

Learning from Hourly Household Energy Consumption: Extracting, Visualizing and Interpreting Household Smart Meter Data

Sam Borgeson^(✉), June A. Flora, Jungsuk Kwac, Chin-Woo Tan,
and Ram Rajagopal

Sustainable Systems Lab, Department of Civil and Environmental Engineering,
Stanford University, Stanford, USA

{sborgeson, jflora, kwjusul, tancw, ramr}@stanford.edu

Abstract. In this paper, we present the *Energy Visualization and Insight System for Demand Operations and Management* platform (VISDOM), a collection of smart meter data analysis algorithms and visualization tools designed to address the challenge of interpreting patterns in energy data in support of research, utility energy efficiency and demand response programs. We provide an overview of how the system works and examples of usage, followed by a discussion of the potential benefits of using VISDOM to identify and target participants whose electricity consumption is best aligned with the goals of efficiency and demand response programs.

Keywords: Information design · Data visualization · Energy · Sustainability · Energy efficiency · Customer segmentation · Machine learning

1 Introduction

By mid- 2014, electricity smart meters had been installed for over 50 million, or 43 %, of US households and were generating in excess of 1 billion data points a day¹ [1]. Smart meter data flows into utilities at a rate approximately three orders of magnitude faster than traditional monthly meter readings, creating significant opportunities for novel forms of analysis and interpretation of energy consumption at a massive scale. In this paper, we present the *Energy Visualization and Insight System for Demand Operations and Management* platform (VISDOM). VISDOM is a collection of smart meter data analysis algorithms and visualization tools designed to address the challenge of interpreting patterns in energy data in support of utility energy efficiency and demand response programs.²

¹ Smart meters record electricity usage at 15 min or hourly intervals, with typical residential configurations defaulting to hourly.

² Energy efficiency programs are a wide array of publically funded programs, typically administered by utilities, designed to reduce customer energy consumption. These programs vary from conveying information, to rebates for energy efficient purchases, to online or in person household energy audits. Demand response programs are similar to efficiency programs but are designed to get customers to reduce their load temporarily during periods of grid stress.

This paper describes each of VISDOM's components, summarizing the algorithms used to characterize customer consumption and the interactive interface for open-ended exploration of the characteristics of large sets of customers. We provide examples of typical usage and discuss the potential benefits of segmenting and targeting customers for participation in energy programs based on the consumption characteristics available through VISDOM.

2 What Is VISDOM?

VISDOM's three main components are customer consumption feature extraction, feature exploration (i.e. interactive plots and mapping), and household segmentation and targeting (i.e. using features to filter and categorize customers). We define features as discrete pieces of categorical or numerical data known or estimated for every customer. They are designed to capture essential characteristics of customers and their energy consumption and can support self-directed learning about consumption patterns and their drivers, academic research into consumption and its correlates, or to identify households that make the best targets for specific energy interventions.

2.1 Feature Extraction

Feature extraction algorithms, implemented in R, [2], iterate through customers, record known features and derive (through statistical summaries) or estimate (through statistical and machine learning techniques) values for others. Known customer features typically include location, rate plan, and billing information. To date we have implemented several dozen derived features, including demand percentiles and variability for different time periods, daily load characteristics, and correlations between demand and external phenomena [3]. Estimated features include parameters from statistical models and machine learning algorithms performing tasks like end use disaggregation, characterizing occupancy patterns and schedules [4], recovering heating and cooling system attributes [3], and clustering and interpreting load shape characteristics³ [5].

Figure 1 provides an illustration of the feature extraction process. Smart meter data and other information known about each customer, location, and building are the raw inputs into feature extraction algorithms. These algorithms output a list of features for each customer, potentially achieving full coverage of a utility service territory. Features can be categorized into the following loose categories.

- **Basic statistics and heuristics**, including mean, median, peak, minimum, variance, etc., computed for different time intervals.

³ A load shape is the pattern of demand, or energy use profile, of a customer over a 24 h period. Load shapes are determined by operational schedules and occupancy, and thus have many applications related to understanding drivers of demand and predicting future outcomes.

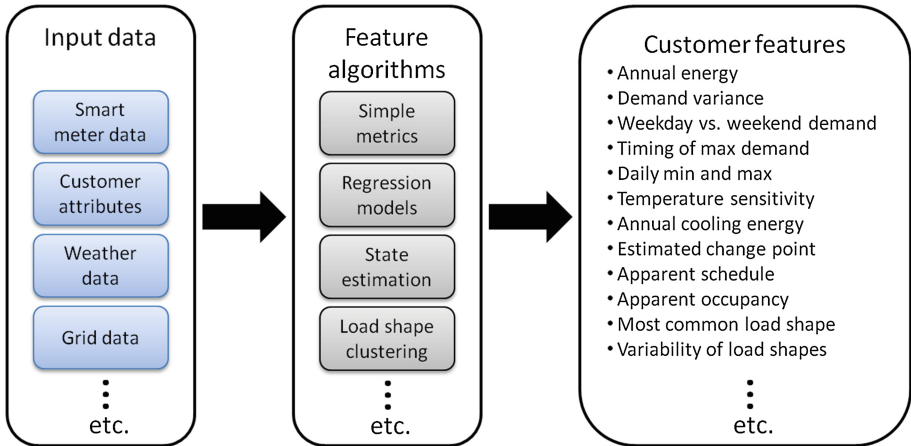


Fig. 1. Diagram of feature extraction, where pre-defined feature algorithms are run against time series meter data, augmented with supporting data, to produce an array of feature values for every customer.

- **Model outputs** of regression, state estimation, end use disaggregation, and other machine learning algorithms, including energy committed to heating and cooling, thermal balance points, occupancy patterns, hourly, day of week, and seasonal variations in consumption.
- **Load shape characteristics**, including clusters of typical consumption patterns, and metrics of load shape variability over time and across customers.

2.2 Interactive Feature Exploration

The VISDOM feature exploration interface allows users to interact with and interpret feature data for large numbers of customers. It is a web-based interactive system, implemented in Python, using Pandas [6] to manage and filter feature data and D3 [7] to create interactive visualizations.

As seen in Fig. 2, the VISDOM interface places customer filtering controls in the left column, visualizations, including maps with zip codes colored by feature values, histograms, scatter plots, cumulative sums, sorted values, and load shapes summaries, in the central column, and visualization controls in the right column. The visual data filters are used to identify specific customers based on their feature values, such as, climate, weather response, baseload, variability, and typical load shapes. Numerical features are filtered using interactive histograms and categorical data is filtered using multi-select boxes. Features drawn from filtered customers are rendered in real time to the interactive data visualizations. They can also be viewed as tabular data or exported as CSV files for further analysis or integration with external systems.

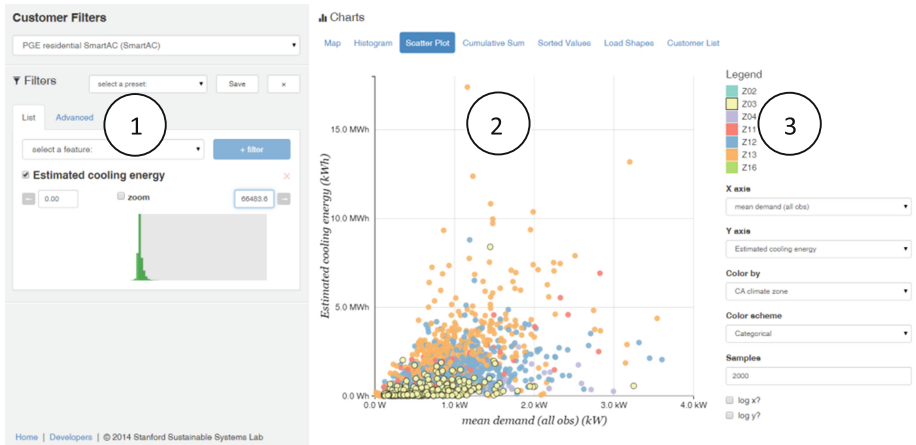


Fig. 2. Illustration of the interactive feature exploration interface. (1) Is the visual filtering interface for selecting a subset of customers with specific feature values, where (a) labels a grey histogram range filter. (2) Is the display area for the selected visualization, a scatter plot in this case, with (c) labeling the other visualizations. (3) Is the legend and the controls specific to the selected visualization, with (b) labeling an interactive feature of the legend.

Figure 2 contains a visualization comparing annual electricity consumption (x-axis) to estimated annual cooling energy (y-axis), colored by California climate zone, using the interactive legend to highlight customers from the central coast (climate zone Z03).

Typical usage moves from left to right in the interface: First, the user defines feature values that identify the group of customers he or she would like to study. Second, the user selects one of the available visualizations of customer features he or she would like to interact with. Finally, the user applies display controls specific to the selected visualization. However, it is expected that users will re-adjust filter controls, cross reference different visualizations, and update visualization control parameters as needed to spot patterns and better understand the underlying customers. Filter settings and entire views, including filter settings, visualization, and visualization settings, can be named and saved for future use or sharing with others.

3 Example Usage

Due to its diverse feature set and flexible interface, the VISDOM tool can be used to study consumption in a variety of contexts. For example: Utilities can use it to tie their infrastructure improvement plans to robust modes of customer demand; Energy Efficiency programs can be focused on the customers best suited to program offers; and city planners can better understand the needs of their citizens. This section provides several examples of VISDOM applied to a data set of 200,000 residential customers in Northern and Central California to illustrate the breadth of potential applications and uses.

3.1 Climate as a Proxy for Cooling

Current best practice for planning air conditioning efficiency programs is to focus on the hottest climates. This sensible approach can potentially overlook significant cooling loads in cooler climates and waste program recruitment resources on very low cooling loads that happen to be in hot climates.

Figure 3a provides a VISO map of average cooling intensity by zip code in the service territory of Pacific Gas and Electric, a California-based utility, with the hottest climate zone, CZ13, highlighted with a red outline. Here we can see that CZ13 does indeed contain significant cooling loads, but a histogram of annual cooling energy by customer in that zone, Fig. 3b, reveals a very wide diversity in cooling energy from one customer to the next. With the ability to estimate cooling energy on a per-customer basis, programs could focus on high usage customers, rather than general geography. This would lead to more effective cooling energy reduction and management programs.

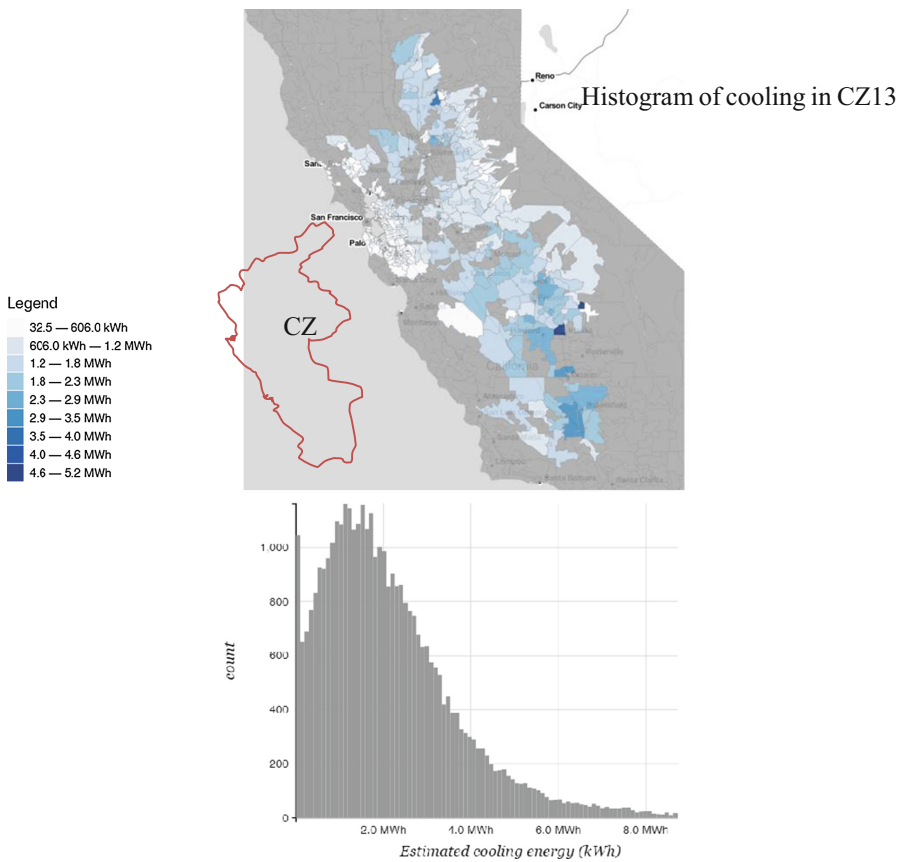


Fig. 3. (a) Map of estimated annual cooling energy for residences in Northern and Central California. (b) Histogram of individual customer annual cooling in the highlighted climate zone, CZ13.

3.2 Load Shape Characterization

Load shapes, or 24 h electricity use profiles, capture within-day variations in household energy consumption that reveal patterns in scheduling and occupancy. VISDOM clusters load shapes by similarity and provides tools for understanding the diversity of load shapes exhibited by individual customers or groups of customers.

The average load shape for each cluster can be thought of as the representative load shape for the whole group. The load shape visualization interface displays the representative load shapes for the clusters that best represent the subset of customers defined by the filter interface.

Figure 4 provides a visualization of the top 6 (by total energy) representative load shapes for residential customers in California’s hot CZ13. It can be observed that most of these shapes display significant afternoon or evening peaks, with one exhibiting a secondary morning/mid-day peak. Afternoon/evening peaks are typically driven by the coincidence of occupancy as residents return home from work and school and by cooling loads, which peak in the afternoon. The second cluster by total energy, which has a flat representative shape, is the first in total members. This means that the load shapes classified into the second cluster tend to be more common, but lower energy than the first. This is best explained by observing that nearly all homes go full days without occupants, so every customer contributes at least a few flat load curves.

Because they lend themselves to behavioral interpretation and reflect patterns in energy demand of significant interest to grid operators and planners, a better understanding of load shapes can translate directly into improved planning and execution of Energy Efficiency and Demand Response programs.

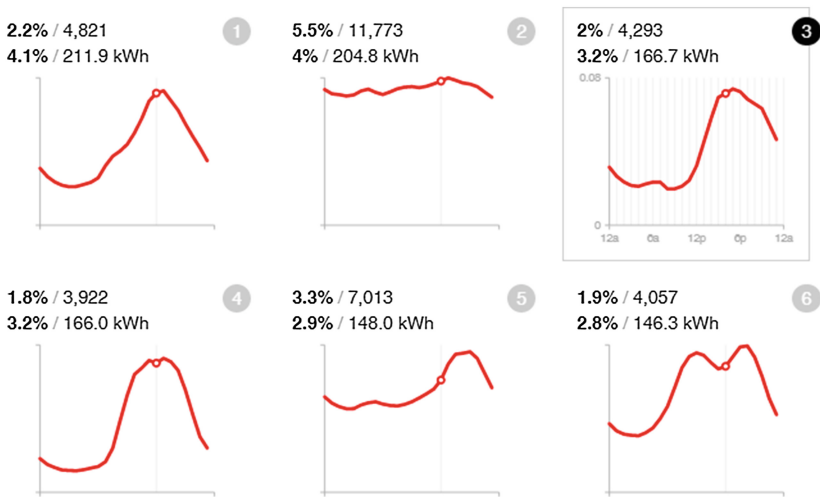


Fig. 4. VISDOM visualization displaying The top 6 load shapes, or daily energy usage profiles, for customers in California’s hot climate zone CZ13.

3.3 Segmentation and Targeting

The customer segmentation and targeting features of VISIBLE make use of predictive models of customer responses to energy efficiency and demand response programs [8] to identify and rank the best customers (targeting) or groups of customers (segmentation) to approach with a specific offer or incentive. Customer responses can be estimated using expert knowledge and opinion, derived from engineering principles, or based on the outputs of models trained on the observed outcomes of past programs. For example, households estimated to use large amounts of air conditioning based on their meter data would likely have the greatest cooling energy (kWh) and demand (kW) savings from participation in programs that improve the efficiency of or control over air conditioning equipment.

Alternately, the features found in VISIBLE can be used as inputs into predictive response models. Features related to temperature sensitivity, cooling energy use, typical patterns of daily energy use, and physical location can be combined with ground-truth program participation and outcome data to generate a propensity score that identifies customers best able to reduce demand during periods when the grid is operating near its physical limits, i.e. when those reductions are most needed and valuable.

Figure 5 illustrates VISIBLE’s cumulative sum visualization. The cumulative annual cooling energy in climate zone CZ13 is calculated from values sorted from highest to lowest (upper red curve) and then again based on random sampling of the population (lower black curve). At the 25th percentile of customers (dashed vertical

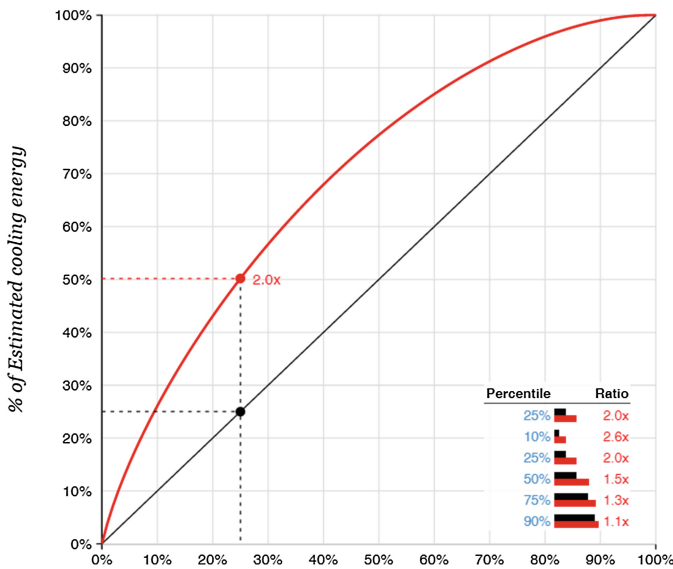


Fig. 5. Cumulative sum of estimated annual cooling for CZ13 sorted from highest to lowest values (upper, red) compared to the cumulative sum of annual cooling selected at random from the population.

line), the sorted cumulative cooling energy is two times the unsorted, indicating a significant potential return on program targeting: 50 % of cooling energy is used by 25 % of the customers, so we could expect that targeted program marketing could provide twice the savings per customer as untargeted.

4 Conclusions

The benefits latent in the extraordinary volume of smart meter data now flowing have only just begun to be realized. We have provided details of design and implementation of VISDOM, a system for extracting meaningful and actionable features from very large samples of smart meter data, including complete samples of customers in a given utility service territory or geographic region. VISDOM allows users to filter, sort and group subsets of customers using feature criteria and to display attributes of filtered customers using generic (i.e. histogram and scatter plot) or task specialized (i.e. cumulative sum and load shape) interactive visualizations.

We have provided examples of how these capabilities can be used to identify and better understand significant variability in cooling loads within hot climates, how customer load shapes (24 h energy use profiles) can be classified into clusters that share similar scheduling and occupancy characteristics, and how the features of individual customers can be used to model their expected response to energy programs so the most responsive customers can be targeted by program recruitment efforts to double program savings.

Collectively, the capabilities of VISDOM can support more precisely targeted and effective Energy Efficiency and Demand Response programs, grid operations based on robust spatial and temporal models of customer demand, and grid planning that accounts for customer demand flexibility when planning large scale infrastructure investments. Each of these applications is an important piece in the puzzle of how to plan, design, and operate the grid in the 21st century.

Acknowledgements. The authors wish to thank Shawn Allen for applying his exceptional skills in browser-based data visualization and interactive design on this project. This research was funded in part by the Department of Energy ARPA-E under award number DE-AR0000018, the California Energy Commission under award number PIR-10-054, and the Precourt Energy Efficiency Center.

References

1. Institute for Electric Innovation, Utility-scale Smart Meter Deployments: Building Block of the Evolving Power Grid, Edison Foundation, September 2014
2. Core, R., Team, R.: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2013)
3. Borgeson, S.: Targeted Efficiency: Using Customer Meter Data to Improve Efficiency Program Outcomes. Dissertation, Berkeley (2013)

4. Albert, A., Rajagopal, R.: Smart meter driven segmentation: what your consumption says about you. *IEEE Trans. Power Syst.* **28**(4), 4019–4030 (2013)
5. Kwac, J., Flora, J., Rajagopal, R.: Household energy consumption segmentation using hourly data. *Smart Grid IEEE Trans.* **5**(1), 420–430 (2014)
6. McKinney, W.: Data structures for statistical computing in python. In: *Proceedings of the 9th Python in Science Conference*, pp. 51–56 (2010)
7. Bostock, M., Ogievetsky, V., Heer, J.: D3: data-driven documents. *IEEE Trans. Vis. Comp. Graph. Proc. InfoVis*, 2011
8. Kwac, J., Rajagopal, R.: Demand response targeting using big data analytics, pp. 683–690 (2013)