An Activity-Based Energy Demand Modeling Framework for Buildings: A Bottom-Up Approach

Rajesh Subbiah

Thesis submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Master of Science

in

Computer Science and Applications

Madhav V. Marathe, Chair

Kristian C. Lum

Achla Marathe

Anil K. Vullikanti

May 3, 2013

Blacksburg, Virginia

Keywords: energy, buildings, activity, regression models, smart grid, demand

An Activity-Based Energy Demand Modeling Framework for Buildings: A Bottom-Up Approach

Rajesh Subbiah

(ABSTRACT)

Energy consumption by buildings, due to various factors such as temperature regulation, lighting, poses a threat to our environment and energy resources. In the United States, statistics reveal that commercial and residential buildings combined contribute about 40 percent of the overall energy consumption, and this figure is expected to increase. 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. In this work, we investigate and model the energy consumption of buildings by taking into account physical, structural, economic, and social factors that influence energy use. We propose a novel activity based modeling framework that generates an energy demand profile on a regular basis for a given nominal day. We use this information to generate a building-level energy demand profile at highly dis-aggregated level. We then investigate the different possible uses of generated demand profiles in different What-if scenarios like urban-area planning, demand-side management, demand sensitive pricing, etc. We also provide a novel way to resolve correlational and consistency problems in the generation of individual-level and building-level "shared" activities which occur due to individuals' interactions.

Dedication

Sri Chengalamma and Sri Nallandavar Thunai

To Almighty Sri Chengalamma and Sri Nallandavar who blessed me

with lovely parents and caring brother

Dedicated to my Parents and Brother

Love you Amma, Appa and Ganesh

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my adviser, Prof. Madhav Marathe, for providing me the wonderful opportunity to work with him. He had faith in me, and supported me at the right times. His immense patience, willingness to help and foresight in research are his greatest virtues that make him the ideal adviser. It was a great privilege to be guided by him.

I owe a lot to Prof. Achla Marathe & Dr. Kristian Lum for their constant guidance and encouragement throughout my graduate life. Their constant support and feedback kept me fascinated and excited with my work. Prof. Achla Marathe's expertise in energy and economic policies, constant supervision and assistance helped me overcome all the hurdles in the completion of this thesis. This thesis would not have been possible without her guidance and support. Dr. Lum introduced me to a variety of statistical concepts that gave a completely new perspective to my work. She patiently listened to my ideas and clarified my naive statistical concepts. I greatly enjoyed working with Dr. Lum and developed special interest towards Statistics. This motivated me to enroll in a statistics course that I passed with an ‘A’ grade.

Furthermore, I am very grateful to Prof. Anil Vullikanti, for insightful comments in my work. He guided me in writing my first paper describing my thesis work.

In addition, I am thankful to my roommates Arvinth, Saran and Sathappan who made my residential experience pleasant. My school friends Hemanth, Spencer, Rushi, Sushrut, Nabeel, Vignesh, Krishna and many others ensured that school was not all work. At NDSSL, I was fortunate to meet, interact and work with great researchers Jae-seung Yeom, Ashwin Aaji, Nidhi Parikh, Gaurav Tuli, and others. I thank my friends living across the globe for their support and encouragement.

I am thankful to my mentors from my previous employment - TCS Innovation Lab, Chennai Prof. Anand Sivasubramaniam, Dr. Arunchandar Vasan and Mr. Rajesh Jayaprakash who encouraged and motivated me to do higher studies. I also, thank my family members for their constant support and encouragement.

Finally, I thank my parents, my brother for their love, support and encouragement. Words are inadequate to express my love for them and I dedicate this work to them.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

In the United States, most of the energy consumed at homes, offices, schools, industries, for vehicle transportation, etc comes from the expense of fossil fuels (see Fig. 1.1). In the recent years, increasing energy demand has intensified the depletion of fossil fuels. This scenario of fast depleting fossil fuels has driven interest in finding new ways to build a sustainable environment with efficient energy production(by use of renewable resources), distribution, and consumption. Towards that end, one of the major technology innovations currently under way is the possible transformation of electrical grids into smart grid. The major objectives of smart grids are - (1) Automated demand response (2) Efficient energy storage and distribution (3) Sustainable energy production. A detailed analysis of smart grid opportunities and challenges are done in [2]. However, this kind of transformation is very complex because it requires a major technology shift, heavy investment on equipments, and people's participation. Thus, it requires a systematic evaluation of different what-if scenarios that identify potential areas for improvement and quantify the impact of any energy policies or strategies. To carry out such analysis we need to develop a modeling framework that captures the various factors influencing energy consumption at a very fine granularity. Such a framework is presented in this thesis for residential and commercial buildings, which account for 40% of the United States total energy consumption. In our modeling framework, we formulate the energy demand as a function of individual and building level activities. Our major hypothesis is that every individual or building will follow a schedule of events, influenced by their demographics. Figure 1.2 shows an example timeline of activities for a single working individual. At the household level, we identify the energy consuming activities, associate appliance usage and calculate energy demand based on the appliance energy rating and duration of activity. We construct the activity-driven dynamic occupancy for commercial buildings and generate per-building demand profile.

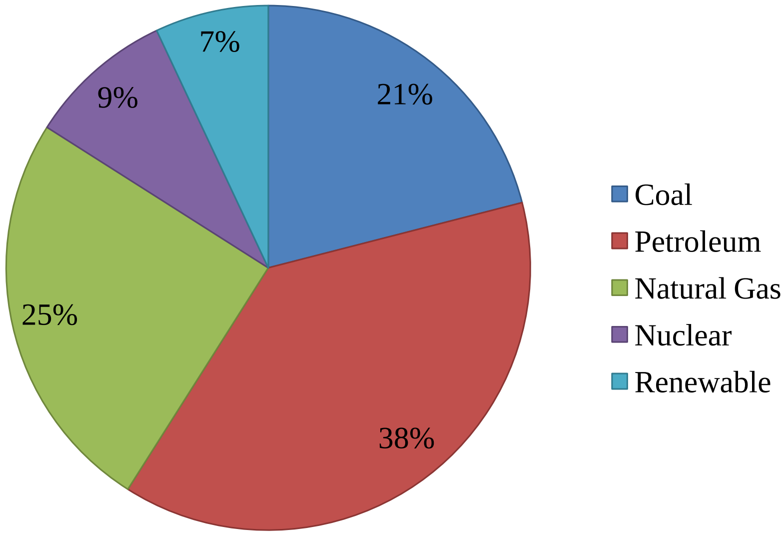

Figure 1.1: United States energy usage by source of fuel used. *Source: US Energy Information Administration* [1]

1.2 Residential Buildings

Background and Significance

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 [3]. 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. We need accurate energy demand forecasting models to understand the feasibility of energy conservation schemes from the context of individual and household level consumption behavior. Residential electricity consumption is influenced by physical factors (e.g. buildings, or infrastructure), social

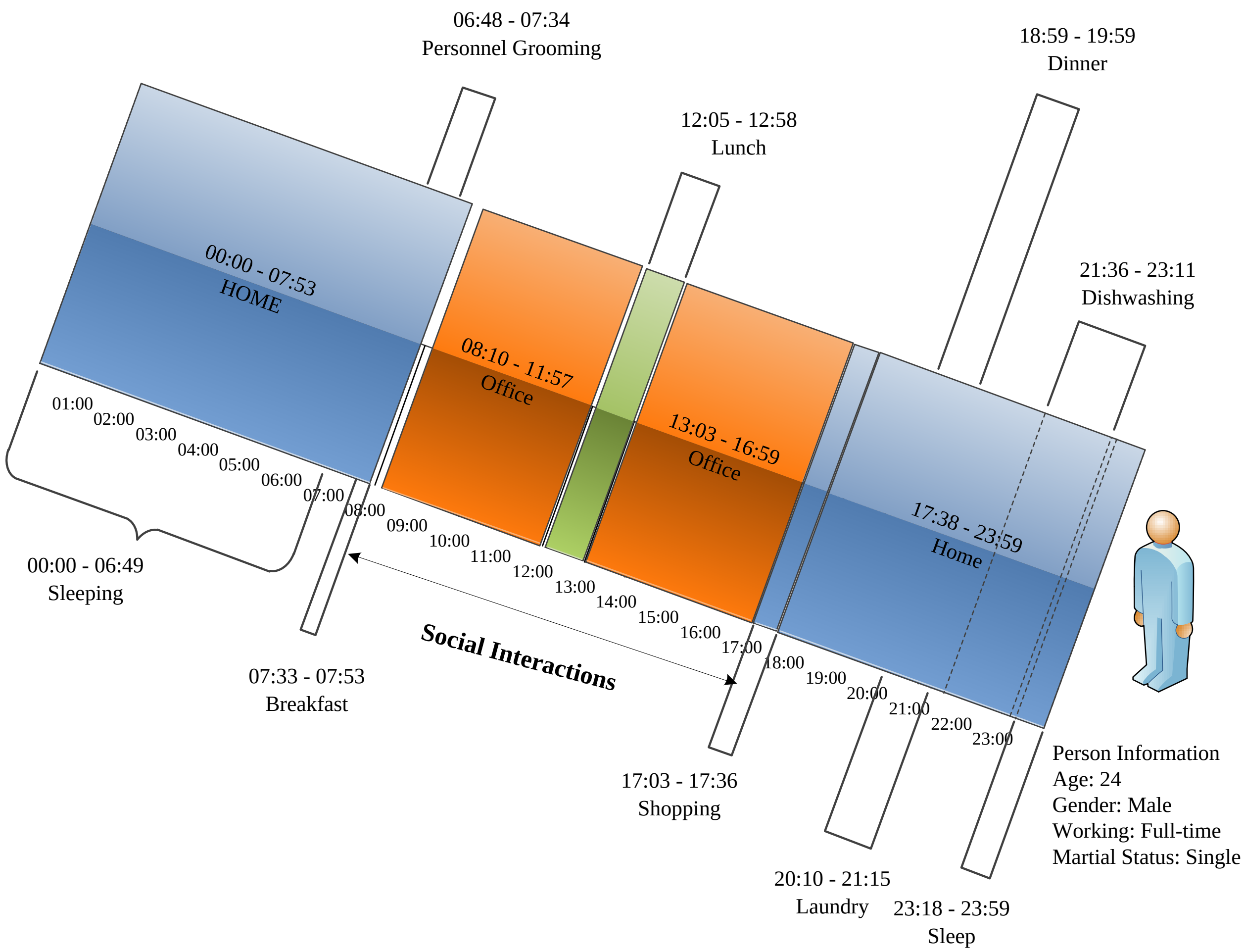

Figure 1.2: Figure shows the timeline of activities for an individual. The modeling framework formulates the energy demand as a function of individual and building level activities

practices (e.g. everyday routines, social interactions, policy interventions), and economic aspects (e.g. market, prices). Thus, it is important to understand these factors by taking into consideration specific contexts. Identification of these factors and their contribution in determining the energy demand is critical for finding ways to influence individual behavior and making them more energy efficient. Therefore, an ideal modeling framework should generate accurate energy demand profiles of households, dis-aggregated to the level of the individual household member by taking into account their social, economic, and behavioral aspects.

Our Contribution

Several studies [4] [5] have looked at modeling residential energy demand using time use data by constructing occupancy patterns. However, there has been little published work [6] [7] 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. For example, all household members may be observed to be simultaneously watching TV– a case in which the energy use due to the television should only be counted once. If one were to use individual diaries without taking the sharing of the activity into account, the energy consumption would be counted once for each household member. 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 household.* 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 associate the active appliance for each activity in the constructed activity sequence and generate the energy demand profile for the entire household. *We propose 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*[8]. We then show the applicability of the detailed demand model in making effective policy decisions like shifting energy intensive activities between peak and off peak hours based on the adaptability of households.

1.3 Commercial Buildings

Background and Significance

In the United States, commercial buildings are responsible for one-fifth of overall energy consumption [9]. Recent forecasts from the U.S. Energy Information Administration (EIA) predict that the commercial sector will be the dominant factor in the increase in energy demand for the years ranging from 2012 to 2035 [1]. Even under extremely stringent energy efficiency policies, the EIA estimates the growth rate of energy demand of the commercial sector to be 0.7 percent per year; floor space is expected to increase by one percent annually[1]. Figure1.3 compares the projected increase in energy demand for the years 2010 through 2035 across different sectors. To manage this rising energy demand, it is of the utmost importance to devise new energy policies and strategies that can help to optimize energy consumption of commercial buildings. Carrying out this kind of optimization requires a modeling framework that takes into account the various contributing factors that influence commercial buildings' energy consumption. Such a model is presented here.

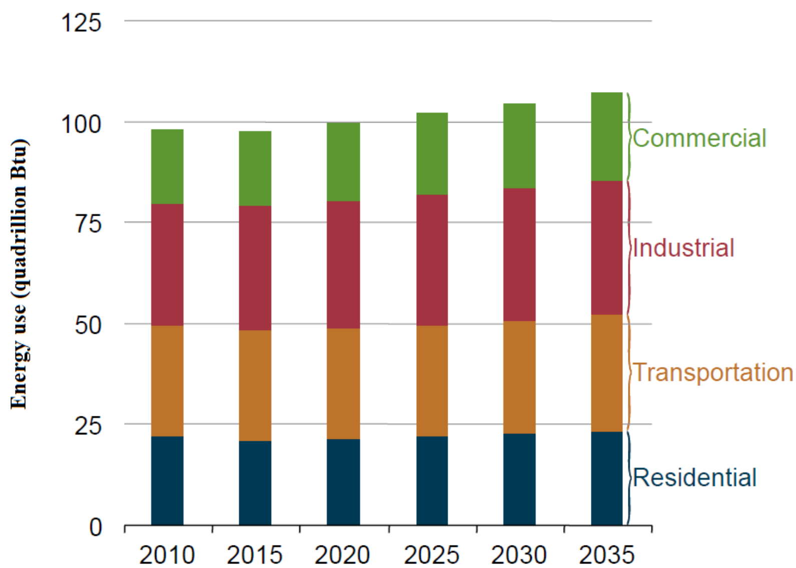

Figure 1.3: Energy use forecast for different sectors. *Source: US Energy Information Administration* [1]

Our Contribution

Several works [10] [11][12] [13] [14] have modeled commercial buildings' energy consumption. However, the main objective of these studies was to establish a benchmark metric, like energy consumed per square foot, and understand what factors influence the benchmark metric. These works are very essential in comparing the energy efficiency of two or more buildings of same type. There are also energy-forecasting tools like DOE-2, eQuest which simulate the energy consumption of a building based on some pre-configured parameters. All these works capture the major factors that influence energy consumption based on space heating, cooling, and lighting equipment. However, these works do not capture realistic dynamics in consumption across the day that is a result of changing occupancy. **We fill this gap by constructing profiles of synthetic individuals' visits based on their day-to-day activity schedule and generating the buildings' realistic occupancy rate at regular intervals.** In this way, we incorporate how the number of occupants at any period of time contributes to the rise and fall in building's energy demand profile. In addition, we also build a statistical model using the US EIA Commercial Buildings Energy Consumption Survey data (CBECS) [15]. The statistical model captures all the significant non-occupancy-related factors that influence the buildings' energy consumption.

1.4 Salient features

Specific salient features of this thesis are as follows:

- A highly dis-aggregated energy demand model capturing the demand profile at person and household level
- A highly scalable integrated demand modeling framework that accounts for social, behavioral and economical aspects of individuals and buildings
- We propose novel statistical based techniques to account for the within-household

activity dependency

- We propose a novel way to associate activities with multiple appliances and generate energy demand
- A high resolution energy demand model for commercial buildings which accounts for people's visit and interaction
- A software modeling framework to conduct and evaluate any energy demand modeling study

1.5 Organization of the thesis

In Chapter 2 and 3, we present our residential and commercial buildings energy demand modeling framework respectively. We start with examining the relevant prior works in the respective areas and summarize the different datasets used by the modeling framework. This is followed by a detailed description of our modeling methodology. We then present our experimental results and each chapter concludes by summarizing our findings. In addition, in chapter 2, we provide an illustrative case study to smooth out the load curve.

In Chapter 4, we describe our highly scalable, modular and extensible energy demand modeling software framework. The chapter begins with exploring the different design objectives required to be incorporated in a flexible and modular framework. Then, it describes our software system in detail by explaining the design and implementation. Finally, the chapter presents a web-based energy demand modeling prototype application used to conduct, analyze demand patterns and evaluate various What-If scenarios.

In Chapter 5, we summarize the different contributions made by this thesis and discuss the possible ways to extend this work.

Chapter 2

Residential Sector

In this chapter[1], we present an overview of our modeling methodology used to generate the demand profile for residential buildings. We construct the demand profile for Washington D.C. data and present results of our experiments which verifies the correctness of our model. We also present an illustrative application which uses our generated energy demand profile for performing peak-demand scaling.

2.1 Prior Works

The prior works in residential energy demand modeling can be broadly categorized into two categories: approaches that model energy consumption at an aggregate level ("top-down"

[1]An extended abstract of this chapter appears in the proceedings of IEEE Innovative Smart Grid Technologies Conference (ISGT) 2013 [16]

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 [17] [18] [19]. 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 [20] [21] [22] [23] 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 [24] provides a fairly exhaustive review of the pros, cons and applicability of various modeling techniques for residential energy consumption.

In this thesis for modeling residential energy consumption, 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 [25], 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.

2.2 Datasets Used

This work uses following data sets:-

1. ***American Time Use Survey (ATUS):-*** In this work, to derive the realistic activity and time schedules of the people we use the American Time Use Survey [25]. The survey contains 24-hour period activity dairies for 13,260 respondents across the US with diversified demographics. Each activity in the survey data embeds activity start time, end time, location, and participating people information. In addition to it, the survey data also contains respondents demographic information (like age, gender, marital status, work status etc). However, for a given household, the survey collects information from only one household member and thus we do not have the complete activity schedule information at the household level.

2. ***D.C. Synthetic Population:-*** To construct the demand profile for the range of an urban city, we use a synthetic data representing Washington D.C. region. The synthetic population is derived from the data gathered from the US Census, the National Household Travel Survey (NHTS), Dun and BradStreet, Land Use, and Navtaq. Details of the methodology can be found in [26, 27]. The synthetic population is statistically representative of the true population of Washington DC at a block group level in the US Census. The synthetic population is embedded with individual and household

level demographic information such as age, gender, race of individuals, and type, size, income of the household.

3. ***US EIA's Residential Energy Consumption Survey (RECS):-*** We use the EIA Residential Energy Consumption Survey- 2009 (EIA-RECS) data to estimate the energy consumed in the household. EIA-RECS is a national survey which collects energy related information from different types of housing units across the country and provides estimates of the energy consumption for the entire United States. We use this dataset to obtain housing unit specific information such as square footage, floor area, wall type etc. and overlay it on the houses in DC. We use the parameters such as household size, household type, household income and regional information to match the synthetic household in DC with the EIA-RECS household.

2.3 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 [27] [26]. Similar characteristics are also available from the ATUS data along with energy related activities. After matching demographics of the individuals in ATUS and the synthetic population, information pertaining to the energy activities in ATUS data is overlaid on to the synthetic individuals. Similarly EIA's Residential Energy Consumption Survey (EIA-RECS) [28] was used to assign building characteristics and appliance information to the synthetic individuals' home locations. The block diagram shown in Figure 2.1 summarizes our overall methodology. Each rectangular box in the diagram represents the input data sets and each rounded rectangular box represents a module in our modeling framework. Notations used are summarized in Table 2.1.

Symbol	Denotes
H_i	Synthetic household i drawn from the synthetic population
P_{ij}	Synthetic household member j from H_i
m	Number of persons in H_i
X_{ij}	Person P_{ij}'s demographic feature set of length n
A_k	Household activity k

Table 2.1: Notations used in residential buildings energy demand model

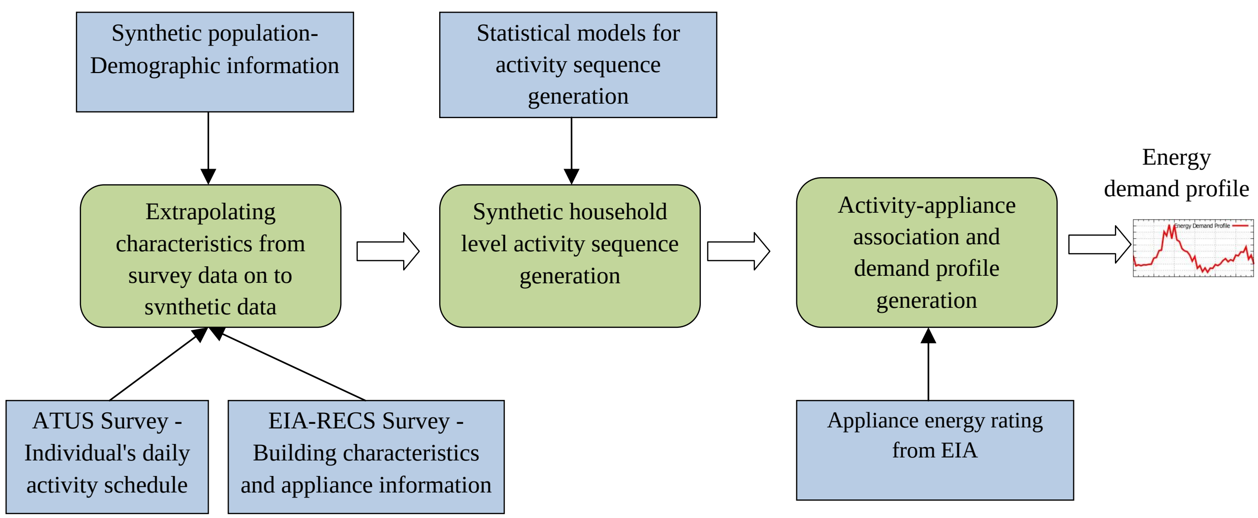

Figure 2.1: Residential energy demand modeling framework

We split all 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 takes a shower or uses the dishwasher. These activities are mainly a function of the household members' daily schedule. For example a person who is working full time will have a completely different schedule of activities compared to a person who is non-working. Similarly a household with children will have a different set of activities and their time of occurrence will differ from a household that has no children.

$E_{Passive}$ is the energy consumed for general maintenance of the house, such as space heating, space cooling, and water heating. This usage mainly depends on the climate and characteristics of the house, namely the type and size of housing unit, fuel used, insulation, wall type etc. and is mostly independent of activities of the residents.

2.3.1 E_{Active} Energy Demand Model

Activity Sequence Generator

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 survey respondent only and not all members of the household. In order to construct a household's energy usage level, we first model each member's daily activity schedule by following the steps given below:

1. Based on [29], we first identify the highest energy consuming activities in a typical household. These are summarized in Table 2.2 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. Sometimes, these activities can also become shared based on the appliance count. For example, if a household has a single TV, multiple individuals simultaneously watching the same TV constitutes a shared activity. If the household has multiple TV sets, it may be an independent activity.

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

Table 2.2: 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 [30] 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

   Only a few people explicitly list activities as "shared" in the ATUS data. Since we have detailed information for only one person in the household, it is ambiguous whether an activity was not performed at all 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 $P_{(1,\dots,m)}$ 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 μ_{ijk}, which represents that member 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 [31]. The next section explains the Logistic regression in detail.

   (b) After calculating $\hat{\mu}_{ijk}$, we used the inclusion and exclusion principle [32] 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. $A_{ik} = 1$. We then determine the time at which the activitiy has taken place by dividing 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 [31].

ii. After calculating the probability for each household member individually, we again use the probability inclusion and exclusion principle [32] 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 [33] as described in Section 2.3.1.

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 of the household.

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.

As such, there are many well tested software implementations available for estimating this model. We use logistic regression for several reasons. One, it is a well established technique for modeling binary data given covariates. Two, the logit link function maps a variable in $\mathbb{R}$ (i.e. $X^T_{ij}\beta$) to $[0, 1]$ and because of this property, logistic regression provides an elegant way to describe relationship between demographic variables and the probability of occurrence of an activity.

Rationale:- We use logistic regression for several reasons. Logistic regression is a well established technique for modeling binary data given covariates. As such, there are many well tested software implementations available for estimating this model. The logit link function maps a variable in $\mathbb{R}$ (i.e. $X_{ij}^T\beta$) to $[0, 1]$. Because of this property, logistic regression describes relationship between demographic variables to the probability of occurrence of an activity elegantly.

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.

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 [29] and the duration over which it is active, we estimate the energy consumption using the formula

$$ApplianceWattage * ApplianceActiveDuration = EnergyConsumed$$

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 2.3: 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 cooking. 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 as shown in Table 2.3.

2.3.2 $E_{Passive}$ Energy Demand Model

We use the EIA Residential Energy Consumption Survey- 2009 (EIA-RECS) data to estimate the $E_{Passive}$ energy consumed in the household. We derive housing unit specific information such as square footage, floor area, wall type etc. from the survey data and overlay it on the synthetic houses in DC. We use the parameters such as household size, household type, household income and regional information to match the synthetic household in DC with the EIA-RECS household. Once a match with similar characteristics is found, the specifics of the housing unit information available from EIA-RECS is overlaid on the synthetic household. This allows us to make realistic estimates of the $E_{Passive}$ energy consumption.

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 [34]. For each synthetic household unit we derive (a) S: Average square footage used for space heating (b) T_p: Household temperature when someone is at home during the day (c) T_a: Household temperature when no one is at home during the day (d) T_n: Household 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 [34], 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 [35], shown in Table 2.4. 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 2.4: Efficiency for various heating equipment

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 [36]. 2.5 shows the list of such activities with their average hot water usage in gallons [36]. 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.

Activity Name	Water Usage
Laundry	7
Dishwashing	6
Shower	10 for 8 min
Cooking	1

Table 2.5: Water usage for Activites

2.4 Experiments and Results

2.4.1 Energy Demand Profiles of DC Population

We use our modeling framework to generate the energy demand profiles of the Washington DC population. We start with the synthetic population of the entire D.C region and then randomly select 5% of the households. This constitutes 62,763 households and 125,268 persons. We implement the statistical modeling algorithms for E_{Active} and $E_{Passive}$ in R using [37] [38]. For each individual in this population, we first generate an activity sequence for an entire day broken into 48 half an hour intervals. This includes estimation of individual as well as shared activities in the household, the times at which the activities are performed, the household members who perform the activity, association of appliances with each activity, the housing unit based passive energy consumption, and the overall energy consumption by each households.

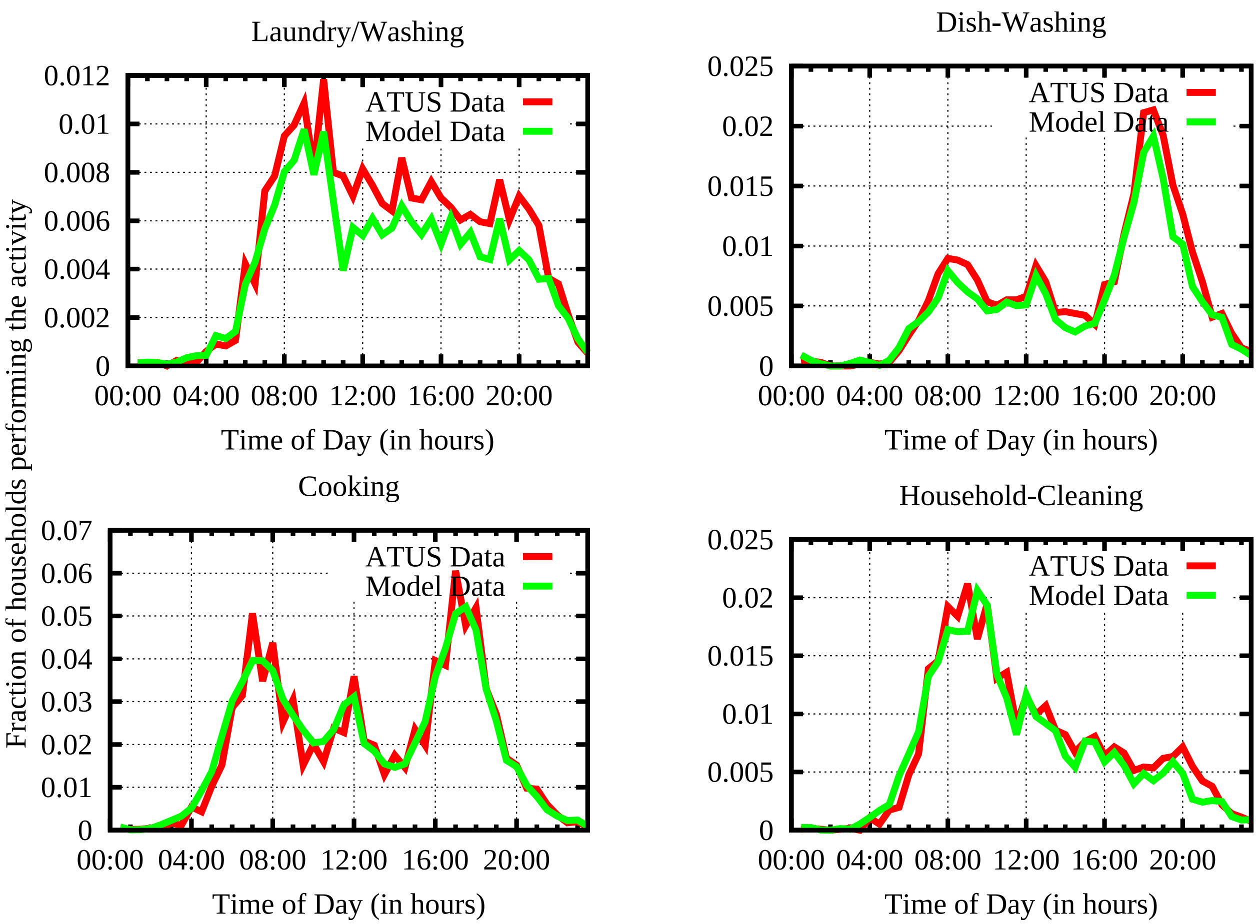

Figure 2.2: Comparison of synthetic data activity occurrence frequency with ATUS data activity occurrence frequency

Figure 2.2 compares our 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.

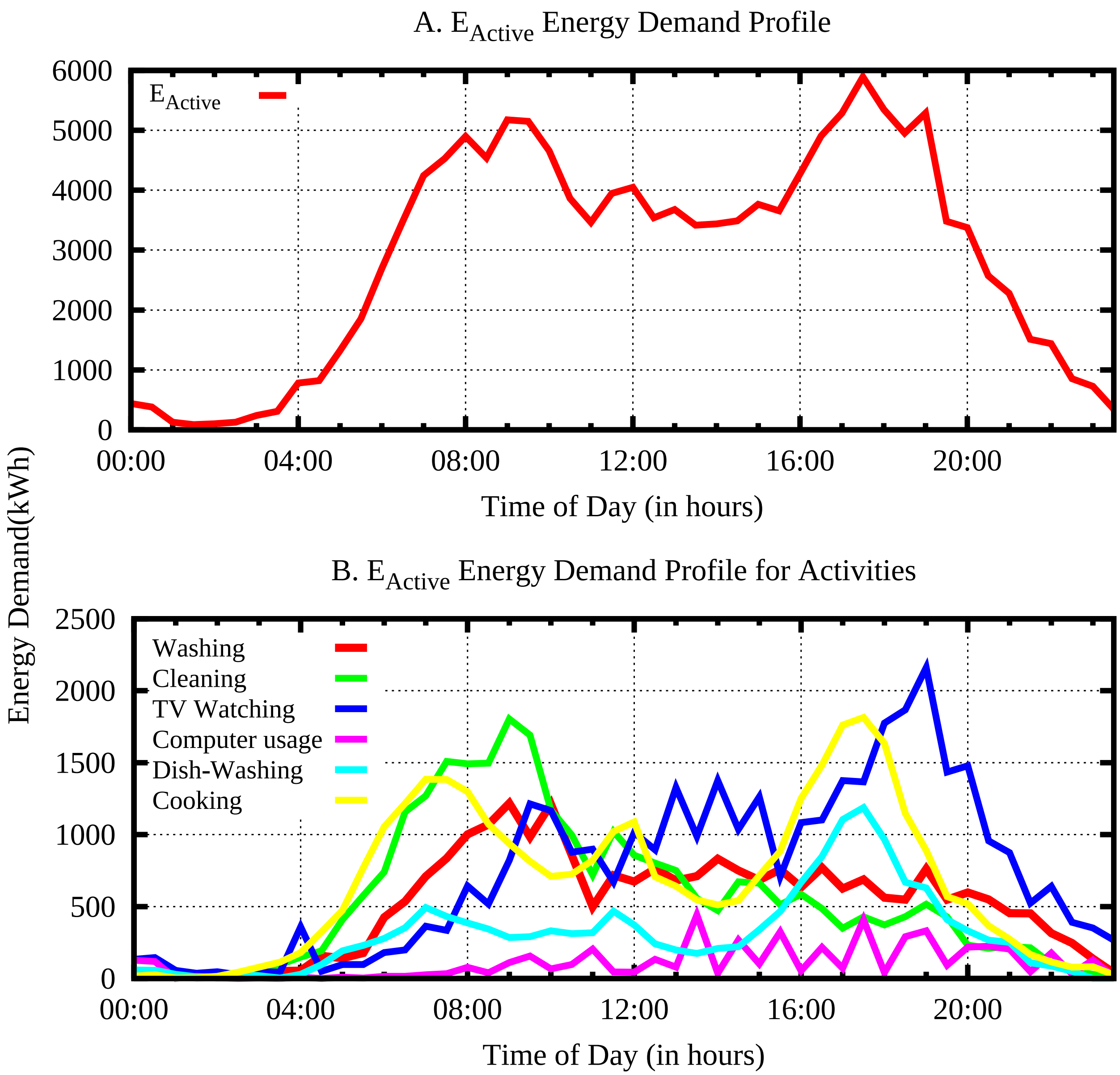

Figure 2.3: (A) Aggregated E_{Active} Energy demand profile (B) Activity Based Energy Demand

For this subset of the Washington DC population, Figure 2.3A shows the aggregate (E_{Active}) energy demand for all the "active" household activities and Figure 2.3B shows the demand pattern for individual activities. The curves show that in the day time, 'washing' and

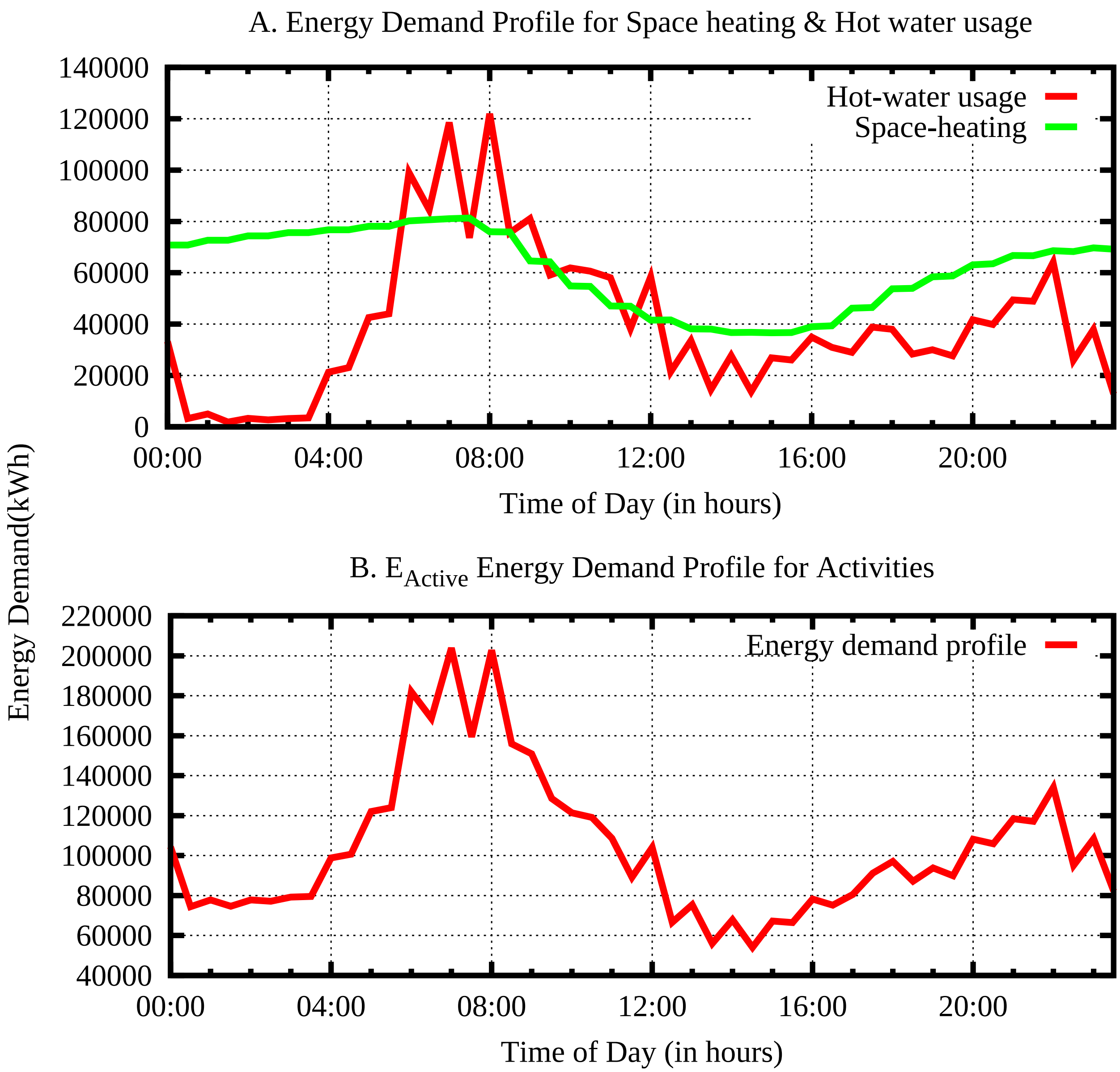

Figure 2.4: (A) Energy demand profile for Space heating and Hot-water usage (B) Total energy ($E_{Active} + E_{Passive}$) demand profile

'cleaning' activities are responsible for the peak in the E_{Active} load curve whereas in the night time, the peak is due to the 'cooking' and 'watching TV' activities. Figure 2.4A shows the estimated energy consumption due to space heating and hot water usage and Figure 2.4B shows the total energy demand which is the sum of E_{Active} and $E_{Passive}$.

2.4.2 Consumption Pattern by Demographics

This section illustrates how we can use this highly resolved demand model to understand the consumption pattern across various household demographics. Figures 2.5 and 2.6 shows the fraction of household performing the laundry and interior cleaning activities. Different curves represent different household sizes i.e. 1, 2, 3 and greater than 4. The results show that for activity "washing", the pattern of consumption across different sized households is the same but the fractions vary.

The employment status of a person in the household influences his/her stay at the household during the peak hours. Since, any employment either full-time or part time has a definite schedule, it influences the person's decision to perform his/her day-to-day activities. To understand its impact on the activity occurrence, we only consider households with two people and categorize them into three groups namely :- (i) Households with both individuals not working(ii) Households with one person working and (iii) Households with both individuals working. Figure 2.7 shows the percentage distribution of households with size 2, in each of these groups. Figure 2.8 shows the washing/laundry activity sequence for each of these three groups. The households, where both individuals are not working, tend to perform these activities during the peak hours. A similar trend exists in case of other activities too. This kind of information can be useful in designing energy policies that account for the demographics

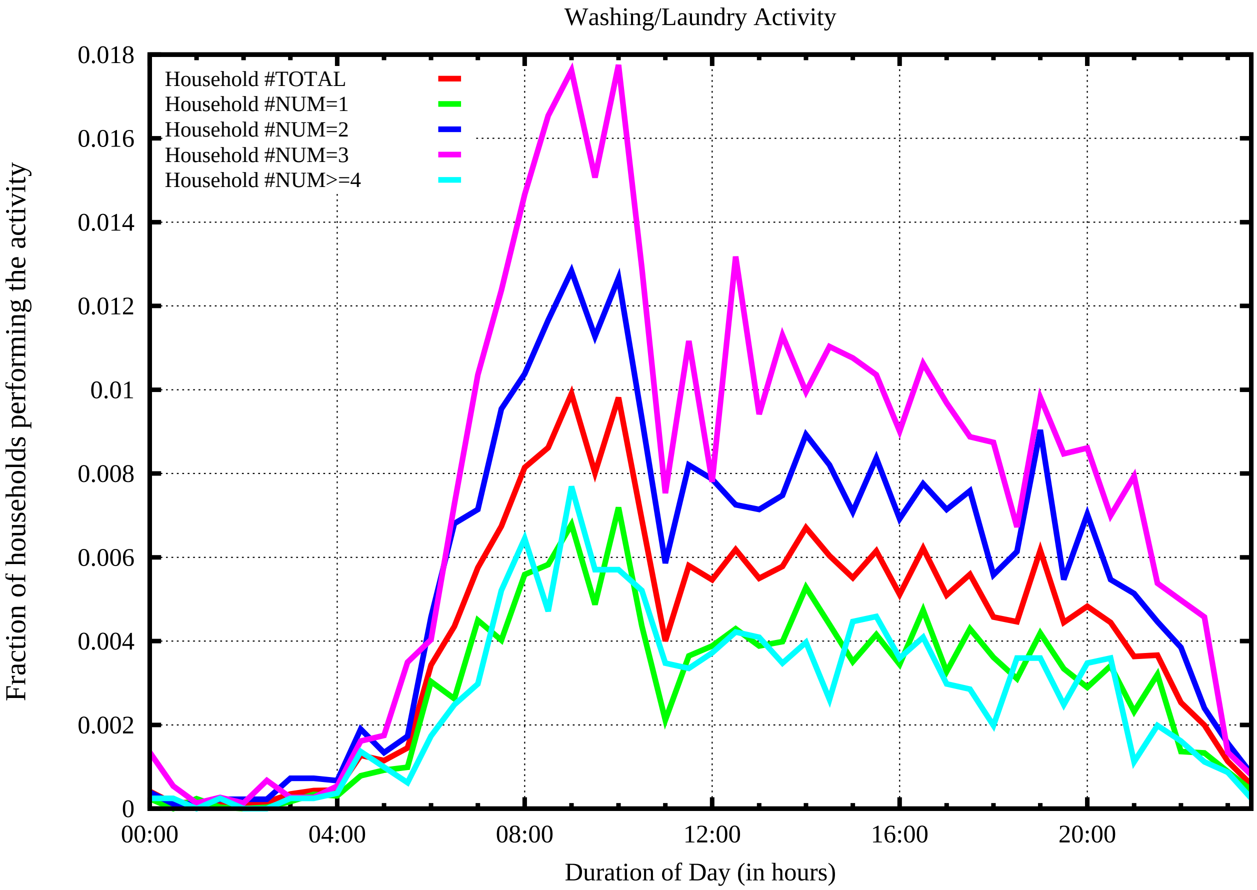

Figure 2.5: Washing activity pattern for various household sizes

of a local region.

2.4.3 An Illustrative Application

Here we illustrate how this detailed demand model can be used to improve energy efficiency. In the absence of new energy sources and energy conservation, energy efficiency is the only option. One way to improve efficiency is to shift the peak time energy consumption to off-peak time; also known as smoothing out the load curve. This will keep the inefficient

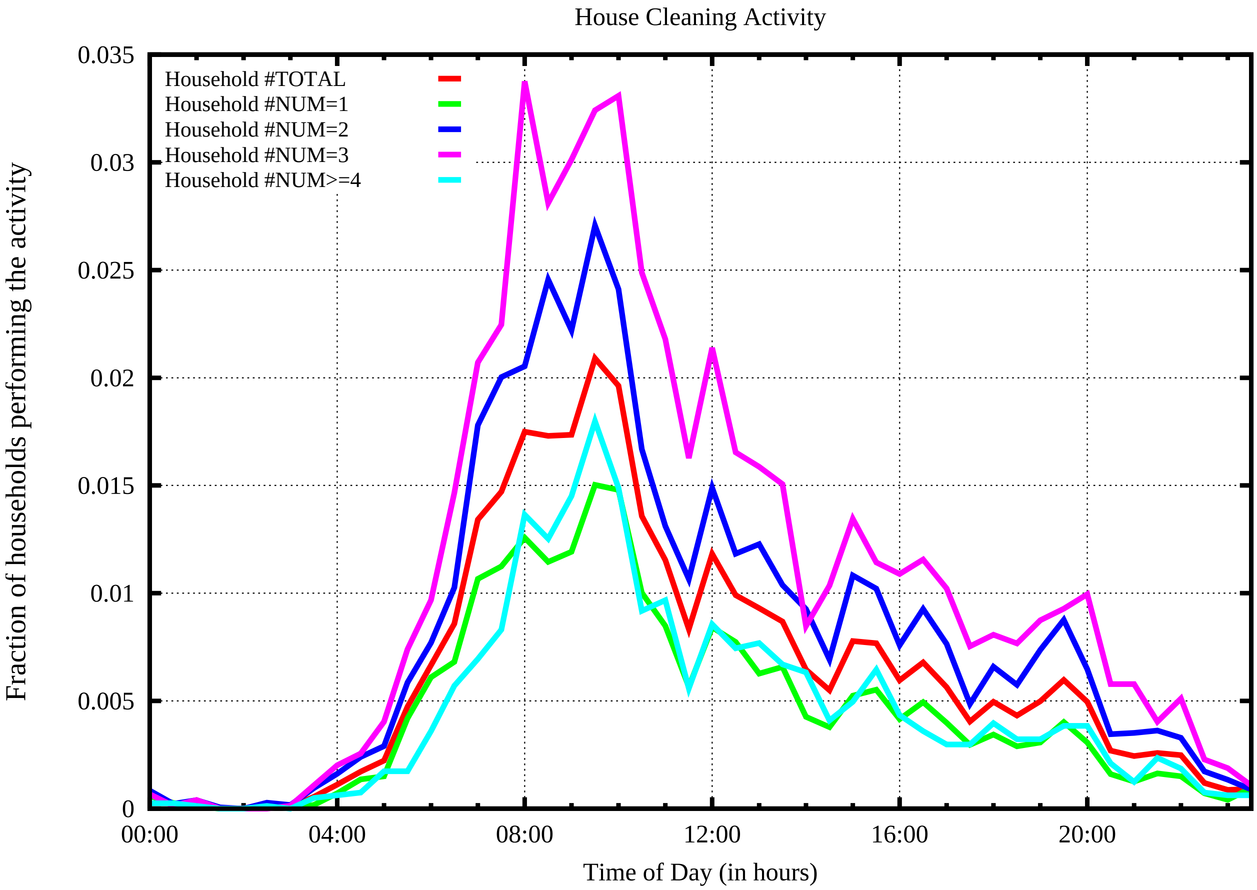

Figure 2.6: Cleaning activity pattern for various household sizes

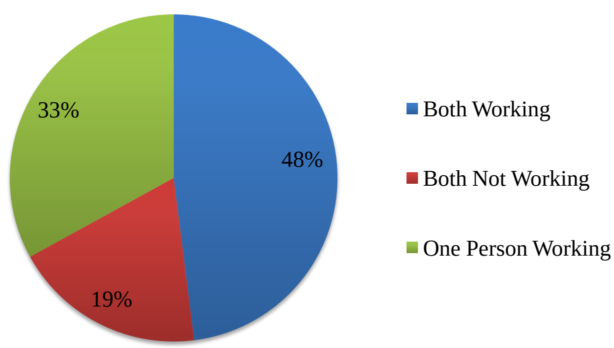

Figure 2.7: Distribution of households based on their employment status for household size = 2

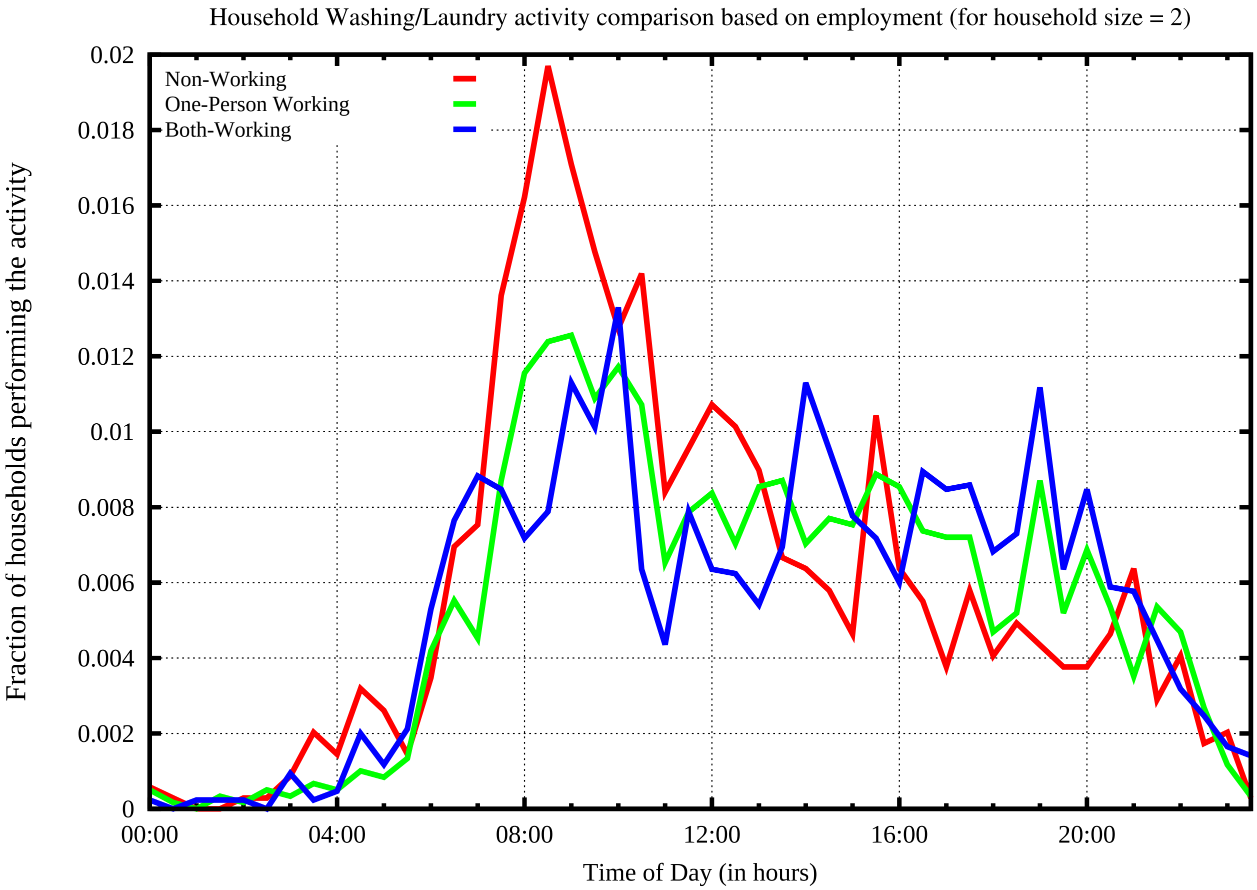

Figure 2.8: Washing/Laundry activity pattern for households with household size = 2, grouped based on their employment status. Figure illustrates that households with both non working persons tend to do energy intensive activities during the peak hours.

generators from coming online to serve the load at peak times. 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. A careful look at these activities shows that about 40% of the households who undertake these activities during the peak hours have at least one non-working adult available in their household. If 50% of these households shift their activities from peak hours to off-peak hours i.e. from 8AM to 11AM to sometime between 11AM to 3 PM, it could smooth out the E_{Active} curve. Figure 2.9 compares the

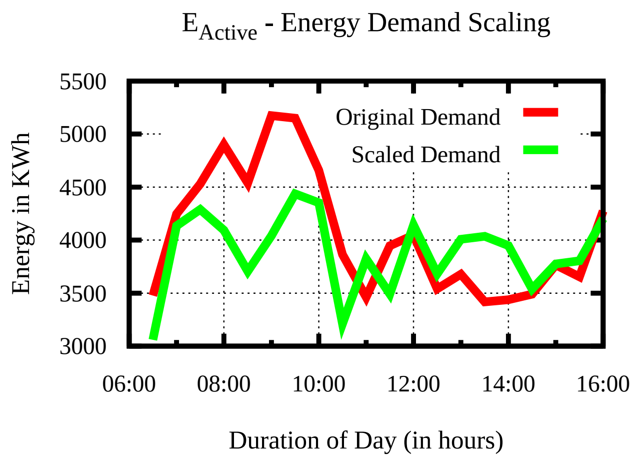

Figure 2.9: An illustrative case study showing the impact of shifting energy intensive activities from peak to off peak hours

scaled demand profile with the actual demand. By moving the timing of these activities, 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 2.9. This is a substantial amount of energy savings at peak time and could mitigate the vulnerabilities that occur when the system is running too close to the edge of capacity.

2.5 Summary

In this chapter, we described a modeling framework which generates highly resolved model to estimate individual and household level energy demand. It models household activity sequences for each member of the household, identifies which activities are shared, which ones occur independently and then maps them to appliance usage and the length of time used. It also calculates energy consumption of each housing unit based on characteristics such as square footage, wall type, floor area, heating fuel type etc. The model is run on a subset of the Washington DC population; demand profiles for each individual and their household is generated for people in this subset.

The results show that 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 with at least one non-working adult, some of the peak time activities were shifted to off-peak time and were performed by the non-working members of the family. This resulted 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.

Chapter 3

Commercial Sector

In this chapter[1], we describe the different works relating to our work and then present an overview of our modeling methodology used to generate the demand profile for commercial buildings.

3.1 Prior Works

According to US DOE's "Buildings Energy Data Book" commercial buildings broadly includes office spaces, educational facilities, food services, retail locations, hospitals, warehouses and storage facilities. In these office spaces, educational facilities and retail locations constitute about 50% of the commercial sector energy consumption [40]. Space conditioning, lighting and water heating at commercial sites represents more than 50% of the total energy

[1]An extended abstract of this chapter is going to appear in the proceedings of ETG Congress 2013,[39]

consumption of the commercial sector [40]. Each of these energy consuming functions vary with number of occupants present in the building. For example, more the number of individuals present, the more hot water gets used. Temperature regulation and lighting also increase with the number of occupants, though probably non-linearly. Estimating the number of occupants will be useful in determining the most efficient equipment settings/configuration that can be set to improve the energy efficiency of the building. Additionally, regardless of occupancy there is some energy consumption due to general building maintenance.

Work by [41] provides a fairly exhaustive review of different energy demand modeling techniques for commercial buildings. Other works have estimated commercial energy consumption as a function of occupancy. [42] [43] focus on determining or estimating the number of occupants in the building. They use sensors to collect occupancy-related data and use mathematical models to estimate the number of occupants. Our approach differs from that described. We propose a modeling framework in which building occupancies are derived from the day to day activity schedules of a synthetic population. Synethetic individuals are assigned activity locations based on geospatial, demographic, and activity scheduling data. As the synthetic individuals move from location to location, they are tracked, giving us dynamic occupancy estimates for each of the modeled buildings.

3.2 Dataset Description

This work uses following data sets:-

1. ***US EIA Commercial Building Energy Consumption Survey 2003:-*** In this work we use the United States Energy Information Administration's Commercial Buildings Energy Survey data (CBECS), which is the most prevalent data source available containing energy information at the building level granularity. This survey data provides information about the annual energy consumption for 5,125 buildings and its physical characteristics(like building type, number of floors, type of wall and etc). It also captures an extensive list of building's characteristics which can probably influence the energy consumption of the building. However, some of the characteristics are applicable only to certain type or group of buildings. So, these special characteristics were not relevant or applicable to the majority of buildings which resulted in many "not applicable" data points. So, in our modeling framework we consider only the most significant and common parameters which influence the energy consumption of the buildings like square footage, operating hours, principle business activity of the building, type of equipment used for space conditioning, water heating, lighting and refrigeration. The table 3.1 lists the different variables that we use in our statistical models and their description.

   In this survey, each building unit has an weight associated with it captured by the variable $ADJWT8$. So, for any building B_i in the survey data we have an weight $ADJWT8_i$. This weight represents that there are $ADJWT8_i$ building units having the same characteristics of B_i. The survey uses this weight for forecasting the energy consumption of the whole United States.

CBECS Variables	**Description**
SQFT8	Building square footage
WKHR8	Building weekly work hours
NWKER8	Number of workers
NFLOOR8	Number of floors
NESLTR8	Number of escalators used
RWSEAT8, PBSEAT8, EDSEAT8, FDSEAT8	Captures seating capacity information at various buildings
FURNP8, BOILP8, PKGHP8, SLFCNP8, HTPHP8, STHWP8, OTHTP8	Percentage of space heated using various heating equipment
PKGCP8, RCACP8, ACWNWP8, CHWTP8, CHILP8, EVAPP8, OT-CLP8	Percentage of space cooled using various cooling equipment
HEATP8	Percentage of space heated
COOLP8	Percentage of space cooled
REGION8 and CENDIV8	Categorical variable capturing census regional and divisonal details
CLIMATE8	Categorical variable representing the area's climatic conditions based on 30 year average number of heating degree days (HDD) and cooling degree days (CDD)
ELEVTR8	Number of elevators used in the building
MAINHT8	Main heating equipment used
MAINCL8	Main cooling equipment used
WTHTEQ8	Main water heating equiment used
RFGWIN8, RFGOPN8, RFGRSN8, RFGCLN8, RFGVNN8	Number of walk-in, open, residential, closed and vending machine based refrigerated units used
PCNUM8	Number of computers used
PCRMP8	Percentage of computer area
SRVNUM8, COPIER8, FAX8, PRNTRN8	Number of servers, coipers, fax machines and printers used
ADJWT8	Building weight
ELBTU8	Building's annual electricity consumption in BTU
MFBTU8	Building's annual fuel usage in BTU

Table 3.1: List of various EIA's CBECS variables used

2. ***D.C. Synthetic Population:-*** To construct the demand profile for the range of an urban city, we use a synthetic data representing Washington D.C. region. The synthetic population is derived from the data gathered from the US Census, the National Household Travel Survey (NHTS), Dun and BradStreet, Land Use, and Navtaq. Details of the methodology can be found in [26, 27]. The synthetic population is statistically representative of the true population of Washington DC at a block group level in the US Census.

3.3 Methodology

We formulate the energy consumption of commercial buildings as a summation of two major components namely an active and passive component. The active component takes into account the social interactions happening inside the buildings due to the people's visit. The social interactions can influence the building's principle activity which might impact the building's energy consumption. For example, the number of people visiting the restaurants might influence its main activity namely 'cooking'. The passive component takes into account all the non-occupancy factors which influence the energy consumption of the commercial buildings. The non-occupancy factors comprises of building's structural (building structure, wall types, square footage), physical (type of equipment used for space conditioning, lighting, refrigeration) and locational characteristics(regional weather conditions).

$$E_{Commercial} \propto f(p) + g(a)$$

$Where,$

$f(p)$ = models the factors which influences passive component of the building's energy consumption

$g(a)$ = models the factors which influences active component (social interactions) of the building's energy consumption

We also define the occupancy O_{it} of any building B_i at time t as the sum of number of workers W_{it} working in the building and the number of visitors/customers visiting the building at t

$$O_{it} = W_{it} + V_{it}$$

Our modeling approach consists of five major steps namely :-

1. Categorizing the commercial buildings into different groups based on their principle activity
2. Building statistical model (incorporates all the influencing factors) for each building group
3. Visit profile constructor for synthetic population
4. Matching and extrapolating features of CBECS building unit onto synthetic locations
5. Generating location based hourly energy demand profile

Following sub-sections explain these steps in detail and we provide an algorithm to implement this framework.

3.3.1 Grouping CBECS buildings based on 'Building's Activity'

We classify the 5,125 CBECS buildings into different classes or groups based on their principle business activity. This approach is similar to the work done by [11]. However, we group the buildings to match and be consistent with our synthetic population. In the synthetic population, individuals have an activity schedule and visit the locations/buildings with a well defined 'purpose'. The purpose of the visit can be Education (School), Purchase/Shop(retail, restaurants), Work (office) or Others. So, we group the CBECS building units into these four groups based on their principle business activity (captured by the survey variables PBA8, PBAPLUS8). The table summarizes the different groups of buildings we use and their related mapping variables.

Building Category	PBA	PBAPLUS8
School	14	28,29
Work	2	2,3,4,5,6,7
Retail	6,15,23,25	42,50,14,32,33
Others	5,8,13	9,10,18,19,35,38,39

Table 3.2: CBECS survey data buildings grouped based on their principle business activity

3.3.2 Statistical model

In the EIA's CBECS data [15] we have the annual energy consumption for the buildings. We use the annual energy usage (annual major fuel used and annual electricity consumed) of the building as the response variable in our regression equation and identify the most significant dependent variables influencing the energy consumption using step-wise variable selection [44]. After identifying the significant dependent variables we fit the data using ***Negative Bionomial Regression***[45] [46]. We carry out this approach for each group of buildings and create models specific to each group.

In each of this model the total energy consumption of the building for a given hour is the response or dependent variable denoted by Y. Then, for each category of buildings we identify the major set of variables which influence the energy consumption of the building and these variables becomes the independent variables in the regeression equation denoted by $\vec{X}$. Each group of buildings have their own unique set of dependent variables. In each of these statistical models we check for interaction [47] between the different types of equipments used for space heating, cooling, water heating, refrigeration and with the building's total operating hours H. If, we will find any significant relationship or dependency we then introduce an interaction between these variable and fit the regression model.

$$Y \sim NB(\vec{X})$$

Rationale:- We use Negative binomial regression because it gave a more realistic fit than the standard Gaussian linear model or the more restrictive Poisson regression model.

3.3.3 Visit profile generator

As mentioned in the earlier section we use a synthetic population representing Washington D.C region. The synthetic population consists of synthetic individuals endowed with demographic and social attributes. A detailed sequence of activities and their respective locations are assigned to each individual based on their demographics such as age, household size, income, and employment status. The types of activities include home, work, school, shop, travel and others. We derive visit profiles for each building based on a location identification scheme, which places the individuals into a suitable building based on their activity type, the building's maximum occupancy, and the distance of the candidate building from the individual's current location and its anchor location i.e. home. For example, if a person is to do "shopping" activity, we assign the person to one of the nearby retail stores using a gravity model. This model consider all candidate buildings for shopping and then assigns a location if it does not exceed the occupancy limit of the location and the distance is within some limits to person's home and from person's current location.

At each location/building in the synthetic population we calculate the occupancy for each hour of the day by summing up the number of synthetic individuals who visit the location for 'Work' and the individuals who visit the location for a special purpose which is associated with the location/building.

3.3.4 Matching synthetic locations with the surveyed buildings

We exploit the common characteristics exhibited by the CBECS buildings and synthetic locations to perform the necessary matching. To identify the suitable match for the synthetic location we filter the CBECS buildings based on census division, region and climatic conditions. Then, since the synthetic population locations contain sub-locations and each sub-location has an occupancy limit associated with it. We use the occupancy threshold has the next criterion to filter the CBECS buildings search space. After performing these steps we have a set of CBECS buildings that match our synthetic location. We use the CBECS building's weight as a criterion to randomly choose/pick one CBECS building for the location. For example we have k buildings, then we sum the weights of each building and divide it with the building's weight to get the probability of the choosing the respective building. We then sample the building's probabilities and choose one building. After, choosing the CBECS building we extrapolate the characteristics seen in the survey building onto the synthetic location.

3.3.5 Energy demand profile construction

For each synthetic location based on the extrapolated characteristics from the CBECS data, hourly calculated occupancy value we use the appropriate building's statistical model and estimate the hourly energy consumption. We then finally aggregate the energy consumption to generate the overall energy demand profile for the region

3.4 Experiment and Results

3.4.1 Statistical Models Implementation and Results

We implement the statistical models using R [48]. For each group of buildings we try to fit the model by selecting the most significant variables which influence or determines the energy consumption of the building. We use step-wise variable selection is implemented using the MASS package [49]. The negative bionomial regression is implemented using the MASS package [49] function "*glm.nb()*". For each the groups of buildings we model the energy consumption using this procedure and select the variables which are most significant. As stated earlier we clearly have two classes of independent/predictor variables namely (i) variables which relating to building's occupancy and affecting the active part of the energy consumption (ii) non-occupancy related variables influencing the passive component. Most of the non-occupancy related variables are common across all the building groups. However, each group has a few unique variables specific to the building group that influence the energy consumption.

In the regression equations, we have few categorical variables whose values need to be treated differently with the other continuous variables. Therefore, we use the R factor function [50] to encode them as factor variables. In this way, we make sure that the modeling equations are functioning properly.

School

In modeling the energy consumption for schools we found that variables 'NWKERS8' and 'EDUSEAT8' contribute to the active component of the energy consumption. The rest of variables influence the passive component of the energy consumption. In the 'non-occupancy' related variables we found that number of PCs (PCNUM8) and number of copiers (COPIER8) have significant influence on the energy consumption. The below negative binomial regression equation shows the list of variables used in modeling school building's energy consumption. Figure 3.1 shows the plot of fitted values versus the actual values; points tightly centered around a 45 degree line indicate a good fit.

$ELBTU8+MFBTU8 \sim (factor(CENDIV8)+factor(CLIMATE8)+factor(WLCNS8)+factor(RFCNS8)+SQFT8+NFLOOR8+NELVTR8+NESLTR8+factor(OPNMF8)+NWKER8+PCRMP8+SRVNUM8+RFGWIN8+RFGOPN8+RFGRSN8+RFGCLN8+RFGVNN8 + PCNUM8 + COPIER8 + EDSEAT8 + H : factor(MAINHT8) + H : factor(MAINCL8)+H : factor(WTHTEQ8), data = school, init.theta = 2.789532743, link = log)$

Office

In office buildings the main influencing occupancy factor is the number of workers. The final regression equation used is as follows:-

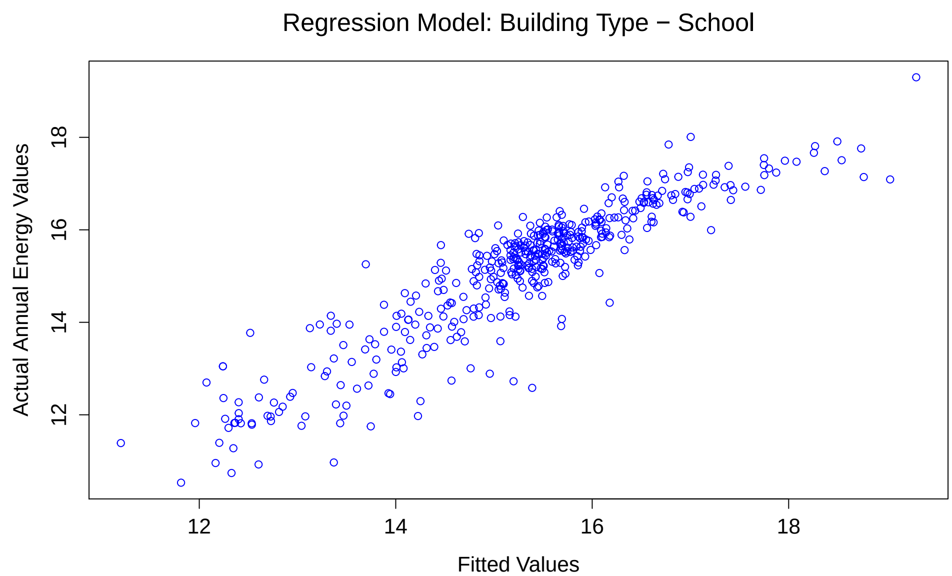

Figure 3.1: Model fitted values vs. Actual values (log scale) for schools

$ELBTU8+MFBTU8 \sim (factor(REGION8)+factor(CENDIV8)+factor(CLIMATE8)+factor(WLCNS8)+factor(RFCNS8)+SQFT8+NFLOOR8+NELVTR8+NESLTR8+factor(OPNMF8)+NWKER8+PCRMP8+SRVNUM8+RFGWIN8+RFGOPN8+RFGRSN8+RFGCLN8+RFGVNN8+PCNUM8+COPIER8+H:factor(MAINCL8)+H:factor(WTHTEQ8), data = trainset, init.theta = 1.593470638, link = log)$

Figure 3.2 shows the fitted values vs. actual values plot.

Retail

In retail buildings, the special measures of occupancy like seating capacity for restaurants and the number of workers feature in the final regression equation. Figure 3.3 shows regression model's fitted values vs. actual values plot and the equation is as follows:-

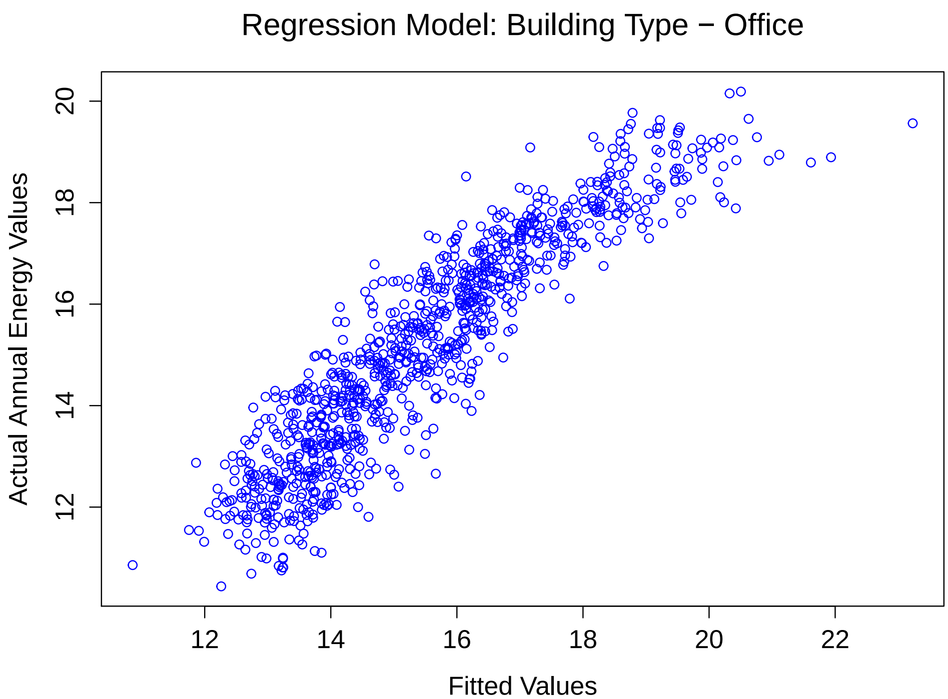

Figure 3.2: Model fitted values vs. Actual values (log scale) for office spaces

$ELBTU8 + MFBTU8 \sim (factor(CENDIV8) + factor(CLIMATE8) + factor(WLCNS8) + factor(RFCNS8) + factor(PBA8) + SQFT8 + NFLOOR8 + NELVTR8 + NESLTR8 + factor(OPNMF8) + PCRMP8 + SRVNUM8 + RFGWIN8 + RFGOPN8 + RFGRSN8 + RFGCLN8 + RFGVNN8 + FDSEAT8 + factor(FACACT8) + NWKER8 + H : factor(MAINHT8) + H : factor(MAINCL8) + H : factor(WTHTEQ8), data = retail_data, init.theta = 2.278492016, link = log)$

Others

The Other building type which includes hospital buildings, warehouse, small and large hotels and etc. have a special some sort of occupancy measure which is include in the regression

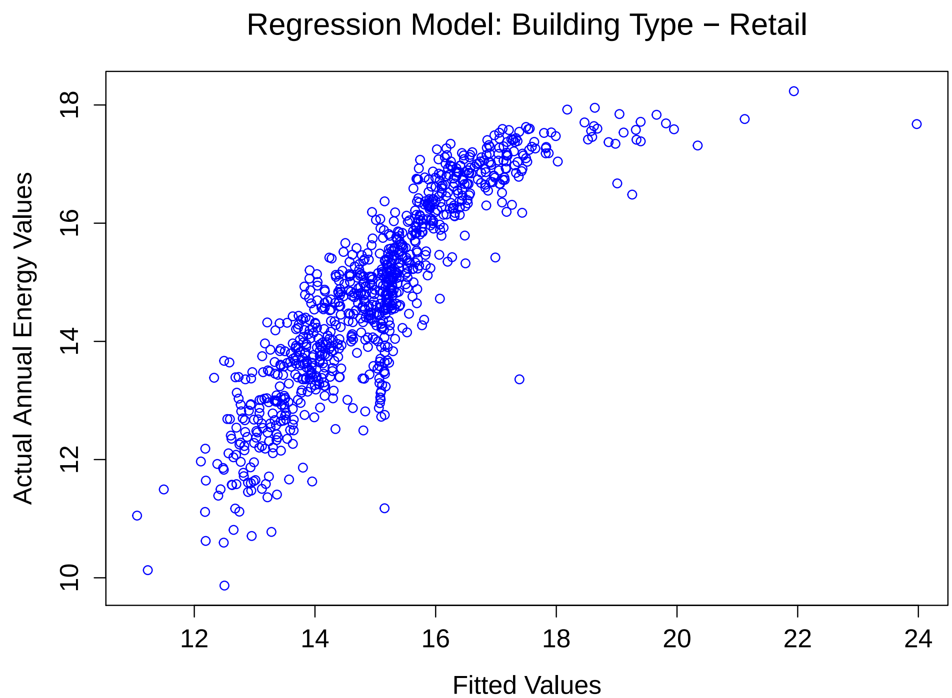

Figure 3.3: Model fitted values vs. Actual values (log scale) for retail buildings

equation. The results of the model is shown in Figure 3.4 and the regression equation is:-

$ELBTU8 + MFBTU8 \sim (factor(PBA8) + factor(REGION8) + factor(CLIMATE8) + factor(WLCNS8) + factor(RFCNS8) + SQFT8 + RFGRSN8 + H : factor(MAINHT8) + H : factor(MAINCL8) + H : factor(WTHTEQ8) + NWKER8 : HCBED8, data = train_data, init.theta = 0.8722299412, link = log)$

3.4.2 Synthetic Locations Energy Demand Profile

The synthetic population of Washington DC consists of 133,901 commercial locations and there are distributed as

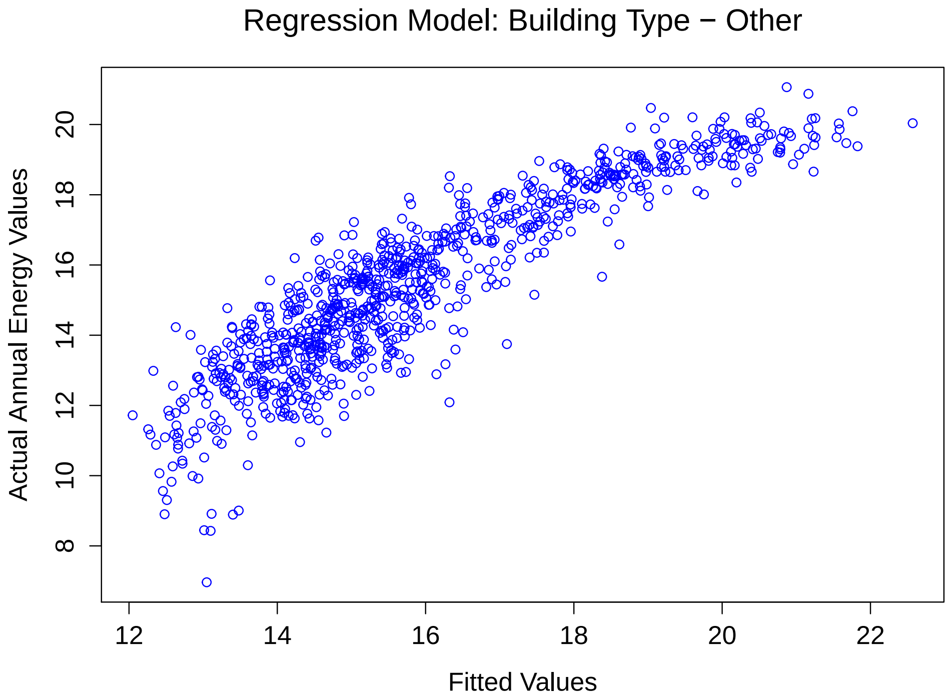

Figure 3.4: Model fitted values vs. Actual values (log scale) for other building types

1. 2,054 School locations
2. 54,067 Work locations
3. 16,334 Retail locations
4. 61,446 Other locations

We construct the hourly occupancy profile for each synthetic location using the visit profile generator. Then we use the occupancy information and estimate the hourly energy demand of the building based on the constructed statistical model.We follow this approach for each building and aggregate the energy demand based on the building type.

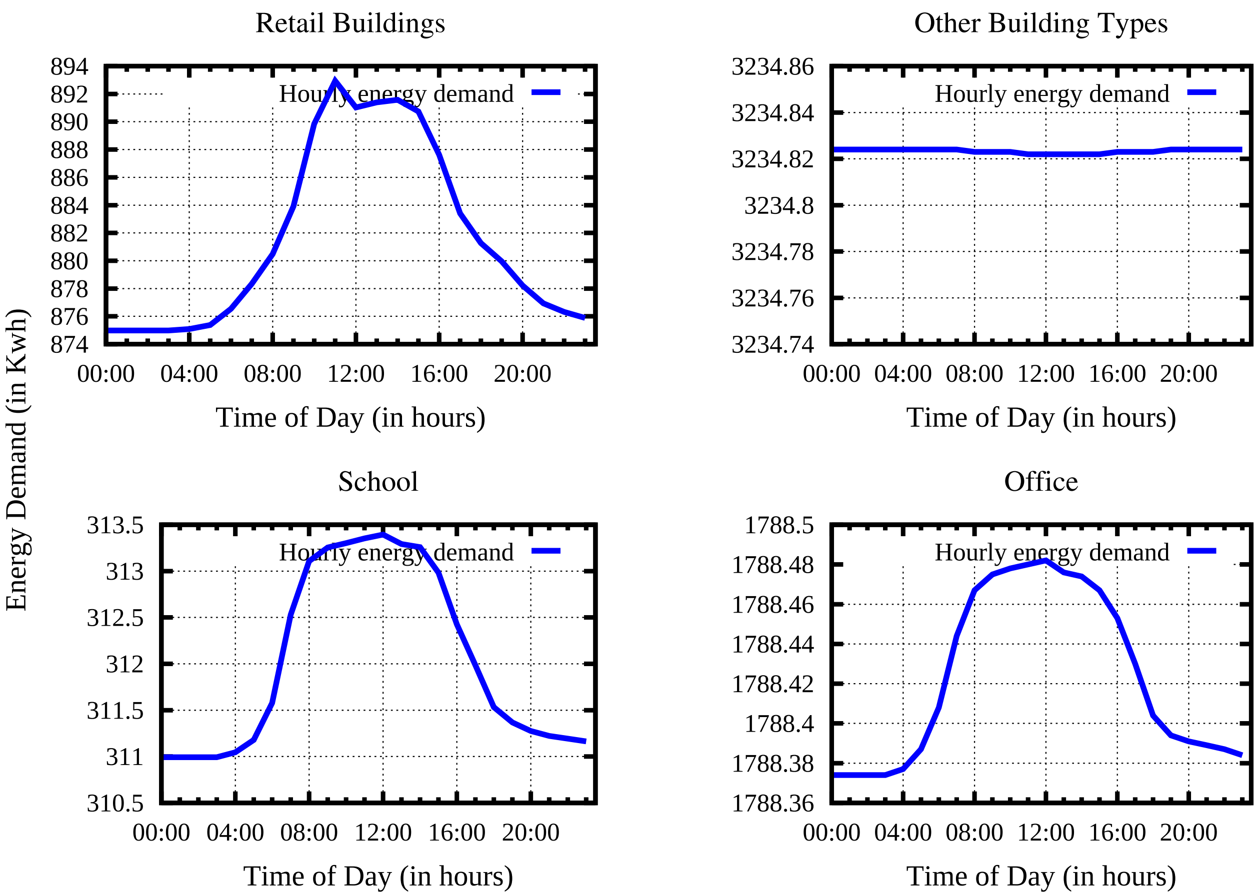

Figure 3.5: Energy demand profiles of various categories of commercial buildings located in Washington D.C. (synthetic data) classified based on their business activity.

Figure 3.5 shows the energy demand profile across various commercial buildings groups over the 24 hour period. The x-axis represents the 24 hour time duration and the y axis represents the corresponding forecasted energy demand in kWh. It is evident that people's visit during the day is having an impact on the energy consumption. This is evident in the retail buildings. Also we were able to estimate the energy consumption due the major activities like space conditioning, water heating, lighting and refrigeration from our regression model. Figure 3.6 shows the total energy demand of all the commercial location in Washington D.C.

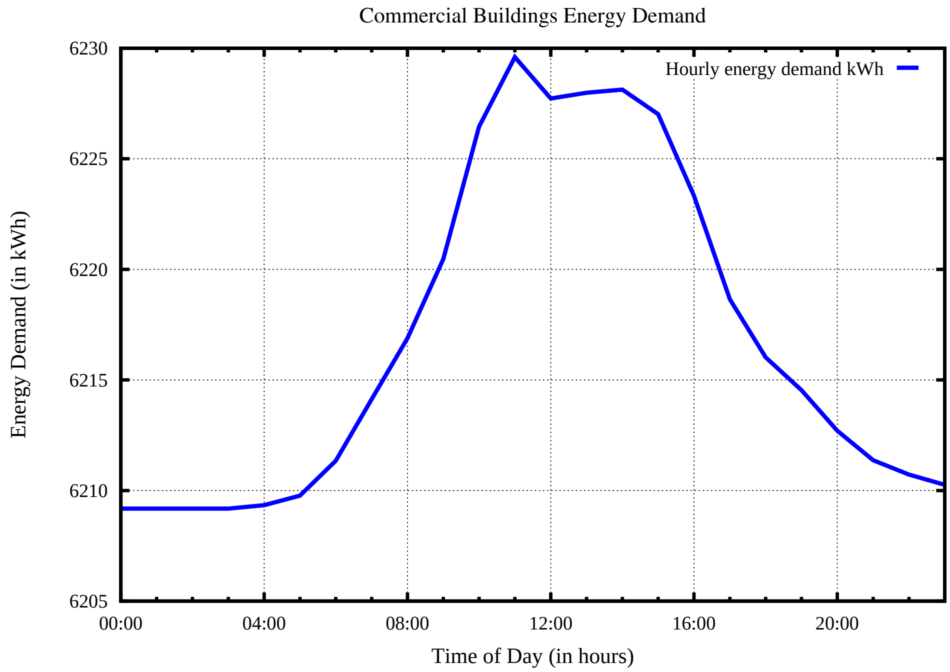

Figure 3.6: Aggregate energy demand profile for all the commercial locations in Washington D.C. (Synthetic Population)

3.5 Summary

We introduce a framework that uses the day-to-day people's activity schedules and derive the visit profiles for each of the commercial locations. The framework uses the people's visit purpose and buildings' visit profiles to construct the realistic occupancy rates for each commercial location at an hourly basis. We then use the hourly occupancy rate to estimate the energy consumption of the building at a very fine granularity. In this way, we accommodate the social interactions happening inside the building while estimating the energy consumption. This helps in evaluating various what-if scenarios relating to energy policies at an urban-city level.

Chapter 4

Energy Demand Modeling Framework

In this chapter, we introduce our software infrastructure used to generate and analyze energy demand profiles for buildings. The framework provides following features: (i) it provides an easy extensible modular framework that supports changes in mathematical models, input datasets, and synthetic data (ii) helps in evaluating different what-if scenarios with minimal modifications (iii) it provides web-based services and seamlessly integrates with any web interfaces; this allows the policy makers who are not modeling experts to realm benefits from our framework. Thus, this chapter describes our software system by focusing into our system design and implementation. The chapter concludes by providing a web-based prototype built on top of our software framework that can be beneficial for energy economists.

4.1 Design and Implementation

Design Overview

Both the residential and commercial buildings energy demand generation involves many steps that are stochastic in nature. In general, each step consists of mathematical modeling process in which we impose the characteristics of an input or training dataset onto a dataset of our interest (coined as synthetic population) and then construct the demand profiles. Therefore, our software design should be able to make the framework extensible and adaptable to different changes. Thus, we might encounter scenarios where we might need to change the modeling algorithms based on the input dataset changes. These changes happen because we choose the best possible model based on the input dataset's characteristics after performing extensive analysis, which formulates our hypothesis well. Thus, if the input dataset changes, then we might need to evaluate the performance of current models on the newer datasets and most likely we might encounter situations where we need to introduce a new model that is more suitable and leverages all the information contained in the new dataset. In addition to that, the design should be flexible enough in interfacing with different What-if scenario modules and visualization tools. Thus, we design for system such a way that these design objectives are satisfied.

Since, the modeling methodology discussed in this thesis formulates the energy demand as a function of individual and building activities. Given a dataset of people and buildings, at an abstract level, we have three major steps to generate the demand profile namely

(i) Embedding information from the training/input datasets (ii) Activity generation (iii) Translating activities into energy demand. Thus, we modularize each of these components into different independent modules. Then, for each module, we identify whether the module is stochastic in nature or not and whether it requires a persistent storage for its results. This helps us in designing the database model for this kind of framework. In addition, for each module, we identify set of APIs that other services can use and set of configuration settings that controls the execution of the module. Since, we interface the modules through APIs; we can easily extend the framework to any data, which contains similar kind of format and parameters. Thus, a software module abstracts major functionality of the framework and provides the framework with the flexibility to support simple extensions to other similar data sets and statistical algorithms.

Each module is implemented in Java and it interfaces with R using Rserve package [51] to carry out the necessary statistical analysis. The statistical models are implemented in R as functional modules. The module written in Java connects to R server session and dynamically loads the statistical modules needed for the current execution in the connected R session. In the connected R session, implemented statistical module is executed and results are passed back to the Java module. We will illustrate this design approach in detail by taking an example module and walk through the different stages of execution. We will use the UML sequence diagram [52] to illustrate the execution process and it is depicted in the figure 4.1. The sequence diagram shows the sequence of stages involved in the module that forecasts the hourly energy consumption of the commercial building embedded with all the necessary characteristics required by the statistical model. Following steps explain the stages of execution in detail:-

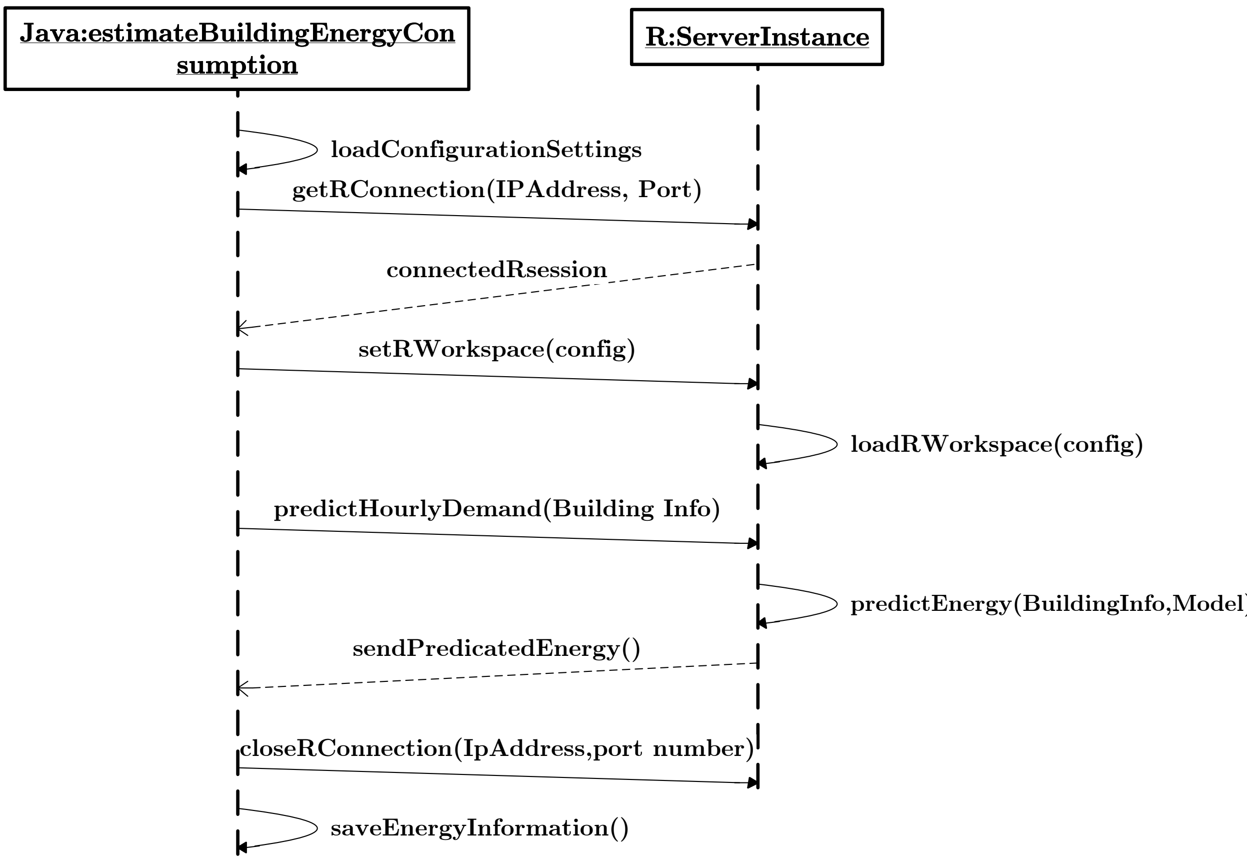

Figure 4.1: Sequence diagram illustrating the stages involved in the module that generates commercial building's energy demand at regular basis.

1. The Java module reads a configuration file; loads the R server details and R statistical module to be executed

2. The Java module sends a connection request to the R server; R server establishes a connection and creates new R session

3. The R server sends R session handle to the Java module

4. The Java module loads the desired R workbench based on the information stored in the configuration file (loaded in the first step)

5. The Java module sends the building's parameters to the loaded R function, which use the statistical model to forecast the energy consumption for the given parameters.

6. The R function returns the forecasted value to the Java function

7. The Java function closes the R session and saves the result

In this manner, the decoupling of statistical models from the main software module allows the system to migrate to different statistical models by just modifying the configuration file without actually changing the Java source code, which needs recompilation. In this way, this design can handle most of the design objectives discussed in the beginning of this section.

Residential and Commercial Buildings Energy Demand Framework

Figure 4.2 shows our system design for residential energy demand-modeling framework. Each box in the diagram represents a software module. We describe the functionalities of these modules, its input parameters, mathematical model and generated output. Table 4.1 summarizes the different parameters used in residential energy demand modeling system and also briefly describes them.

- **Module 1:** This module maps the time use survey individuals onto synthetic individuals

- **Input Parameters:** Time-use survey respondent's demographic information and activity schedule information; Synthetic population's people demographic information (P2)
- **Model:** Classification and Regression Tree implemented in R
- **Output:** Each of the synthetic individual is embedded with the matching respondent's activity schedule information
- **Customization Needed:** The input information is stored and retrieved from a database. Therefore, if there is any change in data, the module needs the data to be stored in a specific format for execution

- **Module 2:** This module maps the RECS buildings' survey characteristics onto the synthetic population's households
  - **Input Parameters:** RECS's buildings demographic information and their characteristics; Synthetic population's households demographic information (household size, household income, location and regional information) (P4)
  - **Output:** Each of the synthetic household is embedded with the matching RECS household's characteristics (equipment information, appliance information and temperature settings)
  - **Customization Needed:** The input information is stored and retrieved from a database. Therefore, if there is any change in data, the module needs the data to be stored in a specific format for execution

- **Module 3:** This module generates the shared activity sequence for a given household
  - **Input Parameters:** Shared activity list & its occurrences in survey data; synthetic household member's demographic information
  - **Model:** Logistic and Poisson regression
  - **Output:** Generated shared activities sequence for the household
  - **Customization Needed:** List of shared activities can be configured by modifying the module's configuration file
- **Module 4:** This module generates the independent activity sequences for given household
  - **Input Parameters:** Independent activity list; synthetic household member's independent activity schedules derived from their matching ATUS individual
  - **Output:** Synthetic household's independent activities schedules
  - **Customization Needed:** List of independent activities can be configured by modifying the module's configuration file
- **Module 5:** For a given household, this module generates the hot-water usage information across a day
  - **Input Parameters:** Synthetic population household's water heating equipment details derived from the survey (P7); Synthetic household's activity sequences

and information about the activities that use hot water (P8) Synthetic population demographic variables (P2)

  - **Output:** Household level hot water demand (in gallons)
  - **Customization Needed:** List of activities that use hot-water can be configured by modifying the module's configuration file

- **Module 6:** For a given household, this module generates the hourly heat flow rate required at keep the household at a desired temperature
  - **Input Parameters:** Synthetic household's derived information - space heating/cooling equipment information, square footage under space conditioning information, thermostat settings, household wall type; Synthetic household In-home activity sequence; Outside hourly temperature from weather data
  - **Output:** Calculated hourly heat loss rate
  - **Customization Needed:** List of In-home activities can be configured, hourly outside temperature information can be changed dynamically

- **Module 7:** Map household activities with appliance and generate energy demand profile for every household
  - **Input Parameters:** Synthetic household's fully constructed activity sequence; Appliance-activity mapping information and its standard energy rating information

  - **Output:** Constructed household-level energy demand profile
  - **Customization Needed:** Appliance power rating information and the activity-appliance mapping information can be configured in the module's configuration file

- **Module 8:** Construct overall energy demand profile by aggregating demand profiles of all households;
  - **Input Parameters:** Energy demand profiles of all synthetic household; Optional policy module information
  - **Output:** Aggregate energy demand profile for all households

Commercial Buildings System Framework

Similarly, the figure 4.3 shows the system design for commercial buildings and the table **??** gives the complete description of model's input and output parameters.

- **Module 1:** This module groups commercial buildings and construct a regression model which represents the energy consumption
  - **Input Parameters:** CBECS's buildings annual energy consumption and their associated characteristics (P1)
  - **Model:** Regression model;

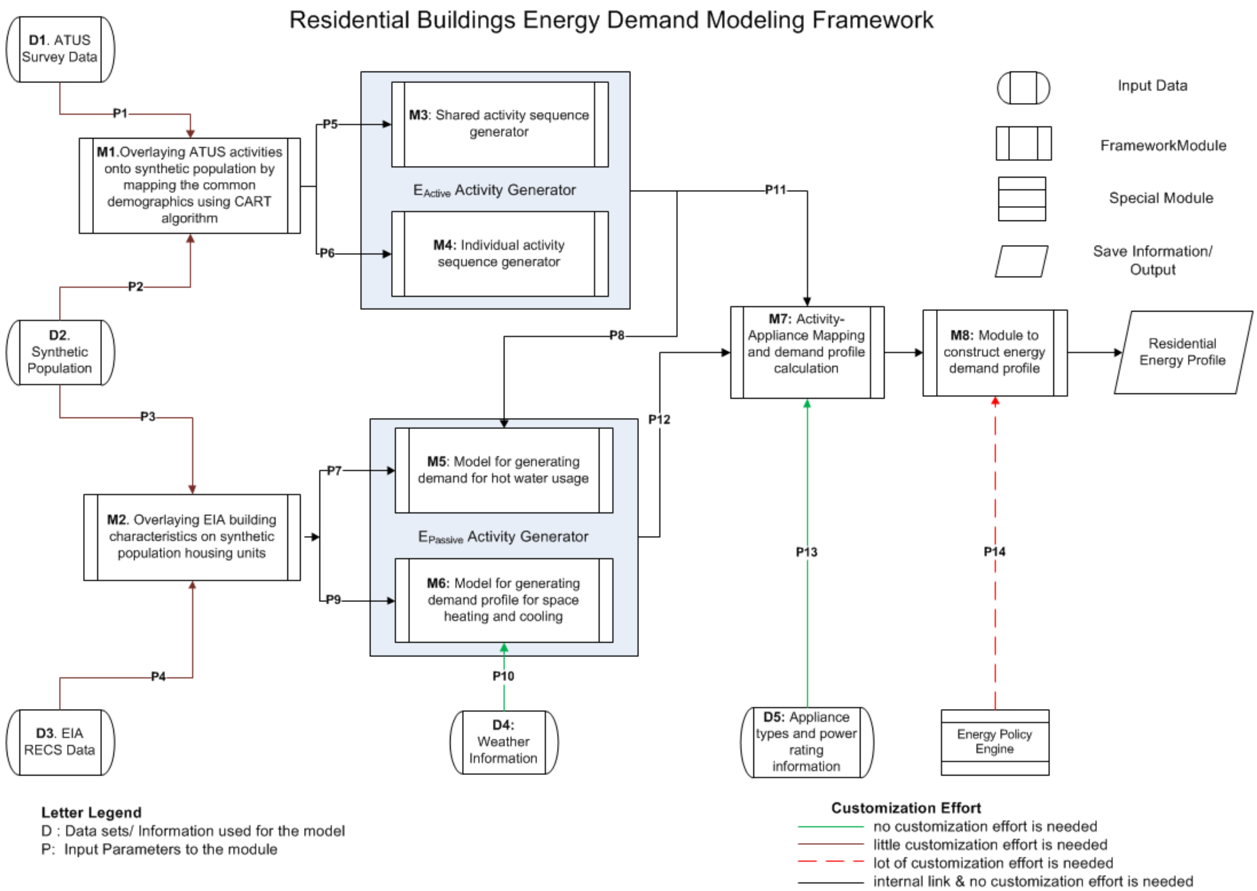

Figure 4.2: A software framework to generate energy demand profile for residential buildings

- **Output:** This module constructs a statistical model for each category of building (grouped based on building's principle activity) and estimates the energy consumption using the model based on the building's characteristics

- **Customization Needed:** If the data-set changes, we might need to review the model and check whether it fits the data well. Therefore, if a data-set changes, this module might need modifications to develop the model which best fits the data

Parameter Number	Description
P1	ATUS respondent's demographic information and his/her activity sequence
P2	Synthetic individual's demographic information
P3	Synthetic household's characteristics (household income, number of household members)
P4	EIA's surveyed household characteristics
P5	Shared activity information, Demographic information, ATUS and Synthetic population mapping information
P6	Individual activity and individual activity information, Demographic information, ATUS and Synthetic population mapping information
P7	Type of water heater used information from EIA
P8	Information regarding activities that consume hot water
P9	Type of space heating equipment used, wall type, floorspace, temperature settings used information from EIA survey
P10	Outside temperature from weather data
P11	$E_{Activie}$ activities information
P12	$E_{Passive}$ activities information
P13	Appliance power rating information from EIA
P14	Policy engine describing 'What-If' scenario for optimizing consumption pattern

Table 4.1: Residential energy demand modeling framework's parameters description

- **Module 2:** This module maps the commercial buildings characteristics onto synthetic population's commercial locations
  - **Input Parameters:** CBECS's buildings characteristics (P2); Synthetic population's commercial location details (P3)
  - **Output:** Each commercial location present in the synthetic population is embedded with a matching CBECS-commercial building's characteristics
  - **Customization Needed:** The common demographic details used for matching are stored in a configuration file and it can be modified.

- **Module 3:** This module constructs the occupancy rate for each commercial location present in the synthetic population (P4)
  - **Input Parameters:** Synthetic population's individual activity details, purpose of their visit, location information;
  - **Output:** Each commercial location in the synthetic population is embedded with the hourly occupancy information
- **Module 4:** This module constructs the hourly energy demand for each commercial location present in the synthetic population
  - **Input Parameters:** Synthetic population's location with their derived characteristics from the CBECS's building mapping (P6), location's occupancy (P7); Appropriate statistical model representing the building's category (P5)
  - **Output:**Constructed hourly energy demand profile for each commercial location
- **Module 5:** Constructs overall energy demand profile by aggregating demand profiles of all commercial locations (P8);
  - **Input Parameters:** Energy demand profiles of all synthetic commercial locations; Optional policy module information
  - **Output:** Aggregate commercial buildings' energy demand profile

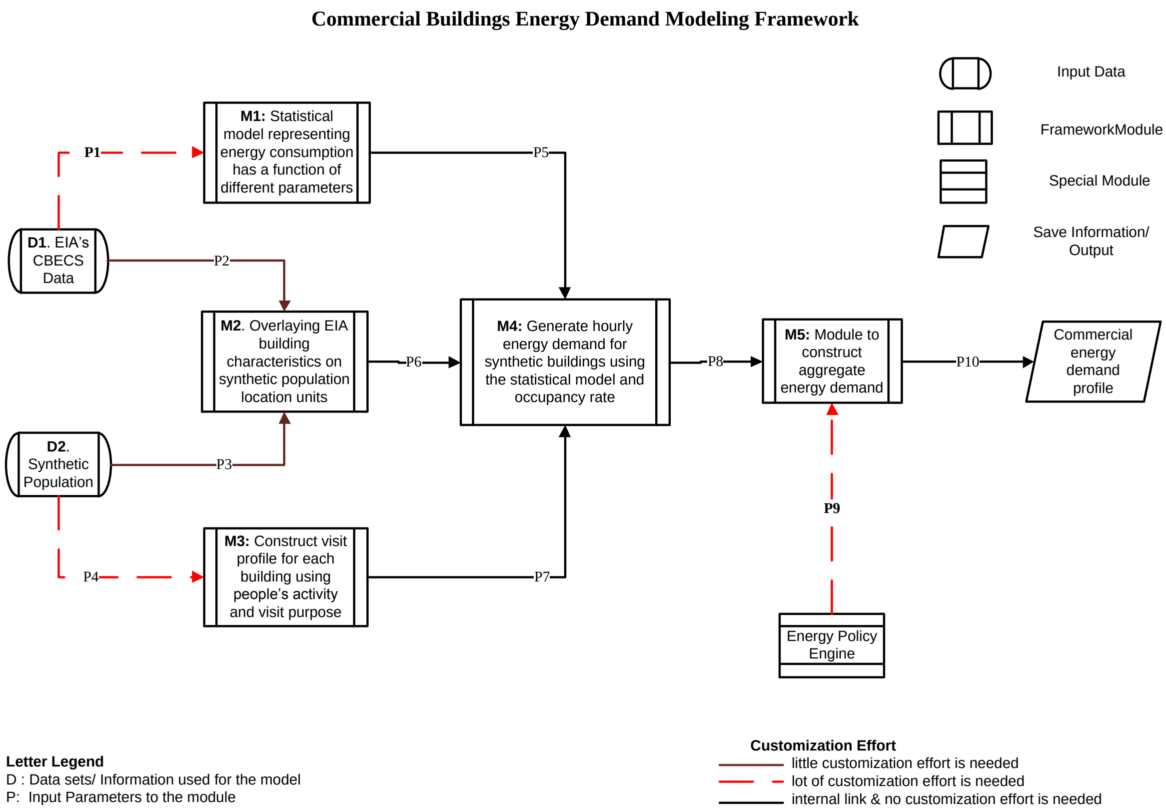

Figure 4.3: A software framework to generate energy demand profile for commercial buildings

4.2 Web-based Energy Demand Modeling Application: A Prototype

Increasing interests in optimizing the current energy systems in both academia and industries has motivated us to develop a web-based system that can expose the benefits of our energy demand-modeling framework. Towards that end, we built a prototype web based system to provide access to our generated energy demand profiles and conduct different analysis. The prototype is built as a J2EE web application and uses Apache Struts framework [53]. We use Enterprise Java Beans (EJB) to interact with our modeling framework to run any new experiment and to retrieve any previously generated demand profile. Figure 4.4 shows the logical block diagram of our prototype. The prototype consists of three layers presentation layer, logic or business layer and data layer. The block diagram shows the flow of information from presentation to data layer via the business layer. The retrieved results are encapsulated into XML structure, this makes it easy to integrate with or migrate to any front end system. Figure 4.5 shows the screen shot of our prototype application in which the front end is developed using Adobe Flex [54].

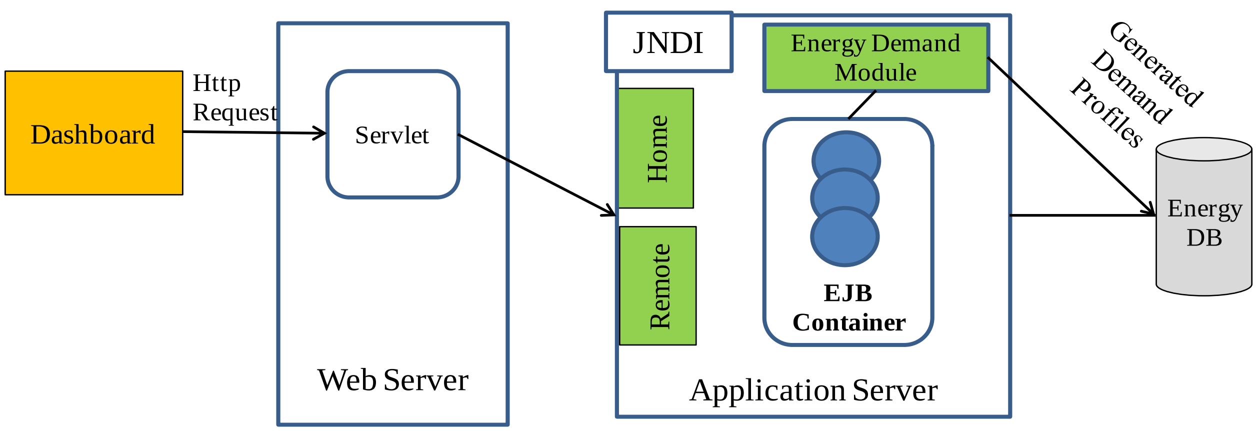

Figure 4.4: Logical block diagram of our prototype system. The diagram illustrates how the three different layers communicates

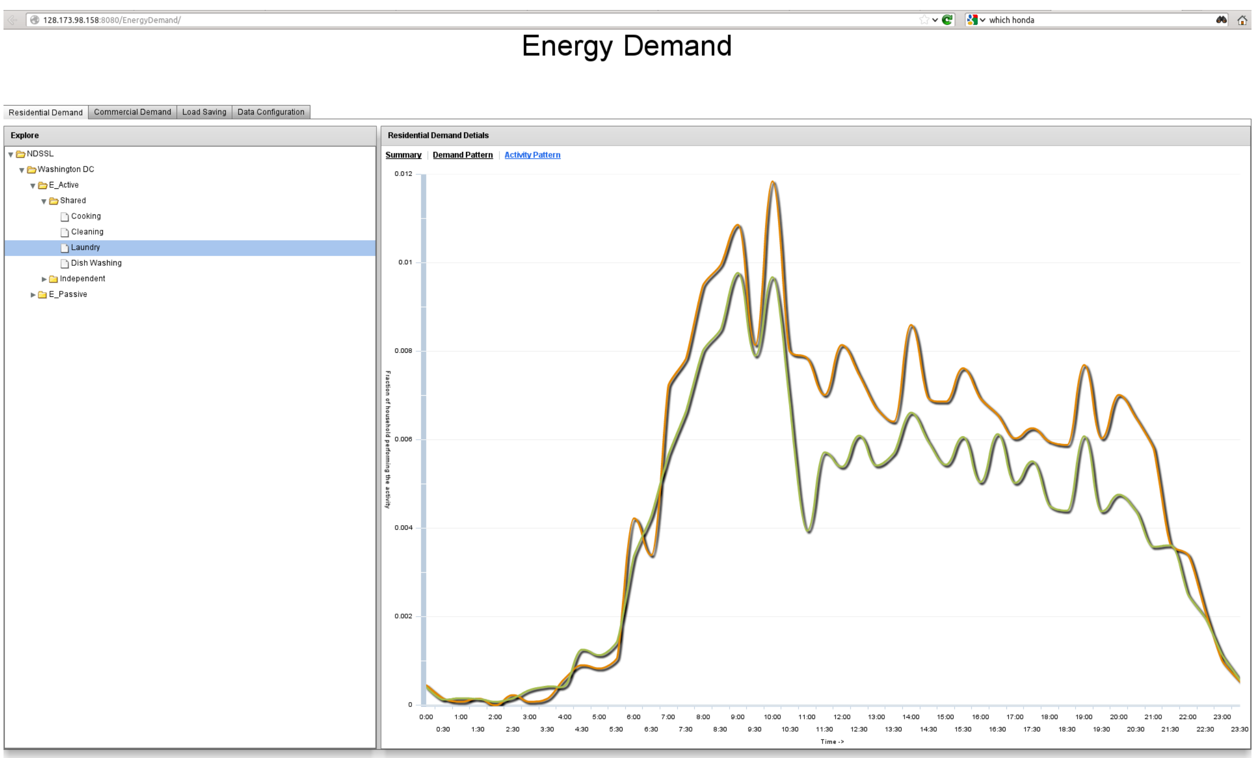

Figure 4.5: Screen-shot of web-based energy demand modeling system (Prototype Version)

Chapter 5

Conclusions, Discussion, and Future work

In this thesis, we presented a highly dis-aggregated energy demand-modeling framework that estimates energy demand profiles based on individual-level and building-level energy-consuming activities. The modeling framework generates energy demand profile at a regular basis by taking into account the physical, behavioral, economical and social factors affecting the energy consumption. The residential energy demand model associates appliance usage for each household activity and calculates energy consumption based on the appliance energy rating and duration of activity. It uses this information to generate a building-level energy demand profile at highly dis-aggregated level. In the residential energy demand model, we provide a novel way to resolve correlational and consistency problems in the generation of individual-level and household level ”shared” activities that occur due to household members' interactions. The commercial energy demand model derives occupancy profiles for different commercial buildings based upon the individuals' activities and their associated locations. It then models how the number of occupants at any period of time contributes to the rise and fall in building's energy consumption. It also incorporates statistical models to capture all the significant non-occupancy-related factors that influence the buildings' energy consumption. This modeling framework can be used to evaluate "What-if" scenarios in urban-area planning, demand-side management, etc. It can also be used to identify potential areas for improvement and quantify the impact of any energy policy or strategy.

5.1 Future Work

In this thesis, we present a methodology to generate a highly dis-aggregated energy demand as a function of individuals' and buildings' activities. The constructed energy demand can complement many of the currently on-going energy systems research. Some possible directions in which this work can be extended are:-

1. Electric Vehicles: The generated energy demand profiles from our model can be used to effectively place the charging stations for electric vehicles. Since, our model captures the visit profile of individuals (economic status, purpose of visit) visiting the locations, their duration of stay, location's spatial information and etc will be helpful in placing the charging stations.

2. Behavior analysis: Some of characteristic features captured by our model will be helpful

in evaluating different behavior oriented What-IF scenarios like peer pressure, adapting to a energy policy, upgrading to energy efficient appliances and etc.

3. Comparative studies: In chapter section 2.4.2, we analyzed the characteristics of energy consumption on household demographics. This work can be extended to study how demand varies across different states (within US), regions, cultures and countries. This kind of comparative study allows us to identify the main influencing factors affecting the energy consumption and help us devise energy policies that make consumption more efficient.

Bibliography

[1] U.S. Energy Information Administration (EIA) , "Annual energy outlook 2012 - with projections to 2035," June, 2012. [Online]. Available: http://www.eia.gov/forecasts/aeo/pdf/0383(2012).pdf

[2] G. Arnold, "Challenges and opportunities in smart grid: A position article," *Proceedings of the IEEE*, vol. 99, no. 6, pp. 922–927, 2011.

[3] U.S. Energy Information Administration, "Share of energy used by appliances and consumer electronics increases in u.s. homes." [Online]. Available: http://www.eia.gov/consumption/residential/reports/2009/electronics.cfm

[4] 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.

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

[6] 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. [Online]. Available: http://ideas.repec.org/a/eee/appene/v87y2010i6p1880-1892.html

[7] 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.

[8] J. Torriti, "Demand side management for the european supergrid: Occupancy variances of european single-person households," *Energy Policy*, vol. 44, pp. 199–206, 2012.

[9] U.S. Energy Information Administration (EIA), "Annual energy review 2011," September 27, 2012. [Online]. Available: http://www.eia.gov/aer

[10] P. Torcellini, M. Deru, B. Griffith, K. Benne, M. Halverson, D. Winiarski, and D. Crawley, "Doe commercial building benchmark models," in *Proceeding of*, 2008, pp. 17–22.

[11] M. Deru, K. Field, D. Studer, K. Benne, B. Griffith, P. Torcellini, B. Liu, M. Halverson, D. Winiarski, M. Rosenberg *et al.*, "US Department of Energy commercial reference building models of the national building stock," 2011.

[12] B. T. Griffith, N. Long, P. Torcellini, R. Judkoff, D. Crawley, and J. Ryan, *Methodology for modeling building energy performance across the commercial sector*. National Renewable Energy Laboratory, 2008.

[13] T. Sharp, "Energy benchmarking in commercial office buildings," in *Proceedings of the ACEEE 1996 Summer Study on Energy Efficiency in Buildings*, vol. 4, 1996, pp. 321–29.

[14] W. Chung, Y. Hui, and Y. M. Lam, "Benchmarking the energy efficiency of commercial buildings," *Applied Energy*, vol. 83, no. 1, pp. 1–14, 2006.

[15] *U.S. Energy Information Administration - Comerical Buildings Energy Consumption Survey (CBECS)*, 2003. [Online]. Available: http://www.eia.gov/consumption/commercial/data/2003/

[16] R. Subbiah, K. Lum, A. Marathe, and M. Marathe, "Activity based energy demand modeling for residential buildings," in *Innovative Smart Grid Technologies (ISGT), 2013 IEEE PES*, 2013, pp. 1–6.

[17] 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.

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

[19] 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. [Online]. Available: http://ideas.repec.org/a/eee/enepol/v29y2001i3p227-236.html

[20] 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. [Online]. Available: http://ideas.repec.org/a/kap/compec/v30y2007i2p125-142.html

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

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

[23] 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.

[24] "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. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364032108001949

[25] *American Time Use Survey*, 2010. [Online]. Available: http://www.bls.gov/tus/

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

[27] 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.

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

[29] *Buildings Energy Data Book.* U.S. Energy Information Administration, 2011.

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

[31] D. G. Kleinbaum, M. Klein, D. G. Kleinbaum, and M. Klein, *Logistic Regression*, 2010.

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

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

[34] National Climatic Data Center, "National Oceanic And Atmospheric Administration," 2011.

[35] U.S. Energy Information Administration, "Heating fuel comparison calculator (excel)," 2012. [Online]. Available: http://www.eia.gov/neic/experts/heatcalc.xls

[36] U.S. DOE - Energy Efficiency and Renewable Energy, "Energy cost calculator for electric and gas water heaters," 2011.

[37] 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

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

[39] R. Subbiah, K. Lum, A. Marathe, and M. Marathe, "A high resolution energy demand model for commercial buildings," Accepted and to appear in the Proceedings of ETG Congress 2013,"Security in Critical Infrastructures today", Berlin, Germany 2013.

[40] E. Efficiency, "Buildings energy data book," *Building Technologies Program, US Department of Energy, Washington, DC*, 2011. [Online]. Available: http://buildingsdatabook.eren.doe.gov/default.aspx

[41] S. Jebaraj and S. Iniyan, "A review of energy models," *Renewable and Sustainable Energy Reviews*, vol. 10, no. 4, pp. 281–311, 2006.

[42] C. Liao and P. Barooah, "An integrated approach to occupancy modeling and estimation in commercial buildings," in *American Control Conference (ACC), 2010*. IEEE, 2010, pp. 3130–3135.

[43] Y. Agarwal, B. Balaji, R. Gupta, J. Lyles, M. Wei, and T. Weng, "Occupancy-driven energy management for smart building automation," in *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, ser. BuildSys '10. New York, NY, USA: ACM, 2010, pp. 1–6. [Online]. Available: http://doi.acm.org/10.1145/1878431.1878433

[44] R. Hocking, "The analysis and selection of variables in linear regression," *Biometrics*, vol. 32, pp. 1–49, 1976.

[45] J. F. Lawless, "Negative binomial and mixed poisson regression," *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, vol. 15, no. 3, pp. pp. 209–225, 1987. [Online]. Available: http://www.jstor.org/stable/3314912

[46] W. Gardner, E. P. Mulvey, and E. C. Shaw, "Regression analyses of counts and rates: Poisson, overdispersed poisson, and negative binomial models." *Psychological bulletin*, vol. 118, no. 3, p. 392, 1995.

[47] P. D. Allison, "Testing for interaction in multiple regression," *American Journal of Sociology*, pp. 144–153, 1977.

[48] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2012, ISBN 3-900051-07-0. [Online]. Available: http://www.R-project.org/

[49] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S*, 4th ed. New York: Springer, 2002, iSBN 0-387-95457-0. [Online]. Available: http://www.stats.ox.ac.uk/pub/MASS4

[50] R Development Core Team, *R Documentation: Factors*, R Foundation for Statistical Computing. [Online]. Available: http://stat.ethz.ch/R-manual/R-devel/library/base/html/factor.html

[51] S. Urbanek, *Rserve: Binary R server*, 2013, r package version 0.6-8.1. [Online]. Available: http://CRAN.R-project.org/package=Rserve

[52] S. W. Ambler, *UML sequence diagram, The Elements of UML (TM) 2.0 Style.* Cambridge University Press, 2005.

[53] T. Husted, C. Dumoulin, G. Franciscus, D. Winterfeldt, and C. R. McClanahan, *Struts in Action: Building web applications with the leading Java framework.* Manning, 2003.

[54] D. Gassner, *Flash Builder 4 and Flex 4 Bible.* Wiley, 2010, vol. 688.