Electricity Markets & Policy
Energy Analysis & Environmental Impacts Division
Lawrence Berkeley National Laboratory

A multi-level load shape clustering and disaggregation approach to characterize patterns of energy consumption behavior

Samanvitha Murthy, Brian F. Gerke, Sarah J. Smith, Aditya Khandekar, Cong Zhang, Richard E. Brown, Mary Ann Piette, Lawrence Berkeley National Laboratory

August 2022

2022 ACEEE Summer Study on Energy Efficiency in Buildings proceedings printed with permission.

The work described in this report was funded by the California Institute for Energy and Environment (CIEE), which is funded by the California Public Utilities Commission (CPUC). Lawrence Berkeley National Laboratory is supported by the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

A multi-level load shape clustering and disaggregation approach to characterize patterns of energy consumption behavior

Samanvitha Murthy, Brian F. Gerke, Sarah J. Smith, Aditya Khandekar, Cong Zhang, Richard E. Brown, Mary Ann Piette, Lawrence Berkeley National Laboratory

ABSTRACT

This study presents representative electrical load shapes, disaggregated to the end-use level, for over 5000 customer clusters across California's residential, commercial, industrial and agricultural sectors. We developed a novel, multi-level load shape clustering approach for residential and commercial sectors leveraging interval meter data for over 350,000 California utility customers collected as a part of the Phase 4 California Demand Response (DR) Potential Study. The clustering approach allowed us to identify typical consumption patterns and categorize customers based on their daily load shape displayed throughout the year. For example, we were able to identify customers with particular energy technologies such as electric vehicles and rooftop solar, as well as building occupancy types such as restaurants, grocery stores and even unoccupied buildings, based solely on whole-building interval data. We then combined the load shape-based clusters with other customer information including building type, climate, geographical area, total consumption and low-income status, to create a set of customer clusters based on both demographics and usage patterns. Total cluster electricity demand was then disaggregated into a wide variety of end-uses using weather normalization and other publicly available end-use load shape datasets. The resulting disaggregated cluster load shapes will be released in anonymized form as part of the Phase 4 DR Potential Study. They will have wide-ranging applications in energy research and policy analysis, including estimation of energy efficiency (EE) and DR potential on the end-use level, time-dependent valuation of EE savings, building stock modeling, and developing customer targeting strategies for EE and DR programs.

Introduction

California has been at the forefront of grid decarbonization efforts, with policies requiring the supply of 100 percent carbon-free electricity by 2045 (SB 100) and greenhouse gas emissions to be reduced to 40% below 1990 levels by 2030 (SB 32). California's power grid is changing rapidly in order to meet these targets; for example, in 2020, non-hydro renewables accounted 28% of total generation in the California Independent System Operator (CAISO) service area (CAISO 2021). Increased grid integration of variable renewable energy (VRE) generation increases the need for flexible resources to balance supply and demand (Mills and Wiser 2015). Demand response (DR) therefore becomes a critical resource to maintain the stability of the system, while also providing additional benefits such as deferment of transmission and distribution system upgrades.

The California DR Potential Study is an ongoing research effort to support California Public Utilities Commission's (CPUC) efforts to enhance the role of DR in meeting the State's resource planning needs and operational requirements. The DR Potential Study uses a bottom-up modeling framework based on detailed demographic and hourly load data of customers

belonging to California's three Investor-Owned Utility (IOUs)[1] in order to estimate California's DR potential. This paper describes the methodology for developing an updated customer dataset into a set of representative customer clusters for Phase 4 of the DR Potential Study. For an overview of Phase 4, see a separate paper in these proceedings (Gerke et al. 2022).

The previous phases of this study used customers' demographic characteristics, location, and total annual consumption to develop customer clusters (Alstone et al. 2017; Gerke et al. 2020). However, this approach did not account for temporal differences among the customers' load shapes, and therefore the resulting aggregation of load generated very general averages that smoothed out the variation in behavior and site occupancy. Load shape-based clustering can improve the specificity of aggregated load shapes by capturing heterogeneity in the consumption patterns, hence also improving the associated estimates of DR potential. Additionally, understanding variations in load shape clusters can improve resource planning and forecasting (Quilumba et al. 2015) and aid in better identifying and targeting customers that are likely to respond during DR events (Smith, Wong, and Rajagopal 2012). For example, households with an evening peak profile, might be good candidates for implementing pre-cooling strategies. The variations in the load profile could be especially useful in structuring energy efficiency (EE) and DR programs by incentivizing specific segments of customers to enroll. Finally, changes in load profiles can provide insights to improve tariff design (Zhou, Yang, and Shen 2013).

Prior work in load shape clustering has been largely been single-sector focused – residential sector (Todd-Blick et al. 2020; Jin et al. 2017; Kwac et al. 2013; Zethmayr and Makhija 2019) or non-residential sector (Nystrup et al. 2021; Luo et al. 2017). Some studies have performed clustering on seasonal load profile data instead of daily load profiles (Rhodes et al. 2014). In addition to finding common patterns, clustering has also been deployed to study similarities in building energy consumption for specific time periods (e.g., during anomalies) and not necessarily the whole year (Divina, Vela, and Torres 2019). Finally, although a common approach to clustering residential and commercial customers has been described (Räsänen et al. 2010), this study only considered buildings that were comparable to households in size and used only 5% of the year's data.

In this study, we have developed a novel multi-level load shape clustering approach using over 350,000 sampled California IOU customers' advanced metering infrastructure (AMI) data for 2019 across residential, commercial, industrial and agricultural sectors. This resulted in 9 residential and 7 commercial load profiles that isolated several typical characteristics of electricity consumption within each sector. We further combined these prototypical load profiles with demographic and geographic characteristics to arrive at a set of 5422 customer clusters across all sectors. Then, we disaggregated these cluster load shapes into a variety of end-uses using weather normalization and other publicly available end use datasets. The weather normalization model was significantly updated since past phases to include residential heating, intraday variability and the capability to account for lag in the reflection of outdoor temperature change in a building load. We also expanded other end-use disaggregation to include existing electric vehicle (EV) charging load from residential and commercial buildings due to an increase

[1] Pacific Gas & Electric (PG&E), Southern California Edison (SCE) and San Diego Gas & Electric (SDG&E)

in EV adoption in California[2] since the previous studies. The resulting set of disaggregated load shapes are key inputs to forecasting future years' load profiles and estimating DR potential.

Some of the boundaries applicable to this study are described as follows. This study only considers customers that are within the three IOUs' service territories. Next, although we received 2018[3] and 2019 hourly load data, clustering and disaggregation were performed only on 2019 data. Load shape clustering was performed only on residential and commercial sectors' customers. We took a simpler approach in clustering and disaggregating the remaining sectors because of fewer samples of customers and the lack of availability of end use load shape data. Finally, since our analysis was based on 2019 data, any future analysis may require additional adjustments to account for COVID-19 related impacts on the load shape.

Analytical Approach

Data Collection and Preprocessing

The DR Potential Study leverages AMI data from the three California IOUs. Phases 1 through 3 used data from 2014; Phase 4 updates the modeling effort with a new dataset reflecting consumption in 2019. Data were requested and collected through a two-stage process via the CPUC. The first stage included demographic data for every single account that was active in 2019; a total of 13.6 million customers. In the second data request, 2018-2019 AMI data was requested for a sample of 3% of accounts (411,000) in the demographic data. More description on the data fields and sampling strategy can be found in a separate paper (Gerke et al. 2022). Each sampled customer was assigned two weights describing the fraction of total customers they represent, in number as well as annual energy consumption, based on the sampling process. We use these weights to construct aggregate cluster load shapes from this sample.

After cleaning and formatting the raw AMI data for the 411,000 sampled customers, we applied a number of preprocessing steps to each time series to develop a final hourly load shape representing total customer demand. These preprocessing steps consist of (1) estimating customer PV generation, for customers known to have rooftop PV, to account for true total energy consumption and (2) filling in estimated consumption data during times that customers' power was shut off due to safety concerns (namely wildfire risk) (Gerke et al. 2022). These final customer load shapes are then used for load shape clustering as described in the next section.

Load shape clustering

The clustering process aims to segment the sampled customers based on similarity in their characteristics (e.g., location, building type, total annual energy use) as well as energy use patterns. The latter part is referred to here as "load shape clustering". We perform this clustering in a multi-step process. The first step, referred to as "Level 1 clustering" develops an identifier for each 24-hour profile in a customer's time series. Next, the cluster centers from Level 1 clusters are grouped further into superclusters based on descriptive analysis. Finally, customers

[2] Battery electric vehicles and Plug-in hybrid EVs' sales increased to 1.9% of total sales of Light Duty vehicles by the end of 2019 (CEC 2021)

[3] 2018 data was primarily requested for the purposes of weather normalization

are once again clustered based on how frequently each supercluster pattern is displayed. We refer to this as "Level 2 clustering". Load shape clustering is done separately for the residential and commercial sectors using K-means clustering algorithm. We begin by normalizing each customer's load data by the daily total in order to ensure that similar shapes are grouped together regardless of the magnitude of the load. We did not perform load shape clustering on the industrial, agricultural and other sector customers due to the lack of specific information beyond their building type to identify their load characteristics.

The objective of K-means algorithm is to minimize the Sum of Squared Errors (Luo et al. 2017). Level 1 clustering involves clustering each day of each customer's demand. In order to decide the optimal number of clusters (N), we chose a sample of 20,000 residential customers. We clustered this data with N=100 clusters such that each customer was assigned 365 cluster values – one for each day's load profile. We then plotted the distribution of the number of cluster shapes required to describe all the load patterns of these customers. We saw that for both residential and commercial sectors roughly N=60 would be sufficient. Beyond this, the division of load profiles would start to become less meaningful for the purposes of DR (i.e., there were an increasing number of qualitatively similar cluster load shapes). In fact, even with N=60, we observed several repetitive load patterns. For example, the algorithm separated load profiles that peak at 8am and 9am into two separate clusters, but they are similar for a DR application.

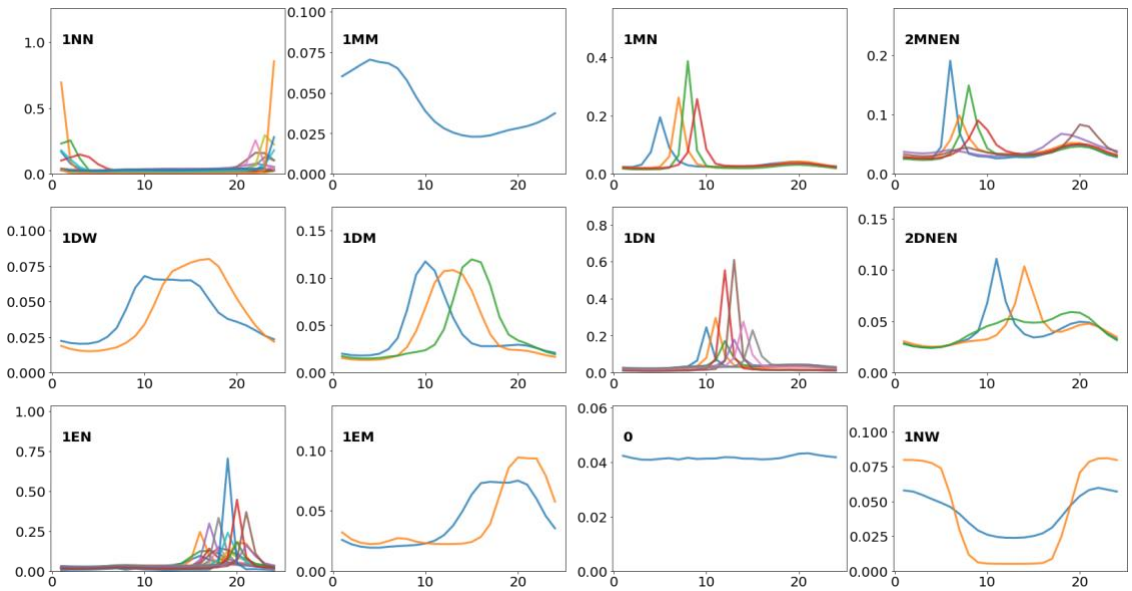

Figure 1. Illustration of reducing 60 residential level 1 cluster centers into 12 superclusters. Each panel in the figure is annotated with text that indicates the supercluster characteristics.

To reduce the dimensionality of the clusters and focus on relevance to DR, we then undertook a qualitative analysis to group the cluster centers further by examining the number of peaks, height and width. From this, we encoded each cluster center using a combination of the number of peaks (0[4], 1, or 2), time of occurrence of peak (Morning, Day, Evening, Night, or

[4] 0 refers to a flat load shape

All[5]) and the width of the peak (Narrow, Medium, or Wide). For example, 2MNEN represents a profile with 2 peaks, the one occurring in the Morning is Narrow and the other occurring in the Evening is also Narrow. This qualitative analysis reduced 60 Level 1 clusters to 12 superclusters in the residential sector (see Figure 1) and 11 in commercial sector.

Next, we computed the frequency of occurrence of each supercluster as a fraction of 365 days for each customer time series. We could represent each customer's full year load pattern using a combination of superclusters. For example, a customer could exhibit 1EN for 50%, 1EM for 25% and 2MNEN for the remaining 25% of the year. We then performed Level 2 clustering on the frequency of appearance of each supercluster. Figure 2 summarizes our approach to determining the optimal number of clusters. We compared cluster centers for different values of K and matched them using sum of least squared differences. We then overlaid the matches on a radar plot and observed the difference in the patterns. If the additional cluster center from the higher K value created a meaningful distinction, we incremented the number of clusters to K+1 and repeated the process until the additional cluster center from the higher K value did not improve the variation in the load profiles. We did this for a range of K values from 4 to 15.

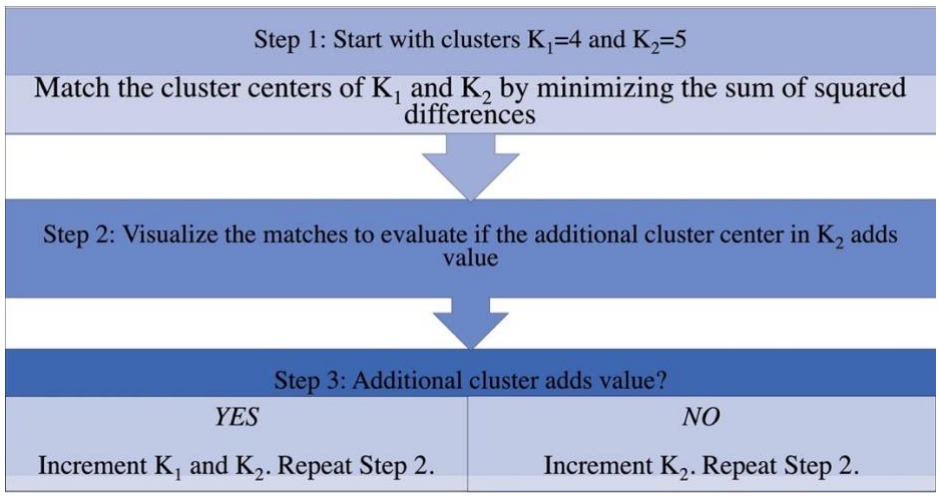

Figure 2. Flow chart describing how the optimal value of Level 2 clusters can be determined. This process is repeated until the incremental value of K does not produce a cluster center that is qualitatively distinct from the previous set.

Figure 3 presents the intermediate results of such an analysis with K=7 and K=8 for the purposes of illustration. The black line in each radar plot shows the cluster centers for K=7 whereas blue and orange lines represent the cluster centers for K=8. The first six radar plots represent highly similar cluster centers. But in the last radar plot, we observe that the center represented by the black line splits into two distinct cluster centers such that the load profiles represented by the blue and orange lines are quite different. The blue cluster center is mostly a combination of flat and double peaking load profiles, whereas the orange cluster center has a combination of load curves with evening narrow peaks. Hence, it makes sense to increment the number of clusters to 8. With this method, we chose 9 residential and 7 commercial load shape

[5] The time of occurrence of peak is assigned as *All* if the peak lasts for over 8 hours in a day

clusters and assigned a meaningful name to each cluster that qualitatively described it. For example, residential cluster with a midnight peak was named as NitePeak. Refer to Figure 4 for residential and Figure 5 for commercial load shape clusters.

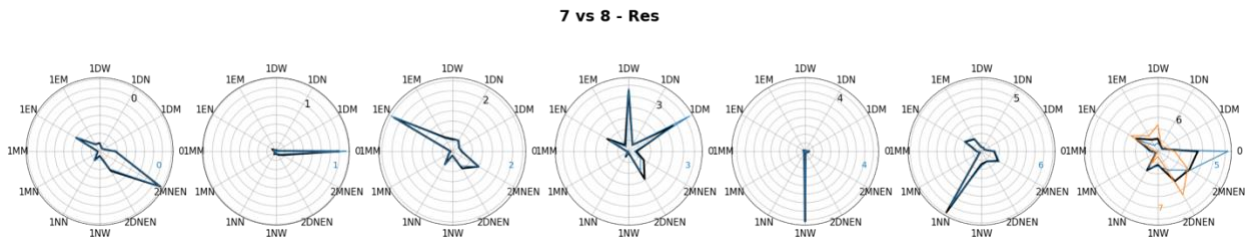

Figure 3. Sample figure comparing Level 2 cluster centers for K=7 and K=8 by overlaying them after minimizing their sum of squared differences. Notice that in the last radar plot, the cluster centers for K=8, depicted in blue and orange, have distinct profiles.

In the final stage of clustering, we assigned a default load shape cluster to the remaining sectors and combined the results of load shape clustering with demographic and geographic characteristics for each customer. For this step, we considered sector, utility, building type, size, climate region[6], receipt of low-income rate subsidy (CARE[7] or nonCARE), local capacity area (LCA), load shape cluster, and quintiles of total annual consumption value. This yielded a set of 5422 clusters across all sectors. For each cluster, we aggregated the 2019 load data of all customers that belonged to the cluster, weighted by their assigned energy use weights from the sampling process, to produce an hourly cluster-level aggregated load shape. Next, in order to release this data publicly, certain anonymization criteria must be met related to the fraction of load represented by any single member of the cluster. To create an anonymized version of the cluster data, we adjust the customer weights recursively and redistribute their surplus to other members in the cluster until the criterion is satisfied.

Disaggregation of cluster load

The hourly load data that we received from the three IOUs, and therefore our resulting cluster load shapes, represent whole-building load. In order to assess DR potential, we disaggregated this data over a variety of end uses. In the previous phases of this study, a limited number of end uses were considered for disaggregating the cluster-level load data. With the advent of smart technologies, retrofits and communicating devices, utility DR programs have also expanded in the recent years to include a variety of end uses. In the phase 4 study, we considered a wide variety of end uses to disaggregate the cluster-level load data. We considered these end uses because they are either already targeted by existing DR programs or are likely to be targeted by future ones. End-use disaggregation described below occurs largely in three steps: (1) disaggregation of temperature-dependent loads for the residential and commercial sectors, (2) disaggregation of EV load in the residential and commercial sector, and (3) disaggregation of all

[6] Climate zones are mapped to climate regions (e.g., hot-dry, marine, cold) in accordance with EE Potential and Goals Study (Sathe et al. 2021)
[7] California Alternate Rates for Energy

other end-uses for each sector independently. The accuracy of the disaggregation depends on the accuracy of the input end-use load shapes and saturation values; since these largely derive from recent and California-specific data sources, our disaggregation should be as accurate as can be computed with present data. In addition, since we used 2019 customer data for the analysis, the load shapes do not account for any COVID-19 induced changes in behavior.

In the first step of the process, we separated temperature and non-temperature dependent loads. This is done at the individual customer level using a temperature normalization model, which determines the correlation between each customer's hourly load data with the hourly outdoor temperature (Alstone et al. 2017) according to the nearest weather station data obtained from California National Oceanic and Atmospheric Administration (NOAA). In this study, the model was updated to include both heating and cooling loads, to consider lags between outdoor temperature and energy demand (to capture the effects of thermal inertia or pre-cooling strategies), and to account for variation in temperature sensitivity across seasons, time of day, and weekdays and weekends. This additional variation allowed us to more accurately model temperature-dependent loads that are based both on outdoor temperature as well as building occupancy patterns. These changes had significant impact on some customer but were found to be unimportant for most. For each time period (e.g., fall weekday evenings), the model computed five parameters to approximate the temperature dependence of a given hour's load to outdoor air temperature. Two of these parameters are the heating and cooling changepoint. The heating changepoint temperature represents the maximum temperature at which heating loads occur, while cooling changepoint represents the minimum temperature at which cooling loads occur. For this analysis, we iteratively tested heating changepoints of 50-70 degrees Fahrenheit and cooling changepoints of 70-90 degrees Fahrenheit for each customer and each time period and chose the model with the best fit. The remaining parameters are the slope of the heating and cooling sensitivity, meaning the amount of additional energy use expected for each degree of temperature difference above or below the changepoint, and the baseline load that occurs regardless of temperature. We applied a threshold to these parameters below which we assumed that there is no space heating or cooling load. A limitation of this approach is that it assigns all load correlated with hourly outdoor temperature to space heating and cooling. Other loads (e.g., water heating, pool pumps) may also correlate with daily average temperature or season, although these are less likely than space conditioning to correlate strongly with the hourly temperature. Figure 3 illustrates how the model disaggregated using a sample customer data for a period of 3 days. The yellow curve in the figure represents the outdoor temperature, while the stacked bars show the total load disaggregated into non-temperature dependent loads (blue) and cooling loads (green) in response to variations in temperature.

EV load disaggregation in the residential and commercial sector was another significant update in the study. With an increase in the EV penetration in California, we were able to disaggregate Level 1 (120 V, can be plugged into a typical building wall outlet) and Level 2 (240 V, faster charging but requiring special installation) charging load using data from EVI-Pro (Bedir et al. 2018). In the residential sector, we used data from the Residential Appliance Saturation Study (Palmgren et al. 2021) from 2019 (RASS) to determine saturation values and average annual consumption for all end uses including EV by each IOU. Specifically, EV-related data in RASS contains information regarding unit energy consumption and availability of Level 1 or Level 2 charging. We used this to determine the average annual consumption. Further,

RASS also contains information about where (e.g., home, office or other location) and how frequently a respondent charges (e.g., once a week, 2-3 times a week). We used this data to determine and allocate different average energy consumption values to those who charge their EVs all the time at home and those who charge it partially at home and other commercial locations (e.g., office). For the latter, we also apportioned their total load between residential and commercial sectors. We used these three sets of information to determine the average annual EV load for Level 1 and Level 2 charging. In order to determine the saturation values of EV in each cluster, we have IOU data indicating if a customer owns an EV as well as if they are on an EV rate. Note that not all EV owners are on an EV Time-of-Use (TOU) rate. Additionally, our analysis also indicated that the number of customers owning an EV as identified by the IOU data is quite underestimated. So, we used California Air Resources Board's (CARB) estimates of EVs on the road in 2019 (CARB 2019) with CEC data on private vehicle ownership (CEC 2015) to estimate the total number of residential and commercial EVs. We then used this data to augment the fractions of customers in each cluster owning an EV as indicated by the IOU data. We used these values as total EV saturation values and capped it at 100%. From this, we assumed that the fraction of customers on an EV rate take advantage of it and have Level 2 chargers and the remaining to utilize Level 1 charging. So, we applied Level 1 and Level 2 EV charging load shapes from EVI-Pro accordingly.

In the commercial sector, we adopted a multi-step approach to disaggregate cluster-level EV load. We restricted EV charging to offices and retail building types only. There are two main sources of EV charging load in the commercial sector – privately owned EVs that may charge partially in a commercial building and commercial fleet EVs that are charged only in commercial buildings. Then, using the CEC estimates of EV (CEC 2015), we computed the ratio of private EVs to commercial fleet EVs and applied commercial charging load shapes from EVI-Pro.

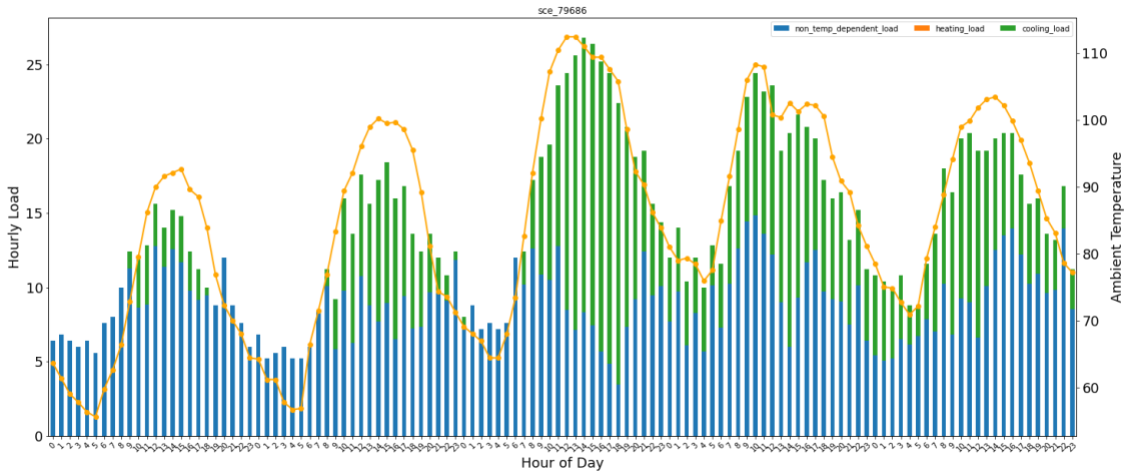

Figure 3. Illustration of results of temperature normalization model for a sample customer for cooling over a period of three days. Notice the lag in the building's cooling load in response to temperature change.

In order to disaggregate other end uses in the residential sector, we were able to subset RASS by utility, building type and climate region. We used load shape data for each building type modeled by ADM associates (Baroiant et al. 2019) for CEC. We then took the product of saturation, annual average energy consumption and the load shape for each hour normalized by

the hourly total load. The product of this intermediate load shape and the non-temperature dependent load is the final disaggregated end use load. We followed a similar approach in the commercial sector using saturation data from the California Commercial End Use Survey abbreviated as CEUS (Itron 2006) and California Commercial Saturation Survey (Itron 2014), and load shapes from ADM associates. For those end-uses where hourly load shape data was not available, we selected a proxy load shape from the set of available load shapes. For data centers, we used the bottom-up energy use modeling of the U.S. data center industry that showed Information Technology (IT) equipment is responsible for 65% of electricity use, cooling for 28%, and other end uses for 7% (Shehabi et al. 2016). We applied these fractions to all hours of the year due to lack of temporally-resolved data on end use behavior.

We used the data from the Manufacturing Energy Consumption Survey (MECS) to disaggregate load in the industrial sector (EIA 2021). First, we mapped the NAICS[8] code of each customer in the cluster to the nearest NAICS code in the MECS database, and calculated the fraction of load across NAICS codes in the cluster. Next, we took the breakdown of electricity consumption across end-uses for each relevant NAICS code and weighted by the cluster-specific fraction. Due to lack of data on temporal usage of these end uses, we applied the final end-use fraction to all hours of the year. For the agricultural sector, we disaggregated demand into pumping and non-pumping loads based on assumed pumping fractions of 100% for crop and water agriculture types (according to NAICS code mapping), 50% for agriculture activities related to livestock or indoor crops, and 80% for other designations.

Results and discussion

The combination of load shape and demographic clustering yielded 5422 clusters across all sectors. The anonymized load shapes after disaggregation will be released as a part of the Phase 4 of California DR Potential Study. However, in this section, we will highlight a subset of the results to show the intricacies captured by our clustering and disaggregation methodologies. We will start by looking at the results of load shape clustering. Figure 4 shows the results of load shape clustering in the residential sector. The top panel shows the Level 2 cluster centers in the form of radar plots. The bottom panel shows the corresponding cluster's average load shape. We chose 9 load shape patterns to represent residential electricity consumption and named each one based on their average load shape.

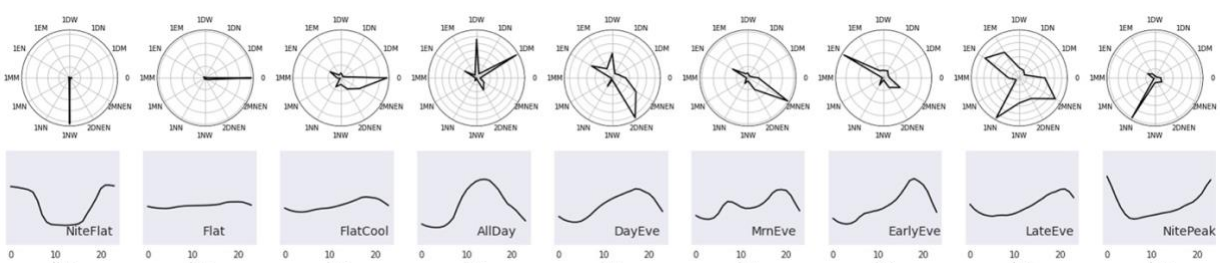

Figure 4. Representation of results from residential load shape clustering. The top panel represents radar plots indicating the cluster centers and the bottom panel represents their corresponding average load shape.

[8] North American Industrial Classification System

The first pattern is called NiteFlat and was found in both residential and commercial sectors (see Figure 5 for commercial load shape clusters). This pattern indicates the presence of the supercluster with a single wide peak at night (1NW) and generally represents vacant buildings or parking lots that have evening outdoor lighting loads and certainly always-on base loads, but no significant occupant-driven loads. Other clusters where 1NW is found in smaller proportion can be interpreted as being intermittently vacant (e.g., vacation homes). The next two patterns indicate Flat and FlatCool load shapes respectively. FlatCool has a combination of flat and some evening peaking load shapes, which further analysis shows to occur in the summer. Thus, FlatCool indicates the presence of cooling load in combination with an underlying flat load shape, demonstrating the power of our clustering methodology to separate homes with and without air conditioning. AllDay indicates peaks that lasts throughout the daytime and can be applicable to homes that are occupied during the day (e.g., families with stay-at-home parents, remote workers, or people outside the labor force). This segment also contains an elevated fraction of single-family homes with rooftop PV, suggesting either that PV adoption may be correlated with high daytime load, or that our modeling approach over-estimated daytime generation for some customers with PV. DayEve and MrnEve represents a typical residential consumption pattern with double peaks indicating a surge in electricity consumption right before and after work hours. EarlyEve and LateEve profiles could indicate a variety of profiles such as occupants working until late night, uncontrolled EV charging, usage of cooking appliances and other plug loads. NitePeak mostly indicates customers responding to EV Time-of-Use rates, which are disproportionally represented in this cluster. This is highlighted in Figure 8.

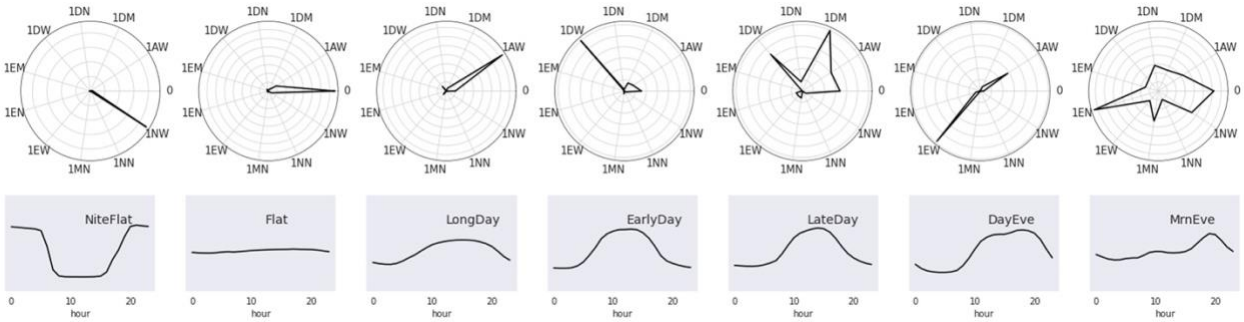

Figure 5. Representation of results from commercial load shape clustering. The top panel represents radar plots indicating the cluster centers and the bottom panel represents their corresponding average load shape.

Figure 5 shows the results of commercial load shape clustering. We identified 7 distinct shapes to describe the consumption patterns in the commercial sector, and named them based on their average load shape. NiteFlat and Flat are similar to the residential load shape clusters with the same names. Figure 6 represents the distribution of building types in each cluster. We can see that the DayEve load shape has a significantly higher fraction of restaurants and other dining facilities than other clusters, reflecting the distinctive operation hours for such establishments, with peaks at midday and in the evening. Next, from Figure 5, we can notice that EarlyDay cluster is made up of mostly single peaking day wide (1DW) load shape with a small percentage of Flat load shape. These are buildings that operate mostly on weekdays and have relatively low loads during the weekend as indicated by the upper panel plot in Figure 7. In comparison, the load profile in the lower panel of Figure 7 exhibits that distinction between weekday and

weekend loads to a much lesser extent. LongDay and LateDay also represent retail buildings and grocery stores that open later in the day. Finally, MrnEve is a double peaking profile with a larger peak in the evening. This represents the load shape of buildings such as fitness centers and performance theaters with a moderate morning load and a higher evening load, as evidenced by the disproportionate representation of the assembly building type in Figure 6.

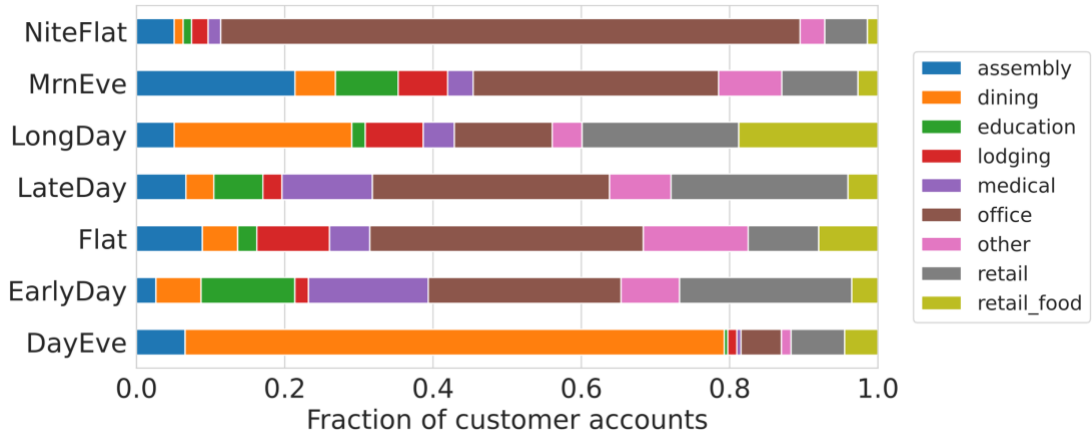

Figure 6. Representation of fraction of customer counts by building type in each commercial load shape cluster

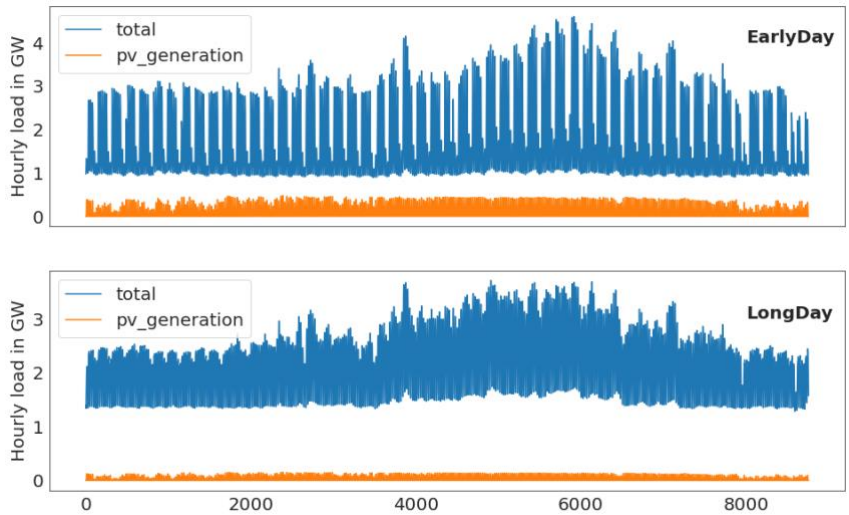

Figure 7. Comparison of annual hourly load of clusters EarlyDay and LongDay. Notice the periodic dip in the consumption in the upper plot indicating consistent changes in weekday and weekend patterns.

Next, let us take a look at some interesting examples of end use disaggregation in both commercial and residential sectors. With the recent increase in the number of EV sales and customers transitioning to TOU rates, specifically EV TOU rates, the load shape clustering was able to clearly distinguish customers who own EVs and respond to the EV TOU rates, since these rates are designed to incentivize customers to charge their EVs starting around midnight. Figure 8 shows the seasonal average load shape of a sample PG&E cluster in the Bay Area with the load shape NitePeak. First of all, this figure shows the numerous end-uses considered for residential disaggregation. Level 1 and Level 2 EV charging loads are indicated in light purple and dark purple respectively. We can observe that the Level 2 EV charging load increases around

midnight and reduces gradually by around 5am. The figure also shows the results of temperature normalization model where we can see heating loads in winter, spring and early hours of fall, and a tiny sliver of cooling load in the summer. This makes sense because cooling need may not be significant in the marine climate region of the Bay Area.

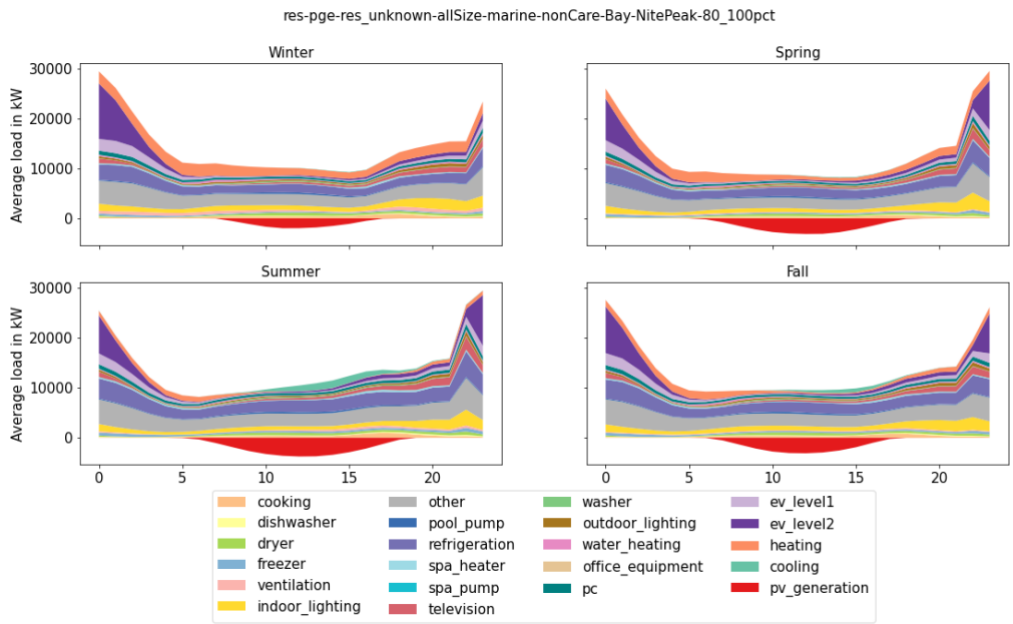

Figure 8. Seasonal daily average load profile of a sample residential cluster with EVs. EV loads are indicated in light purple (Level 1 EV charging) and dark purple (Level 2 EV charging).

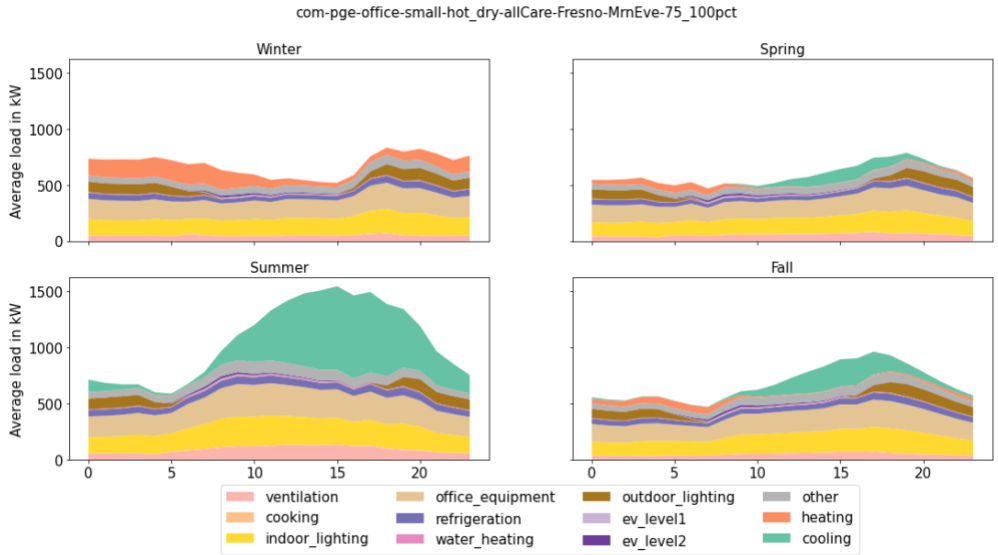

Figure 9. Seasonal average daily load profile of a sample commercial cluster to illustrate the effects of cooling on the overall load shape.

In order to observe the influence of cooling on the overall load shape, let us look at Figure 9. This is the seasonal average load shape of a sample PG&E cluster in the Central

Valley. We can observe that all the non-temperature dependent loads such as cooking and office equipment remain somewhat similar across the four seasons. However, the amount of cooling load in the summer is almost the same as the other end uses combined at its peak. There is also noticeable heating load in winter. Further, in the spring and fall seasons, we notice that the early morning hours have heating followed by the evening need for cooling. Finally, we can observe commercial EV charging load in the middle of the day since this is an office building cluster.

Conclusion

In this study, we described a novel multi-level load shape clustering and load disaggregation methodology for using over 350,000 California IOU customers' smart meter interval data as a part of the phase 4 DR Potential Study. The load shape clustering was performed on residential and commercial customers using the whole building hourly load data for 2019, which yielded 9 residential and 7 commercial unique load patterns that clearly segmented customers with rooftop PV, EV, restaurants and grocery stores. Combining this with the previous phases' demographic clustering yielded 5422 clusters across all sectors. We then disaggregated the cluster-level load data into a variety of end uses in each sector using publicly available datasets. We also updated the phase 2 study's temperature normalization model to include residential heating, account for building's thermal inertia and capture variability in sensitivity of load to outdoor temperature.

The resulting disaggregated cluster load shapes are planned to be anonymized and released as a part of the Phase 4 study. We believe that this dataset can be applied to numerous energy and energy policy research questions. For instance, it can improve the estimates of energy savings from EE measures, EE and DR potential, load planning and forecasting and building stock modeling. Additionally, it can also be applied to study the impact of rate design such as dynamic pricing, which can help identify structural winners and losers at a very granular level.

Acknowledgements

The work described in this report was funded by the California Institute for Energy and Environment (CIEE), which is funded by the California Public Utilities Commission (CPUC). We thank Jean Lamming and Aloke Gupta (CPUC) and Terry Surles and Carl Blumstein (CIEE) for their ongoing support of this research. Lawrence Berkeley National Laboratory is supported by the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

References

Alstone, Peter, Jennifer Potter, Mary Ann Piette, Peter Schwartz, Michael A. Berger, Laurel N. Dunn, Sarah J. Smith, et al. 2017. "2025 California Demand Response Potential Study - Charting California's Demand Response Future. Final Report on Phase 2 Results." 1421800. https://doi.org/10.2172/1421800.

Baroiant, Sasha, John Barnes, Daniel Chapman, Steven Keates, and Jeffrey Phung. 2019. "California Investor- Owned Utility Electricity Load Shapes." Sacramento: California

Energy Commission. https://www.energy.ca.gov/sites/default/files/2021-06/CEC-500-2019-046.pdf.

Bedir, Abdulkadir, Noel Crisostomo, Jennifer Allen, Eric Wood, and Clément Rames. 2018. "California Plug-In Electric Vehicle Infrastructure Projections: 2017-2025." CEC-600-2018-001. California Energy Commission. https://www.nrel.gov/docs/fy18osti/70893.pdf.

CAISO. 2021. "2020 Annual Report on Market Issues and Performance." http://www.caiso.com/Documents/2020-Annual-Report-on-Market-Issues-and-Performance.pdf.

CARB. 2019. "Fleet Database." Sacramento: California Air Resources Board. https://arb.ca.gov/emfac/fleet-db.

CEC. 2015. "Summary of California Vehicle and Transportation Energy." California Energy Commission. https://www.energy.ca.gov/data-reports/energy-almanac/transportation-energy/summary-california-vehicle-and-transportation.

———. 2021. "Vehicle Population in California." California Energy Commission. https://www.energy.ca.gov/zevstats.

Divina, Federico, Francisco Vela, and Miguel Torres. 2019. "Biclustering of Smart Building Electric Energy Consumption Data." *Applied Sciences* 9 (2): 222. https://doi.org/10.3390/app9020222.

EIA. 2021. "2018 Manufacturing Energy Consumption Survey." Washington DC: US Energy Information Administration. https://www.eia.gov/consumption/manufacturing/pdf/MECS%202018%20Results%20Flipbook.pdf.

Gerke, Brian F., Giulia Gallo, Sarah J. Smith, Jingjing Liu, Peter Alstone, Shuba Raghavan, Peter Schwartz, Mary Ann Piette, Rongxin Yin, and Sofia Stensson. 2020. "The California Demand Response Potential Study, Phase 3: Final Report on the Shift Resource through 2030." https://doi.org/10.20357/B7MS40.

Gerke, Brian F., Sarah J. Smith, Samanvitha Murthy, Aditya Khandekar, Cong Zhang, Sunhee Baik, Shreya Agarwal, et al. 2022. "Key Insights from Phase 4 of the California Demand Response Potential Study." In *2022 ACEEE Summer Study on Energy Efficiency in Buildings*. Pacific Grove, CA: ACEEE.

Itron. 2006. "California Commercial End-Use Survey - Consultant Report." CEC-400-2006-005. Sacramento: California Energy Comission.

———. 2014. "California Commercial Saturation Survey." CPU0077.01. California Public Utilities Commission. www.calmac.org.

Jin, Ling, Dorris Lee, Alex Sim, Sam Borgeson, Kesheng Wu, C. Anna Spurlock, and Annika Todd. 2017. "Comparison of Clustering Techniques for Residential Energy Behavior Using Smart Meter Data." In *The AAAI-17 Workshop on Artificial Intelligence for Smart Grids and Smart Buildings*. https://www.aaai.org/ocs/index.php/WS/AAAIW17/paper/view/15166/14673.

Kwac, Jungsuk, Chin-Woo Tan, Nicole Sintov, June Flora, and Ram Rajagopal. 2013. "Utility Customer Segmentation Based on Smart Meter Data: Empirical Study." In *2013 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 720–25. Vancouver, BC, Canada: IEEE. https://doi.org/10.1109/SmartGridComm.2013.6688044.

Luo, Xuan, Tianzhen Hong, Yixing Chen, and Mary Ann Piette. 2017. "Electric Load Shape Benchmarking for Small- and Medium-Sized Commercial Buildings." *Applied Energy* 204 (October): 715–25. https://doi.org/10.1016/j.apenergy.2017.07.108.

Mills, Andrew D., and Ryan H. Wiser. 2015. "Strategies to Mitigate Declines in the Economic Value of Wind and Solar at High Penetration in California." *Applied Energy* 147 (June): 269–78. https://doi.org/10.1016/j.apenergy.2015.03.014.

Nystrup, Peter, Henrik Madsen, Emma M.V. Blomgren, and Giulia de Zotti. 2021. "Clustering Commercial and Industrial Load Patterns for Long-Term Energy Planning." *Smart Energy* 2 (May): 100010. https://doi.org/10.1016/j.segy.2021.100010.

Palmgren, Claire, Miriam Goldberg, Bob Ramirez, and Craig Williamson. 2021. "2019 California Residential Appliance Saturation Study (RASS)." CEC-200-2021-005. California Energy Commission. https://www.energy.ca.gov/publications/2021/2019-california-residential-appliance-saturation-study-rass.

Quilumba, Franklin L., Wei-Jen Lee, Heng Huang, David Y. Wang, and Robert L. Szabados. 2015. "Using Smart Meter Data to Improve the Accuracy of Intraday Load Forecasting Considering Customer Behavior Similarities." *IEEE Transactions on Smart Grid* 6 (2): 911–18. https://doi.org/10.1109/TSG.2014.2364233.

Räsänen, Teemu, Dimitrios Voukantsis, Harri Niska, Kostas Karatzas, and Mikko Kolehmainen. 2010. "Data-Based Method for Creating Electricity Use Load Profiles Using Large Amount of Customer-Specific Hourly Measured Electricity Use Data." *Applied Energy* 87 (11): 3538–45. https://doi.org/10.1016/j.apenergy.2010.05.015.

Rhodes, Joshua D., Wesley J. Cole, Charles R. Upshaw, Thomas F. Edgar, and Michael E. Webber. 2014. "Clustering Analysis of Residential Electricity Demand Profiles." *Applied Energy* 135 (December): 461–71. https://doi.org/10.1016/j.apenergy.2014.08.111.

Sathe, Amul, Karen Maoz, Tyler Capps, Rebecca Legett, Debyani Ghosh, Divya Iyer, Micah Turner, et al. 2021. "2021 Energy Efficiency Potential and Goals Study, Draft Report." San Francisco: California Public Utilities Commission.

https://pda.energydataweb.com/api/view/2531/2021%20PG%20Study%20DRAFT%20Report%202021_Final.pdf.

Shehabi, Arman, Sarah Smith, Dale Sartor, Richard Brown, Magnus Herrlin, Jonathan Koomey, Eric Masanet, Nathaniel Horner, Inês Azevedo, and William Lintner. 2016. "United States Data Center Energy Usage Report." LBNL--1005775, 1372902. https://doi.org/10.2172/1372902.

Smith, Brian A., Jeffrey Wong, and Ram Rajagopal. 2012. "A Simple Way to Use Interval Data to Segment Residential Customers for Energy Efficiency and Demand Response Program Targeting." In *2012 ACEEE Summer Study on Energy Efficiency in Buildings*. Pacific Grove, CA: ACEEE. https://www.aceee.org/files/proceedings/2012/data/papers/0193-000182.pdf.

Todd-Blick, Annika, C. Anna Spurlock, Ling Jin, Peter Cappers, Sam Borgeson, Dan Fredman, and Jarett Zuboy. 2020. "Winners Are Not Keepers: Characterizing Household Engagement, Gains, and Energy Patterns in Demand Response Using Machine Learning in the United States." *Energy Research & Social Science* 70 (December): 101595. https://doi.org/10.1016/j.erss.2020.101595.

Zethmayr, Jeff, and Ramandeep Singh Makhija. 2019. "Six Unique Load Shapes: A Segmentation Analysis of Illinois Residential Electricity Consumers." *The Electricity Journal* 32 (9): 106643. https://doi.org/10.1016/j.tej.2019.106643.

Zhou, Kai-le, Shan-lin Yang, and Chao Shen. 2013. "A Review of Electric Load Classification in Smart Grid Environment." *Renewable and Sustainable Energy Reviews* 24 (August): 103–10. https://doi.org/10.1016/j.rser.2013.03.023.