A Simple Way to Use Interval Data to Segment Residential Customers for Energy Efficiency and Demand Response Program Targeting

Brian Arthur Smith, Pacific Gas and Electric Company
Jeffrey Wong and Ram Rajagopal, Stanford University

ABSTRACT

Interval data from automated metering infrastructure ("AMI") has become widely available to utilities throughout the country. Interval data offers utilities the ability to characterize customers based on load profiles. Customer segments with energy usage profiles that match system-wide demand may be good households to target demand response ("DR") programs because they may have greater potential for shifting loads off peak than current participants in DR programs. Since AMI data is relatively new to many researchers, this paper addresses this key research question: can basic statistical techniques be used to identify customer segments with unique energy use patterns? This paper presents an approach whereby AMI data has identified customer segments with similar load profiles for the potential use of targeting for residential DR programs. Follow-on research will layer attitudinal and demographic variables on the segments and apply them for targeting DR and energy efficiency ("EE") programs.

Use of Segmentation Schemes within the Utility Industry: A Review

There have been a number of attempts to use cluster analysis and similar statistical techniques to identify customer segments with similar behaviors, attitudes and perceptions for targeting of EE and DR programs. Lutzenhiser et al. (2009) and Dethman Associates (2009) have reviewed many of these approaches. Moss (2008) provides a comprehensive overview on the use of market segmentations developed for utility use. There are a number of published resources that provide useful information. For example, EPRI CLASSIFY, developed in the 1990s, is a psychographic segmentation scheme for energy-relevant purchases and energy use which uses a battery of questions to classify residential energy customers according to their "energy needs" on the basis of a battery of 120 questions (or a subset).[1] RESOLVE at University of Surrey has produced a number of papers and reports on environment and lifestyle mapping in the UK, including consideration of energy use.[2]

Since most utility-focused customer segmentations were developed before the widespread availability of AMI data, household energy consumption has been included only at an aggregate level (e.g., monthly-level billing data or total annual energy use). Instead, these segmentations have relied mainly on market survey data and other consumer attributes that

[1] EPRI CLASSIFY, developed in the 1990s, is a psychographic segmentation scheme for energy-relevant purchases and energy use which uses a battery of questions to classify residential energy customers according to their "energy needs" on the basis of a 120-question battery abstract that is available at http://my.epri.com/portal/server.pt?Abstract_id=000000000001016386. CLASSIFY manuals and reports are available online at www.epri.com. There are also more recent reports available for purchase or to EPRI members, e.g., the 2008 report Customer Preference and Behavior.

[2] See http://www.surrey.ac.uk/resolve/publications/index.php.

provide information to suggest that customers fall into particular lifestyle categories. As a result, the personas that embody segment attributes (e.g., "strugglers" who may be young renters and others with low income who use little energy and have little propensity to save; "green idealists" who may be homeowners who user higher amounts of energy and have greater propensity to save) have not been informed by daily patterns of energy use that would be most useful for targeting of residential DR programs. For example, the attitudinal segmentation battery developed by Opinion Dynamics Corporation used the results of a large sample telephone survey as a basis for a segmentation of California household has been useful for targeting the marketing outreach and education program overseen by the California Public Utilities Commission, but this segmentation does not include actual energy use behavior and so may not be particularly useful for DR program targeting.

Use of Residential Segmentations within PG&E

Using a broad range of attitudinal and demographic variables, PG&E's residential customer base has been segmented based primarily on "energy needs" and "energy involvement" variables. Using clustering algorithms, PG&E customers have been assigned to unique segments based on common attributes. Clusters defined by these demographic attributes (e.g., age, household size, presence of children, size and age of home, household income, climate, contact with PG&E) and indicators of engagement with energy and with PG&E (e.g., participation in EE and DR programs, use of utility online tools to pay bills and get information, rate plans) have been identified. Residential customers are assigned to one of approximately 11 clusters. These segments are used within PG&E to:

- Identify the best product or program for a similar group of customers
- Form a basis for creating messaging around relevant attitudes, behaviors and needs
- Determining methods of contact based on common media and channel habits within segments

Interval data: A New Paradigm for Segmenting Customers

AMI infrastructure is being deployed throughout the U.S. and represents an important new data stream for identifying customer segments for targeting of DR programs.[3] Household load shapes (i.e., hourly electric loads measured at the household level) captured by AMI systems reveal significant differences among large groups of customers in the magnitude and timing of their electricity consumption. Interval data provides a means to identify households that have significant potential for DR programs—perhaps even for EE programs as well. Once identified, targeted marketing may produce effective means of recruiting "high potential" customers to DR and EE programs. High potential customers are customers who have discretionary loads, who are likely to be attracted to utility marketing efforts, and who may have high potential to shift loads off peak and/or participate in energy efficiency programs that reduce consumption overall.

[3] For a detailed review, see U.S. Energy Information Administration, November 2011. Smart Grid Legislative and Regulatory Policies and Case Studies

"Smart Meters" supply the information needed to determine key factors that drive peak energy use, notably:

- Whether homes are occupied during the day in summer.
- The likelihood that households have air conditioning ("AC").
- The likelihood that they are using AC during weekday afternoons under different temperature conditions.

For example, homes that are unoccupied during the afternoon and are using AC may be especially good candidates for DR programs targeted at reducing the summer afternoon peak. Once such homes can be accurately identified, then messaging designed to resonate with their attitudinal characteristics can be used to solicit their enrollment in DR programs that are designed to limit AC use during the afternoon.

There are several reasons to think of residential electricity consumption as something that households cause rather than the individuals that comprise them. They are:

- Most of the electricity consumed in a dwelling is consumed collectively by the occupants. AC, heating, lighting, and many major appliances serve the collective group of inhabitants of the dwelling – not any one individual within the household.
- The measurements of electricity consumption that are observable in the case of residential dwellings are made at the household level. It isn't really possible to know which of the household members is responsible for a given unit of electricity consumed in most cases; and the benefit of knowing this for purposes of changing the timing or magnitude of electricity consumption are questionable.
- Many of the processes that affect electricity consumption in households are heavily influenced by social forces that exert powerful influence on the magnitude and timing of electricity consumption (for example, preparing of family dinner).

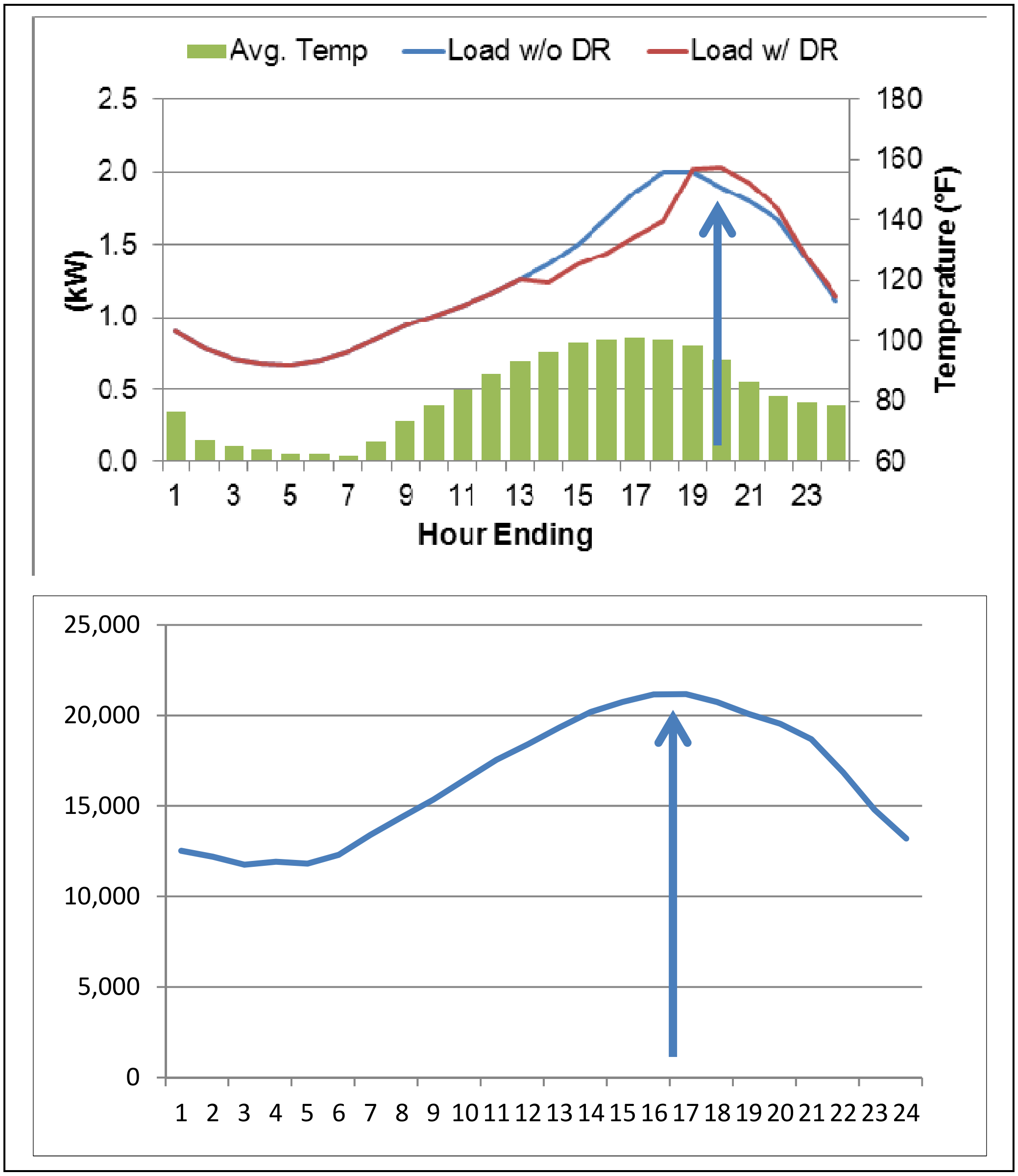
Clearly market survey data sheds light on individual attitudes and behavior and may suggest overall household energy use. Yet since energy consumption occurs at a household level, detailed consumption data as provided by AMI offers great promise as a vehicle to identify targets for DR and EE programs.

Description of the Current Study

Current Challenge for Current Performance of Residential DR Programs

PG&E has been successful in recruiting over 20,000 residential customers to participate in its "SmartRate" DR program. PG&E's historic summer system peak tends to occur between 1:00 PM and 6:00 PM, and the projected load reduction for these DR participant households is designed to align with this peak. Figure One shows a forecasted load shape of the average DR program participant household on a typical event day in the Stockton Local Capacity Area (LCA) as used for transmission planning. This LCA most closely aligns with the geographic area of the sample. The "Load without DR" curve forecasts the average customer's reference load (that is, their load absent DR). The "Load with DR" curve illustrates the expected load reduction during a DR event. If households with peak energy use in the afternoon can be easily identified, they may be good prospects for recruiting for residential DR programs given the typical system load peak that also is shown in Figure One.

Figure 1. Forecasted Summer Peak Day Load With and Without Residential DR (Top) and Pacific Gas and Electric Company Historic System Peak on August 25, 2010 (Bottom)



Demand for current residential customers enrolled in DR programs peaks around 7:30 PM (top chart), about three hours later than the total system peak (bottom chart). Segmentation of residential customers by load shape identifies customers whose peak demand corresponds to total system peak. The willingness and ability of households to shift energy usage off peak must be determined by household surveys, however. Sources: PG&E internal research (top) and California Independent System Operator (bottom).

Hypothesis to Be Tested

The hypothesis to be tested is that electric usage patterns of a large group of residential customers during high-demand summer months are sufficiently constant that they can be segmented into a handful of "signature load shapes." If the hypothesis is found to be true, then the segment that has peak load demands that closely resemble system-wide peak demand may represent a viable target segment for DR programs.

Sample

This study is based on an analysis of 15-minute residential electric usage interval data from the summer of 2011 within the Pacific Gas and Electric Company ("PG&E") service territory. 8,337 households within climate zone[4] 12 (n=6,453) and climate zone 13 (n=1,884) were randomly selected for analysis.[5] These two climate zones were selected because households within these areas represent a significant proportion of summer peak daytime loads in the PG&E service territory (Figure Two shows the location of these two Central Valley climate zones).

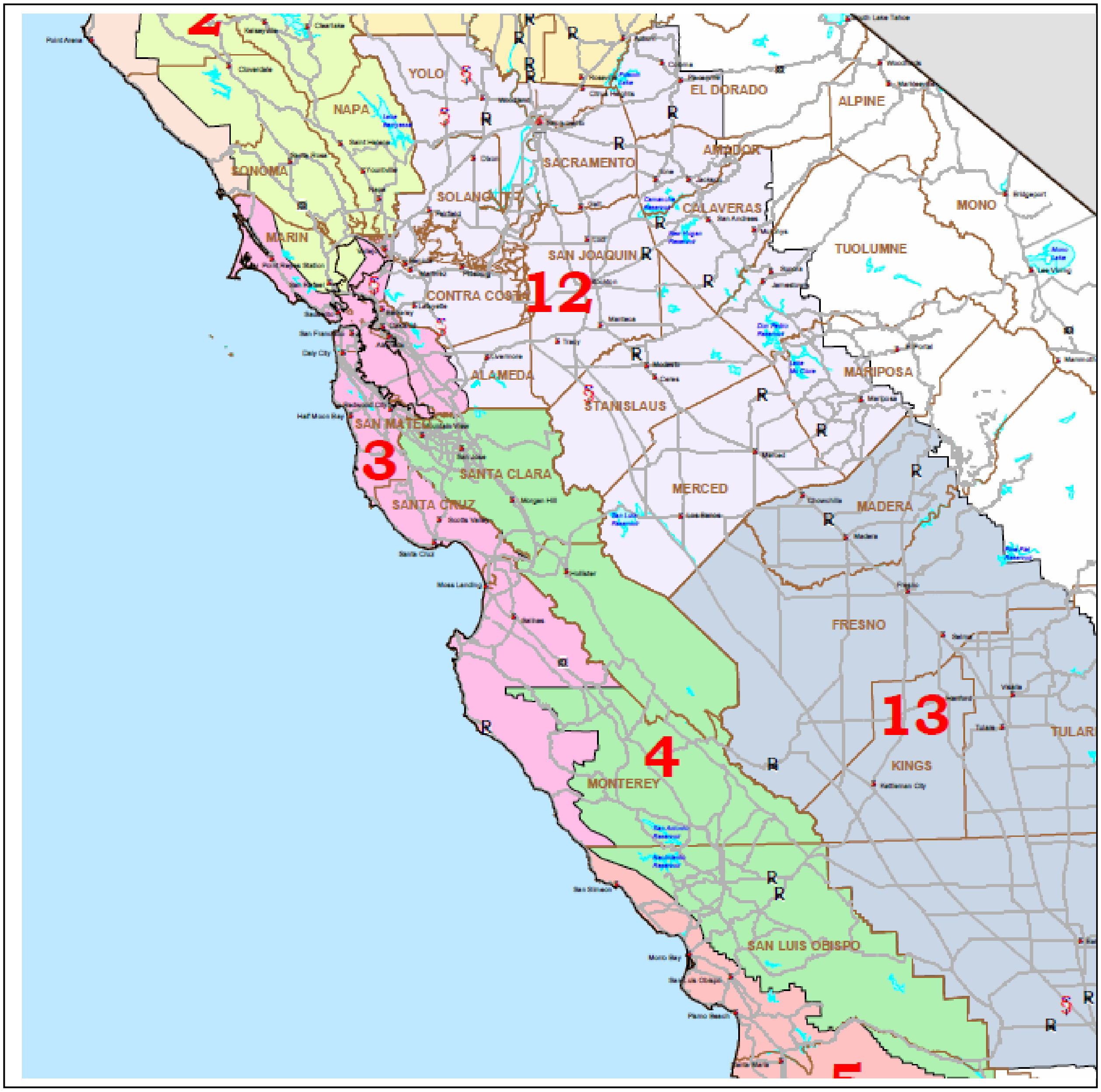
Cities with municipal utilities such as Modesto and Sacramento are excluded from the sample. Climate zone 12 is a highly-populated area situated just inland of the San Francisco Bay Area and has high temperatures that are usually over 100F. The largest cities in climate zone 12 include Concord and Walnut Creek (Contra Costa County), Fairfield and Vallejo (Solano County), and Stockton (San Joaquin County). Climate zone 13 is in California's Central Valley and has high summer daytime temperatures and high summer humidity that leads to higher cooling energy consumption compared to less humid areas of the Central Valley. The largest cities in climate zone 13 include Fresno (Fresno County) and Bakersfield (Kern County). The average household demand in kWh is shown in Figure Three.

---

[4] The Pacific Energy Center's Guide to California Climate Zones and Bioclimatic Design, October 2006 is available at
http://www.pge.com/includes/docs/pdfs/about/edusafety/training/pec/toolbox/arch/climate/california_climate_zones_01-16.pdf

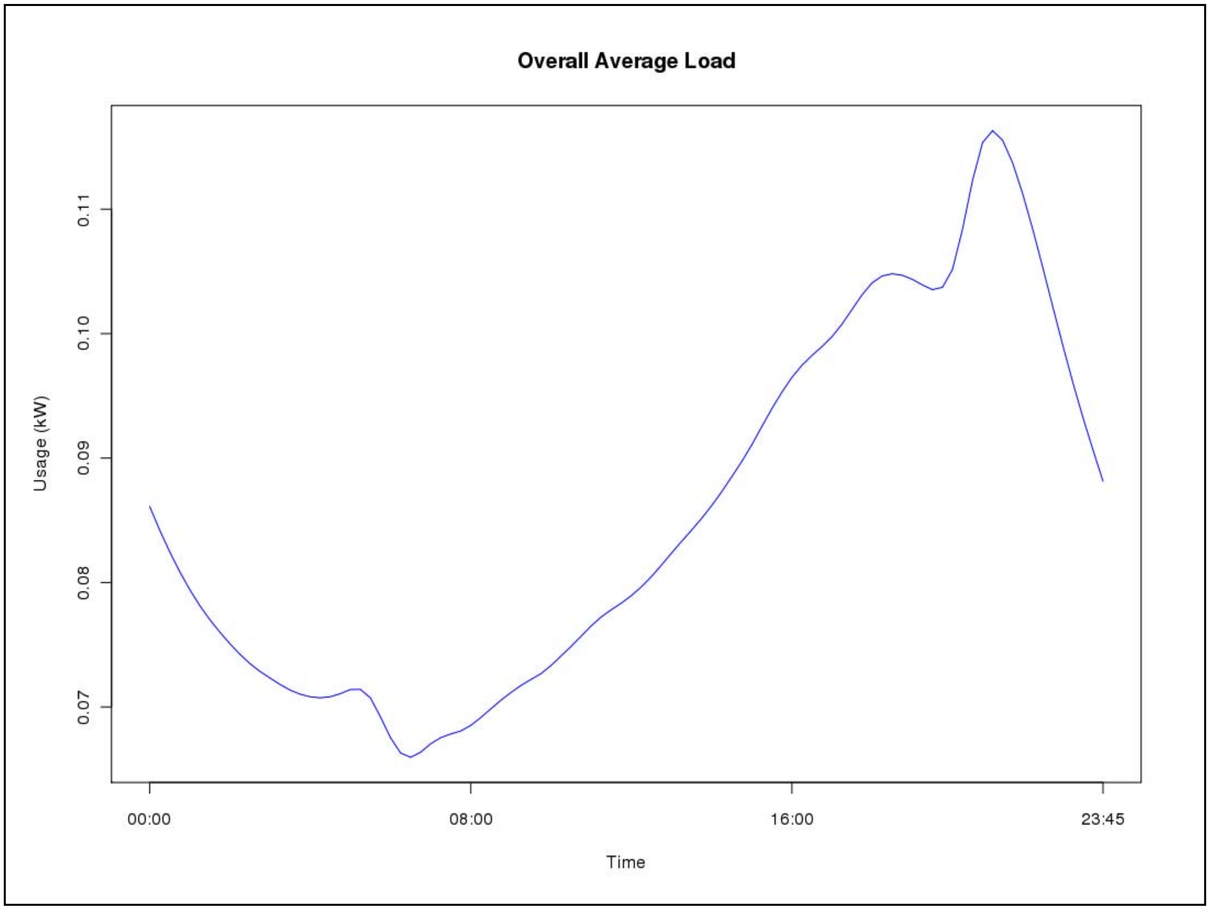
[5] Only a small portion of PG&E's residential customers have 15-minute data. The remaining ones have hourly data. We expect that an analysis of hourly data would yield similar results.

Figure 2. Location of Climate Zones 12 and 13 Used for the Present Study



Source: California Energy Commission

Figure 3. Average Summer Daytime Household Load in the Sample



Shown is the average summer weekday load shape of the entire sample of households used in this study.

Method

Data filtering. First, the data were filtered to include only weekdays (Monday through Friday) during two months of summer (June 1, 2011 through July 30, 2011). The result was 43 weekdays of interval data for 8,337 households that are representative of PG&E residential customers living in climate zones 12 and 13. The data filtering ensured that a relatively homogeneous sample of 15-minute interval data was subjected to the analyses described below.

Energy use entropy. Second, a procedure was developed to identify and remove from the analysis households that showed little if any discernible pattern in day-to-day energy use. The reasoning was that, in order for the clustering algorithm to be successful in identifying groups of households that have similar energy use patterns, it was first necessary to ensure that the households included in the analysis had energy use patterns that were relatively consistent day to day. To accomplish this assessment, we developed a process to measure *energy use entropy*. Entropy is a measurement of "chaos" in a distribution. If the entropy is low (close to zero), then there is little chaos and the probabilities of an energy use pattern falling into a particular segment type are predictable.

The entropy measurement was developed by first performing a K-means clustering on an input matrix where each observation represented one household's load on one particular day. The K-means algorithm was selected because it is highly scalable and can still converge quickly for large datasets. We followed each household and recorded its cluster assignment over the 43-day period. Households that we assigned to a few unique clusters were indicative of consumers who have repetitive and predictable use patterns which are ideal for DR program targeting. We characterized such households as having *low energy use entropy* by virtue of having a relatively similar energy use patterns throughout the days in the sample. Households with low energy use entropy are more desirable targets for DR programs because their load patterns remain relatively consistent (and therefore more predictable) from day to day.

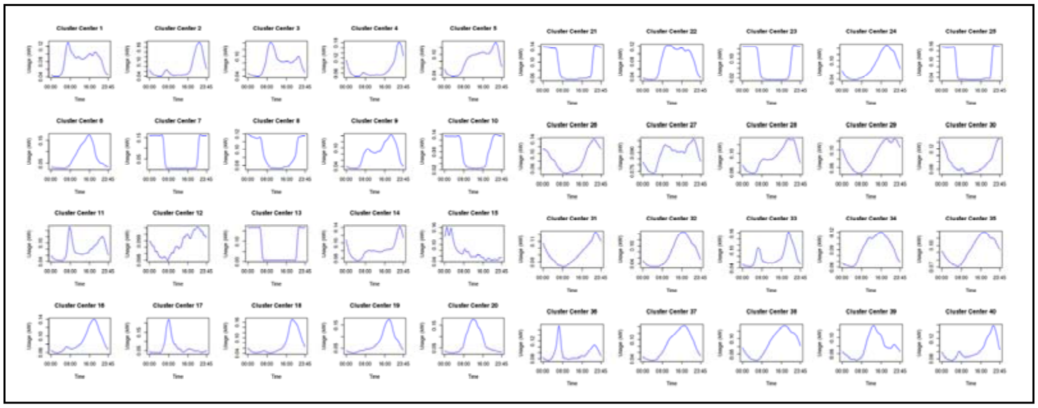
Conversely, households that were assigned to different cluster centers for different days within the study period showed dissimilar (and therefore less predictable) energy use patterns. We characterized such households as having *high energy use entropy* by virtue of having dissimilar energy use patterns throughout the days in the sample. Households with high energy use entropy are less desirable targets for DR programs because their load patterns change significantly from day to day.

As a result of the energy use entropy assessment, the 20% of the households in the study with the highest energy use entropy scores were removed from the analysis. The remaining 80% of the sample showed relatively consistent energy use patterns day-to-day and were subjected to the cluster analysis. The resulting sample included 5,041 households in climate zone 12 and 1,621 households in climate zone 13, for a total of N=6,662 households in the study.

Average load calculation. Data for each household was averaged so that a single load shape that represented average weekday summer use was calculated. This averaging was done to reduce noise in the data and simplify the clustering analysis that was conducted subsequently by making average weekday summer usage of a single household the unit of analysis.

Cluster analysis. The average load shapes for the households were subjected to a K-means cluster analysis using the Euclidean distance. This metric choice allowed us to build a scalable system, but the metric has difficulties identifying clusters that are similar in shape but lag one another in time. In order to properly identify clusters by their shapes, k was generously set to 40. This exposed some redundant clusters, but allowed us to quickly see the full range of shapes, and to consolidate them accordingly through visible inspection. The 40 clusters are shown in Figure Four. Subsequently these clusters were reduced into six basic load shapes.

Figure 4. Initial Cluster Centers Resulting from the K-Means Procedure

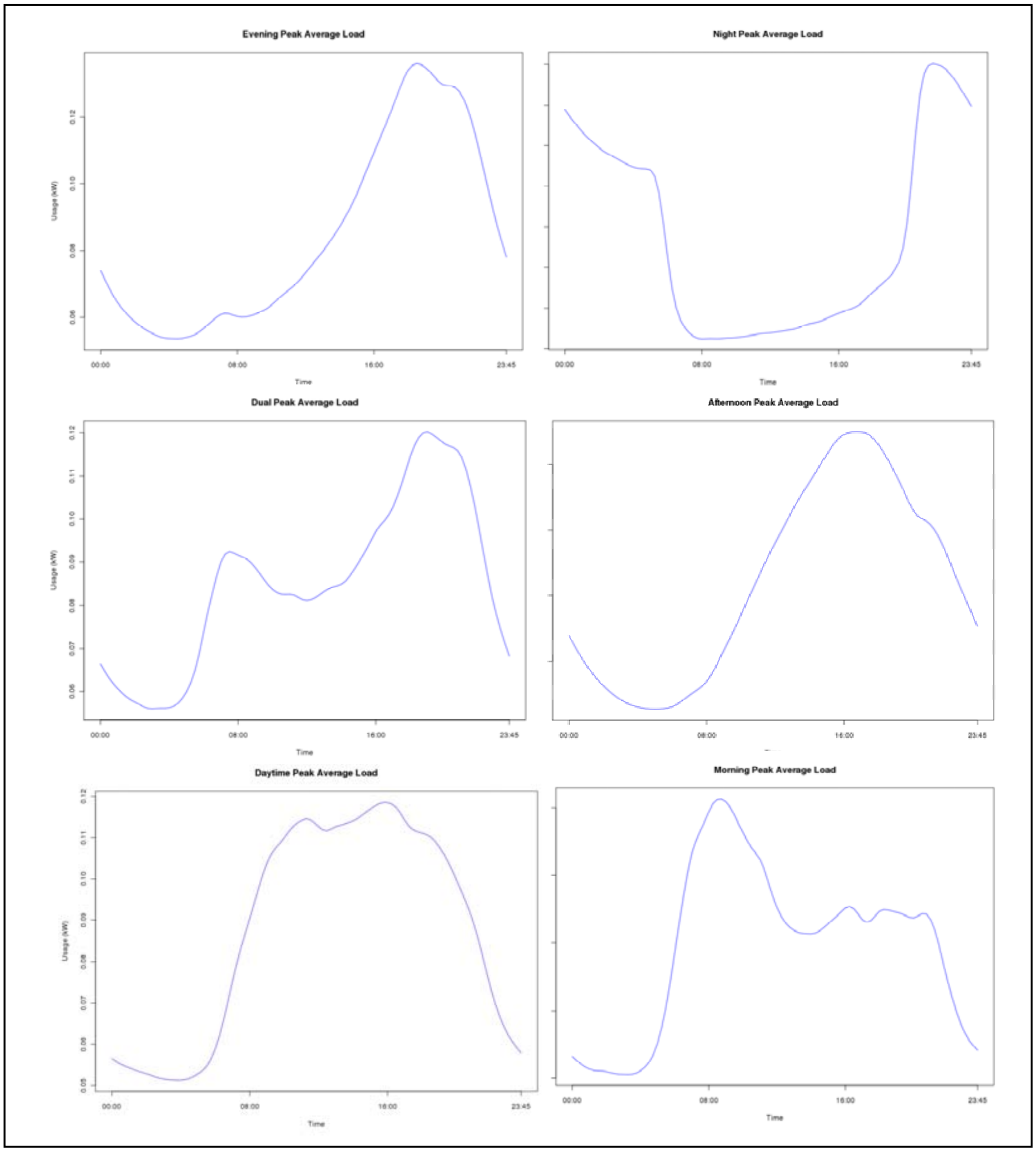


To reduce the number of segments to be analyzed, the 40 load shapes resulting from the clustering were categorized into 6 average load shapes shown in the following figure.

Results

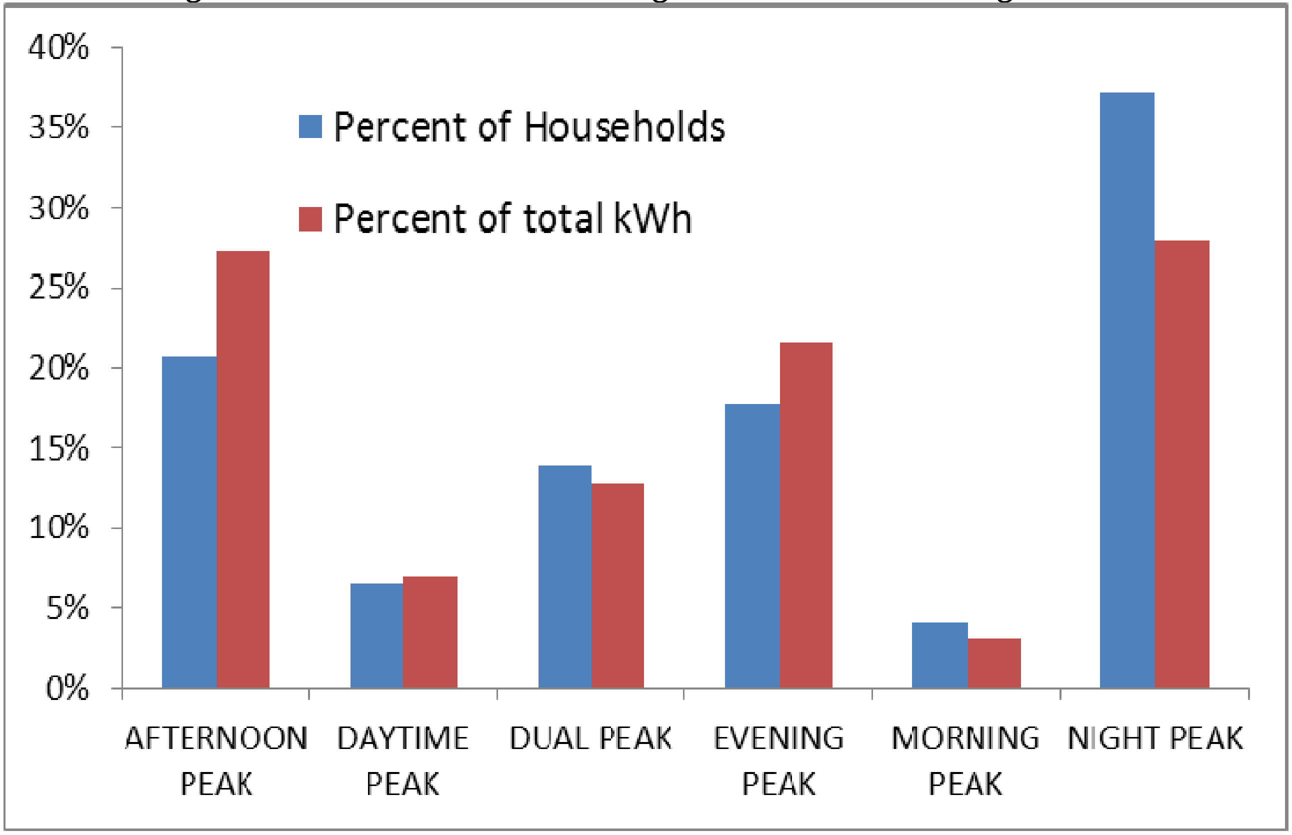
As a result of the clustering procedure, households were classified into "macro" load shapes representing average summer weekday consumption of electricity. The load shapes for each of the final segments is shown in Figure Five, and the percentage of households corresponding to each load shape and their proportion of total kWh demand is shown in the Figure Six.

Figure 5. Average Load Shapes for Households in the Resulting Segments



Load shapes for evening, night, afternoon, morning, daytime, and dual peak segments resulting from the visual assignment of the initial clusters into prototype curves (clockwise from top left-hand graph).

Figure 6. Characteristics of Resulting Six Electric Demand Segments



Households in the "afternoon peak" segment may represent the highest potential for targeting of DR programs given the size of the segment, the proportion of energy use consumed by the segment, and the correspondence of the time of the segment's peak energy use to the time of the overall system peak.

Our interpretation of the targeting potential of the resulting segments, presented in order of potential as targets for DR programs, is as follows:

Afternoon peak. This load shape characterized 21% of the households and 27% of the total electric use of the sample. Households with this load shape show peak usage in the afternoon and have the highest electricity consumption of any segment. Given the similarity of this demand curve with summer peak demand for the system, and given their high overall usage of electricity, this load shape segment offers the most valuable potential for demand reductions from the perspective of the system operator.

Evening peak. This load shape characterized 18% of the sample and 22% of the total electric use of the sample. Households with this load shape show peak usage in the evening and have the second highest electricity consumption of any segment. The load shapes of this segment most are similar to that of the current residential DR program participants. Like the Afternoon Peak group, these households offer significant potential for load shed.

Dual peak. This load shape characterized 14% of the sample and 13% of the total electric use of the sample. It is characterized by dual morning and late afternoon/evening peaks and relatively less usage in the hours in between. This segment may offer some potential for load shed at peak.

Daytime peak. This load shape characterized 7% of the households and 7% of the total electric use of the sample. This load shape is characterized by fairly high demand throughout daytime hours and likely represents households that are occupied throughout the business day. Given the similarity of this demand curve with summer peak for the system, this segment may have potential for DR program targeting. Given high demand throughout the day, this segment may have less ability to shift load off peak, however.

Night peak. This load shape characterized 37% of the households and 28% of the total electric usage of the sample. The load shape of this segment most closely resembles the aggregate residential summer weekday load shape of these climate zones and of the current participants in PG&E's residential DR program. This type of energy use pattern suggests households that tend to be away from home during daytime hours and who have low aggregate loads during business hours but very high demand in the evening and nighttime time periods. It is likely that they do not use AC during the day. Given the dissimilarity of their load shape to the system peak, this segment has low potential for targeting of DR programs.

Morning Peak. This load shape characterized 4% of the sample and 3% of the total electric use. Given the small size of this segment, and the dissimilarity of energy use and peak system use, it has low potential for targeting of DR programs.

Discussion and Next Steps

This preliminary study underscores that the analysis of interval data to understand residential energy load shapes can be accomplished fairly easily with the use of basic statistical techniques that are available in most off-the-shelf statistics applications. In spite of the complex and voluminous nature of interval data, creating a segmentation that characterizes load shapes into meaningful groups can be undertaken using basic software techniques available to utilities of any size with off-the-shelf statistics packages (granted, substantial storage capacity and computational power make the job much easier). The segmentations resulting from this study demonstrate how AMI data can be made actionable.

The present research was a first step in making sense of AMI data to improve program targeting without the need of a large investment in statistical software or consulting services. The next step in this approach is to overlay data from existing household segmentations so that a more nuanced understanding of the segments can be achieved.

Now that residential customers whose peak loads match the total system load have been identified, future survey research involving households within the high potential segments will examine their willingness to participate in DR programs—and their perceived ability to move some of their energy consumption off-peak. This latter point is a critical one, as customers with load shapes where peak time reductions are possible and desirable from a reliability perspective—such as the Afternoon Peak segment—may not necessarily have discretionary load to shed. This important question will be addressed by household survey research planned to be conducted in the future.

We expect that this approach may offer utilities a promising approach for quick, relatively inexpensive research using interval data. Besides the DR application addressed in this initial study, we expect that applications of this technique may be useful for EE program management as well, and this area will be explored in subsequent research.

References

Dethman, Linda and David Thomley. 2009. *Comparison of Segmentation Plans for Residential Customers*. Available at http://energytrust.org/library/reports/091231_Segmentation.pdf.

Lutzenhiser, L. et al. 2009. *Behavioral assumptions underlying California residential energy efficiency programs*. Prepared for CIEE Energy & Behavior Program. Berkeley, CA: California Institute for Energy & Environment. http://ciee-dev.eecs.berkeley.edu/energyeff/documents/ba_ee_res_wp.pdf

Moss, S. 2008. *Market segmentation and energy efficiency program design*. Prepared for CIEE Energy & Behavior Program. Berkeley, CA: California Institute for Energy & Environment. http://ciee-dev.eecs.berkeley.edu/energyeff/documents/MarketSegementationWhitePaper.pdf

Opinion Dynamics Corp. 2009. *Final Segmentation Report*. Available at http://cpuc.ca.gov/NR/rdonlyres/9A3B6444-96AD-4A6D-A392-7588761C3A9D/0/OpinionDynamicsFinalSegmentationReport.pdf