Proceedings of the 2018 Winter Simulation Conference
M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson

SIMULATING RESIDENTIAL ENERGY DEMAND IN URBAN AND RURAL AREAS

Swapna Thorve
Samarth Swarup
Achla Marathe
Youngyun Chungbaek
Eric K. Nordberg
Madhav V. Marathe

Network Dynamics and Simulation Science Laboratory
Biocomplexity Institute of Virginia Tech
Blacksburg, Virginia, 24060, USA

ABSTRACT

Residential energy demand dynamics at household level can be studied through demographic, behavioral and physical characteristics of the household. In this paper, we develop an agent-based model using a bottom-up approach to build disaggregated energy demand estimates at the household level at an hourly interval. A household level analysis is made possible via the use of synthetic populations for the urban and rural areas of Virginia, USA. The energy consumption estimate is based on householders' demographics, their behaviors and activities, ratings of appliances used in energy-related activities, space conditioning fuels, physical characteristics of the home, and weather conditions. Results from the simulation are then validated with actual demand curves from Rappahannock county in Virginia using dynamic time warping. The simulation results show that the model produces realistic energy demand profiles.

1 INTRODUCTION

Approximately 22% of the total energy consumption in the U.S. is attributable to the residential sector (Energy Data Facts 2016). Consequently, residential sector energy usage has been studied widely and is a focus for energy conservation efforts. When the energy requirements of households are assessed it is complex to develop granular level demand profiles due to the sheer number of parameters. This leads to the challenge of scalability of such models when dealing with big data. Ignoring some of the parameters can lead to underestimation of energy consumption for certain activities. Due to these reasons, most prior research work for the generation of demand curves is concentrated at the regional level, city level, and building-stock level. Literature has demonstrated a general lack of availability of real-world energy consumption data. This makes it difficult to understand the underlying changes that affect energy consumption in the residential sector. Hence, validation of such demand profiles becomes challenging. It is also difficult to find 100% match to the demand curve because no two households exhibit identical behavior. National Academies of Sciences, Engineering, and Medicine (2016) have also emphasized that generation of synthetic energy datasets will be key for the future grid in the planning regime, customer feedback, and designing of incentives. Due to lack of availability of realistic data for testing advanced energy models for the smart grid, it is essential to develop this synthetic data. These models should be vigorously validated so that these datasets can be made available to the research community.

1.1 Related Work

Several studies like Wallis et al. (2016), Nguyen and Aiello (2013), Heiple and Sailor (2008) have illustrated that attributes such as climate, dwelling characteristics, demographics, activity schedules, and occupancy patterns are crucial parameters for estimating energy consumption. Swan and Ugursal (2009) present an exhaustive review of modeling techniques used for residential sector energy consumption. Two distinct approaches have been described: top-down and bottom-up. Here, we use a bottom-up methodology to build energy demand profiles at an individual and household level to generate energy demand profiles for a given day. A brief overview of relevant bottom-up models is provided below.

Significant research has been conducted in the European Union for developing energy usage calibration models at household and building stock level. A wide variety of models, including probabilistic-empirical load models, simulations, Markov chain models, and multiple linear regression models, have been developed to explore energy demand at building stock level in urban-rural settings by using a range of attributes such as the year of construction, the heating volume of buildings, the average height of the floor, appliances, weather, national events, and family types (Fonseca and Schlueter 2015; Marszal et al. 2016; Sandels et al. 2014; McLoughlin et al. 2012; Heinonen and Junnila 2014). These studies have been conducted in Switzerland, Ireland, Italy, Denmark, and Sweden. These models generate demand at various scales such as hourly, monthly, or yearly for different seasons. Some models have analyzed annual effects of activities such as cooking, dishwashing, laundry, and cleaning on Dutch households (Bedir and Kara 2017). Studies conducted in China have attempted to explain urban and rural energy consumption profiles by adopting bottom-up modeling techniques (Yang et al. 2015; Zhang et al. 2016; Yu et al. 2018). Fuel type, floor area, climate and effects of demographic transitions on energy consumption are some of the major parameters that were considered in these models. There are only a few agent-based models for energy demand generation (Bustos-Turu et al. 2016; Subbiah et al. 2017). These models show the potential for granular level analysis of energy usage by activities, and conducting impact assessment of low carbon technologies, such as plug-in electric vehicles.

1.2 Our Contribution

1. Major contributions to energy demand generation via agent-based models: (a) finer details such as dwelling type, fuel mixes, equipment efficiencies, urban vs. rural households, and adult occupancy and their energy related activities during the time spent at home are accounted for; (b) improved matching methodology for assigning survey households to synthetic households and survey individual energy activity schedule to synthetic individuals; (c) a shared activity is chosen from individual activities based on appropriate time duration and whether the synthetic individual is home at that time; (d) a workflow is devised for cooking activity by considering three meals corresponding to household members availability, appliances used for these meals and appropriate timings during the day; (e) a model to measure heat-losses in multi-unit dwellings.
2. Our findings indicate that urban areas consume less energy than rural areas in summer and vice versa in winter. Rural areas consume more active energy than urban areas. The heat loss model indicates that single-family detached type of residence consumes the most energy irrespective of the season and area type.
3. The generated synthetic demand curves are validated with real-time data from Rappahannock County in Virginia at the household level using dynamic time warping technique. Findings indicate that 88.5% of the synthetic curves lie within 10% error rate. As far as we know, our paper is the first to perform such rigorous model validation using real-world data.

2 DATASETS

Different datasets are employed to create a comprehensive and integrated energy demand generation model. Table 1 shows a summary of these datasets. The synthetic population provides detailed demographics about individuals and their households, as described in Section 3. The American Time Use Survey (ATUS 2015) provides information on energy-related activities. U.S. Energy Information Administration (EIA) Residential Energy Consumption Survey (RECS) (EIA RECS 2015) provides housing unit-specific information (e.g., floor area, fuel equipment, presence of dishwasher, washer, dryer, electric bulbs, refrigerator). The hourly temperature data are obtained from the North American Land Data Assimilation System (NLDAS 2016). Using the U.S. Census Bureau's urban and rural classification (US-Census 2010), two types of urban groupings are developed : (a) urbanized area and (b) urban cluster.

Table 1: List of data sources.

Dataset	**Description**
Synthetic population	Statistical representation of individuals and households in VA.
ATUS	Contains 24 hour period activity diaries for 13,260 respondents.
EIA-RECS	Structural characteristics of 5686 households in the US.
U.S. Census	Rules for classifying urban and rural areas in the US in 2010.
NLDAS	Hourly climate data from NOAA for North America.
Hourly Load Data	Actual load profiles from Rappahannock county for 100 households.

3 HOUSEHOLD MODELING

The process of constructing synthetic households is described in prior work (Beckman et al. 1996). We briefly recapitulate the process here. The process uses the American Community Survey (ACS) (Mather et al. 0053). ACS is available at various geographical levels, of which the highest resolution is a block group. A block group consists of 600 to 3,000 people. The ACS provides marginal information about variables (e.g. household size, household income) at household level for each block group. The ACS also provides a Public Use Microdata Sample (PUMS), which is a 5% representative sample for a larger region than block group, called a Public Use Microdata Area (PUMA). PUMAs are described by the Census Bureau as "a collection of counties or tracts within counties with more than 100,000 people". Block groups are properly nested within PUMAs, i.e., each block group is present in exactly one PUMA and PUMAs have complete coverage, i.e., there is a PUMA for each block group. These statistical areas are defined for the circulation of PUMS data. Personally identifying information is stripped from PUMS records, such as names and addresses, but they are otherwise complete. In particular, they provide several household-level variables, such as the type of fuel used for heating, which is useful for energy demand modeling.

The synthetic population generation procedure (Beckman et al. 1996; Swarup and Marathe 2017) creates a synthetic representation of each household and the individuals within the household in a region (in our case, Virginia, USA) using ACS data. Marginal distributions about different demographic variables at household level are combined into a joint distribution for each block group using Iterative Proportional Fitting (IPF) (Deming and Stephan 1940; Beckman et al. 1996). The sample data from the corresponding PUMA are used to seed the joint distribution in order to create the correct correlation structure. Next, random draws from the joint distribution are generated and matching records from the sample data are copied into a synthetic population. The number of households so generated match the total number of households in the region under consideration.

Assuming that PUMS data are a representative sample of the true population for each block group, the procedure generates synthetic households that are *statistically similar* to the original population. See (Beckman et al. 1996; Swarup and Marathe 2017) for a discussion on validation of the synthetic populations. The basic technique is now well accepted and widely used in various application areas; see, e.g. Wheaton et al. (2009).

4 ENERGY DEMAND MODELING

An agent-based approach is employed to generate hourly energy demand for a synthetic household. Energy usage by each household is divided into two broad categories, passive energy and active energy. Figure 1 shows the workflow of the model. We introduce the following notation for the model (asterisks are used to designate survey entities):

H_j : synthetic household, R_{j*} : RECS survey household, W_{k*} : ATUS survey household
P_i : synthetic individual, A_{i*} : ATUS survey individual

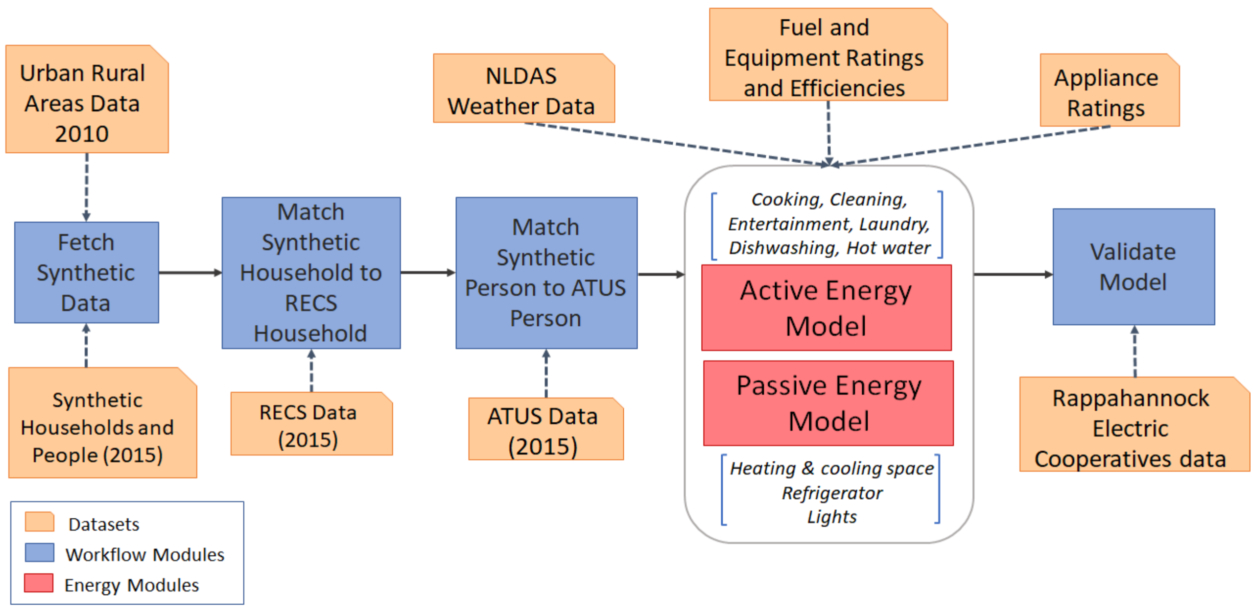

Figure 1: Workflow of the energy demand model.

4.1 Passive Energy Model

The passive energy demand for a synthetic household is modeled as a function of its demographic attributes and its structural characteristics, obtained from the EIA RECS (2015) survey and the synthetic population. The energy consumed by lighting, refrigerator, and space conditioning form the passive energy component. The model used for building the demand profiles for passive energy is enumerated below.

Match RECS household R_{j*} to synthetic household H_j: We first use Recursive Feature Elimination (RFE), which is a backward feature elimination technique that is employed to get the m most important features, to analyze the RECS data. RFE uses a model (e.g., linear regression, SVM or random forest) to select a set of best or worst performing features. The model estimates the importance of a variable by looking at how much prediction error increases when a variable is omitted while all others are left unchanged.

For our setting, the RFE method uses a random forest model. We apply this technique on the RECS data with target variables as urban, rural class labels. Residence type, income, whether house is owned or rented, year the house was built, total number of rooms in the house, number of adults and children, and representative weight of the household as given by RECS survey were considered as input features to RFE. The results showed that residence type, income and representative weight of the household are sufficient to distinguish urban and rural households in the RECS dataset. Based on these features, a best match R_{j*} is found for every H_j. Once the best match is found, energy-related information like square footage, indoor temperature, and water heating equipment are overlaid on the synthetic household.

Assemble the data required for calculating the heat loss: Heating fuel and residence type are taken from H_j whereas square footage, heating and cooling equipment, water heater details and indoor daytime and night time temperatures are taken from R_{j*}. The wall area and ceiling area are derived from the square footage. Standard insulation values for Virginia state for walls and ceiling is used from Insulation Institute (2009). It is assumed that the house is cooled using an air conditioner operating on air source heat pump. This assumption is based on data present in RECS and Heat Pump Systems (). The efficiency values for equipments related to heating and cooling fuels are obtained from the Heating Fuel Comparison Calculator (2016). We assume a simple building model for detached, attached and apartment dwelling types in our passive energy model. For single-family detached units, we assume that all walls of the house are exposed to the external environment. For all other dwelling units like single-family attached and multi-family dwelling units, we consider that only two walls are exposed to heat loss. Apartments are modeled as 4 units on every level. For example, 3 levels are assumed to be present for 12-unit apartment building. Further, each house is modeled as a rectangular box with the ratio 2:3 and the wall area is based on this ratio and the square footage. Ceiling area for a detached unit is same as that of the square footage or floor area whereas the ceiling area for each of the apartments is calculated as 1/4 of the ceiling area. Finally, heat loss calculation for all dwelling types is performed using the Fourier Law. We assume that a constant amount of energy is consumed by the refrigerator and lights for a given day.

4.2 Active Energy Model

Active energy demand for a household is modeled as a function of household members' behavior for a normative day. The active energy demand is simulated with respect to each household members' energy activities, time spent at home, the types of appliances used for the activities, and the duration for which the appliances are used. Active energy demand is simulated only for adults (age 18+) in the household, since the ATUS survey does not include activities for children. We further categorize active energy into 7 activities: personal grooming (showering), cleaning (vacuuming), cooking (breakfast, lunch, dinner), dish-washing, laundry, leisure (watching TV, playing games, watching videos) and computer use. The amount of energy required to heat water for activities like personal grooming, cooking, dish-washing and laundry is also accounted for in the active energy consumption. We do not record any seasonal change in activities since the ATUS survey offers schedules for each respondent for only one day in a year. The detailed procedure for building active demand profiles for households is described below.

Find matching ATUS person A_{i*} for every synthetic person P_i: Every P_i is matched to a best A_{i*} using the fitted values approach described by Lum et al. (2016). The method works by fitting regression models to predict time spent by an individual at locations such as home, work, school, shopping and other locations. Then, a two-step approach is used to match each synthetic individual to a survey individual. Every synthetic and survey person is treated as a record having attributes such as predicted values of time spent at each location, individual demographics and household demographics. Let $D_{A,P}$ be the Euclidean distance between every $P_{i,j}$ and $A_{i*,k*}$. First, person-person distance is calculated to find the closest match A_{i*}, for every P_i in each survey household. Then, the best W_{k*} is chosen for H_j. This between-household distance metric is a minimax over a pairwise, between-individual distance metric, $D_{A,P}$. It is the Hausdorff distance as shown below –

$$D_{V,W}(H_j, W_{k*}) = max_{P_{i,j} \in H_j}\{D_{P,W}(P_{i,j}, W_{k*})\}.$$

$$D_{P,W}(P_{i,j}, W_{k*}) = min_{A_{i*,k*} \in W_{k*}}\{D_{A,P}(P_{i,j}, A_{i*,k*})\}.$$

Each synthetic individual is assigned the activity schedule of the survey individual that is closest in the chosen survey household. For more details refer to Lum et al. (2016).

Build active energy profile for each synthetic household H_j: Personal grooming, computer use and leisure activities are considered as independent activities. Cleaning, cooking, dish-washing and laundry activities are considered to be shared activities within a household (U.S. Bureau of Labor Statistics:American Time Use Survey 2015). First, an activity sequence of energy-consuming activities for every matched A_{i*} is generated and assigned to the synthetic individual P_i. While building the energy activity sequence for a household, the model picks one of each activity–cleaning, dish-washing and laundry–from individual household members' schedules with the longest duration. See Algorithm 1 for details about calculating active energy for a household. Cooking timings and appliance information are captured from Larson (2002). Information on appliances' wattage ratings is obtained from Silicon Valley Power (). Information on hot water usage for activities is obtained from The USGS Water Science School ().

Algorithm 1 Generating active energy profile for a synthetic household

$Input$: H_j and its household members $\{P_{1,j},..,P_{i,j}\}$
$Output$: (E^a_1 ... E^a_{86400}) for H_j household active energy profile in WattSeconds, where E^a_s is the active energy use in the s^{th} second of the day.
procedure GENERATEACTIVEENERGYPROFILE
  Let $Activities_j$ be the generated household energy activity sequence
  $Act_{independent} \leftarrow \{PersonalGrooming, Leisure, ComputerUse\}$
  $Act_{shared} \leftarrow \{Cleaning, Cooking, DishWashing, Laundry\}$
  Let $ApplianceWattage$ be the list of appliance wattages, indexed by activity type
  for each member $P_{i,j}$ in H_j **do**
    $Active_i \leftarrow$ list of energy activities for P_i
    for each activity $Active_{i,k}$ **do**
      if RESOLVECONSTRAINTS($Active_{i,k}$, $Activities_j$) **then**
        $Activities_j \leftarrow$ {Activity type, start time, duration, activity name}
      end if
    end for
  end for
  for each activity G_k in $Activities_j$ **do**
    $energy \leftarrow G_{k,duration} \times ApplianceWattage[G_{k,type}]$
    for each second s from $G_{k,start}$ to $G_{k,stop}$ **do**
      $E^a_s \leftarrow E^a_s + energy$
    end for
  end for
  return (E^a_1 **, ... ,** E^a_{86400})
end procedure

Algorithm 2 Resolve constraints for adding activity to a household activity sequence

$Input$: $Activities_j$ for household j, k^{th} activity $Active_k$ of individual $P_{i,j}$
$Output$: True/False
procedure RESOLVECONSTRAINTS
  $Act_{independent} \leftarrow \{PersonalGrooming, Leisure, ComputerUse\}$
  $Act_{shared} \leftarrow \{Cleaning, Cooking, DishWashing, Laundry\}$
  $addActivityToHousehold \leftarrow False$
  if $Active_{k,type} \in Act_{shared}$ AND $Active_{k,type} \notin Activities_j$ **then**
    $addActivityToHousehold \leftarrow True$
  else if $Active_{k,type} \in Act_{shared}$ AND $Active_{k,type} \in Activities_j$ **then**
    if $Active_k$ has the longest duration for the activity type **then**
      $addActivityToHousehold \leftarrow True$
    end if
  else if $Active_{k,type} \in Act_{independent}$ **then**
    $addActivityToHousehold \leftarrow True$
  end if
  return $addActivityToHousehold$
end procedure

5 RESULTS

Our findings indicate that rural areas consume higher active energy than urban areas (Figure 2). Figure 3 compares the demand profiles of the top five counties/cities in Virginia consuming the highest and lowest active energy. The counties consuming higher energy are all rural areas.

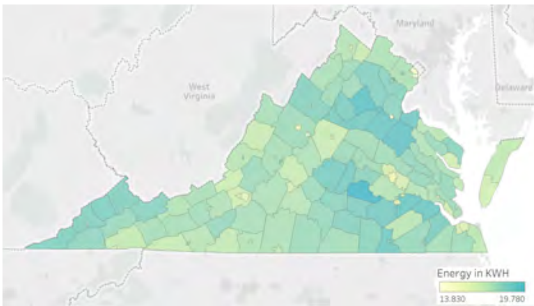

Figure 2: Active energy usage per household (KWH) in urban and rural areas of Virginia. NOVA, Virginia Beach, Richmond, and other urban regions use less active energy than rural areas.

Figure 3: Top five areas consuming highest and lowest active energy per household (KWH) are shown for a normative day. Two distinct groups of energy consumption can be identified. The ones with higher energy consumption are rural areas (Amelia, Culpeper, Caroline, Prince George, and Surry counties).

We observe that overall, rural area households consume more energy in summer than urban areas. For example, the passive energy consumption in cities like Radford, Waynesboro, Harrisonburg, etc., is much less than in rural counties. In winter, however, urban areas cover the full range of energy consumption (highest and lowest). Fairfax city and county are among the highest consumers of passive energy, and cities like Norfolk, Alexandria, and Radford consume the least passive energy in winter (Figure 4). This observation can be attributed to the distribution of fuel mixes, equipment efficiencies of furnaces, air source heat pumps, geothermal heat-pumps, and square footage area to heat and cool the house. The larger the house, the more the energy that is required. Certain areas, especially to the south side of Virginia, consume higher energy than the other counties. This is due to the higher temperatures in those counties. Refer to Table 2 and 3 for further details.

Different residence types are modeled to understand their energy consumption behavior. The single-family detached dwellings consume the most energy, and apartment dwellings consume the least energy (especially ones having more than 20 units), irrespective of weather conditions and whether the house lies in an urban area or rural area. See Figure 5 for demand details. This is largely due to the fact that detached

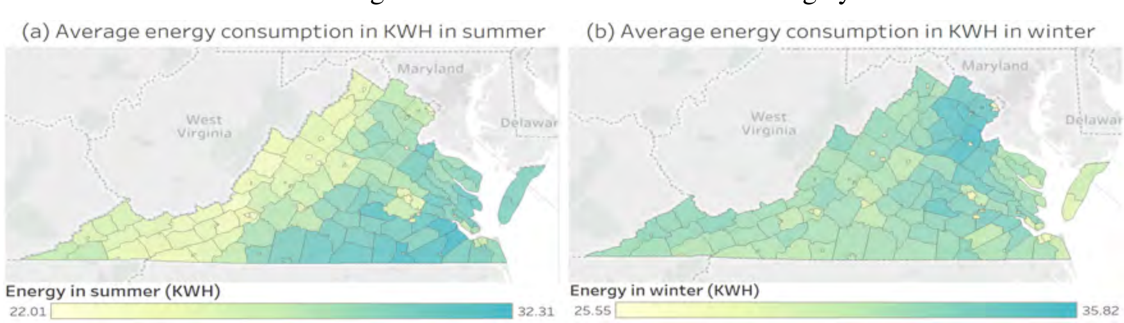

Figure 4: Average household energy consumption (KWH) by variation in seasons. The energy usage is shown for a day in July and February, 2015, representing summer and winter seasons respectively.

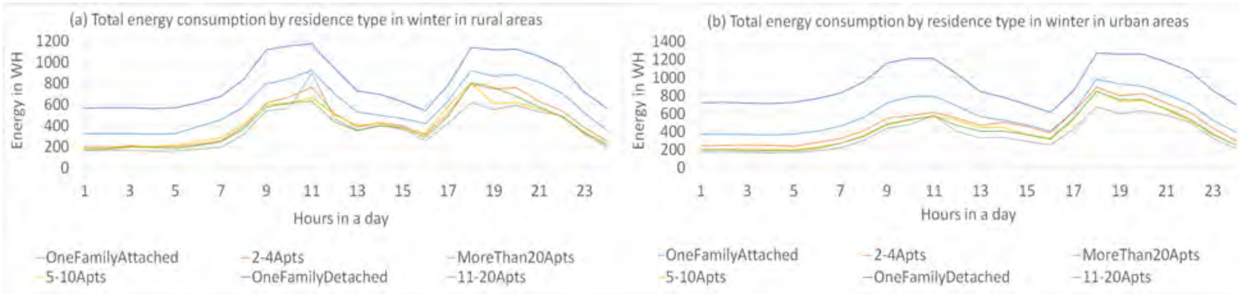

Figure 5: The total energy consumption per household (KWH) in different dwelling types shows that single family detached households consume the most energy in urban and rural areas.

residence types have a higher square footage as compared to other residence types. Also, single-family households have all walls exposed as opposed to multi-unit dwellings. Hence, more energy is required to heat and cool the space. Detached dwellings also have a larger family size as compared to household sizes in apartments. Since urban area detached dwellings have larger square footage, we see that detached houses in urban areas consume more energy than their counterparts in rural areas in the winter season. The information is presented in Table 2.

Table 2: Average values of square footage, household size, household income and heat-cool energy in different seasons in KWH (per household) for a given day.

Residence Type	Sq.ft	Household Size	Income	Heat Cool Winter	Heat Cool Summer
OneFamilyDetached	2974.7	2.68	100180.8	12.4	6.0
OneFamilyAttached	2010.7	2.564	94058.8	6.9	4.5
2-4Apts	1024.0	2.01	42329.3	4.3	2.5
5-10Apts	1042.2	2.07	49024.8	3.6	2.7
11-20Apts	1101.9	2.13	54386.3	3.3	2.3
MoreThan20Apts	1130.5	1.59	73700.8	3.1	1.9

Table 3: Average values of demographic variables and energy usage by areatype. Energy is measured in KWH per household for a given day. Energy usage is specified for four categories, namely, energy used for performing activities, heating water, heating or cooling the house in different seasons.

Area Type	No.Houses	Household Size	Sq.ft	Income	Activities	HotWater	HeatCoolWinter	HeatCoolSummer
U	2090566	2.5	2498.7	95210.0	7.7	8.9	9.7	4.7
R	832820	2.4	2085.9	71703.9	8.6	9.1	9.4	5.2
C	139397	2.3	2196.7	58196.9	8.2	9.6	9.6	4.8

6 VALIDATION

We obtained hourly load profiles of 100 sample households from Rappahannock county, VA, for all days of the year 2016. The sample of 100 load curves were used to validate the 3273 synthetic demand profiles of all the synthetic households generated for Rappahannock county. For each synthetic household hourly load profile, s, we find the closest matching Rappahannock real load profile, r, using dynamic time warping (DTW) (Berndt and Clifford 1994).

Dynamic time warping is a distance measure specifically designed for comparing time series data. Intuitively, it allows stretching and squeezing the time series to find the best corresponding points between them. The DTW algorithm finds an optimal match between two given time series, subject to some constraints, using dynamic programming. The constraints are that every point in each series must be matched to a point in the other series. Further, if a point at time t on one series is matched to a point at time $t+k$ on the other, then any point $t' > t$ on the first series can only be matched to a point at time $t+k+m$, for $m \geq 0$ on the other time series. The appropriateness of DTW for our problem can be explained by a simple example. Consider a real household that has a cooking activity at 4 pm and a TV-watching activity at 7 pm, causing two peaks in the active demand profile. Correspondingly, we might have a synthetic household that has the same two activities, but at 4:30 pm and 6 pm, respectively. Intuitively, these two load profiles are a good match since they will have similarly sized peaks in a small time window (radius). However, standard methods of comparison, such as taking the Euclidean distance between the two time series or calculating Pearson's correlation will not give a good match between the two series. DTW, however, can line up the peaks because it is allowed to compress the part of the time series between the two peaks.

For each synthetic load profile, s, we find the closest Rappahannock load profile, r using DTW with a radius of 3 hours (other radii give similar results, refer to Figure 8). Our results show that 88.5% of the synthetic households had daily energy usage within 10% of the closest matching household from Rappahannock for summer as well as winter profiles generated by the model. We show error rate for summer profiles in Figure 9. Figures 6 and 7 show two matching households' demand curves for a 2016 winter and summer day.

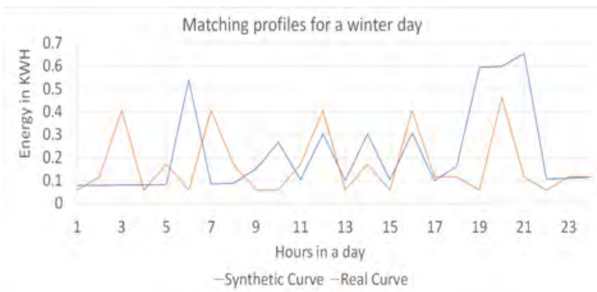

Figure 6: Best real curve match for a sample synthetic curve for winter using DTW and radius 3.

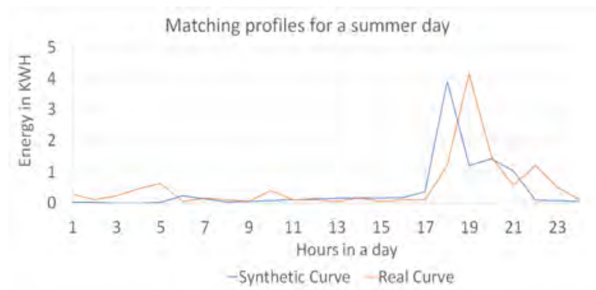

Figure 7: Best real curve match for a sample synthetic curve for summer using DTW and radius 3.

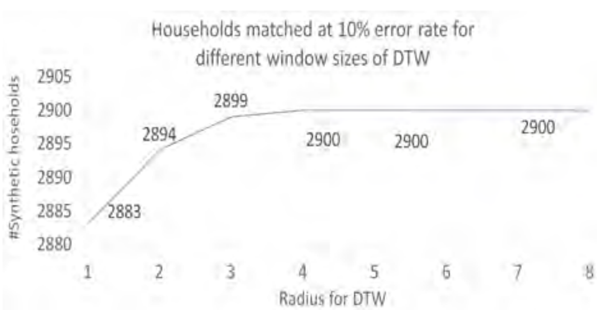

Figure 8: An elbow plot representing the number of synthetic households in Rappahannock county that fall within 10% error rate for different window sizes (radius or w) of DTW matching process. We choose w=3.

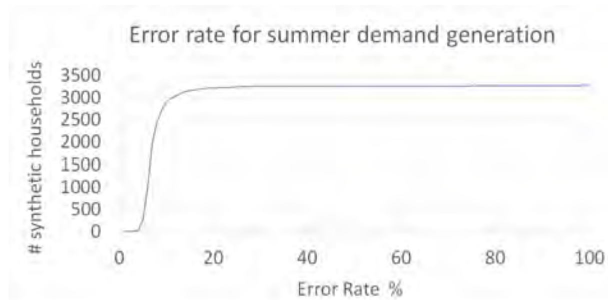

Figure 9: 88.5% of the synthetic households' energy usage in Rappahannock county falls within 10% of the closest matching household from the Rappahannock sample for summer profiles generated by the model.

7 CONCLUSION AND FUTURE WORK

We present a detailed individual-based methodology for generating energy demand profiles for households in urban and rural regions of Virginia, based on a variety of complex drivers, including local climate, household demographics, behavior, occupancy patterns, building stock, fuels, and appliances. Detached households consume the most energy in urban and rural areas. Active energy consumption is higher in rural areas than in urban areas. Urban and rural area passive energy consumption changes by season. A novel approach to validation of the model is presented. The generated synthetic curves show that 88.5% of the synthetic curves lie within 10% error rate, thus presenting good matches to real-time data from real households in Virginia.

The present work can be extended in multiple directions. The availability of detailed energy-use surveys like EIA-RECS, complete hourly temperature data through NASA products like NLDAS, and synthetic populations now make it possible to estimate energy demand at highly resolved levels and scale. Rappahannock is a rural county in VA; we plan to validate the load profiles for more counties and urban areas. This comprehensive study reveals spatiotemporal regularities and patterns that will enable better energy management and efficiency. The model can be useful for a range of stakeholders, including households trying to understand the potential implications of different appliance or energy choices, utilities looking to better forecast the impact of different possible residential trends, and policymakers seeking to assist households in improving their energy efficiency through targeted policies and programs. The model can be extended to study scenarios such as effects of penetration of electric vehicles on electric load (Bustos-Turu et al. 2016), effects of roof-top solar adoption on household consumption, use of energy star appliances, and shifting demand from peak hours to off-peak hours (Subbiah et al. 2017).

ACKNOWLEDGEMENTS

We thank members of National Rural Electric Cooperative Association and Rappahannock Electric Cooperative for sharing anonymized data on energy consumption. This work was supported in part by DTRA Grant HDTRA1-17-0118, DOE Cooperative Agreement DE-EE0007660, NSF grants NSF-NRT DESE DGE-154362, NSF DIBBS ACI-1443054, NSF EAGER CMMI-1745207 and NSF BIGDATA IIS-1633028.

REFERENCES

ATUS 2015. "U.S. Bureau of Labor Statistics:American Time Use Survey". https://www.bls.gov/tus/datafiles_2015.htm. Accessed: March 2018.

Beckman, R. J., K. A. Baggerly, and M. D. McKay. 1996. "Creating Synthetic Baseline Populations". *Transportation Research Part A: Policy and Practice* 30(6):415–429.

Bedir, M., and E. C. Kara. 2017. "Behavioral Patterns and Profiles of Electricity Consumption in Dutch Dwellings". *Energy and Buildings* 150:339–352.

Berndt, D. J., and J. Clifford. 1994. "Using Dynamic Time Warping to Find Patterns in Time Series". In *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*, edited by D. Heckerman et al., 359–370: AAAI Press.

Bustos-Turu, G., K. H. van Dam, S. Acha, C. N. Markides, and N. Shah. 2016. "Simulating Residential Electricity and Heat Demand in Urban Areas Using an Agent-based Modelling Approach". 1–6.

Deming, W. E., and F. F. Stephan. 1940. "On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Tables are Known". *The Annals of Mathematical Statistics* 11(4):427–444.

EIA RECS 2015. "U.S. EIA Residential Energy Consumption Survey". Accessed: November 2017.

Energy Data Facts 2016. "National Residential Energy Facts". https://rpsc.energy.gov/energy-data-facts. Accessed: March 2018.

Fonseca, J. A., and A. Schlueter. 2015. "Integrated Model for Characterization of Spatiotemporal Building Energy Consumption Patterns in Neighborhoods and City Districts". *Applied Energy* 142:247–265.

Heat Pump Systems. https://energy.gov/energysaver/heat-pump-systems. Accessed: October 2017.

Heating Fuel Comparison Calculator 2016. https://ag.purdue.edu/extension/renewable-energy/Documents/ON-Farm. Accessed: May 2017.

Heinonen, J., and S. Junnila. 2014. "Residential Energy Consumption Patterns and the Overall Housing Energy Requirements of Urban and Rural Households in Finland". *Energy and Buildings* 76(Supplement C):295–303.

Heiple, S., and D. J. Sailor. 2008. "Using Building Energy Simulation and Geospatial Modeling Techniques to Determine High Resolution Building Sector Energy Consumption Profiles". *Energy and Buildings* 40:1426–1436.

Insulation Institute 2009. https://insulationinstitute.org/wp-content/uploads/2015/12/VA-2009.pdf. Accessed: November 2017.

Larson, R. B. 2002. "When Is Dinner?". *Journal of Food Distribution Research, Food Distribution Research Society* 33.

Lum, K., Y. Chungbaek, S. Eubank, and M. Marathe. 2016. "A Two-stage, Fitted Values Approach to Activity Matching". *International Journal of Transportation* 4:41–56.

Marszal, A., P. Heiselberg, and O. K. Larsen. 2016. "Household Electricity Demand Profiles A High-resolution Load Model to Facilitate Modelling of Energy Flexible Buildings". *Energy* 103:487–501.

Mather, M., K. L. Rivers, and L. A. Jacobsen. 2005(3). "The American Community Survey". *Population Bulletin* 60:1–20.

McLoughlin, F., A. Duffy, and M. Conlon. 2012. "Characterising Domestic Electricity Consumption Patterns by Dwelling and Occupant Socio-economic Variables:An Irish Case Study". *Energy and Buildings* 48:240–248.

National Academies of Sciences, Engineering, and Medicine 2016. *Analytic Research Foundations for the Next-Generation Electric Grid.*

Nguyen, T. A., and M. Aiello. 2013. "Energy Intelligent Buildings Based on User Activity: A Survey". *Energy and Buildings* 56:244–257.

NLDAS 2016. "North American Land Data Assimilation System (NLDAS) Climate Data". https://ldas.gsfc.nasa.gov/nldas/. Accessed: March 2018.

Sandels, C., J. Widn, and L. Nordstrm. 2014. "Forecasting Household Consumer Electricity Load Profiles with a Combined Physical and Behavioral Approach". *Applied Energy* 131:267–278.

Silicon Valley Power. "Appliance Energy Use Chart". http://www.siliconvalleypower.com/for-residents/save-energy/appliance-energy-use-chart. Accessed: March 2018.

Subbiah, R., A. Pal, E. K. Nordberg, A. Marathe, and M. V. Marathe. 2017. "Energy Demand Model for Residential Sector: A First Principles Approach". *IEEE Trans. on Sustainable Energy* 8(3):1215–1224.

Swan, L. G., and V. I. Ugursal. 2009. "Modeling of End-Use Energy Consumption in the Residential Sector:A Review of Modeling Techniques". *Renewable and Sustainable Energy Reviews* 13(8):1819–1835.

Swarup, S., and M. V. Marathe. 2017. "Generating Synthetic Populations for Social Modeling:Second Tutorial". *Sixteenth International Conference on Autonomous Agents and Multiagent Systems*.

The USGS Water Science School. "How Much Water Does the Average Person use at Home Per Day?". https://water.usgs.gov/edu/qa-home-percapita.html. Accessed: January 2018.

U.S. Bureau of Labor Statistics:American Time Use Survey 2015. "Average Minutes Per Day Men and Women Spent in Household Activities". https://www.bls.gov/tus/charts/household.htm. Accessed: March 2018.

US-Census 2010. "US-Census Urban Areas 2010". https://www.census.gov/geo/reference/ua/urban-rural-2010.html. Accessed: November 2017.

Wallis, H., M. Nachreiner, and E. Matthies. 2016. "Adolescents and Electricity Consumption; Investigating Sociodemographic, Economic, and Behavioural Influences on Electricity Consumption in Households". *Energy Policy* 94:224–234.

Wheaton, W. D., J. C. Cajka, B. M. Chasteen, D. K. Wagener, P. C. Cooley, L. Ganapathi, D. J. Roberts, and J. L. Allpress. 2009. "Synthesized Population Databases:A US Geospatial Database for Agent-Based Models". *Methods Report*.

Yang, Q., M. Liu, C. Huang, Y. Min, and Y. Zhong. 2015. "A Model for Residential Building Energy Consumption Characteristics and Energy Demand:A Case in Chongqing". *Procedia Engineering* 121(Supplement C):1772–1779.

Yu, B., Y.-M. Wei, G. Kei, and Y. Matsuoka. 2018. "Future Scenarios for Energy Consumption and Carbon Emissions due to Demographic Transitions in Chinese Households". *Nature Energy*:109–118.

Zhang, M., Y. Song, P. Li, and H. Li. 2016. "Study on Affecting Factors of Residential Energy Consumption in Urban and Rural Jiangsu". *Renewable and Sustainable Energy Reviews* 53(Supplement C):330–337.

AUTHOR BIOGRAPHIES

SWAPNA THORVE received her B.Tech degree in Computer Engineering from Pune University, India, in 2013. She is currently pursuing her Ph.D. in Computer Science at Virginia Tech. Before joining Virginia Tech, she worked as a software developer at Persistent Systems Limited, India. Her research interests include urban computing, big data and developing interactive systems. She is a IEEE member. Her email address is swapna6@vt.edu.

SAMARTH SWARUP is a Research Associate Professor in the Network Dynamics and Simulation Science Lab at the Biocomplexity Institute of Virginia Tech. He received his Ph.D. in Computer Science from the University of Illinois at Urbana-Champaign. His research interests are in social and behavioral modeling and simulation, resilience and sustainability, computational social science, artificial intelligence, and machine learning. He has published over 60 peer-reviewed conference and journal papers and his work has won multiple best paper awards. His email address is swarup@bi.vt.edu.

ACHLA MARATHE received her M.S. and Ph.D. degrees in economics from the University at AlbanyState University of New York, Albany. She is currently a Professor at the Biocomplexity Institute and at the Department of Agricultural and Applied Economics at Virginia Tech. She is also a faculty of Health Sciences. Before joining Virginia Tech, she worked as a technical staff member at the Los Alamos National Laboratory, New Mexico. Her research interests include modeling of socially coupled systems and economics of disasters. Her email address is amarathe@vt.edu.

YOUNGYUN CHUNGBAEK is a Senior Research Associate in the Network Dynamics and Simulation Science Laboratory at the Biocomplexity Institute of Virginia Tech. She received her Master's degree in Computer Science and her Ph.D. in Educational Research and Evaluation, with a specialty in statistics, from Virginia Tech. Her email address is ychungba@bi.vt.edu.

ERIC K. NORDBERG is a research scientist and senior software engineer at the Biocomplexity Institute of Virginia Tech. Eric received his BS in Biology from Guilford College, his MS in Genetics from The University of Arizona, and his PhD in Genetics, Bioinformatics, and Computational Biology from Virginia Tech. His interests include phylogenetics, usability, and performance tuning. His email address is nordberg@vt.edu.

MADHAV MARATHE is the director of the Network Dynamics and Simulation Science Laboratory, Biocomplexity Institute and a professor in the department of Computer Science, Virginia Tech. His research interests are in network science, foundations of computing, communication networks, computational epidemiology, socially coupled system science and high performance computing. Before coming to Virginia Tech, he was a Team Leader in the Basic and Applied Simulation Science Group that is a part of the Computer and Computational Sciences division at the Los Alamos National Laboratory (LANL). He is a Fellow of the IEEE, ACM, SIAM and AAAS. His email address is mmarathe@vt.edu.