Activity Based Energy Demand Modeling for Residential Buildings

Rajesh Subbiah, Kristian Lum, Achla Marathe and Madhav Marathe
Network Dynamics and Simulation Science Laboratory, Viginia Tech
Email: {csrajesh, klum, amarathe and mmarathe}@vbi.vt.edu

***Abstract*—Statistics reveal that almost one-fourth of the overall power consumption in the US is by residential buildings; this number is only increasing. In order to manage the growing demand for energy, there is a need for energy system optimization, which would require a realistic, high resolution energy-demand model. To achieve this goal, we propose a modeling framework aimed at generating household energy demand profiles based on individual-level energy consuming activities. The model associates appliance usage for each household activity and calculates energy consumption based on the appliance energy rating and duration of activity. We use this information to generate a household level energy demand profile for roughly 63,000 households and discuss the possible uses of generating residential energy demand patterns. This paper provides a novel way to resolve correlational and consistency problems in the generation of individual-level and household-level "shared" activities which occur due to household members' interactions.**

***Index Terms*—Correlation, Demand side management, Logistic regression, Poisson regression, Smart grid.**

I. INTRODUCTION

Twenty-five percent of total energy consumption in the United States is attributable to the residential sector, and that number is expected to rise due to the increased use of appliances and electronic devices [1]. This makes the residential sector an important target group for energy conservation. To analyze any modern energy optimization strategy, accurate energy demand profiles of residential buildings are an important prerequisite. For instance, ongoing transformation of electric grids into smart grids provides the technological basis to implement demand-sensitive pricing schemes aimed at using the electric power infrastructure more efficiently. We need accurate energy demand forecasting models to understand the feasibility of such schemes. Towards that end, we propose a modeling framework which generates accurate energy demand profiles of households based on the household members' activities and also helps us study how the necessary incentives can influence individuals' or households' behavior towards energy usage.

Several studies [2] [3] have looked at modeling residential energy demand using time use data by constructing occupancy patterns. However, there has been little published work [4] [5] for generating energy demand profiles at a detailed household resolution based on the activities performed in each household. These works do not address the inherent time-use dependence within a household due to the sharing of activities among the household members. The national time use survey only takes into account the activity schedule of the individuals who responded to the survey; however, all household members contribute to the total energy load for the household. In order to account for this fact, we propose a detailed demand analysis, dis-aggregated to the level of the individual household member and the appliances used within an individual household.

We fill this gap by proposing a data driven model which takes into account the social, behavioral and economic aspects of individuals and households. We achieve this by building demographic-based individual activity schedules for each household member, while accounting for the within-household dependence due to shared and coordinated activities. We then aggregate the individuals' activities to generate the per household activity schedule. We then show the applicability of the detailed demand model in smoothing the load curve. We present an illustrative study which uses the demand profile to make efficient policy decisions to spread out the activities between peak and off peak hours based on the adaptability of households.

The rest of the paper is organized as follows. Section II examines related work. Section III explains the model and explains how we map appliance usage with activity and generate per household energy demand profile. Section IV explains our experimental results and our illustrative study to smooth out the load curve. Section V concludes.

II. RELATED WORK

Related work in this area can be broadly categorized into two categories: approaches that model energy consumption at an aggregate level ("top-down" approaches) and others that model residential energy demand profiles based on household activities and occupancy patterns ("bottom-up approaches").

The top-down approach models energy consumption as a function of macroeconomic factors, e.g. price and climate, using techniques such as regression over historical averages [6] [7] [8]. These approaches model the effect of long-term changes and macro (system-level) socio-economic and ecological variables on energy consumption. Because this methodology utilizes only aggregate macro-level data, it is relatively simpler to develop.

Bottom-up approaches [9] [10] [11] [12] study the impact of demographics on energy consumption. These works observe that electricity consumption depends heavily on ownership of energy intensive appliances, which, in turn depends on income and the size and composition of the household. A bottom-up approach then estimates energy consumption at the regional and national levels by extrapolating from a representative set of individual households. Work by [13] provides a fairly exhaustive review of the pros, cons and applicability of various modeling techniques for residential energy consumption.

In this paper, we use a bottom-up approach to calculate the per-household energy consumption based on the household members' activity sequence. Using the American Time Use Survey Data [14], we model activity patterns using individual and household level demographic covariates. We then use the parameters obtained from fitting our models to the ATUS data to create new activity diaries for a synthetic population based upon its demographic covariates. We match these activities to the requisite appliances (and their associated energy consumption) to create an energy demand profile for each household.

III. Methodology

In this section, we describe our modeling approach for generating an energy demand profile for the synthetic population representing the Washington D.C. area [15] [16]. Table I summarizes the different data sets used in this study. The synthetic population contains individual and household level demographic information such as type of household, income, number of household members, etc. Using the ATUS data and EIA's Residential Energy Consumption Survey (EIA-RECS) [17] data, we employ a model-based strategy to extrapolate the characteristics seen in the survey data to our synthetic population. Using these characteristics, we construct an activity based energy demand profile.

Dataset	Type	Description
ATUS	Survey Data	Contains 24 hour period activity dairies for 13,260 respondents
EIA-RECS	Survey Data	Contains detailed household level characteristics which affect household's energy consumption
Synthetic Data	Generated data	Contains household level and individual level demographics representing Washington-DC area. It is constructed using census and transportation data

TABLE I
DATA SOURCES USED

We classify the household activities into two major categories and represent the total energy consumption of a household as

$E_{Total} = E_{Active} + E_{Passive}$,

where E_{Active} is the energy consumed due to appliance usage from individual or shared activities, e.g. the energy consumed when a household member uses the dishwasher. These activities are mainly a function of the household's daily schedule. $E_{Passive}$ is the energy consumed for general maintenance of the house, such as space heating, space cooling, and water heating. These activities mainly depend on the climate and characteristics of the house, namely the type and size of housing unit, fuel used, insulation, wall type etc. and are mostly independent of activities of the residents.

A. Energy calculation for Appliance usage

1) Activity sequence generation: The ATUS data consists of the activity diaries of $13,260$ respondents: a 24 hour period detailed description of activities (duration, location, etc.) and the respondent's demographic details. The major limitation of the ATUS data is that it represents the time use pattern of the respondent alone and not the entire household. In order to construct a household's complete daily activity schedule, we followed these steps

1) Based on [18], we select the highest energy consuming activities in a typical household. These are summarized in Table II and appear in the calculation of E_{Active}. Other common but less energy-intensive activities are categorized as well but not included in the calculation of E_{Active}, such as bathing, work, shopping, etc.
2) E_{active} is further refined as follows:
   a) *Shared activities:* These are the activities in which appliance usage is generally shared among all the household members. All shared activities except cooking, are assumed to occur at most once daily.
   b) *Independent activities:* These are the activities in which the appliance usage is not shared. These activities can occur multiple times in a day and are independent for every household member.

Activity Name	Activity Type
Laundry	Shared
Dishwashing	Shared
Computer usage	Independent
Watching TV	Shared
Cooking	Shared
Interior Cleaning	Shared
Checking Email	Independent

TABLE II
ENERGY INTENSIVE ACTIVITIES FROM ATUS

3) To assign an activity sequence to each synthetic individual, we match them with an ATUS survey respondent based upon the similarity of their demographics. Our main objective is to partition the survey data set into smaller data sets defined by the set of n demographic variables represented as $\vec{X} = (X_1 ... X_n)$, so that we can overlay the activity sequence of the ATUS survey respondents on the synthetic population. We use the CART algorithm [19] to construct a binary decision tree. Initially, the complete set of surveyed people are represented as root node of the tree and demographic variables $\vec{X}$ are the splitting variables. At each stage, the algorithm tries to split the node into two groups based on the best possible splitting variable. The algorithm identifies the splitting variables after performing an exhaustive search of all possible combinations. The process is recursive in nature i.e. each node can be split into two child nodes and, in turn, each of these child nodes may themselves be split, forming additional child nodes. The final constructed tree uses marital information as the dependent variable and rest of the variables (gender, employment status, etc.) are used as independent splitting variables.
4) Assigning independent activities
   a) Each household member in the synthetic population is assigned a leaf node based on his/her demographics variables

b) We select a ATUS survey respondent at random from that leaf and assign the activity pattern of the ATUS survey respondent to the synthetic individual.

5) Assigning shared activities

Interior Cleaning, Dishwashing and Laundry

We noticed that few people explicitly listed activities as shared in the ATUS survey data. Since we have detailed information about only one person in the household, it is ambiguous whether an activity was not performed or another household member performed the shared activity and it did not get listed in the survey. Thus, to generate the shared activity sequence for synthetic households, we need to know whether any person in household H_i consisting of people $P_{(1,\ldots,m)}$ have performed a particular activity. If a household has done the shared activity, then we need to find out the most likely time period in which the shared activity was performed and by whom it was performed. The following steps illustrate our approach to assigning a shared activities to households:

a) For each household member, P_{ij} in H_i, we calculate the probability that P_{ij} with the demographic variables $\vec{X}_{ij}$ performs activity k. That is, for A_{ijk} an indicator which takes value one if person j in household i performs activity k and 0 otherwise, we estimate $\hat{\mu}_{ijk} = Pr(A_{ijk} = 1|\vec{X}_{ij})$ using Logistic regression [20]. The next section explains the Logistic regression in detail.

b) After calculating $\hat{\mu}_{ijk}$, we used the inclusion and exclusion principle [21] to calculate the probability, $\hat{\mu}_{ik}$, that *someone* in the household performed activity k. That is, we assume that each individual independently decides on a given day whether or not they will perform activity k. Then, the probability that household i performs activity k is equal to the probability that at least one person in the household decides to do the activity.

c) We then designate household i as having performed activity k with probability $\hat{\mu}_{ik}$.

d) If the activity k has occurred in household i, i.e. $A_{ik} = 1$, then we divide the day into 48 time slots and for each time slot t

i) For each household member, P_{ij} in H_i, we calculate the probability that P_{ij} with the demographic variables $\vec{X}_{ij}$ performs activity k during the time slot t using Logistic regression [20].

ii) After calculating the probability for each household member individually, we again use the probability inclusion and exclusion principle [21] to calculate the probability $\hat{\mu}_{ikt}$ that the household performed shared activity k during the time slot t.

e) Because we calculate this for each time slot independently, some adjustments are required as these 'probabilities' do not necessarily add to one. As stated, we assume that these shared activities occur at most once per day, so given that the activity occurred, we must select exactly one of the time slots in which to place the activity. After calculating the probabilities for all the time slots independently, we re-normalize them so that they add to one, i.e. $\dot{\mu}_{ikt} = \frac{\hat{\mu}_{ikt}}{\sum_{t=1}^{48} \hat{\mu}_{ikt}}$. We then select time slot t with probability $\dot{\mu}_{ikt}$.

We perform the above steps for the all shared activities except cooking and generate the shared activity sequence for each household present in the synthetic population.

Cooking

Unlike other shared activities, cooking can occur multiple times in a day. To model this behavior our main objective is to estimate the number of cooking events that can possibly occur in a household H_i. For each household member, we calculate the expected number of cooking events performed by that P_{ij}, given covariates, using Poisson regression [22].

6) For each household we aggregated the independent activity sequence of all the household members and the shared activity sequence to get the complete activity sequence.

2) Logistic Regression: As described in the earlier section, given an indicator for an activity A_{ijk} (which takes value 1 if activity k was performed by person j in household i and 0 otherwise) and an n dimensional set of demographic variables $\vec{X}_{ij}$ that relate to person ij, our objective is to determine the probability of that person performing the activity,

$$Pr(A_{ijk} = 1|\vec{X}_{ij}) = \frac{1}{1 + e^{\vec{X}_{ij}\beta}}.$$

To obtain the estimated coefficients,

$$\hat{\beta}_k = max_{\beta_k} \prod_{i,j} \pi_{ijk}^{A^*_{ijk}} (1 - \pi_{ijk})^{1-A^*_{ijk}},$$

where $\pi_{ijk} = \text{logit}^{-1}(\vec{X}^{*T}_{ij}\beta_k)$, we fit a logistic regression to the survey data. The (*) notation indicates survey data; unstarred covariates and outcomes refer to synthetic population members.

Then, for each household member in the synthetic population, we matrix multiply his/her demographic covariates by the model coefficients to obtain the probability that he/she performed action k. That is

$$\hat{\mu}_{ijk} = Pr(A_{ijk} = 1|\vec{X}_{ij}) = logit^{-1}(\vec{X}^T_{ij}\hat{\beta}_k).$$

We follow a similar approach for calculating the probability of doing an activity for a particular time slot by calculating the regression coefficients for that time slot separately.

3) Poisson Regression: We use Poisson regression to fit the number of cooking events performed by individuals in household H_i. In this case, the dependent variable is an integer (the number of cooking events performed by each household member). Poisson regression relates the set of demographic covariates to this integer value via the model $C_{ij} \sim Pois(\mu_{ij})$ and $log(\mu_{ij}) = X_{ij}\beta$. By fitting a Poisson regression, we obtain coefficient estimates, $\hat{\beta}$ and from these we sample the number of cooking events for each individual as $Pois(\exp\{X_{ij}\hat{\beta}\})$. To resolve any discrepancies among the household in terms of the total number of cooking events that take place by day, we simply select the maximum number of individual cooking events to be the total number for the household. This is equivalent to assuming that the person who cooked the most times over the day was present at every cooking event. Following this procedure, we use our Logistic approach to find the times at which these events occur.

4) Associating Appliance usage and Energy Demand Calculation: After generating the per household activity sequence, the next major step is to identify and associate appliances to each activity A_k. We assume that these are the appliances which get utilized whenever activity A_k occurs. Using the appliance standard wattage rating from [18] and the duration over which it is active, we estimate the energy consumption using the formula

Appliance Wattage * Appliance Active Duration = Energy Consumed

Activity Name	Appliance Used	Energy rating (watts)	Usage
Laundry	Washer	234	0.45
Laundry	Dryer	670	0.55
Dish washing	Dishwasher	1200	1
Cooking	Microwave	500	.5
Watching TV	Television	220	1
Computer Usage	Computer	160	1
Cooking(Morning)	(Stove, Coffee maker Microwave, Toaster Oven, Blender)	865	(.35,.05 .5, .05 0.0, 0.05)
Cooking(Night)	(Stove, Coffee maker Microwave, Toaster Oven, Blender)	940	(.35,.05 .45, .05 .05, 0.05)

TABLE III
ACTIVITY-APPLIANCE USAGE AND ENERGY RATING INFORMATION

Sometimes, we may encounter an activity which uses multiple appliances. In this scenario, the energy consumed is the sum of energy consumed by each appliance. Also, while calculating the energy for each individual appliance, we need to disaggregate the activity duration to individual appliance level. Since, we do not have the necessary information on this, we use a parameter called "usage fraction" which gives a rough estimate of the duration a single appliance is active. For example, if laundry activity takes 80 minutes and it uses washer and dryer; we associate the usage fraction as 0.45 and 0.55 respectively. This means that in 80 minutes, washer is used for 36 minutes and dryer for 44 minutes. This process becomes difficult for cooking activity because different household use different set of appliances for their the cooking activity. Also, the set of appliances varies based on what meal is cooked (breakfast, lunch or dinner). To approximately estimate the energy consumed due to the cooking activity we associate a generic set of appliances based on the cooking activity time, shown in Table III.

B. Energy Calculation for Common Household Activities

We use the EIA Residential Energy Consumption Survey-2009 (EIA-RECS) data to estimate the energy consumed $E_{Passive}$ due to common household activities. EIA-RECS is a national survey which collects energy related information from different housing units across the country. Then, it estimates the energy consumption for the whole United States. We use the parameters household size, household type, household income and regional information to match the synthetic household with the EIA-RECS households. After filtering the EIA-RECS data by these properties, we pick one housing unit at random based on the weights associated with the EIA-RECS housing unit [17].

1) Space Heating and Cooling: Energy consumed due to space heating and cooling depends on climatic conditions, fuel used, type of heating/cooling equipment used, etc. To approximately calculate the energy consumed due to this activity we pick a day from winter season and gather the hourly weather data for that day from [23]. At each synthetic household unit we derive (a) S Average square footage used for space heating (b) T_p Temperature when someone is at home during the day (c) T_a Temperature when no one is at home during the day (d) T_n Temperature at night (e) Fuel and equipment used for heating (f) Wall type from the EIA-RECS survey data. Using these information and the average hourly outside temperature [23], we use Fourier's law to calculate the Heat loss rate Q as

$$Q = \frac{(Area)*(T_{inside} - T_{outside})}{ThermalResistanceofWall} = \frac{S*\delta T}{R},$$

where T_{inside} depends on the people's occupancy factor and can take any one of these values T_p, T_a and T_n. Since, we have the activity sequence, we sort and order all the activities occurring in the household. Then, we scan for occurrence of any in-house activity in each of the 48 time slots. If we encounter any such activity, then we assume that there is someone present in the house performing the activity. So, we assign $T_{inside} = T_p$ for that time slot and also if the activity encountered is "sleeping" then we assign $T_{inside} = T_n$. Since, the activities that occur in the household and outside the household (like going to work) are complementary to each other, for rest of the time slots we assign $T_{inside} = T_a$.

We assume R, the thermal resistance, is constant throughout the structure, and we use the standard values of R based on wall type: "4inch thick brick" wall with R-value = 4 $\frac{ft^2 \deg Fh}{Btu}$ and a "cellulose fiber" wall with R-value = 3.70 $\frac{ft^2 \deg Fh}{Btu}$.

To keep the household at a desired temperature, we need heating equipment to generate the heat energy required to compensate for the heat loss Q. Generally this kind of equipment has an efficiency parameter, η, which measures the amount of energy that translates to actual work. For example if a household uses a natural gas furnace for their space heating and if the furnace operates at 75% efficiency, then, the furnace needs $\frac{Q}{.75}$ amount of energy to keep the house at the desired temperature. Based on the fuel used in the household, we associate an efficiency value using [24], shown in Table IV.

Based on this information, we calculate the energy required to keep household at the desired temperature on an hourly basis.

Fuel Used	Equipment	Efficiency
Natural Gas	Furnace/Boiler	.78
Natural Gas	Room Heater	.65
Wood	Room heater	.72
Natural Gas	Other	.47
Electricity	Furnace/Boiler	.98
Electricity	Heat Pump	3.3
Fuel Oil	Furnace/Boiler	.78
Kerosene	Room heater	.80

TABLE IV
EFFICIENCY FOR VARIOUS HEATING EQUIPMENT

2) Hot water usage: To estimate the energy consumed due to water heating, we identify activities which require hot water: laundry, dishwashing, showering and cooking. Laundry, dishwashing, taking shower for 8 minutes and cooking requires 7, 6, 10 and 1 gallons of hot water respectively [25]. Based on the type of water heating equipment, electric or gas, we associate the energy factor 0.92 and 0.67 [25] respectively. The energy factor indicates an efficiency measure based on the amount of hot water produced per unit of fuel consumed, which we use to estimate the amount of energy required for each of the activities that consume hot water.

IV. EXPERIMENT AND RESULTS

A. Model Validation and Verification

For model validation, we use a synthetic population of the Washington D.C region. We randomly select 5% of the households i.e. 62,763 households totaling 125,268 synthetic persons. We implement the statistical modeling algorithms in R using [26] [27]. The first experiment verifies the ability of our model to produce realistic activity sequences. For every half an hour interval we calculate the fraction of households performing an activity for both ATUS and synthetic data. Figure 1 compares the model based results for washing, dish-washing, cooking and cleaning activities with the ATUS survey results. As we can infer from these plots our model is able to capture the activity patterns present in the ATUS survey data with less than 5% deviation. Slight deviations between the ATUS data and the model's output are expected, as the synthetic population may have slightly different demographic characteristics than those of the individuals in the survey.

B. Energy demand profile construction

We associate the energy consumption for each activity and construct the energy demand profile for all households. Figure 2A shows the aggregate energy demand profile (E_{Active}) for all the "active" household activities and Figure 2B shows the demand pattern for individual activities. We can infer that during the day-time 'washing' and 'cleaning' activities are responsible for the peak in the E_{Active} load profile. During the night the peak is due to the 'cooking' and 'watching TV' activities. Figure 3A shows the estimated energy consumption due to space heating and hot water usage. Figure 3B shows the aggregate energy demand profile by aggregating both E_{Active} and $E_{Passive}$.

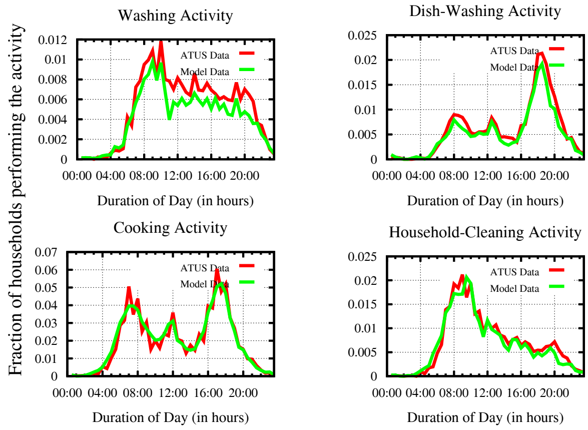

Fig. 1. Comparison of synthetic data activity occurrence frequency with ATUS data activity occurrence frequency

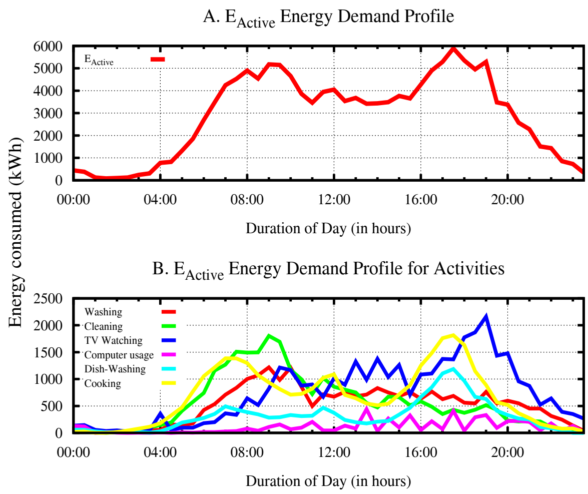

Fig. 2. (A) E_{Active} Energy demand profile (B) Energy demand profile for individual activities

C. Energy Demand Scaling

In this section we will illustrate one possible use of this detailed demand model: smoothing out the load curve. Researchers have been searching for ways to make load more responsive so that the demand can be shifted from peak to off-peak periods. This will keep the inefficient generators from coming online to serve the load. The energy demand profile of E_{Active} shows spikes during the peak hours i.e. 8:00 AM to 11:00 AM. These peaks are mainly due to cleaning and washing activities. After closer examination of these activities, we found that about 40% of the households performing these activities during the peak hours have at least one non-working adult in their household. We select 50% of these households and shift their activities from peak hours to off-peak hours i.e. between 11 AM to 3 PM. Then, we recalculate the energy demand profile. Figure 4 compares the scaled demand profile with the actual. By doing this, we are able to shift about 4.5 MWh from the peak period to the off peak period. These savings represent the difference in the area between the two curves from 8am to 11am in Figure 4. This is a substantial amount of energy savings at peak time and could mediate the vulnerabilities that occur when the system is running too close to the edge of capacity.

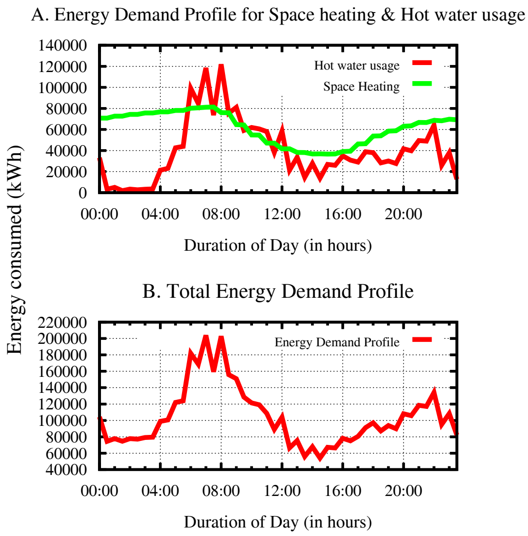

Fig. 3. (A) Energy demand profile for Space heating and Hot-water usage (B) Total energy $(E_{Active} + E_{Passive})$ demand profile

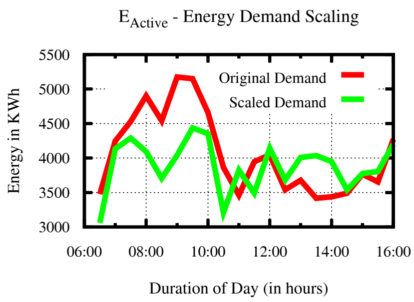

Fig. 4. Energy Demand Scaling

V. CONCLUSION

This research presents a highly disaggregated household energy demand model for the DC region. It models household activity sequences based on individual members of the household, maps them to appliance usage, and generates energy demand at the household level. The model accurately maps the survey data activity sequence on the synthetic households with less than 5% error. An illustrative study shows the applicability of this detailed demand model. For 20% of the households, some of the peak time activities were shifted to off-peak time and were performed by the non-working members of the family. This results in smoothing out the load curve and saving 4.5 MWh at peak time. This model can also help study the necessary incentives for making demand more price responsive and consumption more efficient.

In the future, we plan to investigate scenarios like peer consumption, social norms of consumption and etc. We would like to compare the demand profiles of two completely different regions and study how variation in region, fuel and equipment usage, household and individual demographics impact energy consumption. We would also like to study the magnitude of the effect of regional demographic variation on energy consumption.

REFERENCES

[1] U.S. Energy Information Agency, *Share of energy used by appliances and consumer electronics increases in U.S. homes*, 2011.

[2] I. Richardson, M. Thomson, and D. Infield, "A high-resolution domestic building occupancy model for energy demand simulations," *Energy and Buildings 40 (2008) 15601566*, 2008.

[3] Y. Shimoda, T. Fujii, T. Morikawa, and M. Mizuno, "Residential end-use energy simulation at city scale," *Building and Environment 39 (2004) 959 967*.

[4] J. Widen and E. Wackelgard, "A high-resolution stochastic model of domestic activity patterns and electricity demand," *Applied Energy*, vol. 87, no. 6, pp. 1880–1892, June 2010.

[5] Y.-S. Chiou, "Deriving US household energy consumption profiles from american time use survey data a bootstrap approach," *Eleventh International IBPSA Conference Glasgow, Scotland*, 2009.

[6] J. Min, Z. Hausfather, and Q. F. Lin, "A High-Resolution Statistical Model of Residential Energy End Use Characteristics for the United States," *JOURNAL OF INDUSTRIAL ECOLOGY*, vol. 14, no. 5, SI, pp. 791+, OCT 2010.

[7] E. Hirst, "A model of residential energy use," *SIMULATION*, vol. 30, no. 3, pp. 69–74, 1978.

[8] B. Halvorsen and B. M. Larsen, "Norwegian residential electricity demand–a microeconomic assessment of the growth from 1976 to 1993," *Energy Policy*, vol. 29, no. 3, pp. 227–236, February 2001.

[9] K. Atkins, A. Marathe, and C. Barrett, "A computational approach to modeling commodity markets," *Computational Economics*, vol. 30, no. 2, pp. 125–142, September 2007.

[10] SMRProspectus, "Consumer energy spending and the demographics of over-consumption," 2009.

[11] D. Brounen, N. Kok, and J. M. Quigley, "Residential energy use and conservation:economics and demographics," June 2011.

[12] A. Capasso, W. Grattieri, R. Lamedica, and A. Prudenzi, "A bottom-up approach to residential load modeling," *Power Systems, IEEE Transactions on*, vol. 9, no. 2, pp. 957 –964, may 1994.

[13] L. G. Swan and V. I. Ugursal, "Modeling of end-use energy consumption in the residential sector: A review of modeling techniques," *Renewable and Sustainable Energy Reviews*, vol. 13, no. 8, pp. 1819 – 1835, 2009.

[14] "American time use survey," 2010. [Online]. Available: http://www.bls.gov/tus/

[15] C. L. Barrett, R. Beckman, M. Khan, and et.al, "Generation and analysis of large synthetic social contact networks," in *Proceedings of Winter Simulation Conference (WSC)*, Austin, Texas, 2009, pp. 1003–1014.

[16] R. Beckman, K. Baggerly, and M. McKay, "Creating synthetic base-line populations," *Transportation Research A Policy and Practice*, vol. 30, pp. 415–429, 1996.

[17] "U.S. Energy Information Agency - Residential Energy Consumption Survey," 2009. [Online]. Available: http://www.eia.gov/consumption/residential/index.cfm

[18] *Buildings Energy Data Book*. U.S. Energy Information Agency, 2011.

[19] Breiman, Leo, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*, 1984.

[20] D. Kleinbaum and M. Klein, "Maximum likelihood techniques: An overview," *Logistic regression*, pp. 103–127, 2010.

[21] Thomas Foregger, Kevin Ferguson, *"Principle of inclusion-exclusion"(version 6). PlanetMath.org*.

[22] *Regression Analysis of Count Data (Econometric Society Monographs)*. Cambridge University Press, 1998.

[23] "National Climatic Data Center," 2011.

[24] U.S. Energy Information Agency, *Heating Fuel Comparison Calculator (Excel)*, 2012.

[25] U.S. DOE - Energy Efficiency and Renewable Energy, *Energy Cost Calculator for Electric and Gas Water Heaters*, 2011.

[26] S. Davies and R. core team, *glm: Fitting Generalized Linear Models*. [Online]. Available: http://stat.ethz.ch/R-manual/R-patched/library/stats/html/glm.html

[27] T. M. Therneau, B. Atkinson, and B. Ripley, *rpart: Recursive Partitioning*, 2011. [Online]. Available: http://CRAN.R-project.org/package=rpart