) distribution of the maximum sustained period of overloading (log scale).

to. While prior work has studied other aspects of storage deployed at the home level such as coordinated charging and discharging of distributed energy resources [11], in this work, we focus on the effect of energy storage on energy demand at a home, transformer peak and overloading.

To model energy storage at the home level, we use the following battery specifications. We assume that a battery with a capacity of 13.5 kWh, a peak output capability of 7 kW, and a continuous discharge rate of 5 kW is installed at each home. These specifications follow a popular home battery configuration that is in production today. Therefore, for each day, we assume that a home's energy demand can be shaved by up to 8 kWh (discharging a battery to zero capacity reduces the battery's lifetime, so we assume a maximum discharge of 8 kWh per day). We then shave off the top 8 kWh for each home during the year and generate a new synthetic energy load. We then experiment with different levels of battery penetration across homes in the grid e.g. 10%, 20%, 30%, 40%, 50%. We define penetration as the number of homes in the grid that have an installed battery. For each penetration level, we select the percentage of homes from the grid at random and assume that a battery is present. We then compute the resulting transformer loads using the shaved load in homes that have a battery and the original load for homes that do not have a battery. We repeat each experiment for 20 runs with a different random selection of homes to ensure our results have tight confidence intervals.

Figure 13(a) shows a sample home with the top 8.0 kWh shaved off the total demand of 115.4 kWh during a particular day. The figure shows that load shaving occurs during instances of peak usage. On the other hand, Figure 13(b) depicts the results of varying the battery penetration level from 0% (no batteries at all) to 50%. The figure shows that with only 10% of homes in the grid having a battery, the number of overloaded transformers can be reduced by up to 30%, while the number of heavily utilized transformers can be reduced by up to 16.6%. The figure also shows that increasing the battery penetration to 50% can reduce the number of overloaded and heavily utilized transformers by up to 74% and 60%, respectively.

In summary: *A battery with a daily capacity of 8 kWh installed at 10% of homes can reduce the number of overloaded transformers by up to 30%. This reduction of transformers increases superlinearly, with 50% of home batteries able to reduce transformer overloading in the grid by up to 74%.*

7.2 Storage at Grid Level

Next, we consider the impact of deploying energy storage in the distribution grid, rather than in individual homes. We assume that batteries are installed by the utility next to distribution transformers. Prior work has shown that deploying storage at the transformer level can reduce distribution costs by up to 3.75%, and this makes it a viable option for curtailing costs [35].

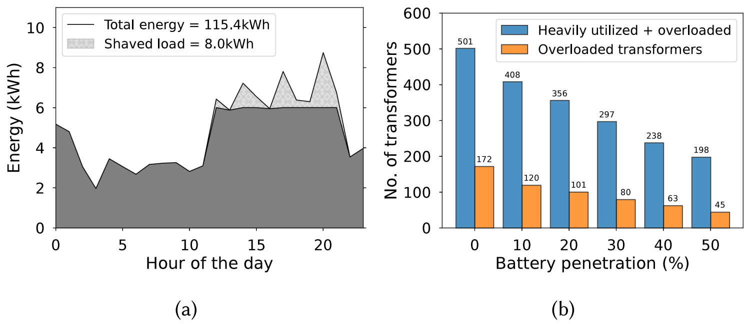
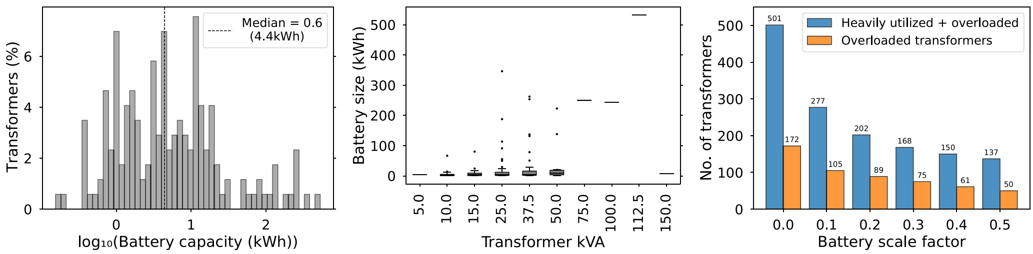


Fig. 13. (13(a)) Original and shaved load for a sample home with an installed battery during a particular day, and (13(b)) reduction in number of heavily utilized and overloaded transformers with increase in home battery penetration.

We begin by addressing the problem of storage sizing, i.e., quantifying the amount of energy storage that is required to mitigate overload at each transformer, and then analyze the impact of installing such storage on peak shaving. Installing energy storage at the transformer level can alleviate transformer overloading by absorbing peak loads thereby increasing transformer lifetime. This in turn can become a key cost saving strategy for utility companies. For example, the cost of replacing a single distribution transformer can vary from $7, 000 [49], while the cost of replacing larger transformers, e.g., a 100 MVA rated transformer, may cost > $2 million [47]. Comparatively, the cost of a small lithium-ion battery system can be as low as $3, 000 [18], which presents considerable savings for the utility. However, since transformers in the distribution grid vary in size, they will need different size batteries to achieve a certain level of load reduction. To calculate the energy storage capacity required per transformer, we propose a simple peak-shaving algorithm that clips the maximum contiguous peak above a given threshold. Specifically, for each transformer, our algorithm scans over its load and computes the contiguous period when the load exceeds the threshold. Our algorithm then computes energy storage capacity by computing the energy above the threshold across the periods, and selects the maximum. Our premise is that the energy storage that can flatten the maximum contiguous peak can also provide energy to flatten the smaller peaks experienced at other periods.

Figure 14(a) shows that energy storage capacity can vary between 1 kWh and 532 kWh. We note that the 90th percentile of energy storage is 50 kWh, which indicates that even a small battery size can dramatically reduce the risk of failures in transformers. In particular, 85% of overloaded and critically overloaded transformers can benefit from an energy storage capacity of 20 kWh or less installed per transformer. Since the battery capacity is a function of the size of the transformer capacity, we plot the distribution of battery size against transformer size. Figure 14(b) shows the median energy storage capacity increases with increases in transformer capacity, with the highest capacity required occurring at 112.5kVA. We hypothesize that such transformers are connected to a large number of homes such as an apartment building. The larger energy storage capacity can be attributed to the higher number of homes that such transformers serve.

We then analyze the effect of adding storage as a function of transformer sizing. We define *battery scale factor* as a 1:1 ratio to a transformer's kVA rating, i.e., for each kVA, what effect would adding 1 kWh of storage to the transformer have on the transformer's overload status. Figure 14(c) shows the result of adding storage using this factor. By adding a 0.1 factor of storage, we are able to reduce the number of heavily utilized and overloaded transformers by up to 45% and 39%, respectively.



(a) (b) (c)

Fig. 14. (14(a)) Energy storage capacity required to limit utilization to no more than 125% across all transformers, and (14(b)) distribution of energy storage capacity needed to limit overloading based on transformer capacities, and (14(c)) reduction in number of overloaded and heavily utilized transformers with increasing battery scale factor.

Lastly, we evaluate the impact of deploying grid-level storage under budgetary constraints. Typically, utility companies have limited budgets for grid upgrades, and it is important to ensure that the limited resources are utilized in a manner that maximizes the gains made in the grid. To model budgetary constraints, we assume that utility companies are limited in the total amount of storage they can deploy across the grid, i.e., in kWh. Next, since transformer replacement cost is directly proportional to the size of the transformer, we assume that deploying storage at a transformer with a higher transmission rating adds more value than deploying at one with a lower transmission rating. This is because decreasing the lifetime of larger transformer incurs a higher replacement cost than a smaller one. However, since the amount of storage that can be deployed is finite, this introduces a tradeoff between the number of transformers at which storage is deployed and the net value gained from such deployment. For instance, is it more beneficial to deploy storage at many small transformers vs one large transformer?

Formally, this problem of distributing finite storage across transformers of different sizes can be formulated as the classic knapsack problem in combinatorial optimization. Here, the amount of storage required by each transformer is the *weight*, the transformer kVA rating is the *profit value*, and the budget limit is the *capacity*. We implement this formulation using our dataset and analyze the tradeoff between number of transformers at which storage is deployed and the overall net gain. Figure 15 depicts the results of this analysis. The figure shows that at lower budgets, the net value gained increases super-linearly as the storage budget increases. However, at higher budgets, the net value gained increases sub-linearly with an increase in storage budget. This is mainly because larger transformers are mostly selected at higher budget allocations. Lastly, the figure shows that deploying ≈25% of the total storage required to eliminate transformer overloading in the grid results in the best net value tradeoff.

In summary: *Energy storage capacity of 20 kWh or less installed at each transformer is capable of reducing the number of overloaded and critically overloaded transformers by up to 85%. This indicates that transformer overloading can be significantly reduced by only a modest investment in grid level energy storage of 20 kWh or less per transformer.*

8 RELATED WORK

In this section, we present related work in the areas of load modeling, load profiling and grid integration with energy storage.

Distribution Grid Network. There have been numerous studies on the distribution network [5, 33, 50]. For instance, [5] studied the grid's resilience to disruptions in the distribution network.

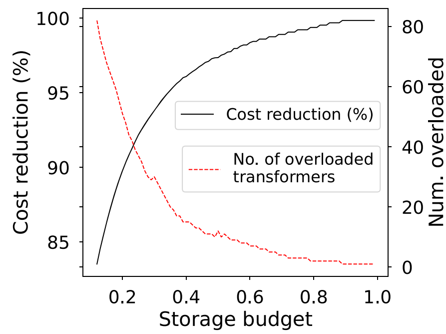


Fig. 15. Tradeoff between the number of overloaded transformers and net value gain under budgetary constraints.

Others have studied the feasibility, or have examined the cost-benefit analysis, of integrating renewables in the distribution network [33, 50]. However, these studies do not analyze the load on distribution edge transformers. Prior work has also studied the impact of load on transformer lifetimes [10, 22, 23, 48, 51]. These approaches provide thermal modeling of transformers, and examine how load and external factors affect transformer lifetime. Our work is complementary to this work, as we provide a broader analysis of the current state of distribution edge transformers in a city over a 1-year period. Prior work has also studied demand patterns at both the household and grid level [4, 26, 31, 46]. These include studies to understand the types of demand profiles for setting power tariffs or enabling demand-response programs [38, 54]. Again, our work differs, as we focus on classifying load profiles across edge transformers, and characterize the current state of the grid to study the effect of emerging technologies, such as integration with energy storage.

Load Modeling. From the utility's perspective, load modeling is necessary to systems planning, operations and maintenance [37]. From a user's perspective, modeling energy consumption enables decision making on energy usage and helps reduce their energy consumption. Prior work on load modeling include both—a user's energy consumption and aggregated load at the grid level. Kolter and Ferreira used monthly data from 6,500 buildings to model and predict energy usage using features of the buildings [30]. Niu et al. presented a load forecasting model for electric utility companies using 7,200 power load recorders in the year 2004 [37]. However, our work analyzes the load both at the grid and individual level with the data from 13,458 electric meters to provide insights into the energy consumption of residential buildings.

Load Profiling. Prior work on customer segmentation includes studies to determine classes of load profiles. Techniques such as k-means, artificial neural networks, hierarchal clustering, HMM, self-organizing map, and so on, have been used to cluster load profiles [4, 14, 17, 31, 46]. Our technique is similar to the one discussed in [31], which selects appropriate number of clusters through an adaptive k-means approach. As discussed in Section 5.2, we employ visual model selection to arrive at an optimal number. Load profiling is determined for a variety of applications in utility companies. One of the common applications of load profiling is determining power tariff structures [38, 44]. Another key area application of load analysis is the field energy justice, where equity and justice in energy usage are analyzed [12, 53]. Other applications of load profiling include consumer-specific demand-response programs [54], creating in-depth customer portfolio [17] and so on. Apart from the load profiling analysis on 220k smart meters installed in California [31] that used hourly data, most datasets used for load profiling are in the hundreds [44]. In contrast, our dataset comprises energy readings from over 13,000 electric meters and at a much finer granularity of 5 minutes.

Energy storage. Prior studies have explored the benefits of using energy storage in conjunction with renewable energy [20, 24, 29, 45]. These studies focus on control policies to meet certain cost objectives. In addition, the use of energy storage has been studied in the context of *load shifting*, where energy storage charges itself during periods of excess generation (or off-peak pricing periods) and discharges when the demand is high [15, 16, 25, 43]. Similarly, prior work has proposed algorithms to shave peaks at both the individual home or the grid level [36, 40, 41]. Again, our work is complementary, as we explore energy storage both at the home and grid level to mitigate the effects of transformer overloading at city-scale.

9 CONCLUSIONS

In this paper, we conducted a wide-ranging analysis on electric consumption of a city-scale dataset. We identified and quantified the general trends/patterns observed in individual homes, distribution transformers, and the city as a whole. We also studied the impact of weather on the aggregate energy demand. We observed that extreme weather events significantly increase energy usage, e.g., by 36% and 11.5% on hot summer and cold winter days, respectively. Further, we decomposed homes and transformers into different groups based on the characteristics of their load profile. Our analysis revealed that 53.6% of homes exhibit bimodal energy usage and 26.2% and 20.1% of homes exhibit unimodal daytime and nocturnal energy usage, respectively. We also found that this directly translates into transformer loads, with bimodal load transformers accounting for 43.9% of the entire grid. We then examined mitigation strategies for reducing transformer overloads using grid-level and home-level energy storage. Our results indicate that both mitigation strategies can reduce overloading in edge transformers. At 50% home battery penetration, we can reduce the number of critically overloaded transformers by up to 63%, and by deploying energy storage of up to 20 kWh per transformer, we can reduce the number of overloaded and critically overloaded transformers by 85%. We expect this work to spur multiple avenues for future work, i.e., an in-depth study into the actual costs and tradeoffs of various mitigation strategies, the impact of transformer overloading on grid reliability, and lastly, the impact of grid reliability on energy justice among various demographics.

REFERENCES

[1] [n. d.]. Transformer Overcurrent Protection. https://www.electricityforum.com/td/utility-transformers/transformer-overcurrent-protection (Accessed on 08/04/2023).

[2] [n. d.]. U.S. Energy Information Administration (EIA). https://www.eia.gov/ (Accessed on 09/04/2020).

[3] 1986. IEEE recommended practice for protection and coordination of industrial and commercial power systems. *ANSI/IEEE Std 242-1986* (1986), 1–592. https://doi.org/10.1109/IEEESTD.1986.81071

[4] A. Albert and R. Rajagopal. 2013. Smart meter driven segmentation: What your consumption says about you. *IEEE Transactions on Power Systems* 28, 4 (Nov. 2013).

[5] R. Albert, I. Albert, and G. L. Nakarado. 2004. Structural vulnerability of the North American power grid. *Physical Review E* (2004).

[6] O. Ardakanian, C. Rosenberg, and S. Keshav. 2013. Distributed control of electric vehicle charging. In *Proceedings of the Fourth International Conference on Future Energy Systems*.

[7] Sean Barker, Aditya Mishra, David Irwin, Emmanuel Cecchet, Prashant Shenoy, Jeannie Albrecht and others. 2012. Smart*: An open data set and tools for enabling research in sustainable homes. *SustKDD, August* 111, 112 (2012), 108.

[8] J. P. Barton and D. G. Infield. 2004. Energy storage and its use with intermittent renewable energy. *IEEE Transactions on Energy Conversion* (2004).

[9] Christian Beckel, Wilhelm Kleiminger, Romano Cicchetti, Thorsten Staake, and Silvia Santini. 2014. The ECO data set and the performance of non-intrusive load monitoring algorithms. In *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-efficient Buildings*. 80–89.

[10] IS Board. 1995. IEEE guide for loading mineral-oil-immersed transformers. *IEEE Std C* (1995).

[11] Phuthipong Bovornkeeratiroj, John Wamburu, David Irwin, and Prashant Shenoy. 2021. VPeak: Exploiting volunteer energy resources for flexible peak shaving. In *Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*. 121–130.

[12] Tim J. Chatterton, Jilian Anable, Jo Barnes, and Godwin Yeboah. 2016. Mapping household direct energy consumption in the United Kingdom to provide a new perspective on energy justice. *Energy Research & Social Science* 18 (2016), 71–87.

[13] Y. Chen, A. Das, W. Qin, A. Sivasubramaniam, Q. Wang, and N. Gautam. 2005. Managing server energy and operational costs in hosting centers. In *ACM SIGMETRICS Performance Evaluation Review*.

[14] Abdolhosein Dehdashti, James Tudor, and Michael Smith. 1982. Forecasting of hourly load by pattern recognition a deterministic approach. *IEEE Transactions on Power Apparatus and Systems* 9 (1982).

[15] A. Di Giorgio, L. Pimpinella, and F. Liberati. 2012. A model predictive control approach to the load shifting problem in a household equipped with an energy storage unit. In *Control & Automation (MED), 2012 20th Mediterranean Conference on*. IEEE, 1491–1498.

[16] B. Dunn, H. Kamath, and J. Tarascon. 2011. Electrical energy storage for the grid: A battery of choices. *Science* (2011).

[17] C. Flath, D. Nicolay, T. Conte, C. Dinther, and L. Filipova-Neumann. 2012. Cluster analysis of smart metering data. *Business & Information Systems Engineering* 4, 1 (2012).

[18] Ran Fu, Donald Chung, Travis Lowder, David Feldman, Kristen Ardani, and Robert Margolis. 2016. *US Solar Photovoltaic System Cost Benchmark: Q1 2016*. Technical Report. National Renewable Energy Lab.(NREL), Golden, CO (United States).

[19] X. Gao, C. A. Schlosser, and E. R. Morgan. 2017. Potential impacts of climate warming and increased summer heat stress on the electric grid: A case study for a large power transformer (LPT) in the Northeast United States. *Climatic Change* (2017).

[20] N. Gast, J. Le Boudec, A. Proutière, and D. Tomozei. 2013. Impact of storage on the efficiency and prices in real-time electricity markets. In *Proceedings of the Fourth International Conference on Future Energy Systems*.

[21] M. H. S. M. K. Hashmi, Seppo Hänninen, and Kari Mäki. 2011. Survey of smart grid concepts, architectures, and technological demonstrations worldwide. In *2011 IEEE PES Conference on Innovative Smart Grid Technologies Latin America (ISGT LA)*. IEEE, 1–7.

[22] J. Hazra, K. Das, A. P. Kumar, B. Narayanaswamy, D. P. Seetharam, and N. Jespersen. 2013. Optimal utilization of power transformers through virtual sensing. In *Innovative Smart Grid Technologies Europe (ISGT EUROPE), 2013 4th IEEE/PES*.

[23] A. D. Hilshey, P. D. H. Hines, P. Rezaei, and J. R. Dowds. 2013. Estimating the impact of electric vehicle smart charging on distribution transformer aging. *IEEE Transactions on Smart Grid* (2013).

[24] L. Huang, J. Walrand, and K. Ramchandran. 2012. Optimal demand response with energy storage management. In *Smart Grid Communications (SmartGridComm), 2012 IEEE Third International Conference on*. IEEE.

[25] David Irwin, Srinivasan Iyengar, Stephen Lee, Aditya Mishra, Prashant Shenoy, and Ye Xu. 2017. Enabling distributed energy storage by incentivizing small load shifts. *ACM Transactions on Cyber-Physical Systems* (2017).

[26] S. Iyengar, S. Lee, D. Irwin, and P. Shenoy. 2016. Analyzing energy usage on a city-scale using utility smart meters. In *Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments*.

[27] J. A. Jardini, C. M. V. Tahan, S. U. Ahn, and E. L. Ferrari. 1997. Distribution transformer loading evaluation based on load profiles measurements. *IEEE Transactions on Power Delivery* (1997).

[28] Paul L. Joskow. 2012. Creating a smarter US electricity grid. *Journal of Economic Perspectives* 26, 1 (2012), 29–48.

[29] Y. Kanoria, A. Montanari, D. Tse, and B. Zhang. 2011. Distributed storage for intermittent energy sources: Control design and performance limits. In *49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*.

[30] Z. Kolter and J. Ferreira. 2011. A large-scale study on predicting and contextualizing building energy usage. In *AAAI*.

[31] J. Kwac, J. Flora, and R. Rajagopal. 2014. Household energy consumption segmentation using hourly data. *IEEE Transactions on Smart Grid* 5, 1 (Jan. 2014).

[32] Chun Sing Lai, Giorgio Locatelli, Andrew Pimm, Yingshan Tao, Xuecong Li, and Loi Lei Lai. 2019. A financial model for lithium-ion storage in a photovoltaic and biogas energy system. *Applied Energy* 251 (2019), 113179.

[33] D. Lew, N. Miller, K. Clark, G. Jordan, and Z. Gao. 2010. Impact of high solar penetration in the western interconnection. *Contract* (2010).

[34] Z. Liu, Y. Chen, C. Bash, A. Wierman, D. Gmach, Z. Wang, M. Marwah, and C. Hyser. 2012. Renewable and cooling aware workload management for sustainable data centers. In *ACM SIGMETRICS Performance Evaluation Review*.

[35] A. Mishra, D. Irwin, P. Shenoy, and T. Zhu. 2013. Scaling distributed energy storage for grid peak reduction. In *Proceedings of the Fourth International Conference on Future Energy Systems*. ACM, 3–14.

[36] A. Mishra, R. Sitaraman, D. Irwin, T. Zhu, P. Shenoy, B. Dalvi, and S. Lee. 2015. Integrating energy storage in electricity distribution networks. In *Proceedings of the 2015 ACM Sixth International Conference on Future Energy Systems*.

[37] D. Niu, H. Shi, and D. Wu. 2012. Short-term load forecasting using Bayesian neural networks learned by hybrid Monte Carlo algorithm. *Applied Soft Computing* 12, 6 (2012).

[38] H. Nizar, Z. Dong, and J. Zhao. 2006. Load profiling and data mining techniques in electricity deregulated market. In *IEEE Power Engineering Society General Meeting*.

[39] U.S. Department of Energy. 2012. Reliability Improvements from the Application of Distribution Automation Technologies - Initial Results. https://www.energy.gov/sites/prod/files/2016/10/f33/Distribution_Reliability_Report_-_Final_Dec_2012.pdf (Accessed on 08/04/2023).

[40] A. Oudalov, D. Chartouni, C. Ohler, and G. Linhofer. 2006. Value analysis of battery energy storage applications in power systems. In *Power Systems Conference and Exposition*. IEEE.

[41] A. Oudalov, R. Cherkaoui, and A. Beguin. 2007. Sizing and optimal operation of battery energy storage system for peak shaving application. In *Power Tech, 2007 IEEE Lausanne*. IEEE.

[42] Giuliano Andrea Pagani and Marco Aiello. 2013. The power grid as a complex network: A survey. *Physica A: Statistical Mechanics and its Applications* 392, 11 (2013), 2688–2700.

[43] P. Palensky and D. Dietrich. 2011. Demand side management: Demand response, intelligent energy systems, and smart loads. *IEEE Transactions on Industrial Informatics* (2011).

[44] B. D. Pitt and D. S. Kitschen. 1999. Application of data mining techniques to load profiling. In *Power Industry Computer Applications, 1999. PICA'99. Proceedings of the 21st 1999 IEEE International Conference*.

[45] J. Qin, Y. Chow, J. Yang, and R. Rajagopal. 2014. Modeling and online control of generalized energy storage networks. In *Proceedings of the 5th International Conference on Future Energy Systems*. ACM.

[46] Teemu Räsänen, Dimitrios Voukantsis, Harri Niska, Kostas Karatzas, and Mikko Kolehmainen. 2010. Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data. *Applied Energy* 87, 11 (2010).

[47] A. A. Romero, E. E. Mombello, and G. Ratta. 2012. An overview on power transformer management: Individual assets and fleets. In *2012 Sixth IEEE/PES Transmission and Distribution: Latin America Conference and Exposition (T&D-LA)*. IEEE, 1–7.

[48] M. J. Rutherford and V. Yousefzadeh. 2011. The impact of electric vehicle battery charging on distribution transformers. In *Applied Power Electronics Conference and Exposition (APEC), 2011 Twenty-Sixth Annual IEEE*.

[49] John Shugars. 2002. Replacing distribution transformers a hidden opportunity for energy savings. *Teaming for Efficiency: Commercial Buildings: Technologies, Design, Performance Analysis, and Building Industry Trends* 3 (2002), 351.

[50] T. Stetz, F. Marten, and M. Braun. 2013. Improved low voltage grid-integration of photovoltaic systems in Germany. *IEEE Transactions on Sustainable Energy* (2013).

[51] G. Swift, T. S. Molinski, and W. Lehn. 2001. A fundamental approach to transformer thermal modeling. I. Theory and equivalent circuit. *IEEE Transactions on Power Delivery* (2001).

[52] Ark Tsisserev. 2020. Overloading of transformers–What are the code requirements - *IAEI Magazine*. https://iaeimagazine.org/columns/canadian/overloading-of-transformers-what-are-the-code-requirements/ (Accessed on 08/04/2023).

[53] John Wamburu, Emma Grazier, David Irwin, Christine Crago, and Prashant Shenoy. 2021. Towards equity in energy efficiency analyses. In *Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*. 259–263.

[54] T. Wijaya, T. Ganu, D. Chakraborty, K. Aberer, and D. Seetharam. 2014. Consumer segmentation and knowledge extraction from smart meter and survey data. In *SDM*.

[55] Alexander Zeh, Marina Rau, and Rolf Witzmann. 2016. Comparison of decentralised and centralised grid-compatible battery storage systems in distribution grids with high PV penetration. *Progress in Photovoltaics: Research and Applications* 24, 4 (2016), 496–506.

Received 16 January 2023; revised 4 August 2023; accepted 17 January 2023