Better Understanding Customers: Developing SMB DNA to Improve Customer Interactions and Catalyze Positive Behavior Changes

Anne-Lise Laurain, Tingwen Bao, Pawel Zawadzki, Kevin Johnson, Pacific Gas & Electric Company

ABSTRACT

The small and medium businesses (SMB) segment has been underserved by Energy Efficiency (EE) programs due to the heterogeneity of the nature of their businesses. The SMB load segmentation project initiated by Pacific Gas & Electric Company (PG&E) uses advanced data analytics to extract features from smart meter interval data and then cluster customers into similar groups based on their usage behaviors. The load shapes for each SMB customer are coded to six representative load characteristics (genes): seasonality, weekly profile, peak hour, active hours, base load, and load factor. When combined with customer demographic information, including NAICS code, this "DNA" can be used to reveal gaps in EE portfolios, allow EE program managers to tailor their existing products and design new ones, and assist the customer targeting effort for demand response (DR) programs which would increase the effectiveness of marketing. The complex data analytics used help to succinctly describe the customers allows colleagues from non-analytical lines of business within PG&E to quickly and easily develop a better understanding of how customers use energy, how the utility can have a positive impact on their bills, and the electricity grid's resiliency.

INTRODUCTION

Advanced Meter Infrastructure and Interval Data

PG&E was one of the first electric utilities to deploy Advanced Meter Infrastructure (AMI) and create customized solution with the help of interval data. AMI was first rolled out in 2006 (PUC 2006) making it possible for PG&E to track and monitor gas and electric consumption at the individual customer level in a real time. Since then both real time and historical load data have become valuable resources for various lines of business. Example use cases include:

1. Evaluation of the effectiveness of energy efficiency, demand response, and distributed energy resources (DER) programs by comparing the customer load before and after adoption
2. Providing insight to grid maintenance and upgrade planning through forecasting the potential for peak time load shift due to changing customer demographics
3. Identification of DER customers or other customers with appliances like HVAC, by analyzing load patterns. Targeting those customers with programs/ products that can reduce their bill and benefit the electricity grid
4. Enhanced customer engagement by creating web tools for customer to track and manage their own energy use and costs (My Energy portal)

5. Identification of opportunities to delay or defer system maintenance through the identification of targeted non-wire alternatives (TDSM)

Small and Medium Businesses

Small and medium sized businesses comprise a very important customer sector for PG&E; SMB's contribute up to 22% of PG&E's overall revenue (PUC 2016). Despite this significant revenue ratio, the SMB segment is so diverse that it has been difficult for the utility and its Energy Solutions and Services (ES&S) account reps to develop a comprehensive understanding of SMB customers. A number of qualitative research studies have been conducted to provide better understanding of the size, industry composition, energy usage patterns and utility engagement preferences of SMB customers over the past several years.

Last year, PG&E initiated a quantitative effort to create more behavior-based segmentation (termed "cohort"). Six cohorts have been created, accounting for 66% of the SMB population, based on a wide range of customers' demographics and attitudinal variables. These variables include: PG&E product penetration and program participation, bill amounts, frequency of interaction, energy savings potential, etc. These cohorts have been used internally to:

- Provide guidance for product and program targeting
- Determine best communication channel and customize messaging for customer outreach

As PG&E develops a better understanding of the SMB customer segment via this cohort information, an increasing desire for energy usage information has arisen internally. To fulfill this desire for quantitative information, especially on energy usage, PG&E's Data Analytics and Governance (DAG) team began to develop a solution that was not only informative, but also more user-friendly to use than load curves. The goal is to summarize load shape data and visualization into actionable snippets that are easy to understand for utility colleagues in every line of business, no matter their technical backgrounds. As a result, the information could be used along with other customer segmentations to enhance PG&E's service to SMB customers, without requiring every utility employee to learn how to read and extract features from the load curves.

Past Studies on Customer Load Shapes and Segmentation

Creating, understanding and segmenting customer energy use profiles have attracted great interest in both utilities and academia. Past load shape studies have analyzed both customer and appliance level usage patterns. Joseph et al. created representative daily load shapes by season for various appliances, like refrigerators, by looking at 1,050 residential customer's load data from three IOUs in California (Eto et al. 1990). Another research team, Kwac et al., determined the 16 load shapes most frequent for residential households. This team also introduced the idea of energy entropy to measure the volatility of customer's load (Kwac, Flora and Rajagopal 2014). In this analysis, customers were grouped into 16 segments through multi-dimensional segmentation based on the quantile of both their energy usage and energy entropy.

Like the above mentioned work, past studies, have been focused on residential customers, who are a more homogenous group than SMBs. SMB, which we believe has the most heterogeneous load shapes but with the highest potential of energy savings, is not well explored to our knowledge. In addition, although a typical load shape analysis approach is good at answering what the customer groups like one attribute at a time, it is not the best method to uncover all aspects for every single customer in one analysis. To improve the understanding of SMB customers'' load shapes, the DAG team created the "SMB DNA" with the goal of:

- Simultaneously providing all load characteristics that are crucial for the utility
- Making the data easy to understand and use by analysts and business decision makers
- Allowing for comparison between customers and customer groups and usefulness as inputs for follow up analysis
- Being flexible enough to meet the needs of different lines of business and yet customizable upon request

METHODOLOGY

Data Source

The SMB DNA analysis uses customer interval data collected through PG&E's automated metering infrastructure (AMI). The population of analysis is currently limited to PG&E's small and medium commercial/industrial customer with active electric services and whose smart meters are enabled throughout 2015.

Unlike most industries that define customer size by their revenue, profit, or head counts, PG&E, as a utility, defines customer size based on energy usage. The customer size definition used in this analysis has been widely used within EE, DR, and Energy Solutions and Services (ES&S) teams in PG&E. It is defined as shown in Table 1 below:

Table 1. Definition of customer size

Customer Size	Definition
Small	annual electric usage < 40,000 kWh AND annual gas usage < 10,000 therms
Medium	any usage between small and large
Large	annual electric usage > 500,000 kWh OR annual gas usage < 250,000 therms

To generate DNA characteristics the DAG team used 2015 electricity interval data for 355,169 smart meters (service points, SP). This data covers commercial and industrial customers in 19 Divisions (see map on figure 1) and across 27 NAICs (North American Industry Classification System) at NAICs second level.

Figure 1: Map of the Divisions

Determining DNA Characteristics

Each line of business cares about different aspects of the load shape. For example, DR program managers want to know about a customer's ability to shift load during a very specific set of system peak hours. EE program managers, on the other hand, want to identify those customer with a relatively high base load compared to their peak demand or compared to businesses in the same sector.

To fully capture those needs, internal interviews and check-ins were conducted at several stages of this study to identify and understand load characteristics critical to each line of business. As a result, six load characteristics were determined to be the most informative—Annual Seasonality, Weekly Pattern, Peak Hour, Active Hours, Base Load, and Load Factor . These six factors could be considered the "genes" of a SMB customer's DNA.

DNA Genes

The DAG team generated these customer load shape characteristics, or genes, using SQL query language and the following methodology. Two pretreatments were used before generating genes:

- Append customer service start date to the interval data if the customer has started after Jan 1st 2015.
- Create a "Low Usage" flag to denote all customers whose monthly usage is less than 260 kWh for each of the 12 months in 2015. These customers are unlikely to benefit from EE and DR programs.

These two data pretreatments are designed to help easily identify SMB customers that would be a poor fit for EE and DR programs. New SMB customers with less than 1 year tenure would then be excluded from analysis that needs DNA to be representative of a full year of usage. This also screens out SMB customers with really low usage as a business. Those are normally vacant premises that would hardly be of interest for EE and DR programs.

After these pretreatments, DNA has been generated for all left customers based on the following definition:

Annual Seasonality. Assessing energy usage across the standard four seasons (spring, summer, fall, winter, see table) allows for changes in energy usage to be tracked against both general weather and economic cycles. Relatable examples may include a ski lodge that would receive a "winter-spring" label or a manufacturing facility that ramps up production ahead of the holidays, that would receive a "fall –winter" label. To be labelled as active during seasons requires having four consecutive months with an average kWh usage over 80% of monthly average kWh usage over the entire year. Customers have some active months but do not strictly cover any 'calendar season' (see Table 2) are defined as "No Clear Season".

Table 2: Definition of the Seasons

Winter	January 1st to March 31st
Spring	April 1st to June 30th
Summer	July 1st to September 30th
Fall	October 1st to December 31st

Weekly Pattern. Weekly patterns identify the day or string of days when a business is actively using electricity. Only days during the above mentioned active seasons are taken into consideration. Active days must be consecutive; otherwise the weekly pattern is classified as "Other". For example, a traditional elementary school will not have its summer usage included in the weekly pattern calculation, which would be defined as "Mon-Fri." To be labelled as active on a weekday requires having an average daily kWh usage over 80% of daily average kWh usage over course of a month.

Active Hours. Active hours represents the time of day when a customer is actively using electricity—at least 80% of the average hourly kWh within active season and weekdays. The active hours of a customer is represented by a start time to end time. If a customer has multiple active time slots within the same day, their active hours will be shown as "Multiple Time Slots". Note that the active hour period is not necessarily the open-for-business hours, for example a bakery may have active hours from 4 AM to 2 PM during which they are making pastries, but be open from 8 AM to 6 PM to serve customers.

Peak Hour. The peak hour is defined as the hour with maximum occurrence of the daily maximum across the 24 hours of an active day. Its occurrence should be over 20% of total occurrence to be considered as "peak hour". If no hour has over 20% occurrence for a customer, the peak hour will be defined as "No Recurrent Peak".

Base Load. Two base loads have been defined in this analysis—Active Base Load and Inactive Base Load. An Active Base Load is the median consumption during Active Hours. This reflects the level of kWh necessary to keep a business running on a normal business day. On the other hand, an Inactive Base Load is the median consumption during non-active hours. The Inactive Base Load reflects the level of kWh needed during off hours. Note that DG customer who regularly export electricity on active days do not have their Base Load defined.

Load Factor. Load factor is the ratio of maximum kWh usage divided by the active base load. This ratio ranges between 0 and1. The lower the load factor, the more volatile a customer's load is. In addition, the load factor also sheds light on the potential for shifting load during peak hours.

RESULT AND DISCUSSION

General SMB DNA Trends

These six characteristics have been generated for all SMB customers. See general trends for each characteristic below. As previously stated, load DNA is defined by a business' kWh consumption and may not reflect "open for business" hours. These results should be used in concert with other information in order to develop a more holistic understanding of the customer. For example, a "summer" user could either be a business only actively running in summer or a business with a relatively inefficient HVAC system.

Annual Seasonality. The dominant seasonality for SMB is "All-Year", which indicates the business has been open and running all year round. It characterized 32.7% of the total non-low-usage SMB population. All Year was then followed by "summer", "No clear Season", "winter-spring-summer", and "spring-summer" which each independently accounted for around 9% of the population.

Weekly Pattern. The dominant weekly pattern is "All-Week", which indicates the business is actively using electricity seven days a week. It characterized 45% of the total non-low-usage SMB population. All-Week was followed by "Mon-Fri" which characterized 29% of the total population and "Mon-Sat" that characterized 11% of the population.

Active Hours. The most common active hour period is "12:00am-11:59pm" which characterized 9.6% of the non-low-usage SMB population. "12:00am-11:59pm" customers are those who have relative steady electric usage throughout the active days. It then followed by "8:00am-5:00pm", "8:00am-6:00pm", and "7:00am-5:00pm" which characterized 4.3%, 3.2%, and 3.1% of the population respectively. When combined the top 15 active hour periods represent 42% of the SMB active user population, of these 15 all but one start between 7AM and 9AM and end between 4PM and 9PM.

Peak Hour. 47.8% of the total population did not have a peak hour that occurs over 20% of the active days. Beyond this population segment, the most common peak time is 3pm (6.5%), followed by 4pm (4.2%), 2pm (3.7%), and 8pm (3.6%). Peak hour is distributed relatively evenly throughout normal business hours with three peaks, one at in the early morning (10am), one in the early afternoon (3pm), and one in the early evening (8pm) (Figure 2).

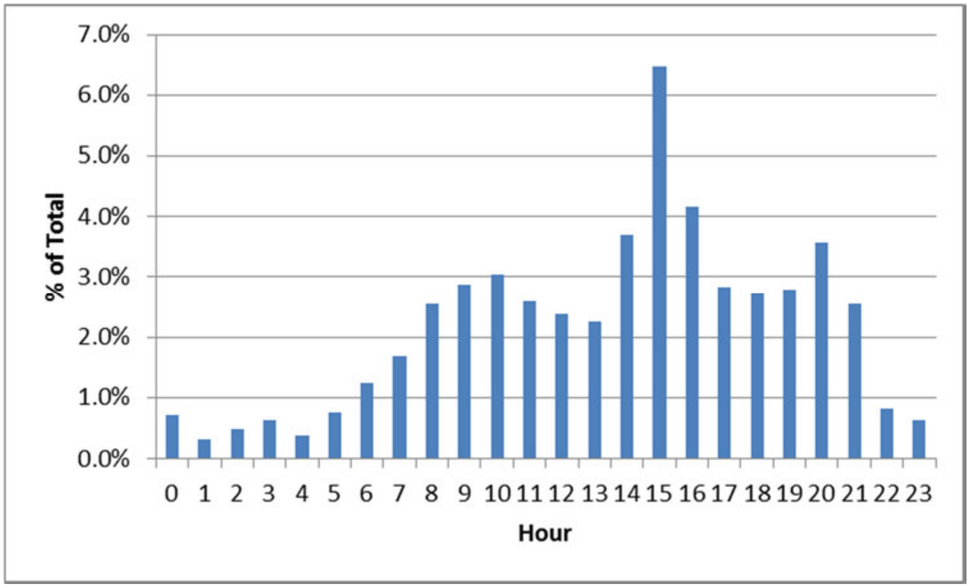

Figure 2. Distribution of SMB Peak Hours

Figure does not contain Low Usage SMB customers, customers with less than a full year of interval data, or customers without a recurrent peak hour

Base Load. The most common Active Base Load is around 2.5-10 kWh which characterized over 40% of the SMB active non-DG customers. A second common Active Base Load cluster range is between 0.5 and 2 kWh, which characterized another 35% of customers. The majority of customers were characterized with an inactive base load less than one kWh, 57% of the active non-DG population. This indicates that most businesses have relatively low kWh need during their off hours. An outlier customer with higher active and/or inactive base load compared to their industry peers could indicate a less efficient appliance which might be a great target for EE products.

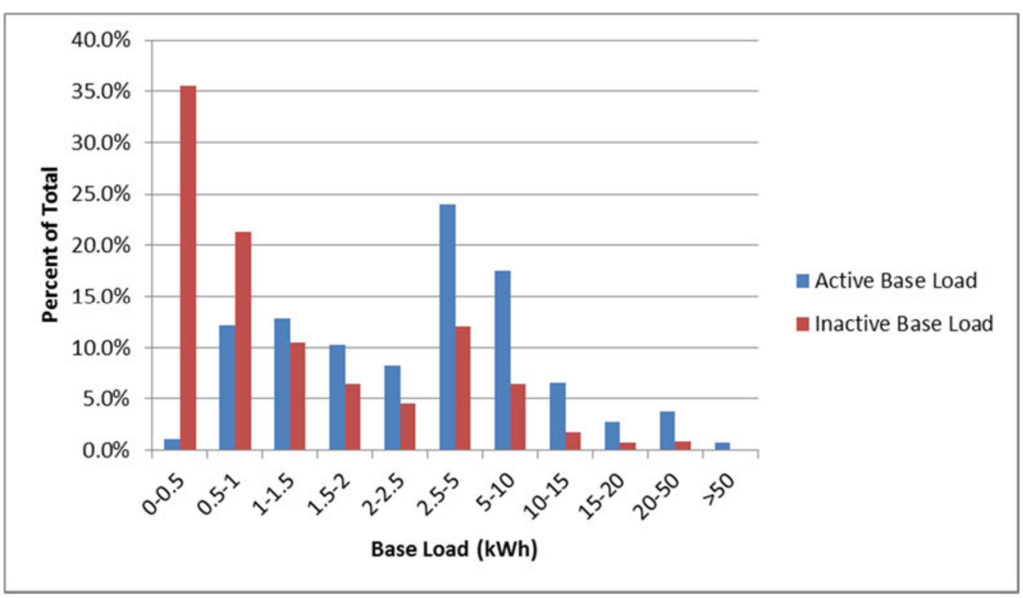

Figure 3. Distribution of Active Base Load and Inactive Base Load

Figure does not contain Low Usage customers, customers with less than a full year of interval data, or DG SMB customers

Load Factor. Most SMB non-low-usage customers had a load factor between 0.2-0.5, see Figure 4. This means that the customer's maximum kWh is about two to five times higher than the average kWh usage in their active hours. Because there are a number of reasons for a low load factor, load factor should be used in combination with other DNA characteristics in order to provide accurate guidance and should only be compared against customers in the same industry and general location.

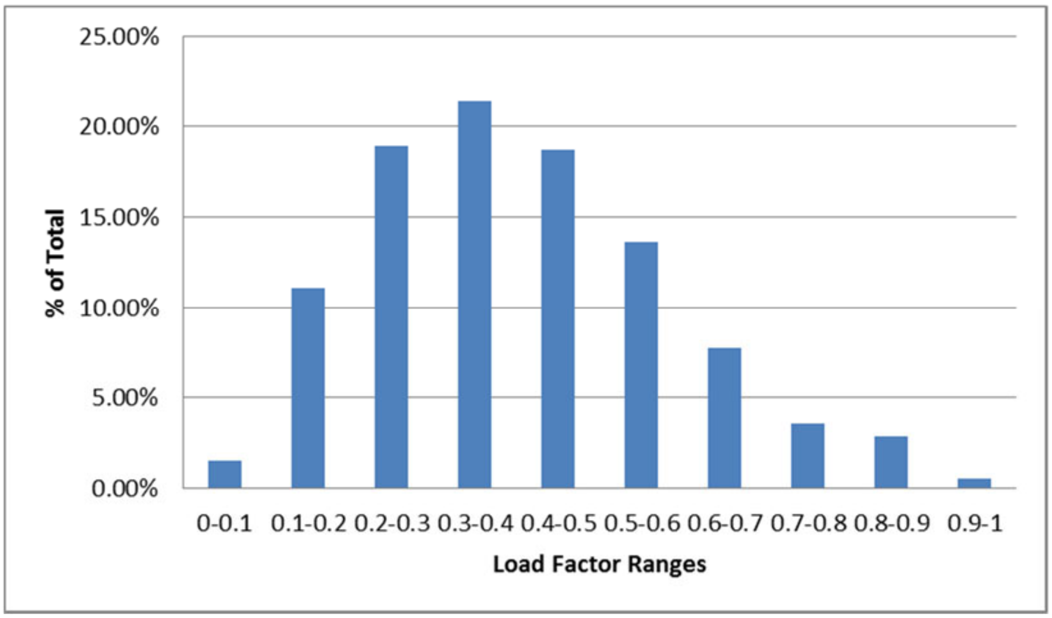

Figure 4. Distribution of SMB Load Factor

Figure does not contain Low Usage customers or customers with less than a full year of data

Diversity of Industry Segments DNA

Determining customers' DNA in mass makes it possible to understand the homogeneity and heterogeneity of energy usage patterns within and between industry segments. This enables the utility to better understand the unique electric usage behavior of each industry and further customize EE solutions and customer services based on each industry's norms.

As an example, Table 3 and Table 4 below compare the load shape characteristics of the Educational Services (NAICs 62) and Accommodation & Food Services (NAICs 72) industry segments. Their most common active days and peak hours are different from each other, but consistent with the nature of their business. Nearly half of the Educational Services open Monday to Friday whereas around 79% of the Accommodation & Food Services open seven days a week (Table 3). In terms of their peak hour, Educational Services are more likely to peak in early morning and early afternoon, during school hours, whereas Accommodation & Food Services' peak at noon and dinner time when most people have their meal (Table 4).

Table 3. Comparison of 3 most common weekly patterns between Educational Services and Accommodation & Food Services

	Naics2=62		Naics2=72	
	Educational Services		Accommodation & Food Services	
Top 3 Weekly Pattern	Mon_Fri	49.70%	All_Week	78.80%
	All_Week	16.20%	Other	10.70%
	Mon_Sat	13.80%	Mon_Sat	6.90%

Table 4. Comparison of the peak hours between the Educational Services and Accommodation & Food Services industry sector

Peak Hour	Naics2=62 Educational Services	Naics2=72 Accommodation & Food Services
0	0.4%	1.2%
1	0.2%	0.2%
2	0.1%	0.2%
3	0.2%	0.2%
4	0.1%	0.4%
5	0.2%	0.9%
6	1.2%	1.1%
7	1.9%	1.5%
8	6.2%	2.2%
9	3.4%	3.0%
10	2.2%	3.8%
11	2.7%	4.6%
12	1.8%	6.4%
13	2.6%	5.5%
14	5.0%	3.0%
15	3.2%	3.3%
16	5.5%	3.1%
17	4.3%	6.5%
18	4.2%	10.0%
19	3.1%	8.0%
20	2.0%	6.5%
21	0.6%	3.6%
22	0.1%	0.9%
23	0.2%	0.6%
No_Recurrent_Peak	48.8%	23.3%

While usage characteristics are generally homogeneous within an industry segment, at an individual customer level however, there can be significant differences within the same industry segment. For example, although the dominate weekly pattern is the same for all 3 subsectors of the Accommodation & Food Service industry segment, shown in Table 5 (79%, 69%, and 85% respectively). The peak hour of each subsector is different. As shown below, unlike "Food Service Places" which peak during lunch and dinner time, "Alcoholic Beverage Drinking Places" commonly peak at both dinner time and midnight when people are more likely to consume alcohol? In contrast, "Traveler Accommodation" businesses peak in the morning and again around 9pm when most travelers are preparing for and settling down from their day.

Table 5. Comparison of the peak hours within the Accommodation & Food Service industry sector

Peak Hour	Naics 7224 Drinking Place (Alcoholic Bev)	Naics 7220 Food Services-Drinking Places	Naics 7211 Traveler Accommodation
0	14.2%	0.6%	0.7%
1	1.9%	0.1%	0.1%
2	0.8%	0.1%	0.0%
3	0.2%	0.1%	0.1%
4	0.1%	0.1%	0.1%
5	0.1%	0.3%	0.4%
6	0.3%	0.5%	1.3%
7	0.5%	0.9%	1.9%
8	0.3%	1.2%	5.2%
9	0.2%	2.2%	2.7%
10	0.8%	4.3%	0.5%
11	0.8%	6.4%	0.5%
12	0.9%	8.0%	0.9%
13	0.9%	7.1%	0.4%
14	1.4%	2.9%	0.6%
15	1.7%	2.5%	0.8%
16	3.3%	3.2%	0.7%
17	6.8%	7.6%	1.4%
18	7.8%	13.0%	2.1%
19	5.9%	11.3%	2.6%
20	6.5%	6.2%	11.8%
21	3.4%	1.5%	18.5%
22	1.6%	0.6%	2.5%
23	4.2%	0.3%	0.7%
No_Recurrent_Peak	35.5%	19.0%	43.3%

Industry segments have long been used as one of the major tools for product and program targeting. Table 5, however, contests the accuracy of utilizing the sector averages compared to subsectors. By categorizing load via the DNA method, PG&E is able to better understand energy use behavior in both aggregate and at the individual business level. Based on this understanding, each utility line of business is able to choose the most suitable level of detail, tailor the campaign messages, and focus on the customers most likely to benefit from a product or program.

Next Steps

This analysis is the first step into analyzing SMB customer's usage patterns. Some of the immediate next steps are to:

1. Make this load characteristic information easily available by creating an internal tool with SMB DNA information and customer profile
2. Better understand the homogeneity and heterogeneity of SMB's load shape for each industry sector by analyzing the representative load shapes and outliers
3. Advocate for customer targeting based on load and demographic characteristics

After the DNA characteristic have been tested and further proved by the immediate use above, some of the long term efforts are:

1. Determine the DNA characteristics for the rest of PG&E's customer segments (residential, large commercial, industrial, agricultural, government, etc.). The type of characteristics will be adjusted according to the nature of the customer group and the need from each utility line of business
2. Keep customer DNAs update to date and in a database that is well maintained

Acknowledgement

We would like to thank our colleagues from Customer Energy Solution and Marketing who provide valuable insights and feedbacks that greatly assisted this analysis. We would like to acknowledge JP Dolphin and Ye Lin for comments that greatly improved the manuscript. We are also immensely grateful to Grace He and Timotej Gavrilovic, for their wisdom and support, the project will not be successful without them.

REFERENCES

Public Utility Commission. 2006. Decision 06-07-027

Public Utility Commission. 2016. Advice Letter: 4946-E-A

Eto, J. H., Ruderman, H., Golan, A., Heinemeier, K. E., and Wood, D. J. 1990. *Analysis of California Hourly End-Use Residential Electricity Load Shapes.* LBNL. eaei.lbl.gov/publications/analysis-california-hourly-end-use

Kwac, J., Flora, J., and Rajagopal, R. 2014. "Household Energy Consumption Segmentation Using Hourly Data." IEEE Transactions on Smart Grid 5 (1): 420-430.