

Analyzing Energy Usage on a City-scale using Utility Smart Meters

Srinivasan Iyengar, Stephen Lee, David Irwin, and Prashant Shenoy
University of Massachusetts Amherst

ABSTRACT

Understanding the energy usage of buildings is crucial for policy-making, energy planning, and achieving sustainable development. Unfortunately, instrumenting buildings to collect energy usage data is difficult and all publicly available datasets typically include only a few hundred homes within a region. Due to their relatively small size, these datasets provide limited insight and are insufficient for analyses that require a larger representation, such as an entire city or town. In recent years, utility companies have installed advanced electric and gas meters, i.e., “smart meters” that enable energy data collection on a massive scale. In this paper, we analyze such a dataset from a utility company that includes energy data from 14,836 smart meters covering a small city. We conduct a wide-ranging analysis of the city’s gas and electric data to gain insights into the energy consumption of both individual homes and the city as a whole.

In doing so, we demonstrate how city-scale smart meter datasets can answer a variety of questions on building energy consumption, such as the impact of weather on energy usage, the correlation between the size and age of a building and its energy usage, the impact of increasing levels of renewable penetration, etc. 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 observe that 700 homes are highly energy inefficient as its energy demand variability is twice that of the aggregate grid demand. Finally, we study the impact of increasing level of renewable integration in homes and show that solar penetration rates higher than 20% of demand increases the risk of over-generation and may impact utility operations.

CCS Concepts

•General and reference → Empirical studies; •Information systems → Data analytics; Clustering;

Keywords

Electricity, Grid, Cluster Analysis

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

BuildSys ’16, November 16 - 17, 2016, Palo Alto, CA, USA

© 2016 Copyright held by the owner/author(s). Publication rights licensed to ACM. ISBN 978-1-4503-4264-3/16/11...\$15.00

DOI: <http://dx.doi.org/10.1145/2993422.2993425>

1. INTRODUCTION

The building sector contributes an estimated 40% of the energy and 70% of the electricity consumed in the United States each year [4]. As a result, there is a significant interest in understanding and optimizing building energy usage. Recently, a number of inexpensive consumer/utility-grade 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 [4]. 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 [7, 8, 15]. Public datasets, including Pecan Street, the ECO dataset, the UMass Smart* 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.

With the increasing deployment of smart electric and gas meters across their entire customer base, utility companies are well-positioned to conduct more comprehensive larger scale energy data studies. We argue that utility-scale studies can reveal more detailed insights across a larger population. In addition, as a centralized administrative domain, utilities have the ability to leverage these insights to increase energy-efficiency at large scales, e.g., by implementing energy-efficiency programs. In this paper, we present a city-wide analysis of energy usage data from electric and gas meters over a 15 month period. The anonymized dataset is gathered from a small city in the New England region of the United States using 14,836 electric and gas meters deployed across the local utility’s entire customer base. In conducting our analysis, our goal is to understand the type of information utilities and government can learn from such large-scale smart meter deployments, which are becoming increasingly common. Specific examples of questions we seek to answer are below.

1. What is the distribution of electricity and gas usage across the customer base? How is electricity and gas usage influenced by weather, seasons, time of day and the day of the week?

2. What types of daily load profiles exist across the customer base and how do these load profiles relate to the daily routines within households?
3. How does electricity and gas usage change with the size of the home and age of the home? Are newer homes more energy efficient and, if so, by how much?
4. What is the distribution of energy-inefficient homes across the customer base? What fraction of customer homes are candidates for energy-efficiency improvements?
5. How do renewable sources, such as rooftop solar, impact the demand seen by the grid, as well as the energy fed back into the grid? How does the overall city-wide demand profile change with increasing penetration of renewables?

Our analysis answers these questions and provides a representative example of how utilities can leverage smart meter data to - i) improve their customer's energy-efficiency, ii) access impact of renewable integration, iii) discover specific customer load profiles and iv) quantify impact of weather etc. For example, the distribution of electricity and gas usage across homes can determine energy usage outliers, i.e. homes that use significantly more or less energy than other homes. Homes that use significantly more energy can be targeted for in-home energy audits or encouraged to participate in energy-efficiency programs. Likewise, homes that use significantly less energy can be inspected for malfunctioning meters or potential energy theft. In conducting our city-scale data analysis, this paper makes the following contributions:

Aggregate Demand Analysis. We first analyze the aggregate grid demand across all homes 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.

Individual Home Demand Analysis. We next analyze the load profiles of individual homes. Specifically, we decompose homes into different groups based on the characteristics of their load profile, which is based on the daily routines and patterns, as different groups will be amenable to different energy-efficiency optimizations. We also quantify the impact of building age, size, and type of heating system on energy usage. We then identify homes that are good candidates for energy-efficiency programs based on its variability in energy usage based on the weather. Such energy-efficiency programs might offer financial assistance for improving HVAC efficiency, e.g., by installing better insulation.

Impact of Renewables. Finally, we examine the impact of increasing levels of solar penetration on the grid, including ramp-up flexibility and the potential for over-generation. Specifically, we show that increasing solar penetration will require a significant increase in peaking power plant capacity (as much as $2.8\times$) and that solar penetration rates higher than 25% of aggregate demand increases the risk of over-generation and may impact utility operations.

2. BACKGROUND

Our paper focuses on analyzing and characterizing the energy usage across an entire city, both at the level of individual homes and at city-scale, using data collected from smart meters. A smart meter is a utility-deployed meter that enables fine-grained metering of energy usage and reporting in real-time—compared to traditional meters that are read manually, and report consumption at a coarse-grained level of once a month. Some smart meters have wireless

Num. of residential customers	11,431
Electric meters	11,186
Gas meters	3,650
Electric meter granularity	5 minutes
Gas meter granularity	1 hour
Duration	15 months

Table 1: Key characteristics of the data set

reporting capabilities that enable real-time usage reporting, as well as outage notifications.

The dataset used in this study was gathered from a small city in the New England region of the United States using a deployment of 14,836 electricity and gas meters from 11,431 residential customers; commercial customers are not included in the dataset and are not part of our present study. The electricity meters monitor and report usage at a resolution of 5 minutes, while the gas meters monitor and report usage each hour¹.

We analyze usage data for a 15 month period from October of 2014 to the end of 2015. Note that some customers have two meters, one for gas and one for electricity. Consumers that depend on other sources of fuel for heating (e.g. oil, propane, electricity) have only one meter for electricity. The dataset is anonymized and thus the mapping of usage data to specific customer addresses is not known. However, the dataset includes anonymized information about the approximate age of each residential building (e.g., decade the home was built) and the approximate size of the home. Table 1 depicts the key characteristics of the data set.

In addition to electricity and gas data, we also supplement our dataset with additional types of information.

1. **Solar generation data.** Some customers have rooftop solar that can net meter their energy back to the grid. The utility has deployed meters to monitor the electricity generated by these solar installations. Our dataset currently includes solar data for only two such anonymized homes.

2. **Weather data.** We gathered weather data at one-hour granularity for the city from the Weather Underground website².

Below, we analyze this city-scale dataset to answer questions about aggregate usage; the distribution of energy usage across customers; the relationship of energy usage and weather and building characteristics. Later, we discuss the different types of customer segments that we observe in our dataset. Further, we study the impact of age and size of the building on the energy consumption. Following this, we present our analysis in finding energy inefficient homes. Finally, we study the impact of locally generated renewables on the energy demand.

3. ENERGY USAGE: TEMPORAL, SEASONAL AND WEATHER IMPACTS

We begin our analysis by examining 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 week, seasons, and the weather on energy usage.

3.1 Temporal and Seasonal Analysis

We first analyze the distribution of electricity and gas usage across residential customers. Figures 1(a) and (b) depict the histogram of

¹The gas meters have smaller batteries than the electricity meters, such that more fine-grained metering is infeasible since it would reduce the battery life to less than 7 years (life of the meter).

²Weather dataset API: <https://www.wunderground.com/>

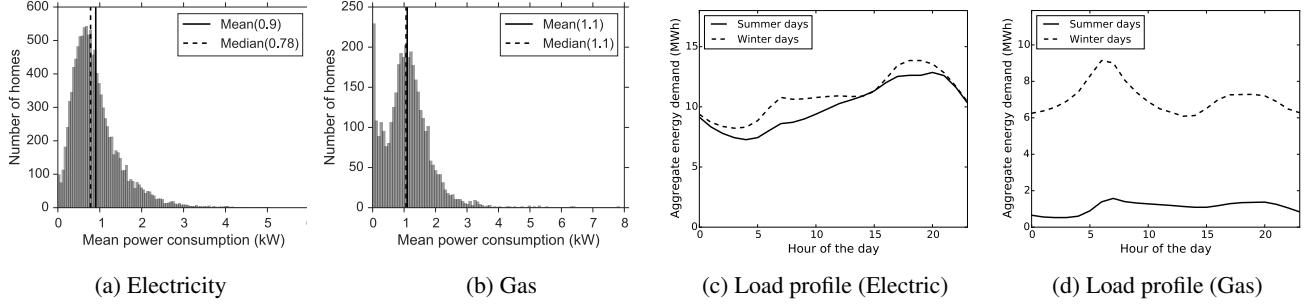


Figure 1: (a) and (b) Average electricity and gas consumption distribution of residential buildings in the year 2015. (c) and (d) Aggregate load profile of electricity and gas meters during summer and winter.

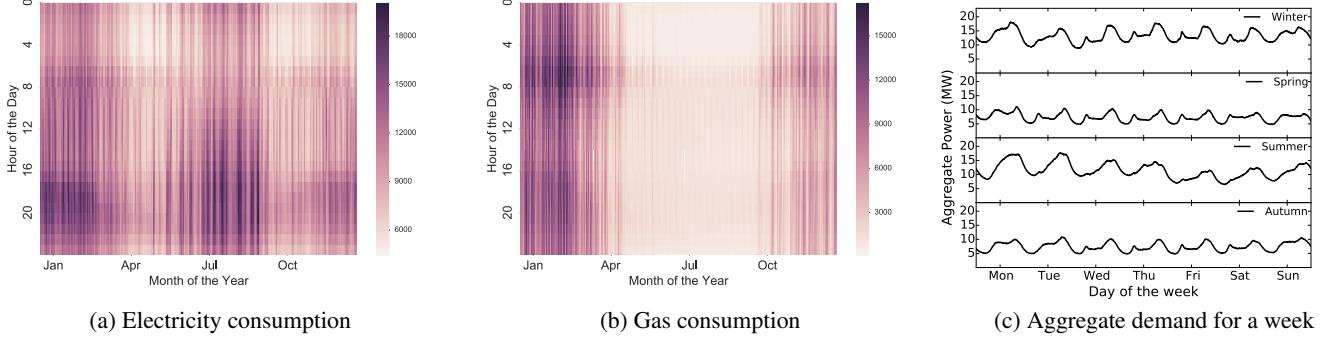


Figure 2: (a) and (b) Aggregate electricity and gas consumption from all residential buildings in the year 2015. (c) Periodicity in aggregate demand at the grid level for a representative week in each season.

average electricity and gas usage. Figure 1(a) shows a mean power consumption of 0.9 kW and the distribution shows a long tail where the 99th percentile of the consumption is 3.1 times the mean. The mean consumption is lower than the average usage of 1.24 kW reported for a typical US household [4]. Figure 1(b) plots a similar histogram for average gas power consumption and depicts an average consumption of 1.1 kW and the 99th percentile of consumption is 2.9 times the mean. Figures 1(c) and (d) 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. Thus, unless stated otherwise, winter days are defined as the period from Oct 2014 to Apr 2015, whereas summer days are defined as the period from May 2015 to Sep 2015.

Figure 1(c) depicts the electricity demand which varies between 7.2 MWh and 13.8 MWh — approximately a factor of 2 difference between off-peak and peak hours on average. The summer electricity demand shows two peak periods — a morning peak around 8 a.m. and an evening peak around 7 p.m. — presumably due to higher occupancy during the hours and usage of cooling equipment. The winter demand shows increases over the course of the day and a peak around 7 p.m. — a 25% and 10% increase during morning and evening peak hours over the same for the summer peak hours, respectively. 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. However, due to the proliferation of gas in some homes, the increase in electricity consumption is not markedly higher than the peaks observed in summer.

The gas demand shows the opposite behavior, as depicted in Figure 1(d). The winter energy demand shows the same morning and evening peaks, as shown by the electricity demand in summer —

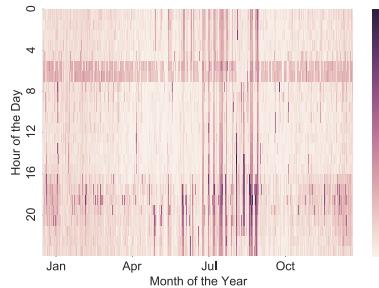
presumably due to the operation of heating equipment. Unlike the electricity demand profile, the morning peak is more prominent than the evening peak. Further, the winter demand varies from 6 MWh to 9.1 MWh and is nearly seven times that of an average summer gas demand. Gas usage is low in summer months since it is only used for cooking and hot water and not for heating homes.

Figures 2(a) and (b) are heat maps showing aggregate usage of electricity and gas for each hour (shown on the y-axis) of each day over the course of the year 2015 (shown on the x-axis), where darker colors indicate higher energy usage. In Figure 2(a) the aggregate electricity usage reveals the same morning and evening peaks are seen in Figure 1(c). 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. In Figure 2(b) reveals higher energy usage due to the use of gas heaters in colder months (Oct to Apr).

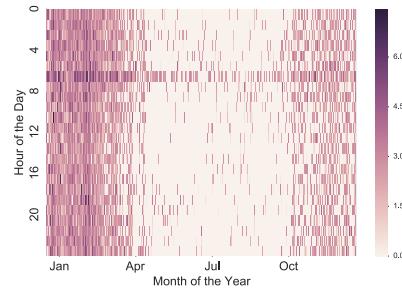
Figure 2(c) demonstrates the aggregate electricity demand over the course of an entire week for four seasons: winter, summer, fall, and spring. The figure shows that time of day effects for each day — the overall pattern repeats across seasons, although the magnitude of the peaks and the average usage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

Figure 3(a) and (b) are heat maps showing electricity and gas usage from an individual home respectively. In Figure 3(a) electricity usage pattern reveals clear peaks during morning and evening hours over the entire year with somewhat higher usage during summer evenings. Figure 3(b) reveals higher usage in winter months than summer months. The figure also reveals higher usage during the morning (around 7 a.m.) throughout the year — presumably due to the need for hot water for showers.

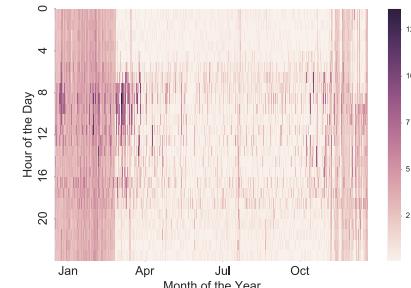
Figure 3(c) depicts a home with no gas meters with electricity providing heat during the winter. The figure shows higher electricity demand in winter for electric heating and also shows higher



(a) Home A: Electric meter data



(b) Home A: Gas meter data



(c) Home B: Electric meter data

Figure 3: (a) and (b) Energy consumption of a home using gas heater. (b) Energy consumption of a home using electric heater.

	Day Type	Summer (MWh)	Winter (MWh)
Electric	Business Days	246.2	260.47
	Non Business Days	228.78	263.69
	All Days	241.19	261.39
Gas	Business Days	26.18	166.28
	Non Business Days	26.13	177.55
	All Days	26.16	169.5

Table 2: Summary of the average aggregate daily energy demand at the grid level.

morning and evening peaks. It also reveals higher usage for a few days in August — presumably due to higher cooling demand.

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 gas demand is higher in winter due to the use of electric ACs and gas heaters, respectively.

3.2 Weather Influence Analysis

Next, we study the impact of outside temperature on the energy usage — both during winter and summer. Figure 4(a) and (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 4(a) shows a strong negative correlation of -0.81 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 4(b) shows a moderate positive correlation of 0.54 between temperature and electric usage in summer. Since the summer of 2015 was mild with a mean temperature of 65°F (18.3°C), there was less cooling demand. Further, due to generally milder summers, many homes do not have AC or run them infrequently, leading to a moderate (rather than high) correlation to outside temperature. The figure also shows that winter demand varies from 165.4 MWh to 367.9 MWh, while summer demand varies from 158.5 MWh to 356 MWh. In addition, the average absolute day-to-day change in the aggregate electricity demand is 5 MWh and 19.4 MWh in winter and summer, respectively.

Figure 5 shows similar results for gas — a strong negative correlation of temperature and gas usage is observed in winter, while a weak correlation is observed in summer. Since winters in the northeast US are harsh, with an average temperature of -0.5°C, all homes have heaters leading to the high correlation of gas energy usage to the temperature in colder months. The gas demand varies from 27.4 MWh to 336.7 MWh in winter, while it is mostly flat with an average demand of 26.1 MWh in summer. Further, the average absolute day-to-day change in the aggregate gas demand is 25.7 MWh

Events	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

Table 3: Criteria for extreme weather events.

and 1.4 MWh in winter and summer, respectively. Table 2 summarizes the average aggregate demand for both electric and gas meters across the two seasons for business and non-business days.

We also observe a strong positive correlation of 0.96 between the gas and electric aggregate demand during winter. This is because of low temperatures in winter that trigger use of electric and gas heaters in different households. However, in summer, we observe a moderate negative correlation of -0.63 between the gas and electric aggregate demand.

Figure 6 compares average daily summer and winter usage to “extreme” weather days (e.g. hot summer day to an average summer day). Table 3 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 6(a) shows demand can be 36% higher on a hot summer day than that on an average summer day — due to increased AC usage. The significant rise in usage during hot summer day may also be attributed to the relatively milder summer observed in that region in the year 2015. 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 and gas usage with temperature in winter. (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.

4. CUSTOMER SEGMENTATION: 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. 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 cus-

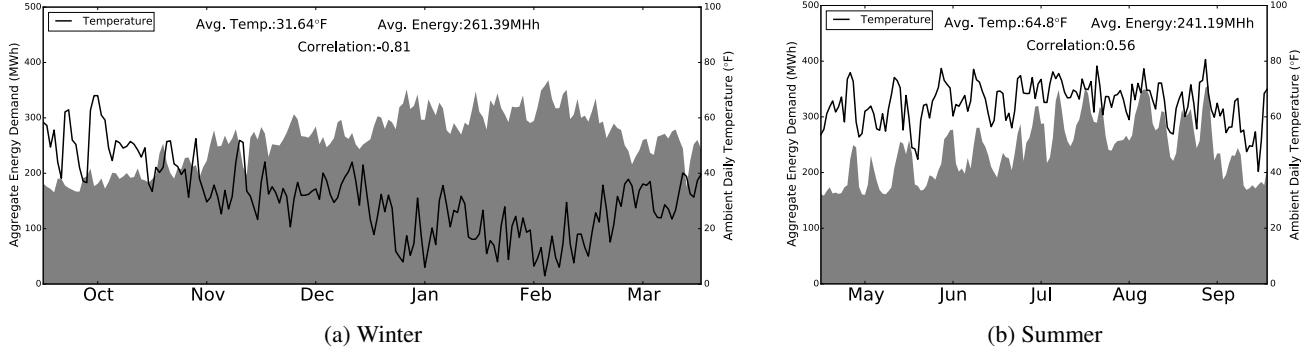


Figure 4: Variations in daily aggregated electricity consumption at the grid level for the entire year.

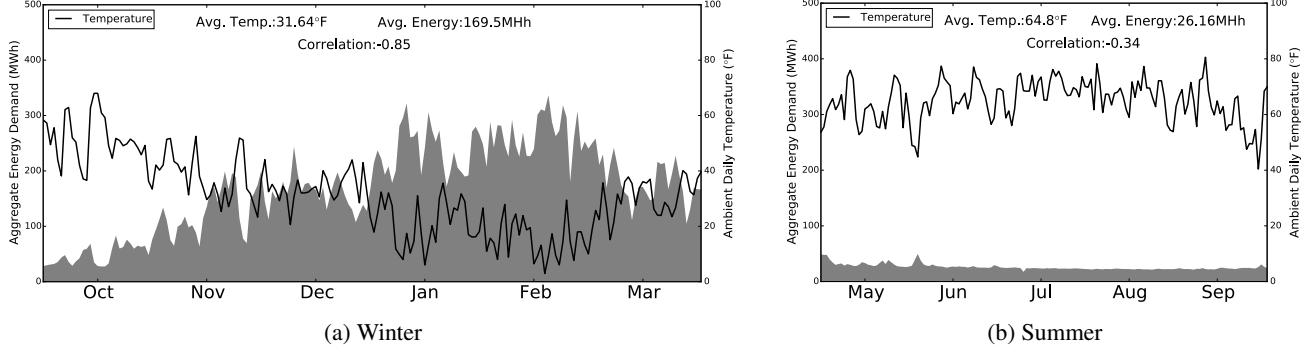


Figure 5: Variations in daily aggregated gas consumption at the grid level for the entire year.

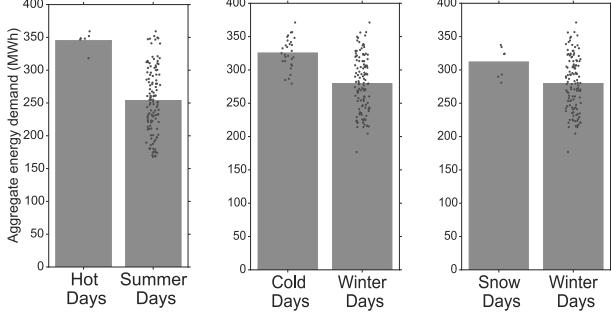


Figure 6: Electricity consumption for extreme weather events.

omer 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 the year 2015, 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 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.

Figure 7 shows the 8 clusters (customer segments) that resulted

from our analysis. The lightly shaded lines are the profiles of each home in present in that cluster and the bold line represents the centroid of the load pattern within each cluster. Broadly, there are 4 clusters that are bimodal with 2 peak usage periods of varying degrees and 4 clusters that are unimodal with a single peak usage period over the course of the day.

Table 4 summarizes the key characteristics of the customer segments within each cluster that include cluster type, peaks observed, the number of homes along with mean and standard deviation of daily energy consumption in kWh. As shown, around 6080 homes (54.4% of total) exhibit bimodal usage, 2953 homes (26.5% of total) exhibit unimodal daytime peak usage while 2143 homes (19.1% of total) exhibit "nocturnal" usage.

Figure 7(a)-(d) depict the four bimodal clusters. Figure 7(a) and (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 the 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.

Figure 7(e)-(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. Figure 7(g) and (h) represents nocturnal homes where the off-peak period occurs in the late morning or

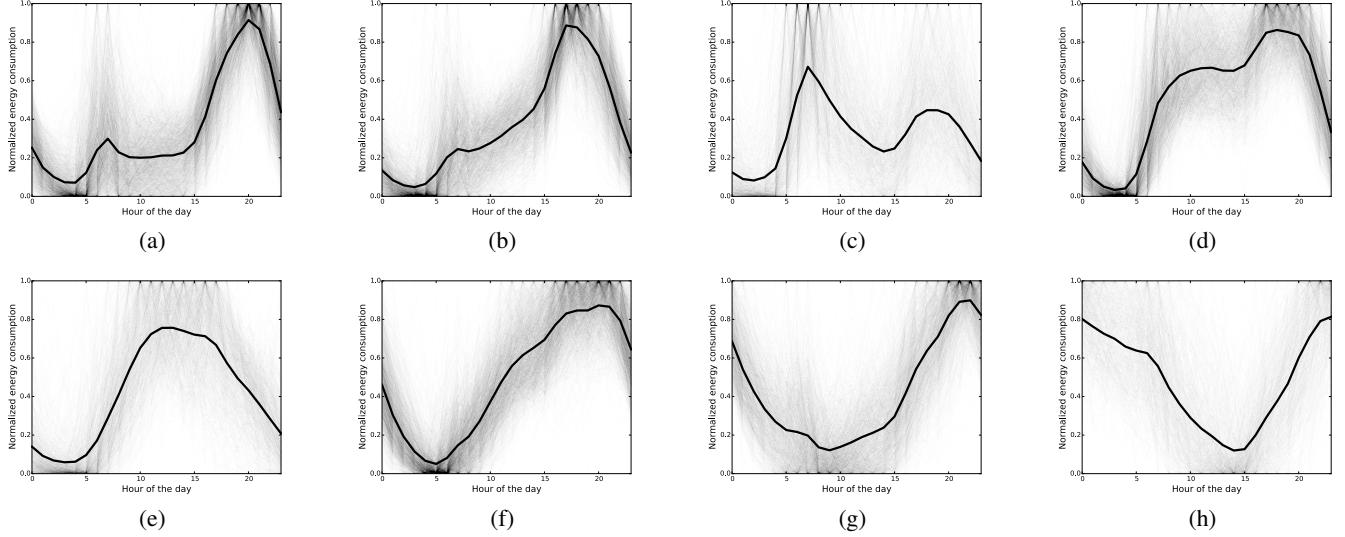


Figure 7: Load profile clusters of buildings observed across the entire dataset. Clusters (a) to (d) show a bi-modal distribution with two visible peak demand hours. Clusters (e) to (h) show a uni-modal distribution with peak at different time of the day.

Cluster	Type	Peaks	Homes	Mean	Std
a	Bi	7am, 9pm	1847	21.17	20.8
b	Bi	7am, 5pm	1773	22.24	22.06
c	Bi	7am, 6pm	813	24.3	28.55
d	Bi	12pm, 6pm	1647	25.35	24.63
e	Uni	12pm	956	20.06	21.56
f	Uni	9pm	2007	22.14	19.89
g	Uni	11pm	1410	18.87	16.26
h	Uni	12am	733	17.09	16.9

Table 4: Key characteristics of each customer segment.

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

In summary: Our customer segmentation reveals how the energy profiles correspond to their daily routines, with 54.4% of homes exhibit bimodal energy usage, whereas, 26.5% and 19.1% of homes exhibit unimodal daytime and nocturnal energy usage, respectively.

5. IMPACT OF AGE AND SIZE ON ENERGY USAGE

The overall energy consumption of a residential building depends on a variety of factors, such as the number of residents, the number and type of appliances in the home, the size and age of the building, the material used for construction, and so on. Typically, larger a home, the more energy is needed for lighting, appliances and to heat or cool its premises. Similarly, with improving building construction technologies in recent decades, newer homes tend to be more energy-efficient than older homes. Of course, these factors may not always determine energy usage of a home. For instance, two homes of the same size may have vastly different energy consumption depending on the occupancy patterns and appliance usage within them. An older home that has been renovated and modernized may actually be quite energy efficient despite being an old home. In this section, we study the correlation between the age and size of a home and its overall energy usage to quantify how energy usage rises with increasing home sizes and increasing age.

Although our data set is anonymized, only a subset of the dataset includes coarse grained information on the age of the home, in terms of the decade it was built, and the size of each home (e.g., less than 1000 sq. ft, 1000-1500 sq. ft, 1500-2000 sq. ft, 2000-3000 sq. ft and greater than 3000 sq. ft). We use this data to study how the energy usage varies with increasing age and size and also how it varies within each “bucket”.

Figure 8(a) shows the annual energy consumed (Gas + Electric) for the year 2015 by homes within each age and size group. As shown in the figure, within each age group, as the house size increases, so does its energy usage. We observe that a typical 2000-3000 sq. ft. home consumes, on an average, 26.8% more energy than a small 1000 sq. ft home. Each error bar in this bar graph also depicts the variance in energy usage (95% confidence interval) observed in each group. We note that the variance in very old homes—those built in the 19th century, i.e., before 1900—exhibit high variance. As some of these homes may have undergone extensive remodeling, some of these old homes are substantially more efficient than others that have not undergone such renovations.

The figure also depicts how energy usage of homes within the same size group varies with age. Typically within each size group, the energy usage reduces when we move from pre-1900 homes to the 1950-1975 built homes, indicating that relatively newer homes consume less energy than an older home within a size group. However, within each size group, there is an uptick in annual energy usage in more after 1975. Although such an increase in energy consumption for the newest homes may seem surprising, we attribute it to newer homes having more electrical appliances than older homes. For example, it is now typical for a new home in the Northeastern U.S. to include air conditioning, while many older homes lack an AC due to the mild climate. Thus, while the newest homes are more energy efficient based on building construction technology, they also have more appliances and HVAC systems that contribute to an increase in their energy usage when compared to similar sized homes that were built earlier.

Figure 8(b) the annual energy consumed (Gas + Electric) for the year 2015 in homes that use gas or electricity for heating. Like before, the figure shows that the average energy consumption, in general, increases as the size of the home increases regardless of

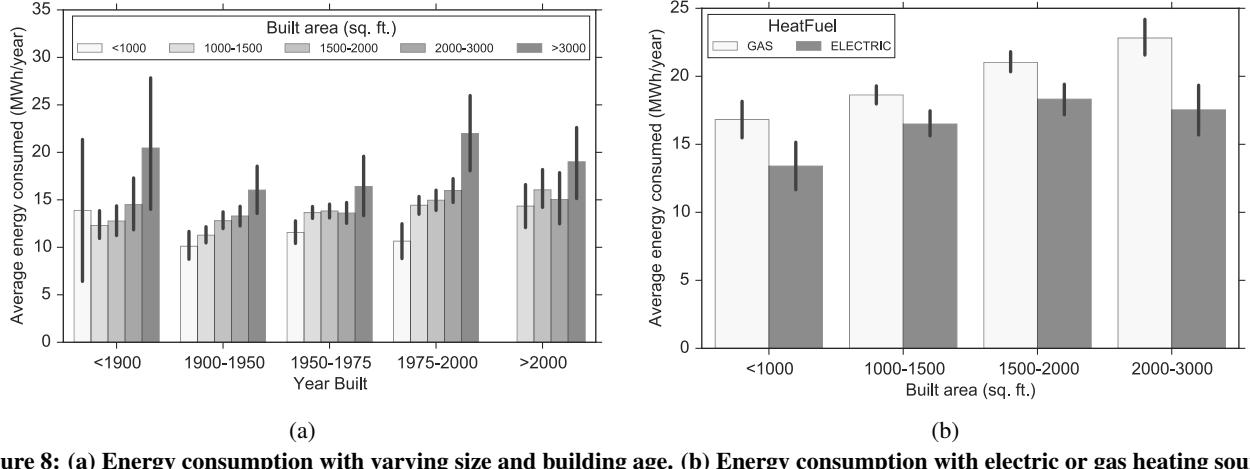


Figure 8: (a) Energy consumption with varying size and building age. (b) Energy consumption with electric or gas heating source.

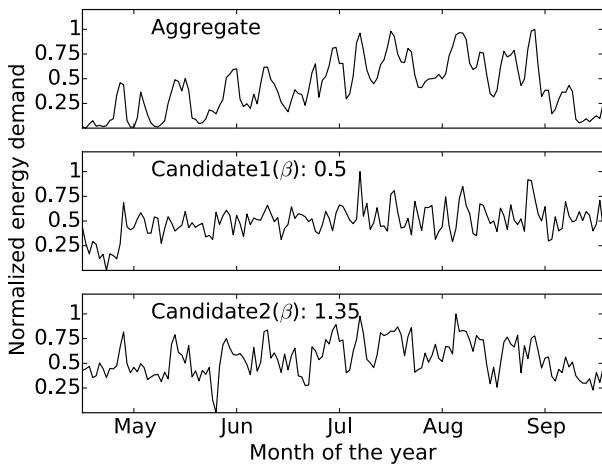


Figure 9: Illustrative example to show variability of two homes in comparison with the aggregate energy demand.

heating fuel type. Interestingly, the figure also reveals that homes that use gas for heating consume more energy than homes that use electric heaters — presumably because more homes that use gas are centrally heated, while homes with electric heaters may use room-specific heating rather than coarse-grain zone-specific heating. Other factors such as the efficiency of gas and electric heaters may also contribute to higher consumption. The efficiency of electric heaters is higher than gas heaters, as they convert are 100% of electric energy to heat.

In summary: (1) There is a higher variance in energy usage within a particular size group for very old homes depending on the amount of remodeling over the years. (2) Relatively newer homes tend to consume less energy within each size group, but the newest homes exhibit an increase in energy use presumably due to a larger number of electrical loads and appliances. Within each age group, energy consumption increases with home size.

6. ENERGY EFFICIENCY ANALYSIS

In section 3, we established the influence of weather on energy consumption of residential buildings. However, the weather influence is variable across homes as its energy efficiency may differ due to the type of home insulation used, HVAC systems etc. Also, inefficient homes are more sensitive to weather changes due to poor heating/cooling in these homes. In this section, we study the

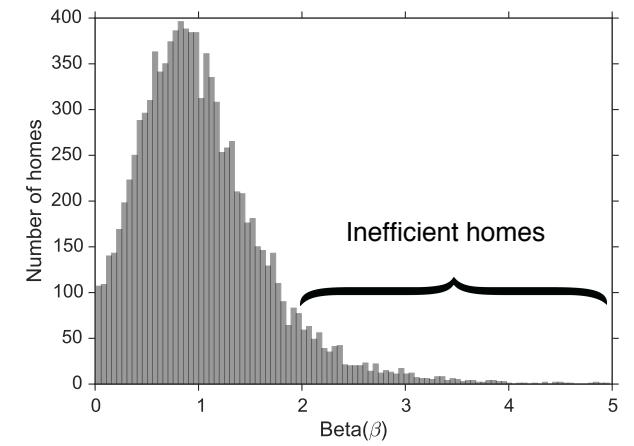


Figure 10: Frequency distribution of Beta for all the homes. Homes with $\beta > 2$ show high variability compared to the aggregate demand and can be targeted for energy audits.

variability in energy consumption of individual homes and suggest methods to determine inefficient ones. Here, we compare energy demand variability in individual homes with that of the aggregated energy demand of residential homes. Comparing with the aggregate energy demand gives us insight into how sensitive an individual home is to weather changes relative to an average home. Further, this information can be leveraged to conduct targeted energy efficiency audits of residential homes.

To measure the variability in energy usage of a candidate home relative to that of the aggregate energy demand, we introduce a metric: *Beta* ($\beta_{candidate}$) given by the equation below:

$$\beta_{candidate} = \frac{Cov(\Delta E_{candidate}, \Delta E_{aggregate})}{Var(\Delta E_{aggregate})} \quad (1)$$

where, function *Cov* and *Var* are the covariance and variance, respectively; $\Delta E_{candidate}$ and $\Delta E_{aggregate}$ are the relative difference in energy usage of the candidate home and the aggregate demand between subsequent days.

The *Beta* ($\beta_{candidate}$) is a standard metric used in time-series analysis and commonly used in the finance domain to measure the correlated relative variability. In particular, the β metric is used to measure the volatility of stocks (individual homes, in our context) with respect to a benchmark index (aggregate demand, in our context). For instance, a $\beta = 1.5$ of a home suggests that when the

aggregate energy demand increases by 10%, the demand from this home increases by 15%. Similarly, a $\beta = 0.5$ of a home suggests that when the aggregate energy demand increases by 10%, the demand from this home increases by only 5%.

Here, we focus on the months with higher temperatures, i.e. May to Sep. During these months, variation in daily energy use of a home is dominated by HVAC systems such as ACs, AHUs, etc. We do not consider data from gas meters, as their consumption during these months is not influenced by weather (see Figure 5(b)). Figure 9 illustrates the effectiveness of β in identifying homes with larger variability. This figure shows the normalized energy demand for the (i) aggregate of all homes, and (ii) two candidate homes from May to Sep 2015. As shown in the figure, *Candidate1* is less volatile over the period compared to the aggregate demand with $\beta_{Candidate1} = 0.5$. Thus, we conclude that *Candidate1* is less sensitive to weather changes (and thus more efficient) compared to the average aggregate energy profile. Conversely, we find that *Candidate2* follows the grid usage more closely albeit with a slightly larger variation. With $\beta_{Candidate2} = 1.35$, *Candidate2* is a relatively inefficient home. Although, it must be noted that this metric may not identify *all* inefficient homes. For example, in summer, an inefficient home without an AC will have a lower *Beta* than an efficient home with one. However, a very high *Beta* (> 1) indicates that a home is more sensitive than an average home.

Figure 10 shows the frequency distribution of *Beta* for all the homes in the dataset. We find that more 1850 homes have a $\beta > 1.5$ and more than 700 homes have a $\beta > 2$. Higher the β value, greater the sensitivity of the home to outside weather conditions. Essentially, these homes have a higher potential for energy efficiency improvements through actions such as custom retrofits, changing HVAC systems etc. A similar analysis for winter months would involve using the aggregate demand from the gas and electric meters.

In summary: With $\beta > 2$, over 700 homes have twice the variability than the aggregate energy demand and can possibly benefit from energy efficiency programs.

7. IMPACT OF RENEWABLE SOURCES

We examine the intermittency from rooftop solar installations from electric meters in our dataset across different seasons. Further, we illustrate the impact on daily electricity profiles of a household due to solar adoption. Finally, we present a case-study where we report the impact of increased solar penetration on grid stability.

7.1 Rooftop Solar generation analysis

Figure 11(a) shows the power generation trace for a day in April 2015 from one of the solar energy smart meters. During this week, we observe three days (April 16, 17 and 20) when the PV panel generates power at its max capacity of 8 kW due to clear sunny weather conditions. However, on April 18, 19, and 22, we observe that the solar energy is highly variable due to frequent cloud cover. During these days, we observe the solar power generated varying between 2 kW to 8 kW. We also observe little solar generation on April 21 due to an overcast weather. The maximum power recorded on this day is lower than 1kW. Thus, even for days in the same season, the power generation can be highly variable.

As discussed earlier, the average load profiles for homes have similar energy usage patterns, albeit the magnitude of the energy consumption varying on a day-to-day basis. Further, the load profiles of residential buildings with a rooftop PV installation have higher variations due to the added stochasticity in solar generation. We illustrate this through Figure 11(b), which shows the electricity consumption along with solar generation for a household on Jan 6, 2015. The house has a relatively stable power consumption of

around 2-4 kW for most of the day. The spotted region in the figure exhibit the *net* demand seen by the utility. We observe that solar energy is primarily generated between 8 a.m. to 5 p.m. Further, we notice that from 11 a.m. to 4 p.m., the home, for the most part, is a net supplier of energy with a peak net generation of around 4 kW. The area marked in gray exhibits the solar energy net metered to the grid. Again, the amount of energy net metered is quite variable for the reasons discussed earlier in this section.

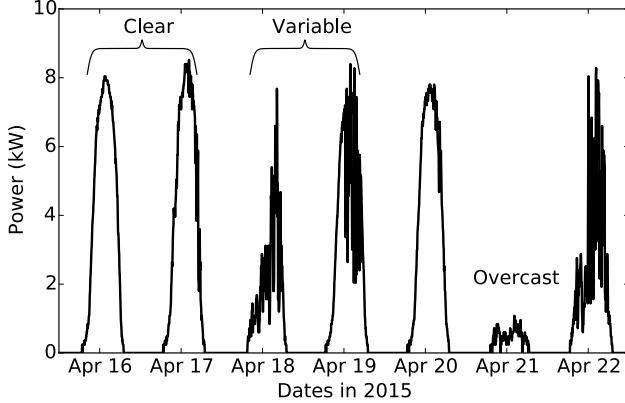
Figure 12(a) and (b) shows the average daily energy generated by the same PV panel installation. The scatter plots show all the samples for select days. Figure 12(a) compares the energy generated during all the summer days with rainy summer days. In summer, the daily energy generated vary between 8 kWh to 58 kWh with an average of ≈ 38 kWh. On rainy summer days, the daily energy generation is between 9 kWh to 32 kWh with an average of 23 kWh. Figure 12(b) shows the average daily energy consumption for the different set of months in a year representing each season. Due to inclement weather in the Northeast U.S. during winter months of Dec-Feb and shorter day lengths, the average energy generated is 12 kWh with quite a few days recording lower than 5 kWh. During autumn (Sep-Nov), the average daily energy generation is around 27 kWh and is more than double the winter months. During spring (Mar-May) and summer (Jun-Aug) months, the PV panel generates a daily average of 36 kWh and 39 kWh respectively. However, the range of energy generation during spring is higher than that of summer. This is because of late winter weather with shorter day lengths that contribute to lower energy generation in March. On the other hand, longer day lengths with moderately lower temperatures contribute to a higher energy generation in May.

7.2 Case Study: Impact of distributed solar generation

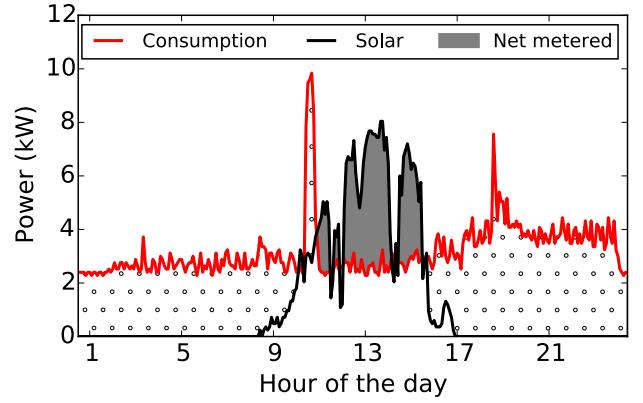
Utility companies have greater operational control over the power supply using traditional sources of power generation such as hydro and thermal generators. However, variability and high-penetration of solar have introduced new challenges in maintaining the power output. Using this case study, we seek to address the following questions: (i) How does net demand change with varying levels of solar integration? (ii) How quickly will the utility need to ramp-up to meet the demand? (iii) Does the utility run the risk of overgeneration at current demand levels?

To simulate the impact of increased solar penetration on the power grid, we combine our solar power trace for the entire year and assume its average to be the solar output from a single home. We then compute the net energy demand after addition of solar for varying levels of solar penetration in residential homes. Figure 13 shows the net demand with solar energy observed on an average day. As seen in the graph, the shape of the net energy demand, known as the *duck curve*, resembles a sitting *duck* for a high solar penetration level. The *duck curve* can be divided into four regions, with two alternating periods of ramp up and ramp down of power supply. Due to the shape of the energy demand, utility companies run into two problems: *ramping-up* of power when solar power is unavailable, and *overgeneration* of solar power in the afternoon.

Ramp-up flexibility. Usually, peaking plants are provisioned after exceeding base load power capacity to meet the peak demand. Integrating solar to the grid adds additional load to peaking plants that it may not be provisioned for. The figure shows the ramp-up capacity needed for different solar penetration levels when the grid is integrated with solar. We observe two ramp-up periods, the first starting at 4:00 a.m and the other at 1 p.m. Further, the ramp-up capacity needed in the first period (between 4am and 7am) is much lower than the second half (between 1pm to 7pm), and re-

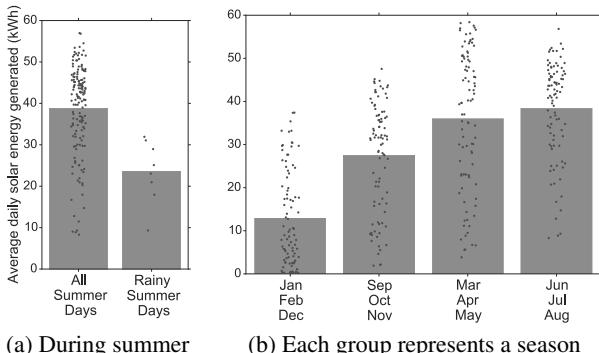


(a) Solar generation trace for a week



(b) Solar generation and electricity consumption trace for a day

Figure 11: Solar generation trace from one of the solar meters from the data set.



(a) During summer (b) Each group represents a season

Figure 12: Solar energy generated on (a) an average day in summer and (b) during different seasons in the year 2015.

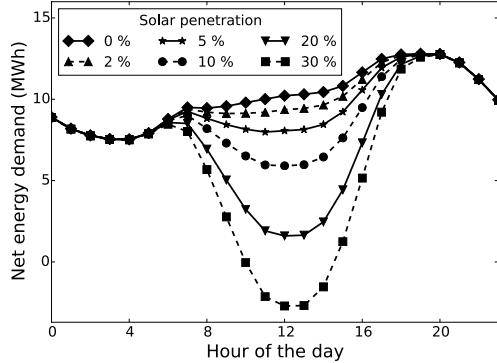


Figure 13: Impact of solar penetration on power grids.

quire higher capacity peaking plants. In particular, with 5% solar penetration in homes, the capacity of peaking power plants needed to serve the first ramp-up period almost doubles during the second ramp-up period. Also, when solar penetration level increases from 0% to 2%, the ramp-up capacity increases by 29.1%. With a high solar penetration of 20%, the increase in ramp-up capacity is almost 2.8 times that of current ramp-up capacity.

Overgeneration. The risk of overgeneration occurs during the periods of high solar output when the supply exceeds the net demand causing an imbalance in the grid. Further, most power generators (such as hydro and thermal) have minimum operating levels worsening this imbalance. Intervention in real-time energy market

may balance out the surplus energy by dropping spot prices to low or even negative values. In the past, prices of spot prices were negative for a short period in Denmark and Germany due to surplus in solar energy [9]. Depending on the utilities' minimum operating levels and distribution portfolio, the graph indicates that it may run the risk of overgeneration in the afternoon. Since the solar output is maximum in the afternoon, it is expected to observe a drop in the net demand. With a solar penetration of approx. 25%, if the current demand continues for the next few years, the net demand runs the risk of being negative.

8. RELATED WORK

In this section, we present some related work in the areas of load modeling, load profiling and grid integration with solar. However, most of the work discussed below do not analyze a city-scale data from multiple aspects.

Load Modeling. From the utility's perspective, load modeling is necessary to systems planning, operations and maintenance [11]. From a user's perspective, modeling energy consumption enables decision making on energy usage and help reduce their energy consumption. Prior work on load modeling include both — a user's energy consumption and aggregated load at the grid level. Kolter et. al. used monthly data from 6,500 buildings to model and predict energy usage using features of the buildings [7]. Niu et. al. presented load forecasting model for electric utility companies using 7200 power load recorders in the year 2004 [11]. However, our work analyzes the load both at the grid and individual level with the data from 14,836 electric and gas meters to provide insights into the energy consumption of residential buildings.

Load Profiling. Prior work on customer segmentation include studies to determine classes of load profiles. Techniques such as k-means, artificial neural networks, hierachal clustering, HMM, self-organizing map etc. have been used to cluster load profiles [8, 6, 3, 1, 14]. Our technique is similar to the one discussed in [8], which selects appropriate number of clusters through an adaptive k-means approach. As discussed in section 4, we employ visually model selection to arrive at an optimal number. Load profiling is determined for a variety of applications in utility companies. One of the common application of load profiling is determining power tariff structures [12, 13]. Other applications of load profiling include consumer-specific demand-response programs [15], creating in-depth customer portfolio [6] and so on. Apart from the load profiling analysis on 220k smart meters installed in California [8] that

used hourly data, most datasets used for load profiling are in the hundreds [13]. In contrast, our dataset comprises energy readings from over 14,836 electric and gas meters and at a much finer granularity of 5 minutes.

Grid Integration with Solar. Numerous studies have been done to analyze the impact of renewable integration with grid [9, 10, 5]. Lew et. al. studied the imbalance in voltage due to a high level of rooftop PV penetration for a city [10]. Prior studies also include analysis of transient stability and steady-state stability in the grid due to rooftop PV penetrations [5].

9. DISCUSSION AND FUTURE WORK

Our paper differs from prior studies discussed above in terms of - i) the number ($>14k$) and types (Electric and Gas) of smart meters, ii) location (Northeast US), and iii) the breadth of analysis (energy efficiency, customer segmentation, renewable integration) etc. Our work, larger than most studies on electric metric data, also presents an analysis of gas consumption that is primarily used in homes for providing heat during winter. Further, the geographical region of our study (Northeastern US) is climatically quite different than many other regions where a similar study (e.g California [8]) has been conducted. Northeastern US experiences extreme winter temperatures that could be as low as $-25^{\circ}F$ with mild summers. Consequently, the daily load profiles and yearly consumption patterns differ significantly. However, the techniques used in this paper are not tuned to work with just our dataset and thus can be leveraged for similar studies for datasets from other geographies.

As discussed in section 8, there exist multiple papers studying smart meter data. However, most studies restrict their analysis to just a specific problem such as identifying distinct load profiles, studying the impact of renewables etc. Nonetheless, the breadth of analysis discussed in the paper is not only more comprehensive but also presents the interplay of different factors (impact of weather on renewable energy or energy consumption; or impact of renewable adoption on customer and utility load profiles). Thus, a comprehensive analysis, like ours, is more insightful.

In future, we plan to conduct a deeper analysis involving demographic information, appliance usage etc. We would like to compare and augment the research work with similar studies presented in the literature [1, 2].

10. CONCLUSION

In this paper, we conducted a wide-ranging analysis on electric and gas consumption of a city-scale dataset. We identified and quantified the general trends/patterns observed in individual homes and the city as a whole. We also studied the impact of weather, age and size of the building 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 into different groups based on the characteristics of their load profile to identify daily routines and patterns. Our analysis reveal that 54.4% of homes with bimodal energy usage and 26.5% and 19.1% of homes exhibit unimodal daytime and nocturnal energy usage, respectively. We also observed that while bigger homes exhibit higher energy usage, older homes show higher variance in energy usage depending on the renovations done over the years. In addition, our analysis show that 700 homes are highly energy inefficient as its energy demand variability is twice that of the aggregate energy demand. We also studied the impact of increasing level of solar penetrations in homes and show that the rates higher than 20% of demand increases the risk of over-generation and may impact utility operation.

11. ACKNOWLEDGEMENTS

We would like to thank the anonymous reviewers and our shepherd Rishee Jain for their feedback. This research was supported in part by NSF grants IIP-1534080, CNS-1405826, CNS-1253063, CNS-1505422, and the Massachusetts Department of Energy Resources. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the funding agencies.

12. REFERENCES

- [1] A. Albert and R. Rajagopal. Smart meter driven segmentation: What your consumption says about you. *IEEE Transactions on Power Systems*, 28(4), Nov 2013.
- [2] N. Batra, A. Singh, and K. Whitehouse. Gemello: Creating a detailed energy breakdown from just the monthly electricity bill. In *KDD*, August 2016.
- [3] A. Dehdashti, J. Tudor, and M. Smith. Forecasting of hourly load by pattern recognition a deterministic approach. *IEEE Transactions on Power Apparatus and Systems*, 9, 1982.
- [4] U.S. Energy Information Administration. (visited on May 2016). <https://www.eia.gov/>, 2016.
- [5] S. Eftekharnejad, V. Vittal, G. Heydt, B. Keel, and J. Loehr. Impact of increased penetration of photovoltaic generation on power systems. *Power Systems, IEEE Transactions on*, 28(2), 2013.
- [6] C. Flath, D. Nicolay, T. Conte, C. Dinther, and L. Filipova-Neumann. Cluster analysis of smart metering data. *Business & Information Systems Engineering*, 4(1), 2012.
- [7] Z. Kolter and J. Ferreira. A large-scale study on predicting and contextualizing building energy usage. In *AAAI*, 2011.
- [8] J. Kwac, J. Flora, and R. Rajagopal. Household energy consumption segmentation using hourly data. *IEEE Transactions on Smart Grid*, 5(1), Jan 2014.
- [9] D. Lew, L. Bird, M. Milligan, B. Speer, X. Wang, E. M. Carlini, A. Estanqueiro, D. Flynn, E. Gomez-lazaro, N. Menemenlis, A. Orths, I. Pineda, J. C. Smith, L. Soder, and P. Sorensen. Wind and Solar Curtailment Preprint. 2013.
- [10] D. Lew, N. Miller, K. Clark, and G. Jordan. Impact of High Solar Penetration in the Western Interconnection Impact of High Solar Penetration in the Western Interconnection. 2010.
- [11] D. Niu, H. Shi, and D. Wu. Short-term load forecasting using bayesian neural networks learned by hybrid monte carlo algorithm. *Applied Soft Computing*, 12(6), 2012.
- [12] H. Nizar, Z. Dong, and J. Zhao. Load profiling and data mining techniques in electricity deregulated market. In *IEEE Power Engineering Society General Meeting*, 2006.
- [13] B. Pitt and D. Kitschen. Application of data mining techniques to load profiling. In *Power Industry Computer Applications, 1999. PICA'99. Proceedings of the 21st 1999 IEEE International Conference*, 1999.
- [14] T. Räsänen, D. Voukantis, H. Niska, K. Karatzas, and M. Kolehmainen. 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.
- [15] T. Wijaya, T. Ganu, D. Chakraborty, K. Aberer, and D. Seetharam. Consumer segmentation and knowledge extraction from smart meter and survey data. In *SDM*, 2014.