University of Louisville

ThinkIR: The University of Louisville's Institutional Repository

Electronic Theses and Dissertations

8-2020

Mathematical models and analysis for demand side management in residential electricity distribution networks.

Arnab Roy
University of Louisville

Recommended Citation

Roy, Arnab, "Mathematical models and analysis for demand side management in residential electricity distribution networks." (2020). *Electronic Theses and Dissertations.* Paper 3504.
https://doi.org/10.18297/etd/3504

MATHEMATICAL MODELS AND ANALYSIS FOR DEMAND SIDE

MANAGEMENT IN RESIDENTIAL ELECTRICITY DISTRIBUTION

NETWORKS

Arnab Roy
B.Tech., West Bengal University of Technology, 2012
M.Tech., Indian Institute of Technology, Roorkee, 2014

Dissertation submitted to the Faculty of
the J.B. Speed School of Engineering,
of the University of Louisville
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy
in
Industrial Engineering

Department of Industrial Engineering,
University of Louisville
Louisville, Kentucky,

August, 2020

MATHEMATICAL MODELS AND ANALYSIS FOR DEMAND SIDE

MANAGEMENT IN RESIDENTIAL ELECTRICITY DISTRIBUTION

NETWORKS

Arnab Roy
B.Tech., West Bengal University of Technology, 2012
M.Tech., Indian Institute of Technology, Roorkee, 2014

Dissertation Approved on

July 14, 2020

By the following Dissertation Committee

_____

Dr. Lihui Bai, Chair

_____

Dr. Monica Gentili

_____

Dr. Jason Saleem

_____

Dr. Michael McIntyre

ACKNOWLEDGEMENTS

ABSTRACT

MATHEMATICAL MODELS AND ANALYSIS FOR DEMAND SIDE
MANAGEMENT IN RESIDENTIAL ELECTRICITY DISTRIBUTION
NETWORKS

Arnab Roy

July 14, 2020

Development of smart grids along with communication technologies have led to the increased attention and adoption of demand side management (DSM) in the residential sector. Among various DSM schemes, demand response (DR) is a market-based mechanism to shave peak electricity consumption at the system level. In the past decade, the academia has seen a growing literature studying load management methodologies for residential consumers. A typical demand response program has three important facets: the energy cost, comfort of the consumers and overall system efficiency. In this dissertation, we investigate and develop models for effective load control to minimize energy cost and for understanding electricity consumer behavior so as to best design DR schemes. Participation in a real-world field demonstration not only stimulated our motivation for these studies, but also provided us with real-world data to validate the developed models and analyses. This in fact makes the dissertation distinct from current academic literature.

We first develop a control algorithm for Heating Ventilation Air-Conditioning (HVAC) systems in households during a peak period. The dynamic programming based model can determine the optimal temperature set-points of a thermostat given the lower and upper limits of temperature that household feels comfortable and the desired duration of the control. The temperature limits act as a quantitative metric for the comfort level of consumers. The objective is to minimize the energy consumption. The model is particularly suitable for DR programs with critical peak pricing, in which a higher electricity rate occurs during the peak period. When deployed separately during the peak and adjoining two periods, the model can keep the inside temperature within the given limits while consuming minimal energy during the peak period. This ensures that the HVAC system would have minimal usage during the peak period as the temperature is kept within the limits. In addition, we show that alternative start and end times of the control algorithm can be tested for each home. Analyses of the alternative options provide us with information about the insulation of the building. We perform computational experiments with real-world data to show the efficacy of the proposed methodology.

Second, we propose a mixed-integer linear fractional programming (MILFP) model to optimally deploy the dynamic programming based HVAC controllers among a pool of homes in a staggered fashion. Doing so, the model aims to flatten the demand curve over time thus maximizing the load factor for the entire distribution network. In addition, we develop a reformulation of the MILFP model into an MILP model which significantly reduces computational time for medium-scale instances. Furthermore,

for large-scale instances, excessive computational times by general purpose solvers motivate us to develop a customized bi-section search algorithm. Our extensive computational experiments conclude that the customized algorithm is able to solve real-world as well as randomly generated instances in reasonable CPU times.

In another effort, we study the behavior of consumers when subject to dynamic pricing under a DR program. We model the price-responsive behavior with utility functions and develop a bi-level programming model to estimate the coefficients of such a function utilizing consumption data from advanced metering infrastructure (AMI) from the field demonstration project mentioned previously. The upper level objective is to minimize the estimation error between the measured data and the optimum consumption while the lower level is for each household/consumer to maximize their total utility of energy consumption. We propose a trust-region algorithm to solve the non-linear bi-level utility estimation (BLUE) model after employing linear and quadratic approximation for the upper and lower level objective function, respectively. A mathematical property of the optimal solution is exploited to develop a cut that has significantly improved the computational time. Numerical experiments with real world data are conducted to validate the proposed models. In addition, we show the strong positive correlation between the utility coefficients and the widely used price elasticity property.

Finally, this dissertation also presents several empirical models to assess the effect of smart technologies on electricity consumption under a demand charge dynamic pricing rate. The models developed here were being utilized in the aforementioned

demand response pilot study. We present a statistical test based model to estimate the change of coincident load of residential consumers with the installation of efficient appliances including heat pump water heaters, smart thermostats, and battery storage units. The method utilizes a day matching algorithm to pair days with similar weather conditions. The consumption data from the two paired up days are used to conduct a paired t-test to evaluate the statistical significance of the changes. The results reveal that insulation plays an important role in energy savings along with battery systems.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

In modern society, electricity is a driver for economic activity and significantly influences the growth of a nation. The power grid provides the necessary infrastructure for the delivery of electric power from the generating plants to the consumers. It is an intricate network of generation plants, transmission lines, transformers, substations and more. The power grid in the U.S. have gone through a series of improvements throughout the decades and in its modern form connects approximately 150 million consumers with more than 3000 utility companies. Alongside the grid, the U.S. electricity market has also evolved substantially through deregulation and restructuring aimed towards efficiency and reliability. Demand has also surged with the proliferation of electrical devices in our daily lives. In the 21st century, we live in a digital world surrounded by connected devices. At the same time society has become more aware of the environmental impact of fossil fuels to generate electricity. This has resulted in generation diversification and efforts to minimize the carbon footprint of the industry through energy conservation and optimal deployment of assets. The power grid infrastructure is in need for an overhaul to accommodate these changes and also support the growth of digital technology dependent on it.

A primary concern in the development of grid infrastructure is the issue of reliability of service. This stems from the fact that electrical energy is expensive to store with current technology and hence a balance between supply and demand has to be maintained at all times. The issue is particularly magnified during peak periods of the system which generally occurs 5% of time in a year. Generating plants tackle this by maintaining expensive auxillary units and the high cost is reflected in the electricity spot markets. The situation will only be exacerbated as more of the residential sector adopts electric Heating Ventilation and Air Conditioning (HVAC) systems over natural gas ones (U. S. Energy Information Administration, Annual Energy Outlook, 2020). In response to extreme weather conditions, heavy use of HVAC systems for space cooling/heating inside buildings is the most influential factor in assessing and managing systems peak load (Samadi et al. (2010),Muratori et al. (2013)).

A particular way to address the aforementioned concerns is through real-time communication between the consumer and the provider. The next-generation electricity grid is enabled with a two-way communication system between different entities and is equipped to cope with the increased demand and complexity of the digital age. The new technology, named smart grid, has enabled utility companies to better manage the load with minimum disruption to the consumer. Modern control and communication technologies have enabled the integration of distributed generation (DG) into the system paving the way for renewable energy sources (RES) to play a vital part in the reliability of the grid (Calderaro et al., 2011). Understandably, the academia has been

greatly interested in investigating and developing smart grid solutions. Research in this field has been broad including simulation framework for different scenarios and optimization models to increase the security of the system. Bigerna et al. (2016) provides a detailed review of the pertinent works in this area.

The development of advanced metering infrastructure (AMI) is at the heart of the new communication system between the consumer and the utility and have enabled the latter to remotely monitor and measure usage, connect or disconnect the service, detect malpractice and outages, and track voltage fluctuations. In particular, AMIs have equipped the utility companies to introduce load management programs in the residential sector. The residential sector in the U.S. accounted for about 21.21 quadrillion Btu of electric energy in 2019 (U. S. Energy Information Administration, Annual Energy Outlook, 2020) or about 20% of the total generating capacity.

Utilities pay a premium in the electricity market to satisfy the demand during the peak periods. As generation methods gets diversified with the addition of RES it will get difficult to modulate the supply to follow a certain pattern according to demand (Strbac, 2008). Maintaining reserves to supply during peak periods results in significant increase of the cost. Smart grid enables the utility firms to shift from the traditional approach of supplying all the required demand even at a premium to controlling the demand by engaging the consumers to respond to the state of the system. Demand side management (DSM) refers to load management techniques carried out with the objective of balancing the energy supply and demand during peak periods. Objective of such programs is to shape the load curve as desired and

can be accomplished through load shifting, valley filling, peak reduction and more (Gelazanskas & Gamage, 2014).

One way to accomplish the balance between supply and demand is to de-incentivize the consumption during peak periods. Demand response (DR) schemes aim to achieve that by transferring the non-uniform rate paid by the utilities in the electricity market to the end-users through dynamic pricing models during system peaks or when the system reliability is compromised. Albadi and El-Saadany (2007) lists three ways in which the utility companies can achieve an intentional modification of demand through customer response. The consumer can reduce their peak period consumption in response to the higher price at the cost of sacrificing their comfort or they can shift the load to off-peak periods. The latter can however be done only for certain activities such as dishwashers or washer and dryers. Thirdly, the consumers can also participate in the generation side by using small generating capabilities such as solar power or batteries and can supply power to the grid after satisfying their own demand. A number of time separated pricing programs have been proposed and implemented: real-time pricing (RTP), critical peak pricing (CPP), day ahead pricing (DAP), time-of-use pricing (TOU) etc.

Incentive based programs (IBP) are a classical component of market based DR strategies. These programs can be executed in tandem with dynamic pricing schemes such as critical peak pricing (CPP) where the consumers are subjected to a higher rate during the peak period. Smart grids have facilitated the ability to control consumer appliances by the utilities. Through direct load control (DLC) utilities can operate

remotely certain connected appliances such as HVAC systems and water heaters of participant households. The literature in this area includes the modelling of such programs for residential buildings as load management systems (see Section 2.1) with varied objectives.

This dissertation centers around models developed for the implementation and analysis of demand response programs such as DLC and CPP in the residential sector. The motivation of such a study stemmed from participation in a real-world field demonstration DR project for residential consumers in midwestern U.S. with a municipal utility company. In total 3000 homes participated in the project, of which, 330 homes were installed with various smart technologies such as connected thermostat and efficient heat pump water-heater systems. Real-time consumption could be monitored and collected in small intervals with AMIs installed in all of them.

Our first set of contribution is towards the development of a DLC framework for residential consumers considering both the comfort of consumers and reliability of the system. A novel dynamic programming (DP) based model is proposed to determine the optimal temperature set-points of a thermostat during peak consumption period. The model is particularly suitable for IBP combined with CPP undertaken by utility companies. Extension to multiple homes enables evaluating alternative control durations for each household. This application can provide information about the building insulation with respect to feasible control durations. Next a mixed integer linear fractional programming (MILFP) model is developed for a control scheduling problem

(CSP), i.e., an optimal scheduling of the temperature set-point algorithm for a pool of homes. The objective is to maximize the system load factor to flatten the demand curve. An efficient reformulation-linearization method is proposed to transform the MILFP in to a mixed-integer linear programming (MILP) model for its efficient solution. However, for large instances, even the reformulation in conjunction with the state-of-the-art solvers such as CPLEX still encounter computational deficiency due to lack of memory. To overcome this obstacle, we develop a bi-section search algorithm to solve the MILFP model within a modest amount of time even for large instances. Furthermore, we develop a Lagrangian relaxation approach to estimate the upper bound for the nonlinear mixed integer problem in order to assess the optimality of the solutions from the proposed bi-section search algorithm. Overall, the temperature set-point algorithm wrapped by the MILFP model provides a DLC framework for utility companies seeking to maximize the load factor while not affecting the comfort level of consumers. Application of the proposed models on real-world data provides numerical evidence of their efficacy.

Second, we propose a novel bi-level optimization model to estimate the utility function coefficients representing price-responsiveness of electricity consumers in a dynamic pricing DR program. The upper level objective is to minimize the estimation error between the measured data and the optimum consumption while the lower level is for each household/consumer to maximize their utility of consuming energy. The model is developed at a per consumer level and we demonstrate that the coefficients of a utility function can be used to assess the price-responsiveness of DR consumers.

Furthermore, we propose a trust-region algorithm (TRA) for solving the bi-level utility estimation model, based on a computationally efficient reformulation. In addition, we show a mathematical property of the optimal solution, which can be used as a cut in the algorithm. Extensive computational experiments demonstrate the validity and efficiency of the proposed TRA, especially showing substantial reduction on computational time when cuts are added to the re-formulation of the bi-level program. Our post optimization analysis demonstrates a positive correlation between the utility coefficients calculated by the bi-level model and the commonly used price elasticity in economics, which further validates the bi-level modeling approach.

Our third set of contribution is towards the analyzing the effects of smart technologies and dynamic pricing, specifically CPP, on the consumption pattern of residential homes. A day-matching algorithm is developed to pair days with similar weather conditions. The matched days are then utilized in a paired t-test to compare the energy consumption as well as the coincident load before and after a DR program in a pilot community. Coincident load is defined as the consumption of a particular household during the peak period. The metric gives a measure of the contribution of each home towards the peak. These empirical models are developed in conjunction with the utility company for the DR project and have been applied to the collected real-world data.

We organize the dissertation as follows. Chapter 2 reviews the literature of the different aspects of DR that we have addressed in this work. Next, in Chapter 3, we present the DP based temperature set-point estimation problem (TSEP) along

with the analysis of application to multiple homes. Chapter 4 presents the MILFP

formulation for scheduling the TSEP for multiple homes. A bi-section search algorithm

is developed to solve the proposed MILFP model efficiently. Chapter 5 presents the

utility coefficients estimation problem for price-responsive consumers. An empirical

study of smart energy technologies in residential consumers is presented in Chapter 6.

In Chapter 7 we offer concluding remarks and proposed future extensions of the work.

CHAPTER 2

LITERATURE REVIEW

2.1 Residential load control and scheduling

A pivotal component of a DR scheme is the control strategy or the load scheduling program, which enables the end-users to make decisions on the usage pattern of appliances. Researchers from varied domains have investigated the topic with the operations research community having a significant contribution (Tsui & Chan, 2012). Operations Research in this domain is mainly applied to determine the optimal parameters such as power allocation, load scheduling, price signals, and temperature set-points under DR. The review of literature presented below is mainly concerned with the application of OR methodologies. Aside from several simulation studies (e.g., Kirschen et al. (2000), Venkatesan et al. (2012)), we mainly focus on relevant literature divided into two broad categories: exact optimization models and heuristics.

2.1.1 Exact optimization

Varied optimization techniques including dynamic programming, linear and non-linear programming, mixed integer programming and optimization under uncertainty have been used in control and scheduling decisions for DLC. A popular objective of

the developed models is the minimization of cost. Hsu and Su (1991) and Yang and Huang (1999) combine dynamic programming (DP) with unit commitment problem to minimize the system production costs. The latter work considers large air-conditioning units and incorporates fuzzy logic in the DP model to address uncertainty. Similarly, Huang et al. (2004) design a fuzzy DP model to schedule interruptible loads with the ability to modify them real-time due to forecasting errors through an adaptive control strategy. The objective of the model is to minimize the energy consumption during peak period. Laurent et al. (1995) model direct load control of electric water heater load to reduce the maximal peak load without shifting the peak. The overall problem is formulated as a linear program and is solved using a column generation technique. Ha et al. (2006) develop a DP based energy management system (EMS) to minimize the consumer cost considering total power constraints and the user comfort.

Another section of literature is dedicated towards DLC scheduling with particular attention to customer comfort. Pedrasa et al. (2009) model scheduling of interruptible loads through a multi-objective optimization problem concentrating on customer convenience solved with binary particle swarm optimization. Minimizing customer discomfort is the main objective in the DLC framework designed by Ramanathan and Vittal (2008) through a stochastic optimization problem. Rigorous sensitivity analysis is conducted with constraint sets and parameters to analyze demand-side management programs. Mohsenian-Rad and Leon-Garcia (2010) use linear programming to model the user behavior when balancing between minimizing cost and operation of appliances so that consumer comfort is not compromised. H. Chen et al. (2014) take a game

theoretic approach to model DSM scenarios for selfish consumers aiming to schedule their electricity load with instantaneous billing.

Some works have incorporated system features when designing DLC programs for residential buildings. Integration of DLC with interruptible load management is modelled by Huang and Huang (2004) which can provide reserves for ancillary plants. A fuzzy dynamic programming model is developed for the scheduling of DLC. A genetic algorithm is proposed by Yao et al. (2005) to determine an optimal DLC schedule so that the curtailed load at each interval is minimized to lower revenue loss of utility companies. Consumers are considered in groups and the shedding time for each such group is leveled. Setlhaolo et al. (2014) formulate the load scheduling problem with system constraints as a mixed integer nonlinear optimization problem with time-of-use as the pricing scheme. A specialized branch and bound method solves the model shaving 25 % of the cost.

A host of works such as Strickler and Noell (1988), Laurent et al. (1995) and Espinosa (1987) have studied the problem with the objective of peak load reduction. Ashok and Banerjee (2003) study co-generation of power from different sources under a load management program with an optimization framework. The model can reduce a significant amount of peak load from the system. Aghaei and Alizadeh (2013) and J. Chen et al. (1995) coordinate DLC with the unit commitment problem to reduce the peak load alongside reducing system costs. Wei and Chen (1995) study the energy allocation of air-conditioning system through a multi-pass DP model and apply the method in a real-world environment to reduce peak load.

Alongside deterministic models, scholars have also concentrated on the development of models incorporating uncertainty in different aspects. Conejo et al. (2010) schedule hourly load of consumers as a response to uncertain real-time pricing with robust optimization. The model is solved hourly with the objective to maximize the consumer utility. In Jia and Tong (2012) a two-stage stochastic optimization program aims to maximize the conditional value at risk. The effectiveness of adding renewable energy to the supplier side is demonstrated. Yi et al. (2013) apply optimal stopping rules to a stochastic optimization model to schedule appliances in real-time under demand response. The scheduling is done in two parts - scheduling of operation followed by the power allocation for every appliance. Dorini et al. (2013) study the response of residential customers to real-time pricing (RTP) in DR. They tackle the energy consumption as stochastic finite response models in a form of a chance constrained optimization.

While the above-mentioned works model the problem as a single level optimization, others have used bi-level models to represent the hierarchy in decision making in DR. For example, the Stackelberg relationship between the retailer and the consumer under a demand response program is modeled by Zugno et al. (2013) as a bi-level optimization program. The upper level as the retailer's problem of maximizing profits and the lower level as the consumer's problem of minimizing electricity costs and discomfort. Meng and Zeng (2013) use a Stackelberg game formulation to model the DR relationship between consumers and retailers. The leader's problem is solved by a genetic algorithm while the follower's linear programming problem with an analytical

method. In Safdarian et al. (2014), the upper level aims to flatten the total power profile at the system level while the lower level aims to minimize expense of the customers. A proposed distributed algorithm scheme is proposed to reach the Nash equilibrium. In Jalali and Kazemi (2015), residential load scheduling is modeled with multiple electricity suppliers (utility companies) as a two non-cooperative game, one for the supplier side and one for the customer side.

2.1.2 Heuristics

Due to high complexities of the DR optimization models, many researchers choose to employ metaheuristic and learning algorithms instead of formulation-based exact solution methods. Yao et al. (2005) develop a genetic algorithm (GA) to schedule the load of a group of customers. The algorithm is composed of several GAs working in succession and following a master problem. C.-R. Chen et al. (2013) uses a genetic algorithm to optimize the power usage by customers with a new power rate charging scheme. Ghazvini et al. (2015) design an incentive based DR program by modelling the retailer's decision-making problem as a multi-objective short-term optimization with GA as the proposed solution method. Integer genetic algorithm has also been used by Kinhekar et al. (2016) in their study of a DR scheme with the goal of fitting the energy consumption curve to a target curve set by the utility company.

Apart from genetic algorithms, other heuristics algorithms such as swarm optimization, search algorithms, learning algorithms are also used to schedule residential load under DR programs. For example, Armas and Suryanarayanan (2009) propose a heuristic

to schedule a distributed energy resource (DER) system consisting of photo-voltaic cells, storage systems, and the micro grid. The density estimating robust algorithm determines the control actions of the inverter connected to the grid. Logenthiran et al. (2012) use a heuristic based evolutionary algorithm (EA) while Iqbal et al. (2016) develop a binary particle swarm optimization for managing multiple devices. A multi-objective optimization problem in allocating daily energy for a distributed energy resource system is tackled with particle swarm optimization by Graditi et al. (2015).

In O'Neill et al. (2010) a novel learning algorithm using Markov decision processes is proposed for residential EMS. Q-learning schemes have been developed by Kara et al. (2012) and Wen et al. (2015) for optimal scheduling of thermostatically controlled loads. Dusparic et al. (2013) propose a multi agent based reinforcement learning approach to schedule loads from peak periods to periods with high availability of renewable sources. Finally, recent advancements in batch reinforcement learning is utilised in Ruelens et al. (2017) for scheduling thermostatic load such as heat pump water heater and HVAC.

2.2   Modeling of consumer behavior

While demand response programs are gathering momentum around the world, researchers and practitioners increasingly realize the importance of understanding electricity consumers' behavior. Under modern economics theory, consumers maximize their utilities in economics-related activities, whether on rational or bounded rational basis.

Hence, it is interesting to study the response of such consumers to dynamic pricing. Some researchers have used the hypothetical notion of utility maximization while others use increasingly available energy consumption data (e.g., AMI data, wi-fi connected devices data) to study consumer utility. For example, Conejo et al. (2010) employ the notion of maximizing consumers' total utility in optimal load scheduling in response to RTP. Li et al. (2011) propose a demand response strategy for different categories of appliances (including plugged-in hybrid vehicles) by maximizing the utility of each. A distributed algorithm is developed to determine optimal pricing assuming that each consumer maximizes their utility. They show that there exists an optimal price under which consumer equilibrium will yield system-wide maximum benefits. Jalali and Kazemi (2015) use consumer payoff models to develop a load-scheduling scheme involving more than one supplier. The model involves two games simultaneously: the supplier side and the customer side with a distributed algorithm to achieve Nash equilibrium.

The consumer utility company interaction is modeled as a reverse Stackelberg game by Ratliff et al. (2014). Their method determines the utility function of the consumers and suggests the incentives for changing the consumption behavior. On the other hand, Saez-Gallego et al. (2016) develop a market-bidding problem to determine the characteristics of a cluster of price-responsive consumers with external factors (e.g., weather conditions). The proposed bi-level model sets the upper level objective as minimization of the estimation error between the measured data and the optimum consumption, while the lower level objective as maximizing consumer utility. Finally,

an optimal pricing algorithm is developed by Samadi et al. (2010) that maximizes the utility of all the consumers while keeping the total consumption below generating capacity.

A certain section of the literature have concentrated on the price elasticity of residential consumers. A supply demand market model is developed by Bompard et al. (2007) to examine the effect of demand elasticity when there is strategic price bidding on the supply side to manipulate the market. The study reveals for low competition levels demand elasticity has a significant role to play on market outcomes. The profitabality of efficient generation plants increases with increasing elasticity and by providing incentives towards installing new technologies. Ericson (2011) studies the response of consumers when provided a choice between tariffs. Welfare measures are considered to develop a choice function and analysis reveals that consumption pattern does not have a significant influence in selection of the pricing rate. The study also shows that critical peak pricing (CPP) rate is popular with consumers who have some form of energy management system installed. Impact of price eleasticity of demand by consumers is explored by Thimmapuram and Kim (2013) in an agent based simulation model of the Korean electricity market. The consumers equipped with smart grid technologies benefit with increasing price elasticity. Additionally, Kirschen et al. (2000) model consumer behavior with a matrix of elasticities between different consumers. The authors show the benefit of the information while scheduling generation and rates.

2.3 Energy consumption models

Proposed demand response frameworks in literature in general require a model for energy consumption by the HVAC system and the effect of change of consumption. For HVAC systems this effect can be noticed in the inside temperature of the building. Since this work is focused on the data from HVAC energy consumption, we focus this literature review on such models which use similar data or has similar objective. Such models can be broadly divided into two categories, physics based and data-driven methods. There are also grey-box techniques which employ data analysis to determine the parameters for the physics based modeling. Physics based models are mostly deductive and continuous while being static or dynamic in nature. Data-driven models are inductive and discrete and can be different data mining algorithms, statistical models, state-space, and stochastic models.

2.3.1 Consumption models

In most physics based models, thermal resistance networks are usually used to model the system. Electrical network components are then used to model the attributes of the HVAC system like resistors and capacitors. One important class of physics based models uses time dependent differential equations. For example, Tashtoush et al. (2005) model the HVAC system as a dynamic model and then use it in a subsequent PID controller scheme. The inside temperature of the room is modeled as zone temperature using energy and mass balance equations. Muratori et al. (2012) present a thermodynamic model studying the energy consumption of the HVAC systems with

simulation results in MATLAB (Higham & Higham, 2016).

Another physics based method for modeling energy consumption of the HVAC system is through as an equivalent thermal (ETP) model. Katipamula and Lu (2006) study different HVAC control strategies for demand curtailment through simulating the HVAC system as an ETP model. Xu, Wang, and Huang (2010) use a lumped parameter model to describe the zone temperature of a building in a constant air volume (CAV) air-conditioning system. This model is then used in robust model predictive control scheme. On the other hand, Wemhoff and Frank (2010) has worked on modeling the HVAC system as a lumped parameter model which takes into account fluid and energy transport, and thermodynamics.

Data mining algorithms are being used extensively in HVAC modeling with the improvement of high performance computing. Kusiak and Li (2010) build predictive models for the cooling load of the air handling unit (AHU) using various algorithms including neural networks, boosting tree, support vector machines, and random forest with multi-layer perception neural networks being the most accurate. J. Chen et al. (2011) model the ground source heat pump through artificial nerve neural networks, which treats the system as a black box and considers the interactions between as inputs and outputs. Kusiak and Xu (2012) study the energy consumption of a HVAC system through a non-linear autoregressive model with dynamic ANNs. Tang et al. (2014) model the short term prediction of HVAC load using clustering algorithms and multi-layer perception neural networks. Lixing et al. (2010) and Xuemei et al. (2010) utilize support vector machines (SVM) to forecast building cooling load. In addition,

the former incorporates ant colony optimization to determine the SVM parameters while the latter use autoregressive integrated moving average model (ARIMA) to build a linear model alongside the non-linear SVM model.

In literature, statistical models for HVAC systems have consisted of multivariable regression models, autoregressive models such as ARIMA and ARX, finite impulse response models, and output error models. Among them, ARX and ARIMA considers both input and output information and thus is more suitable for a closed loop control system. Dong et al. (2013) examine the relationship between the AHU fan speed and the power consumption using a second order equation through curve fitting. Virk and Loveday (1994) develop a regression model of the room temperature based on weather factors including temperature and humidity, energy consumption, and zone temperature. J. Ma et al. (2011) develop auto-regressive exogenous (ARX) models for HVAC energy usage and temperature for different zones in the building; while Huh and Brandemuehl (2008) study the compressor power and capacity using autoregressive integrated models (ARI).

Finally, there are also works in the realm of stochastic models for HVAC systems. Zlatanović et al. (2011) develop a stochastic formulation of building fans using Gaussian probability density functions. They noted that pdf of a cooling process requires a considerable amount of data, which is often hard to obtain. Furthermore, Y. Ma et al. (2015) develop a stochastic model predictive control scheme using finitely supported pdfs from historical data for load forecasting.

2.4   Demand response empirical studies

We end this chapter with the literature on empirical studies conducted on demand response. Most of these works are conducted by government agencies to assess the viability of the program.

Works in this area consist of reports by different utility companies. The U.S. Department of Energy Office of Scientific and Technical Information have conducted multiple studies such as Shen et al. (2012) and Hale et al. (2018) to gauge the effect of demand response programs in electricity markets. The former study covers the effect of DR in high potential electricity markets with distributed resources while the second one looks into addressing the increasing energy demand through demand response. Stoll et al. (2017) simulate 14 different DR scenarios in the Florida market and conclude that such programs overall would reduce the cost of production by reducing the low-load hours of generators. A comprehensive study to assess the contribution of DR programs in the U.S. has been conducted by Cappers et al. (2010). They observe that participants in organized wholesale markets become better in load curtailment and cost savings with experience while less developed wholesale markets are interested in adopting the programs.

CHAPTER 3

DYNAMIC PROGRAMMING FOR TEMPERATURE SET-POINT ESTIMATION

3.1 Introduction

In this chapter, we focus on the development of a control scheme for HVAC systems for households. In pricing based DR programs such as CPP and TOU, consumers are charged a higher rate during the system peak. This period is determined through forecasting methods by the utility companies or other agencies and are conveyed to the customers ahead of time. For example, Figure 3.1 shows a typical time-of-use rate structure from the Pacific Gas and Electric Company in California. The period marked in yellow is the forecasted peak-period while the rest is off-peak period. Consumers are given incentives, e.g. monetary credits, to shift their load to the off-peak periods.

Figure 3.1: Example pricing structure from PGE Calif.

Forecasted peaks are typically between 3 to 5 hours and any consumption during this period is charged at a much higher rate than the standard off-peak rate. Residential customers can reduce the penalty of higher cost by keeping their electric consumption close to the baseline referred to as the non-shiftable loads such as refrigerators, lighting, television etc. In certain DLC programs, utility companies can manage the consumption through load control algorithms executed via connected devices. To this end, HVAC systems are particularly suitable. This is because, on the one hand, they contribute significantly to the total energy consumption and on the other hand, their thermal storage property enables them to be controlled by maintaining the desired temperature requirements. However, user's comfort is an important trade-off here as it is closely related to the preferred inside temperature of the household.

One way to keep the inside of a home around the preferred temperature for a certain period with the minimal usage of air-conditioning is pre-cooling. Pre-cooling is referred to cooling the home before the peak period during summer to the lowest preferred temperature such that minimal energy is consumed during the peak period. This method has been utilized in many works such as Yin et al. (2010) and Katipamula and Lu (2006). However, the optimal thermostat set-points has to be determined so that the temperature constraints are not violated, and the energy consumption is kept at a minimum. Predictive control schemes such as Model Predictive Controllers (Mayne et al. (2000)) usually use a reference temperature point in their designs. Multiple studies have used a static set-point, the customer preferred temperature as the reference point. Avci et al. (2013a) have used a simple algorithm related to Real

Time Pricing to estimate the optimal set-points based on customer preferences.

We propose a dynamic programming (DP) based temperature set-point algorithm for thermostats. The algorithm ensures that the inside temperature is maintained within some preferred bounds while minimum energy is being consumed. We begin with the statistical learning models for the residential HVAC systems employed in the optimization problem. Subsequently, we present the temperature set-point estimation problem (TSEP) as a DP model. The model has the provision for testing alternative start times for the control actions. The feasibility of such alternative start times can be used to identify the insulation quality of the building. Finally, we present the numerical experiments conducted with real-world data to demonstrate the advantage of using such a DLC framework.

3.2 Models and formulations

3.2.1 Statistical learning models for HVAC energy consumption

The temperature set-point estimation problem (TSEP) in Section 3.2.2 requires predictive models for HVAC power and inside temperature of the residence. In literature, there are numerous works on the development of sophisticated HVAC models for prediction purposes (see Section 2.3). Consumption of electricity inside the residence is dependent on multiple factors, some which are general (e.g. weather factors such as temperature, dew-point, humidity) while others are specific (e.g. area of the household, insulation, and consumer schedule). In this case, the TSEP (Section 3.2.2) utilizes a parametric prediction model (James et al., 2013) for both HVAC average

power and inside temperature of the residence. As such, there is no preference to the type of model used; parametric models comprise of a large family of learning models such as linear regression, naïve Bayes, perceptron, and neural nets with fixed architecture and any one of them can be used interchangeably as per need and performance.

After careful analysis of the data in our real-world study, we have considered linear regression to be the learning method of choice to predict the HVAC energy in each time period. Difference between the inside temperature and the temperature set-point of the HVAC system is considered as an independent variable in the model along with outside temperature for each period. In Section 3.3.1, we present the numerical evidence for the choice of such variables along with the prediction accuracy. The motivation is to build a basic model which can be developed from the data readily available. We want to stress that, functionality of the control method does not depend on the learning models as long as they are parametric and we welcome any extension of this work by using different modeling techniques for the HVAC load including adding uncertainty about weather.

t : *Time period t*

T_i^t : *Inside Temperature of the home between $t - 1$ and t*

T_o^t : *Outside Temperature between $t - 1$ and t*

T_s^t : *Thermostat Set-Point Temperature at time t*

P^t : *Average power consumed by the HVAC unit between $t - 1$ and t*

The predictive model for the HVAC average power can be written as:

$$P^t = \alpha_0 + \alpha_1(T_i^t - T_s^{t-1}) + \alpha_2 T_o^t \tag{3.1}$$

where $\alpha_{0,1,2,3}$ are the linear regression parameters determined from the collected data.

The second predictive model utilized by the TSEP is for the inside temperature of the household at each time period. The DP model for TSAP requires a recursive equation to determine the state of the system, in this case, the inside temperature (see Section 3.2.2). To this end, we observe that inside temperature of a household at any time period is significantly correlated to the temperature in the previous periods (Section 3.3.1). Auto-regressive model with exogenous variables (ARIMAX) (Cui & Peng, 2015) particularly suits our purpose. Outside temperature and HVAC average power are considered as the exogenous variables. We present the numerical justification of such choices and the accuracy results in Section 3.3.1. Again, in this case also, any parametric model would suffice for TSEP. Using the notations defined above, the inside temperature ARIMAX model with a constant term, a single lag-term, and parameters $\gamma_{0,1,2,3}$ can be defined as:

$$T_i^t = \gamma_0 + \gamma_1 T_i^{t-1} + \gamma_2 T_o^t + \gamma_3 P^{t-1} \tag{3.2}$$

We proceed to defining the TSEP for a single household and it's property to test alternative control start times.

3.2.2 Temperature set-point estimation problem

Every unique home has a preferred inside temperature which can be taken as a metric for satisfaction. Considering a range of $\pm k$ from the preferred, we keep the inside temperature in the model to be constrained between $T_p + k$ and $T_p - k$. To achieve this goal during the peak-period the home has to be pre-cooled first. Let $\mathbb{T}^{-1}$ and $\mathbb{T}^0$ be the defined as the set of the intervals of pre-cool and the peak periods respectively. In addition, $\mathbb{T}^{+1}$ is taken as set for intervals post-cool period after the peak where the HVAC system is still controlled with the optimal set-points to avoid shifting the peak. Hence, $\mathbb{T} = \bigcup_i \mathbb{T}^i$ for $i = \{-1, 0, +1\}$ are regarded as the DLC periods of the HVAC system.

We design the control problem as an optimization model with the objective of minimizing the HVAC energy consumption while keeping the inside temperature within the preferred bounds. Dynamic programming is particularly suitable because the overall problem for each control-period can be decomposed into sub-problems for each time interval. The decision variable in our model is the thermostat set-point temperature at time interval t, T_s^t. The state variable is the inside temperature of the home at t, T_i^t, the disturbance is the outside temperature T_o^t while each stage is t. Table (3.1) list the notations used in the DP model along with the details of the model. Preferred temperature T_p is used to determine the bounds for the inside and set-point temperature.

The objective function of the model is to minimize the total energy consumed

Table 3.1: Temperature set-point estimation problem notation

$Notation$	Description
t	Time interval
T_s^t	Thermostat set-point temperature (Decision Variable)
T_o^t	Outside temperature between time $t-1$ and t (Disturbance)
T_i^t	Inside temperature between time $t-1$ and t (State Variable)
P^t	Average power consumed by the AC between time $t-1$ and time t
J_t^*	Optimum energy consumed by the AC between time 0 and time t
T_p	Preferred temperature of the consumer
$\mathbb{T}$	Control period of time intervals t

by the HVAC until time T. The objective at time t is given by the Equation (3.3). On the boundary at $t = T$ i.e. at the end of the period, the objective measures the minimum energy consumed over the entire control period.

$$J_t^*(T_i^t) = \min_{T_s^t}(P^t + J_{t-1}^*(T_i^{t-1})) \quad \forall t \in \mathbb{T} \tag{3.3}$$

The average power at each time interval t is given by the Equation (3.1) while the recursive relationship for the state variable T_i^t is expressed by the Equation (3.2). T_i^t is constrained within T_{max} and T_{min} which can be defined as $\pm k$ from T_p. The temperature set-point estimation problem (TSEP) is defined as:

$$TSEP: \quad J_t^*(T_i^t) = \min_{T_s^t} \left(P^t + J_{t-1}^*(T_i^{t-1}) \right)$$

subject to :

$$T_i^t = \gamma_0 + \gamma_1 T_i^{t-1} + \gamma_2 T_o^t + \gamma_3 P^{t-1}, \forall t \tag{3.4}$$

$$P^t = \alpha_0 + \alpha_1(T_i^t - T_s^{t-1}) + \alpha_2 T_o^t, \quad \forall t \tag{3.5}$$

$$T_{min} \le T_i^t \le T_{max}, \forall t \tag{3.6}$$

$$T_{min} \le T_s^t \le T_{max}, \forall t \tag{3.7}$$

The recursive equation (3.4) characterizes the state of the system at every period while (3.5) calculates the average power consumed at the particular state. The above DP model is applied in three stages of the control period: $\mathbb{T}^{-1}$, $\mathbb{T}^0$, and $\mathbb{T}^{+1}$. In the pre-cooling period, initial inside temperature is assumed to be at the preferred point T_p. From this point, the inside temperature is allowed to drop to $T_p - k$ which becomes T_{min} in this case while T_{max} remains equal to T_P. After pre-cooling the home, temperature is allowed to rise to $T_p + k$ which is the T_{max} and $T_p - k$ is the T_{min}. The third stage is the post-cooling phase. In this stage, the temperature is brought back to the preferred point after being allowed to rise in the peak period. So, $T_p + k$ become T_{max} and T_p is the T_{min}.

3.2.3   Alternative control durations

The TSEP is applied in 3 separate control periods consequetively in the proposed DLC framework. Let $\mathbb{T}^{-1}$, $\mathbb{T}^0$, and $\mathbb{T}^{+1}$ be the set of intervals of the pre-cool, peak, and

post-cool period respectively. Thus, the total DLC period can be defined as $\mathbb{T} = \bigcup_i \mathbb{T}^i$ for $i = \{-1, 0, +1\}$ and $\mathbb{T}^i \cap \mathbb{T}^j = \varnothing$ for $i, j = \{-1, 0, +1\}$ and $i \neq j$. Let the total number of intervals in the three periods be T^{-1}, T^0, and T^{+1} repectively. Therefore, we can define the the three periods as: $\mathbb{T}^{-1} = \{1, ..., T^{-1}\}$, $\mathbb{T}^0 = \{T^{-1}+1, ..., T^{-1}+T^0\}$ and $\mathbb{T}^{+1} = \{T^{-1} + T^0 + 1, ..., T^{-1} + T^0 + T^{+1}\}$.

The primary goal of application of TSEP is to lower the inside temperature to $T_p - k$ and to T_p in the pre and post-cool period respectively by consuming the least amount of energy. In the case of some homes, this can be achieved within less than the allocated time intervals in the control periods of $\mathbb{T}^{-1}$ and $\mathbb{T}^{+1}$. In such cases, the TSEP can be started from an time interval later than 1 in case of pre-cool and can be stopped at an time interval earlier than T in the post-cool. However, the lowest preferred temperature cannot be reached for some of the start time or end time intervals as the resulting control period would be too short. Hence, those start times are deemed as infeasible and can be determined by the infeasiblity of TSEP due to constraints (3.6) for such a control period. Each home h will have an unique set of feasible start and end time intervals denoted by $\mathbb{S}_h$ and $\mathbb{E}_h$ for pre and post-cool respectively where $\mathbb{S}_h \subset \mathbb{T}^{-1}$ and $\mathbb{E}_h \subset \mathbb{T}^{+1}$. However, every home needs a minimum number of time intervals to cool down. Let, t^0 be the number of intervals in the minimum required period. Therefore, the latest feasible start time of TSEP in the pre-cool period would be $T^{-1} - t^0$ and the earliest end time would be $T^{-1} + T^0 + t^0$.

Based on the size of $\mathbb{S}_h$ and $\mathbb{E}_h$ a home h can have a number of alternative feasible control durations. Analysis of feasible control durations provides us information about

the insulation condition of the building. Homes which can be cooled down in fewer time intervals have better insulation since the heat from outside has little effect on the cold air inside. Indeed, in Section 3.3.3, we show the correlation between the feasible options of homes and their insulation quality.

Application of the TSEP is independent of each household. But solving the problem for multiple home simulataneously can shift the peak from the event period to the adjoining pre-cool and post-cool periods. This can happen when multiple homes are being cooled simultaneously for the same duration. Hence, application of TSEP should be considered on a system-wide basis rather than on an individual household level. This raises the question of selection among the feasible control durations for each household such that the system reliability is not compromised. We address this concern by proposing a model for optimal scheduling of TSEP for multiple homes in the next chapter.

3.3 Computational results

3.3.1 Statistical learning models for HVAC energy consumption

Data from a demand-response field project with a municipal utility company in midwest U.S.A are used to conduct numerical experiments. The undertaking involved installing approximately 300 consumers with advanced metering infrastructure (AMI) and subjecting them to dynamic pricing. The residents were intimated beforehand of a 3 hour window where the electricity price would be significantly higher ($11.65 per kWh) than the standard rate they are accustomed to ($0.0065 per kWh). The
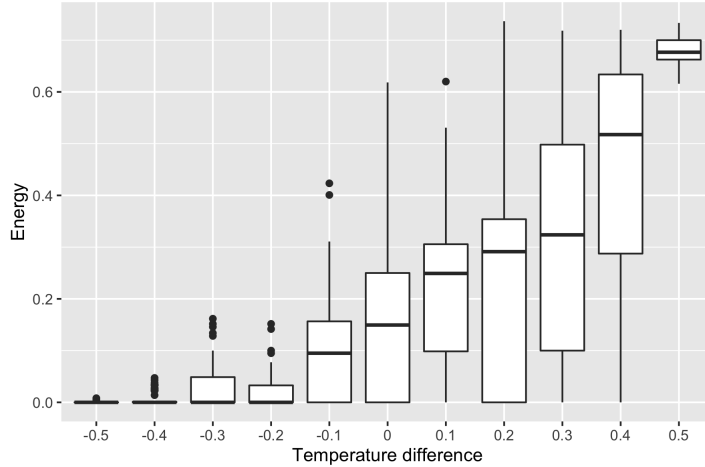
Figure 3.2: Box and whisker plot of HVAC energy consumption at different temperature difference levels

frequency of these peak-periods are particularly higher in the summer because of the increased usage of air-conditioning. The learning models are built on data collected before the implementation of the dynamic pricing.

The inside temperature of the household, the set-point temperature, the HVAC status from the programmable thermostats, and outside temperature from the local weather station are collected at 15 minute intervals. This is because changing the set-point at a more frequent rate might damage the equipment. Additionally we observe that the HVAC systems operation in these households fluctuate within the -0.5°F and 0.5°F range of the temperature difference between the set-point of the thermostat and the inside. Beyond, 0.5°F , the AC remains on for the next 15 mins while remains off below -0.5°F. The power rating of the AC reveals that the average power consumption for a time interval of 15 mins is 0.75kWh when operating continuously. Figure 3.2 shows the interquartile range of average HVAC power for such a home with respect to the temperature difference.
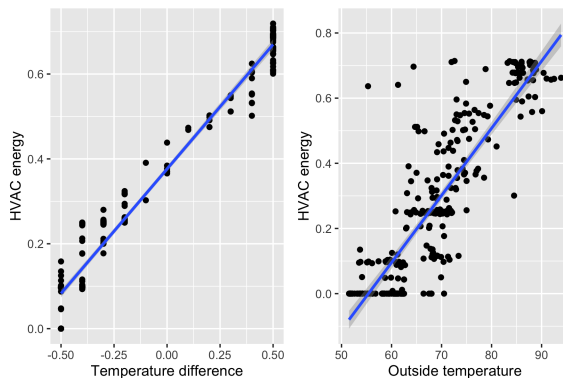
(a) Load vs. temperature difference and outside temperature
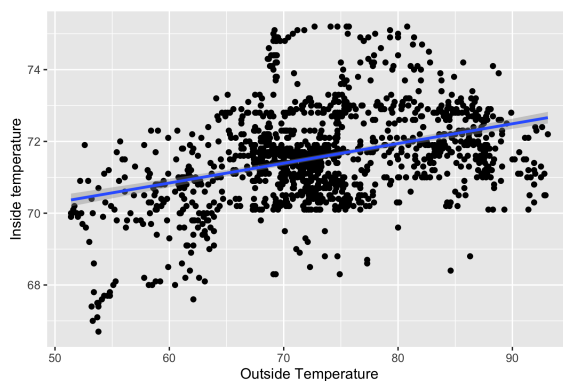
(b) Inside vs. outside temperature. The green line represents the linear relationship.

Figure 3.3: Dependent vs. independent variables considered in the predictive models

As explained in Section 3.2.1, we use a linear regression and an ARIMAX model to predict the HVAC average power and inside temperature of the households. In Figures (3.3a) we see the linear relationship between HVAC load and difference between thermostat set-point and inside temperature along with the increasing linear trend of inside temperature with outside temperature in plot (3.3b). The motivation here is to first develop a simple model with information easily available at the grid level without installing any other equipment.

Average Pearson's coefficient, over a sample size of 300 homes, of the factors
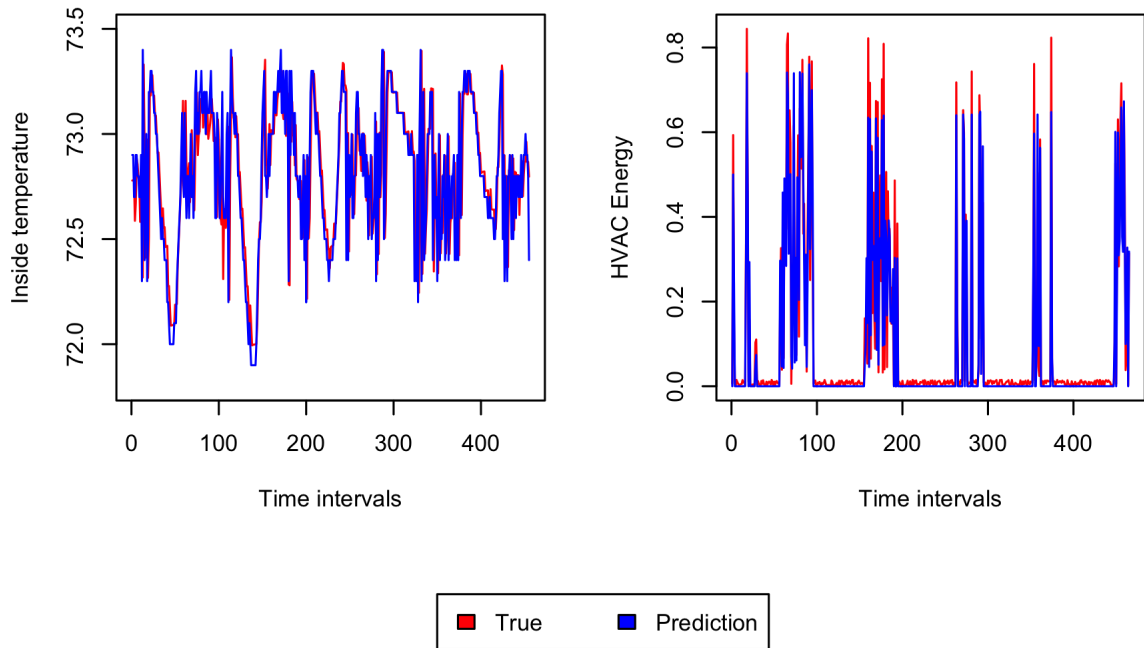
Figure 3.4: Out-of-sample(OOS) performance of the learning model

with respect to the dependent variable are 0.81 and 0.75 for temperature difference and outside temperature respectively for HVAC energy consumption. Values of the correlation coefficient for the inside temperature model are 0.85, 0.77, and 0.8 for inside temperature lag term, outside temperature, and HVAC load respectively. Average R^2 and the out-of-sample OOS R^2 are 0.78 and 0.64 for the HVAC load model and 0.88 and 0.74 for the inside temperature model respectively. These values indicate that the model, in spite of being simple, can provide satisfactory predictions. Figure (3.4) demonstrate the OOS performance of the models on the data of a single home over 5 days (480 time intervals).

3.3.2 TSEP for a single home

In this section, we proceed to present the results of the application of the TSEP on real-world instances. In the DR field project, the participant homes were categorized into three classes based on the installation of smart technologies. In Table (3.3), we define the three categories along with the ID of the homes selected as a representation of each category. Detailed results in terms of the inside temperature, HVAC energy, and the set-point are presented.

The model is applied on a summer day which is forecasted to have a peak consumption period of 4 hours. This constitutes the peak period $\mathbb{T}^0$ with each control interval t being equal to 15 minutes. Adjoining this period, we have the pre-cool ($\mathbb{T}^{-1}$) and post-cool ($\mathbb{T}^{+1}$) control periods each consisting of 3 hours or 12 time intervals. So, $T^{-1}, T^{+1} = 12$ for $\mathbb{T}^{-1}$ and $\mathbb{T}^{+1}$ repectively and $T^0 = 16$ (see Section 3.2.3). Here we deploy the TSEP for the whole of pre-cool and post-cool control periods for a single control duration. We also consider three instances depending on the value of the parameter T_p, the preferred temperature (see Table 3.2). In the base case, we consider $T_p = 75°F$, in the high tolerance one 77 °F and the case of low tolerance, $T_p = 73°F$. The tolerance margin k is taken to be 2°F. The model is applied on three different categories of homes with increasing levels of installed smart technologies. The categorization is explained in Table (3.3) with efficient water heaters and HVAC systems installed in all of them.

Figures (3.5) show the progression of HVAC energy, set-point temperature, and

Table 3.2: Example Cases

Case	Preferred temperature(°F)	Minimum temperature(°F)	Maximum temperature(°F)
Base	75	73	77
High tolerance	77	75	79
Low tolerance	73	71	73

Table 3.3: Home Category

Home ID	Home Type	Equipment Installed
1	Ultra	Insulation, programmable thermostat, battery storage system
2	Advanced	programmable thermostat, battery storage system
3	Basic	programmable thermostat

inside temperature throughout the three control-periods for the base case for Home 1. The model is capable of keeping the inside temperature within the tolerance bounds. HVAC energy consumption during the high price peak period is kept at a minimum avoiding heavy financial penalties. We compare the total energy consumed by application of the TSEP with the situation where no load control algorithm is applied. Under no external control the average power is determined with Equation (3.5) with $T_s^t = T_p$ $\forall t$. From Table (3.4), the DP solution yields approximately 36% of energy savings when compared to the situation with no control in the base case. Similar results are portrayed in the high tolerance (Figure 3.6) and the low-tolerance cases (Figure 3.7) for Home 1. Energy consumption in the high tolerance case is the minimum while in case of low tolerance, the AC remains practically on for the whole of pre and post-cool control periods in case of all the three sample homes. However, the energy consumption during the peak period is minimal (Figure 3.7a) thus fulfilling the objective. Considerable energy savings are calculated for all the three instances of each home as shown in Table (3.4). High tolerance case yields
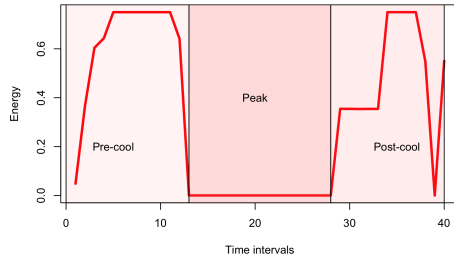
(a) Energy consumed by the AC
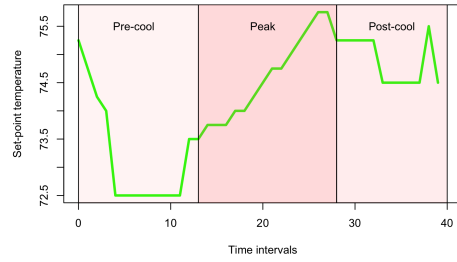
(b) Thermostat set-point temperature
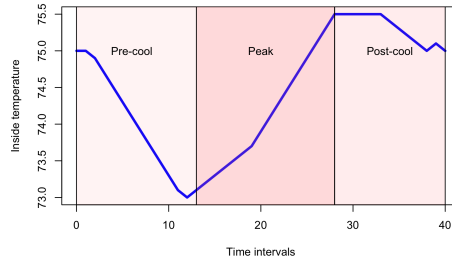
(c) Inside temperature

Figure 3.5: TSEP results of Home 1 for base case

the highest energy savings as the AC has to operate less due to the high preferred inside temperature. The low tolerance case yields the minimum energy savings but the numbers are significantly lower in case of advanced home 2 and basic home 3. This can be attributed to the better weatherproofing provided to Home 1 which helps to insulate the home against the heat outside keeping the inside temperature low. Figures (3.8) and (3.9) present the HVAC energy, set-point temperature, and the inside temperature of Homes 2 (advanced) and Home 3 (basic) respectively for the low tolerance case. The profiles indicate that the TSEP solution can maintain the inside temperature of the homes within the specified bounds while consuming minimum energy during the peak period even when the preferred temperature is low on a summer day.
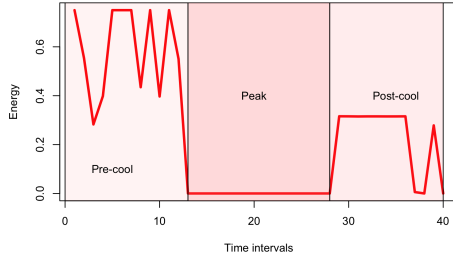
(a) Energy consumed by the AC
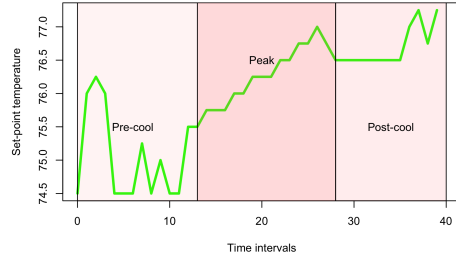
(b) Thermostat set-point temperature
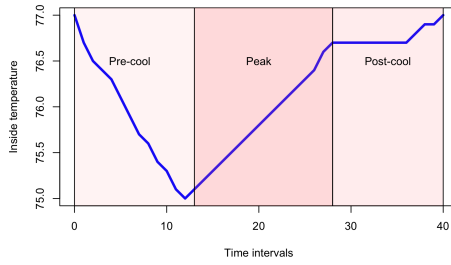
(c) Inside temperature

Figure 3.6: TSEP Results of Home 1 for high tolerance case
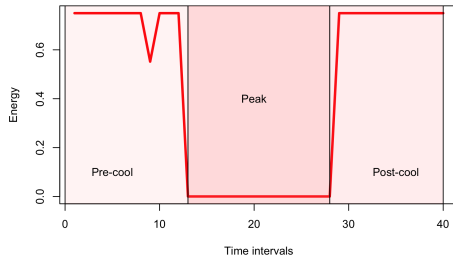
(a) Energy consumed by the AC
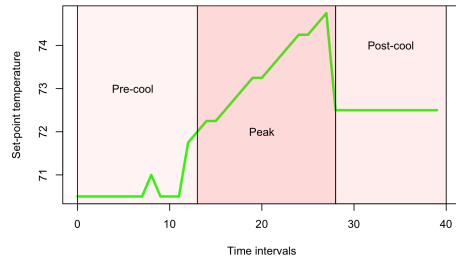
(b) Thermostat set-point temperature
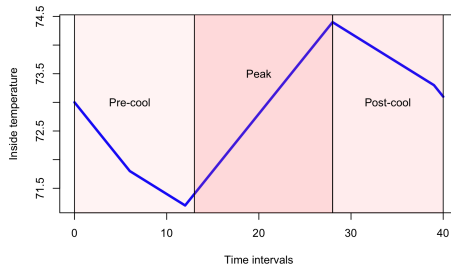
(c) Inside temperature

Figure 3.7: TSEP Results of a Home 1 for low tolerance case
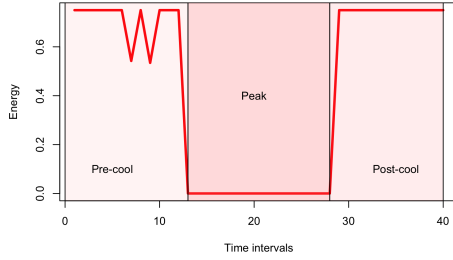
(a) Energy consumed by the AC
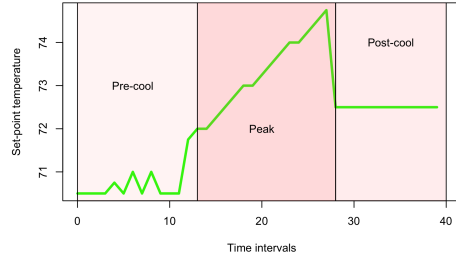
(b) Thermostat set-point temperature
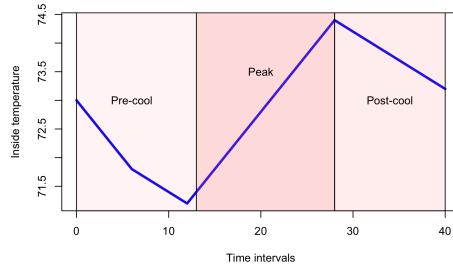
(c) Inside temperature

Figure 3.8: TSEP results of Home 2 for low tolerance case
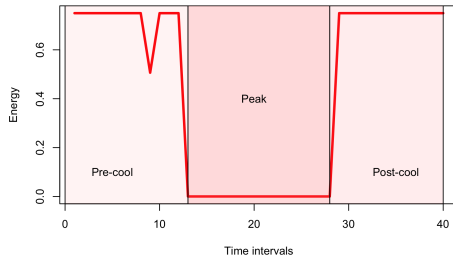
(a) Energy consumed by the AC
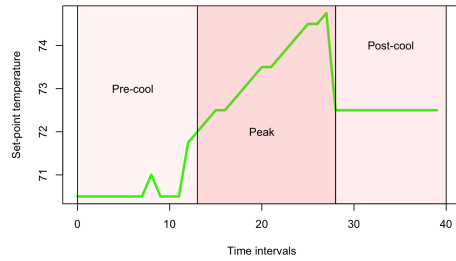
(b) Thermostat set-point temperature
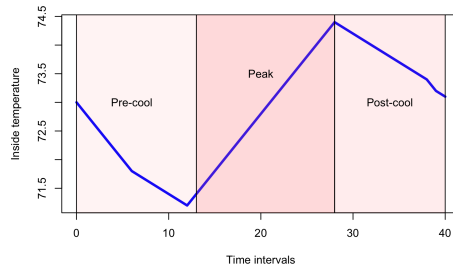
(c) Inside temperature

Figure 3.9: TSEP results of Home 3 for low tolerance case

Table 3.4: Comparison of TSAP results vs. no-control

HVAC energy consumption in base case (kWh)			
Home ID	TSAP	No control	Difference (%)
1	13.42	20.91	35.82
2	13.48	18.53	27.26
3	12.45	16.92	26.44
HVAC energy consumption in high tolerance case (kWh)			
Home ID	TSAP	No control	Difference
1	9.93	20.36	51.25
2	9.99	16.84	40.63
3	9.93	16.63	39.26
HVAC energy consumption in low tolerance case (kWh)			
Home ID	TSAP	No control	Difference
1	17.80	21.64	17.72
2	17.58	18.61	5.53
3	17.76	18.17	2.29

3.3.3 TSEP for multiple homes with alternative control durations

Encouraged by the results of TSEP on a single home for a single control duration, we proceed to present the results of applying TSEP on multiple homes and assessing alternative control periods. Initial data analysis of inside temperature and HVAC load revealed that the homes under study in general need at least 1 hour to cool down to their lowest preferred temperature. Thus, t^0 defined in Section 3.2.3 is equal to 4 time intervals for this set of results. Since $T^{-1}, T^{+1} = 12$ and $T^0 = 16$, we can define $\mathbb{S}_h = \{1\} \vee \{1, 2\} \vee, ..., \vee \{1, 2, ..., 8\}$ and $\mathbb{E}_h = \{29\} \vee \{29, 30\} \vee, ..., \vee \{29, 30, ..., 40\}$ for home h. The size of $\mathbb{S}_h$ and $\mathbb{E}_h$ can range between 1 and 9 depending on the number of feasible start and end times for home h. We select a sample size of 132 homes which made a selected cutoff in terms of the accuracy of the learning models.
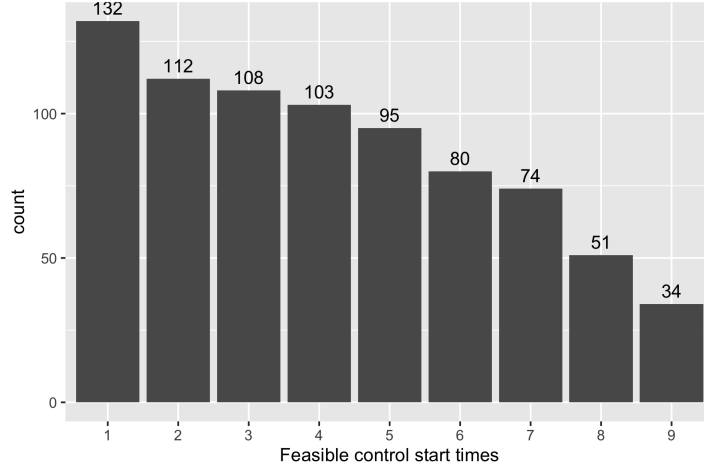
Figure 3.10: Categorization of homes according to their feasible control options

Included in the barplot in Figure (3.10) is the cumulative distribution of the number of homes with respect to the size of their set of feasible start times $\mathbb{S}_h$ during pre-cool. Homes having the larger $\mathbb{S}_h$ can be cooled down within lesser time intervals. These homes are presumed to have better insulation than others. We categorize the homes according to its number of feasible control start times (size of $\mathbb{S}_h$) and compare each category by HVAC load under no control. We also define thermal capacity q which gives the average HVAC energy consumed by the household for a degree difference between the set-point and internal temperature. The average is taken over all the homes which have the same number of feasible start times of TSEP. Homes having better insulation willl have a lower value of thermal capacity q.

$$q = \frac{\text{Average HVAC energy consumed}}{\text{Average difference between thermostat set-point and inside temperature}}$$

Figure (3.11) shows the relationship between the categories of homes and average thermal capacity over all the homes with the same sized of $\mathbb{S}_h$. The slope of the curve
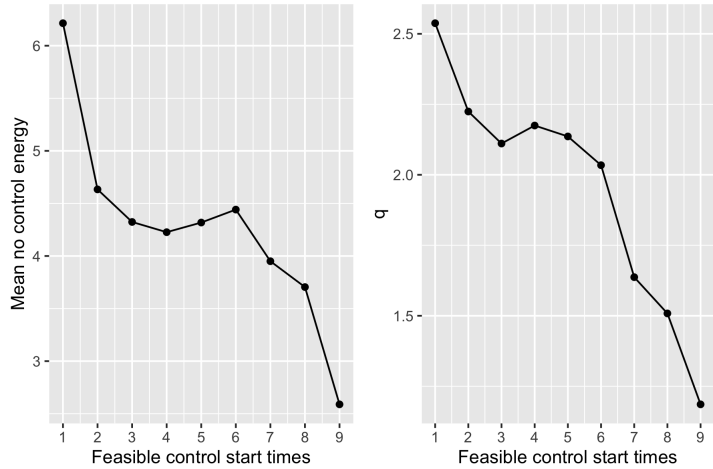
Figure 3.11: Relationship of feasible TSEP start times with thermal capacity q and no control HVAC load

follows our assumption that homes with larger number of feasible start times have a lower q and thus better insulation. In addition, these homes have also consumed less energy under no control as shown in Figure (3.11).

These results indicate that applying the TSEP to multiple homes can give us valuable insights about their insulation and thermal capacity. They can be used as a validation tool for other methods for determining the insulation quality of a building or can be used in other studies such as clustering a large pool of homes using this metric.

We offer some remarks to summarize the numerical results presented above. We demonstrate the efficacy of application of the TSEP for direct load control of residential HVAC systems during a peak period. Extending the problem to multiple homes with alternative control durations helps us to gain knowledge about the thermal capacity or the insulation quality. Residential consumers can be clustered according to their feasible start times of TSEP and we observe the relationship between such

groups and their HVAC energy consumption.

3.4 Conclusions

In this chapter, we propose a direct load control framework for HVAC systems applicable to a pool of residential consumers. We consider the control strategy under critical peak pricing and executed through smart thermostats which can be operated remotely. A dynamic programming based model (TSEP) is developed to estimate the optimal temperature set-points of such a device. The model is capable of maintaining the inside temperature of households within a certain preferred range while minimizing the HVAC energy consumption with the set-points as decisions. When applied the TSEP control algorithm to homes in a field demonstration DR project, computational results suggest that significant amount of energy is saved by the model with the maximum savings observed in homes with advanced weatherproofing. This shows that insulation plays and important role in electric energy consumption. We apply TSEP to multiple homes and determine the set of feasible control durations for each home. There is a strong correlationship between the thermal capacity of a home and the number of feasible control durations associated with the home. This provides valuable information about the insulation of the building which can be used in many applications (e.g. clustering of homes by their insulation quality).

CHAPTER 4

AN HVAC CONTROL SCHEDULING MODEL WITH MIXED-INTEGER LINEAR FRACTIONAL PROGRAMMING

4.1 Introduction

In the previous chapter, we propose a DP based model to estimate the optimal temperature set-points of a thermostat in a direct load control (DLC) program. The temperature set-point estimation problem (TSEP) is modeled at a single household level. Due to relatively low computational overhead, it can be scheduled for multiple homes simultaneously. However, cooling several homes at the same time may lead to a secondary peak or may just shift the peak from the forecasted peak period (Muratori et al. (2014)). This is undesirable as it does not alleviate the original concern of load variation over time and the resulting economic deficiency at the system-level. In our view, scheduling of any load control algorithm must include system load characteristics to maintain peak to average ratio or a relatively high load factor of the grid.

In the case of TSEP, we show in Section (3.2.3) that each home can have alternative feasible start and end times for the control actions. Therefore, when

applying TSEP to multiple homes one can leverage the start and end times of such controllers for each home in order to maintain system-level load factor. This will be in direct contrast to a naïve approach where the start and end times are selected which consume the minimum total energy for each residence. The motivation for this approach is that since each household consumes the infimum energy among all the feasible options, collectively, the total energy of the pool would not be high. We name this policy as the 'minimum energy' option. However, this scheme does not guarantee flattening the load curve.

In this chapter, we present a mixed integer linear fractional programming (MILFP) model for a control scheduling problem (CSP) in deploying TSEP for a pool of homes. The decision variables are binary variables to indicate the start time interval and end time interval while the objective is to flatten the demand curve to better allocate the resources and increase reliability. We also present a reformulation of the MILFP model in to a MILP problem to solve medium-scale instances with reasonable CPU times by general-purpose solvers. However, to reduce the time for large-scale instances, we propose a bi-section search algorithm to solve the fractional programming model. A Lagrangian dual approach to estimate the bounds of the model is presented to evaluate evaluate the solution quality by the proposed bi-section search algorithm. Our computational experiments suggest that jointly the TSEP in Chapter 3 and the CSP in this chapter provide an efficient solution to a DLC framework for utility companies to shave peak load for their residential customers.

4.2 Models and formulations

4.2.1 Control scheduling problem

Utility companies are interested in scheduling load control policies which help to flatten the load curve and avoid shifting of peak demand. An indicator to describe the consumption characteristics over a period of time is load factor. It is defined by,

$$\text{Load Factor} \; = \; \frac{\text{Average Load}}{\text{Maximum load in given time period}}$$

Load factor can give a measure of how much of available resource is being used at a certain point in time. A value of > 0.75 indicates that the load profile of system is balanced and can be achieved by distributing consumption over time. A higher value also demonstrates that capacity is not idle most of the time thus lowering costs. Utility companies, hence, are interested to maximize the load factor of its' pool of customers.

4.2.2 Control schedule problem

This motivates us to look into the scheduling problem with the objective of maximizing the load factor of the collection of homes under direct load control. The resulting optimization problem would have the binary decision variable x where $x = x_{h,s} \cup x_{h,e}$ indicating the start time interval $s \in \mathbb{S}_h$ and the end time interval $e \in \mathbb{E}_h$ of a home h. Table (4.1) defines the parameters used in the model,

The total energy consumed by each home during the peak period would be the same for every alternative start times because the initial and final state of TSEP is

Table 4.1: Control schedule problem parameters

Notation	Description
$\mathbb{H}$	Set of homes h under direct load control
$\mathbb{S}_h$	Set of feasible start time intervals s for home h
$\mathbb{E}_h$	Set of feasible end time intervals e for home h
$\mathbb{T}^i$	Set of time intervals of pre-cool, peak, and post-cool for $i = -1, 0, +1$ respectively
$\mathbb{T}$	Direct load control period consisting of T intervals; $\mathbb{T} = \bigcup_i \mathbb{T}^i$ for $i = \{-1, 0, +1\}$
$E_{h,s}^{-1}$	Total energy consumed by home h if TSEP is applied from interval s in pre-cool and peak period
$E_{h,e}^{+1}$	Total energy consumed by home h if TSEP is applied until interval e in post-cool period
$P_{h,s,t}$	Average power consumed by h if TSEP is applied from interval s during time interval $t \in \mathbb{T}^{-1}$
$P_{h,e,t}$	Energy consumed by h if TSEP is applied until interval e during time interval $t \in \mathbb{T}^{+1}$
a	Maximum power among all homes consumed during $\mathbb{T}$

the same when applied during the peak period whose length is forecasted. Thus, this

parameter does not depend on the decision of start or end times. We have combined

the energy consumed during the peak and that in pre-cool for start time s in to a single

parameter $E_{h,s}^{-1}$. Also it is noteworthy that for intervals where $t < s$, the average power

is determined using the Equation (3.5) where the preferred temperature is taken as

the set-point temperature. Below we present the control scheduling problem (CSP).

$$CSP : z^* = \underset{a,\ x}{\text{maximize}} \quad \frac{\sum_{h,s} E_{h,s}^{-1} x_{h,s} + \sum_{h,e} E_{h,e}^{+1} x_{h,e}}{Ta} \tag{4.1a}$$

$$\text{s.t.} \quad a \geq \sum_{h,s} P_{h,s,t} x_{h,s}, \quad \forall t \in \mathbb{T}^{-1} \tag{4.1b}$$

$$a \geq \sum_{h,e} P_{h,e,t} x_{h,e}, \quad \forall t \in \mathbb{T}^{+1} \tag{4.1c}$$

$$\sum_{s} x_{h,s} = 1, \quad \forall h \in \mathbb{H} \tag{4.1d}$$

$$\sum_{e} x_{h,e} = 1, \quad \forall h \in \mathbb{H} \tag{4.1e}$$

$$a \geq 0 \tag{4.1f}$$

$$x \in \{0,1\} \tag{4.1g}$$

Contraints (4.1b) and (4.1c) define a as the supremum of the average power consumed by every home at each interval of the control period. Constraint (4.1d) ensures that each home is assigned to a single feasible start and end time of TSEP.

The linear relaxation of CSP obtained through relaxing the binary restriction contraints 4.1g and 4.1g is a linear fractional program(LFP). Bazaraa et al. (2013) have studied this problem in detail and have shown that the objective function is both pseudoconcave and psudoconvex. An advantage of this property is that CSP can be solved to global optimality by solution methods which can accomodate pseudoconvex or pseudoconcave objective functions. Branch-and-bound algorithms and extended cutting-plane methods (Wolsey, 1998) for non-linear programs can be utilized for such a purpose (You et al., 2009).

MILFPs are in general non-convex mixed integer non-linear programs (MINLP) and are NP-hard. Presence of discrete variables along with non-convexity makes them computationally intractable for large scale scenarios (Tawarmalani & Sahinidis, 2013). We envision application of CSP on a large pool of homes and each having multiple start and end time options. Charnes and Cooper (1962) develop an efficient reformulation strategy from LFP to linear program which has been widely used for problems with continuous variables. In the next section, we present a reformulation of CSP to a mixed integer linear program (MILP) by extending the Charnes-Cooper transformation integrating with Glover's linearization scheme (Glover, 1975). Yue et al. (2013) have implemented a similar approach to solve a batch process scheduling problem in chemical plants. This enables the application of efficient MILP solution methodologies such as branch-and-cut (Wolsey, 1998) to solve the CSP.

4.2.3 Reformulated control schedule problem

The proposed reformulation procedure is a two-part scheme. The first takes cues from the Charnes-Cooper transformation and introduces a new continuous variable $u = \frac{1}{a}$ which helps us to transform the fractional objective function to a bilinear function.

The equivalent MINLP model of CSP (CSP-NL) is presented below:

$$CSP - NL : z^*_{NL} = \underset{u, \ x}{\text{maximize}} \quad \frac{1}{T}(\sum_{h,s} E^{-1}_{h,s} x_{h,s} + \sum_{h,e} E^{+1}_{h,e} x_{h,e})u$$

$$\text{s.t.} \quad \sum_{h,s} P_{h,s,t} x_{h,s} u - 1 \leq 0, \quad \forall t \in \mathbb{T}^{-1}$$

$$\sum_{h,e} P_{h,e,t} x_{h,e} u - 1 \leq 0, \quad \forall t \in \mathbb{T}^{+1}$$

$$\sum_{s} x_{h,s} = 1, \quad \forall h \in \mathbb{H}$$

$$\sum_{e} x_{h,e} = 1, \quad \forall h \in \mathbb{H}$$

$$u \geq 0$$

$$x \in \{0, 1\}$$

The bilinear terms $x_{h,s}u$ and $x_{h,e}u$ give rise to the non-linearlity in the above problem. Following Glover's linearization technique (Glover, 1975), these terms can be linearized by introducing a number of auxillary variables and constraints. The set of auxillary variables introduced are defined as $w_{h,s} = x_{h,s}u \ \forall h, s$ and $v_{h,e} = x_{h,e}u \ \forall h, e$ and these enable us to transform CSP to an equivalent MILP (CSP-MIP) given below,

$$CSP - MIP : z_{MIP}^* = \underset{w,\ u,\ v,\ x}{\text{maximize}}\quad \frac{1}{T}\sum_{h,s} E_{h,s}^{-1} w_{h,s} + \sum_{h,e} E_{h,e}^{+1} v_{h,e}$$

$$\text{s.t.}\quad \sum_{h,s} P_{h,s,t} w_{h,s} - 1 \le 0, \quad \forall t \in \mathbb{T}^{-1} \qquad (4.3\text{a})$$

$$\sum_{h,e} P_{h,e,t} v_{h,e} - 1 \le 0, \quad \forall t \in \mathbb{T}^{+1} \qquad (4.3\text{b})$$

$$\sum_{s} x_{h,s} = 1, \quad \forall h \in \mathbb{H} \qquad (4.3\text{c})$$

$$\sum_{e} x_{h,e} = 1, \quad \forall h \in \mathbb{H} \qquad (4.3\text{d})$$

$$w_{h,s} \le u, \quad \forall h, s \qquad (4.3\text{e})$$

$$w_{h,s} \le M x_{h,s}, \quad \forall h, s \qquad (4.3\text{f})$$

$$w_{h,s} \ge u - M(1 - x_{h,s}), \quad \forall h, s \qquad (4.3\text{g})$$

$$w_{h,s} \ge 0, \quad \forall h, s \qquad (4.3\text{h})$$

$$u \ge 0, \qquad (4.3\text{i})$$

$$v_{h,e} \le u, \quad \forall h, e \qquad (4.3\text{j})$$

$$v_{h,e} \le M x_{h,e}, \quad \forall h, e \qquad (4.3\text{k})$$

$$v_{h,e} \ge u - M(1 - x_{h,e}), \quad \forall h, e \qquad (4.3\text{l})$$

$$v_{h,e} \ge 0, \quad \forall h, e \qquad (4.3\text{m})$$

$$x \in \{0, 1\} \qquad (4.3\text{n})$$

Constraint sets (4.3e) and (4.3j) and (4.3g) and (4.3l) force $w_{h,s}$ and $v_{h,e}$ to take

the value of u whenever $x_{h,s}$ and $x_{h,e}$ is equal to 1 respectively for all $h \in \mathbb{H}$, $s \in \mathbb{S}_h$

and $e \in \mathbb{E}_h$. While sets (4.3f) and (4.3k) impose $w_{h,s}$ and $v_{h,e}$ to be equal to 0 when

$x_{h,s} = 0 \ \forall h, s$ and $y_{h,e} = 0 \ \forall h, e$ respectively. M is a sufficiently large number and ideally should be equal to the upper bound of u which gives the tightest formulation of the problem. The following two lemmas prove that CSP-MIP is a reformulation of MILFP CSP.

Lemma 1. *If $\{w^0, v^0, u^0, x^0\}$ is a feasible solution set for the problem CSP-MIP, then $\{x^0, a^0\}$ is a feasible solution set for CSP such that $a^0 = \frac{1}{u^0}$, $w^0_{h,s} = x_{h,s}u^0 \ \forall h, s$ and $v^0_{h,e} = x_{h,e}u^0 \ \forall h, e$ and vice versa.*

Lemma 2. *If $\{w^*, v^*, u^*, x^*\}$ is a global optimal solution set for the problem CSP-MIP, then $\{x^*, y^*, a^*\}$ is a global optimal solution set for CSP such that $a^* = \frac{1}{u^*}$, $w^*_{h,s} = x_{h,s}u^* \ \forall h, s$ and $v^*_{h,e} = x_{h,e}u^* \ \forall h, e$ and vice versa.*

Proof: CSP-MIP is a MILP problem and is derived from MINLP problem CSP-NL through exact linearization which makes these two problems equivalent. CSP-NL is a transformation of CSP via variable substitution and hence equivalent. By extension, there exits a one-to-one mapping between the solutions of CSP-MIP and CSP. ∎

MILP problems are much more tractable than their non-linear non-convex counterparts. General purpose solvers such as CPLEX 12.5 (Manual, 1987) using effcient algorithms e.g. branch-and-cut are capable in solving large scale instances of such problems within a desired tolerance. It should be noted that, global optimizer BARON 15.6.5 (Tawarmalani & Sahinidis, 2004) can be used to solve CSP without any necessary changes. In Section 4.4.1, we show that though BARON is capable of solving smaller instances of CSP within a considerable CPU time limit but is outperformed by CPLEX solving CSP-MIP for medium sized instances for 132 homes with at miximum 9 feasible start or end times. However, for large scale instances these methods may run into memory problems as in a worst-case scenario all the binary variables have to be investigated. Indeed, application of CSP-MIP to a medium scale instance takes an excessive amount of time when solving with CPLEX. We present the result of such applications in Section 4.4.1. To overcome this obstacle, we develop an efficient

algorithm based on bi-section search to solve CSP within a reasonable amount of time even for large instances.

4.3   Bi-section search algorithm to solve CSP

We observe that in CSP there are no interactions between the control options i.e. $x_{h,s}x_{h,e}$ $\forall s \in \mathbb{S}_h, e \in \mathbb{E}_h \ \forall h$. This motivates us to decompose CSP for this instance in two separate problems for computational efficiency as following:

$$\max_{x,\ a}\{\frac{1}{T}\sum_{h,s} E_{h,s}^{-1} x_{h,s}|(4.1b),(4.1d),(4.1f), x_{h,s} \in \{0,1\} \ \forall h,s\} \qquad (4.4)$$

$$\max_{x,\ a}\{\frac{1}{T}\sum_{h,e} E_{h,e}^{+1} x_{h,e}|(4.1c),(4.1e),(4.1f), x_{h,e} \in \{0,1\} \ \forall h,e\} \qquad (4.5)$$

Let problem (4.4) and (4.5) be labelled as *pre*-CSP and *post*-CSP respectively.

The following property helps to design the efficient bisection search algorithm,

Lemma 3. *If $\{x_{h,s}^0, a^0\} \ \forall h, s$ is a feasible solution set for the problem pre-CSP, and $\{x_{h,e}^0, a^0\} \ \forall h, e$ is a feasible solution set for the problem post-CSP then $\{x^0, a^0\}$ is a feasible solution set for CSP such that $x^0 = x_{h,s}^0 \cup x_{h,e}^0$.*

Lemma 4. *If $\{x_{h,s}^*, a^*\} \ \forall h, s$ is a feasible solution set for the problem pre-CSP, and $\{x_{h,e}^*, a^*\} \ \forall h, e$ is a feasible solution set for the problem post-CSP then $\{x^*, a^*\}$ is an optimal solution set for CSP such that $x^* = x_{h,s}^* \cup x_{h,e}^*$.*

Proof: The above two lemmas follows from the non-existance of the term $x_{h,s}x_{h,e}$ $\forall s \in \mathbb{S}_h, e \in \mathbb{E}_h$ and $\forall h$ in CSP. ∎

We observe that either of *pre*-CSP or *post*-CSP can be transformed into an integer programming (IP) problem by taking a as a parameter. Hence, we have the following CSP-IP subproblems,

$$\max_{x}\{\frac{1}{T}\sum_{h,s} E_{h,s}^{-1} x_{h,s}|(4.1b),(4.1d), x_{h,s} \in \{0,1\} \ \forall h,s\} \qquad (4.6)$$

$$\max_x \{\frac{1}{T}\sum_{h,e} E_{h,e}^{+1} x_{h,e} | (4.1c), (4.1e), x_{h,e} \in \{0,1\} \; \forall h,e\} \tag{4.7}$$

a is defined as the maximum energy consumed by all the homes over all the control intervals. There exist efficient algorithms which can tackle even large instances of IP problems (Wolsey & Nemhauser, 1999). Our search algorithm takes advantage of this by solving instances of CSP-IP for different values of a.

Bi-section search involves solving the problem iteratively for values of a parameter calculated at the bi-section of a defined interval. The algorithm efficiently strives to move the bounds of the interval closer until an artibtrary small width is achieved which contains the optimal a^* of CSP. The value of a for each sub-problem is taken at the mid-point of the interval.

The maximum energy consumed by all homes together, a, appears in CSP in the denominator of the fractional objective function and does not interact with the other decision variable x. This means a solution of (4.6) and (4.7) for any value of a greater than it's lower bound will give us a feasible solution to CSP. We utilise this idea to provide a direction to our search algorithm.

Remark. *To converge to a^* corresponding to optimal solution z^* of CSP, the upper bound for search interval of a is decreased if the incumbent subproblems (4.6) and (4.7) are feasible while the lower bound is increased if any one of the incumbent subproblems are infeasible.*

Care should be taken about the selection of the upper bound for the initial interval of a to make the algorithm converge within a reasonable number of steps. One such upper bound for a can be the incumbent peak of the system i.e. the peak load that has been already attained without the need for extra resources. Utility

providers can be interested in maintaining the same peak which utilizes all their active resources while not implementing ancillary ones. A trivial initial lower bound of a can be 0, however, a tighter interval can be implemented by having some insights (see Section 4.4.1) about the instance data. A summarization of the algorithm is given below,

---

Algorithm 1 Control Scheduling Algorithm (CSA)

1: **procedure** Control Scheduling Algorithm
2:     Intialize the algorithm with bounds for a as $[a_{LB}, a_{UB}]$ and maximum number of iterations as N
3:     Set iteration counter k to 0
4:     **if** k = 0 **then**
5:         Solve (4.6) and (4.7) with a_{UB} to get incumbent solution z' with $x^k = x^k_{h,t} \cup x^k_{h,e}$
6:     **else**
7:         At each iteration k do,
8:         $a^k \leftarrow a_{LB} + (a_{UB} - a_{LB})/2$
9:         Solve (4.6) and (4.7) for z^k with x^k
10:        **if** (4.6) or (4.7) is infeasible **then**,
11:            $a_{LB} \leftarrow a^k$
12:        **else**
13:            $a_{UB} \leftarrow a^k$
14:        **end if**
15:        **if** $a_{UB} - a_{LB} \leq \epsilon$ OR,   ▷ Check stopping criteria
            $z^k - z' \leq \rho$ **then**
16:            Stop,
17:            $x^* \leftarrow x^k$
18:            $a^* \leftarrow a_k$
19:            $z^* \leftarrow z^k$
20:        **else**
21:            $z' \leftarrow z^k$
22:            $k \leftarrow k + 1$
23:            Go to Step 6
24:        **end if**
25:     **end if**
26: **end procedure**

4.3.1 Lagrangian dual for CSP upper bound

Our numerical experiments in Section 4.4 show that CSA is able produce the same solution as solving CSP-NL with global optimizer BARON 15.5.6 for moderate sized instances. However, for larger instances BARON fails to converge while the proposed algorithm terminates within a reasonable amount of time. To gauge the solution quality for such instances it is desirable to determine an upper bound to the objective function. In this section, we present a Lagrangian dual approach which provides a tighter bound than linear relaxation in our experiments (see Section 4.4.2).

We derive from Lemma 1 that an upper bound to the optimal solution z_{MIP}^* of CSP-MIP will also be an upper bound to z^* of CSP. Also following Lemma 4, we can have corresponding subproblems *pre*-CSP-MIP and *post*-CSP-MIP.

$$z_-^* = \max_{w,\,x,\,a}\Big\{\frac{1}{T}\sum_{h,s}E_{h,s}^{-1}w_{h,s}|(4.3a),(4.3c),(4.3e)\text{ to }(4.3i),(4.3n)\Big\} \qquad (4.8)$$

$$z_+^* = \max_{v,\,x,\,a}\Big\{\frac{1}{T}\sum_{h,e}E_{h,e}^{+1}v_{h,e}|(4.3b),(4.3d),(4.3i)\quad\text{to}\quad(4.3n)\Big\} \qquad (4.9)$$

Upper bound to these subproblems can be determined by either dualizing one of the constraints and solving the corresponding Lagrangian relaxation or relaxing the binary conditions and solving the corresponding linear program. We construct a lagrangian relaxation CSP(λ) of CSP-MIP subproblems by dualizing the corresponding

constraint sets (4.3a) and (4.3b) with the set of multipliers λ_t where $t \in \mathbb{T}^{-1} \cup \mathbb{T}^{+1}$.

$$\text{pre-CSP}(\lambda) : z^*_{D-}(\lambda) = \underset{w,\ u,\ x}{\text{maximize}} \sum_{h,s} \left(\frac{E^{-1}_{h,s}}{T} + \sum_{t \in \mathbb{T}^{-1}} \lambda_t P_{h,s,t} \right) w_{h,s} - \sum_{t \in \mathbb{T}^{-1}} \lambda_t$$

$$\text{s. t.} \quad (4.3c), (4.3e) \text{ to } (4.3i), (4.3n)$$

$$\text{post-CSP}(\lambda) : z^*_{D+}(\lambda) = \underset{v,\ u,\ x}{\text{maximize}} \sum_{h,e} \left(\frac{E^{+1}_{h,e}}{T} + \sum_{t \in \mathbb{T}^{+1}} \lambda_t P_{h,e,t} \right) v_{h,e} - \sum_{t \in \mathbb{T}^{+1}} \lambda_t$$

$$\text{s. t.} \quad (4.3d), (4.3i) \quad \text{to} \quad (4.3n)$$

From Lemma (4) it further follows that, $z^*_{MIP} = z^*_- + z^*_+ \leq z^*_{D-} + z^*_{D+}$ and consequently from Lemma (1) we extend the relationship to $z^* \leq z^*_{D-} + z^*_{D+}$. The best such upper bound to z^* is obtained by combining,

$$z^*_{D-} = \min_\lambda z_{D-}(\lambda)$$

$$z^*_{D+} = \min_\lambda z_{D+}(\lambda)$$

There are several methodologies in literature to solve the Lagrangian dual problem. Subgradient optimization (Wolsey & Nemhauser, 1999) is a proven effective way to iteratively determine the dual. The method starts from an initial multiplier λ^0 and determines iteratively subsequent sequences solving pre-CSP(λ) and $post$-CSP(λ) at each iteration until some stopping condition is met. We follow the update rule proposed by Goffin (1977) which has shown to converge to z^*_D under certain conditions. The update rule is defined for any iteration i as,

$$\lambda^{i+1}_t = \max\left\{0;\ \lambda^i_t + \frac{\rho^i f_t(\bar{z} - z_D(\lambda^i))}{\sum_{t \in \mathbb{T}^{-1}} f^2_t}\right\} \tag{4.12}$$

where,

$$\rho^i \quad \text{satisfies} \quad 0 < \rho^i \leq 2,$$

$$\bar{z} \quad \text{is an upper bound of} \quad z_D^*$$

$$f_t = \sum_{h,s} P_{h,s,t} w_{h,s}^* \lambda^i - 1, \quad \forall t \in \mathbb{T}^{-1}$$

$$= \sum_{h,e} P_{h,e,t} v_{h,e}^* \lambda^i - 1, \quad \forall t \in \mathbb{T}^{+1}$$

$$\mathbb{T}^{-1} \cap \mathbb{T}^{+1} = \varnothing$$

The design of the subgradient algorithm can incorporate stopping rules such as lower limit on the difference of λ between iterations i.e. $(|\lambda^{i+1} - \lambda^i| \leq \epsilon)$, or an upper limit on the number of iterations. In some applications, the value of ρ^i is reduced periodically to 0. The stopping rules along with the choice of the parameters $(\rho^i, \bar{z}, \lambda^0)$ can differ by implementation. These conditions sometimes lead to an approximation of z_D^* which is acceptable in most applications.

Let $z_{D-}(\lambda^*)(z_{D+}(\lambda^*))$ be the best approximation for $z_{D-}^*(z_{D+}^*)$ and can serve as an upper-bound to z^*. We proceed next to present the results of the numerical experiments conducted with the proposed models.

4.4   Computational results

We solve the two corresponding IP problems (4.6, 4.7) at every iteration of CSA. Similarly two corresponding MILPs (4.8, 4.9) are formulated and the upper bounds are derived following the method in Section 4.3.1. The solutions are combined to give

the upper bound for the CSP. Number of binary variables for each problem (both CSP-MIP and CSP-IP) and for each period, pre and post-cool, will be equivalent to the number of homes times the alternative start and end times of TSEP here equal to 18. We will present the results in combined form for ease of understanding for the reader. In addition, to the real-world data, we also present results for simulated data to observe the performance of CSA for larger instances.

4.4.1 Control scheduling results for real-world data

We begin by presenting the performance of the bi-section search algorithm in terms of the load factor of a pool of homes. Stopping conditions for CSA are taken to be $\rho = 1 \times e^{-6}$ and $\epsilon = 1 \times e^{-5}$. The solution quality of each problem for the same instances are included in Table (4.3). In terms of load factor, (Table 4.2), we observe that CSA is able to produce the solution provided by the general purpose solvers for the first three instances while out performing them in CPU time. In case of 75 homes, it converges to the same solution as CPLEX solving CSP-MIP to a 0.00% optimality gap (Table 4.3) while BARON runs 'out of resource' and ends up with an inferior solution. In case of the largest instance involving 132 homes, CSA provides a superior solution compared to the other methods while reducing the computation time significantly.

From Table (4.3), we observe that the upper bounds provided by the Lagrangian dual model CSP(λ) (LD gap) are close to the ones provided by CPLEX for each instance. The subgradient optimization algorithm is run for 200 iterations for all the

instances with parameter settings $\rho^i = 0.025 \quad \forall i$, $\lambda^0 = 0$ and $\bar{z}$ is taken as intial upper bound calculated with the start and end times for each home which consume the minimum energy. This shows that the Lagrangian dual upper bound can be used as a metric to gauge the solution quality given by the bi-section search algorithm. For CSP-NL, the relative improvement metric (δ_r) (Sahinidis, 2017) is reported as the gap.

Table 4.2: CPU times

Number of	CSP-MIP(CPLEX)		CSP-NL(BARON)		CSP(CSA)
Homes	Load Fator	CPU Time	Load Factor	CPU Time	Load Factor
15	0.6273	7854.32	0.6273	16785.23	0.6273
25	0.6224	12956.47	0.6223	22458.75	0.6224
50	0.6139	23751.25	0.6138	63741.23	0.6139
75	0.6093	46489.69	0.5818	out of resource	0.6093
132	0.6097	100000.05	0.5527	out of resource	0.6102

Table 4.3: Solution quality

Number of	CSP-MIP		CSP-NL		CSP
Homes	CPLEX gap	LD gap	BARON gap	LD gap	LD gap
15	0.00%	0.23%	0.00%	0.23%	0.23%
25	0.00%	0.42%	0.00%	0.42%	0.42%
50	0.00%	0.65%	0.00%	0.66%	0.65%
75	0.00%	0.87%	8.96%	5.36%	0.87%
132	0.89%	1.01%	13.48%	10.27%	0.93%

A naïve policy to schedule TSEP over a pool of homes is to select the start and end time of TSEP for each home which consumes the minimum total energy. Another one can be starting and ending the application at the beginning and end of the pre and post-cool periods respectively for all homes. However, this does not incorporate system reliability. We demonstrate in Figure (4.1) and in Table (4.4) the advantage of applying CSP to determine the optimal schedule over using alternative policies

such as scheduling with minimum energy, all period control, and without direct load control during the CPP period. The load profile demonstrates that the solution of both CSP-MIP and CSA is able to flatten the curve for the pre-cool and post-cool period while in the no control scenario significantly more HVAC consumption occurs during the peak period (periods 12 to 28). CSA is able to provide a better solution for the larger instance of 132 homes and is reflected in the load curve in Figure (4.1). The load profile of the homes when no direct load control is applied is much 'flatter' compared to the control methodologies. However, there is close to nil curtailment during the high cost peak period. This translates to a better overall load-factor (Table 4.4) but a high penalty in terms of total cost (Table 4.5) in USD. Application of TSEP ensures that minimal energy is consumed during the peak period while maintaining the comfort level inside the household. Table (4.5) includes the cost difference between applying no DLC framework and application of different scheduling policies of TSEP. All the scheduling techniques offer significant cost savings during the 10 hours long total direct control load period. Flattening the load profile helps the utility companies in making educated decision about the allocation of resources. Low consumption during the peak period results in a worse load factor but ultimately helps in demand management.

4.4.2   Control scheduling results for simulated data

The next batch of experiments are designed to demonstrate performance of the model and the algorithm for large instances. To accomplish this, we generate data for a set of homes of size 500, 1000, 1500, 2000, 2500, and 3000. The uniform distribution

60

Table 4.4: Load factor comparison between scheduling policies

Homes	All period control	No control	Minimum Energy	CSP
15	0.35	0.81	0.37	0.63
25	0.39	0.84	0.40	0.62
50	0.39	0.82	0.39	0.61
75	0.39	0.84	0.38	0.61
132	0.38	0.86	0.39	0.61

Table 4.5: Total cost comparison between scheduling policies

Number of Homes	No control total cost	vs. No control		
		CSP	Minimum energy	All period control
15	1445.17	92.52%	92.80%	92.78%
25	1881.62	94.13%	94.35%	94.32%
50	4353.95	96.30%	96.49%	96.48%
75	6640.82	96.72%	96.98%	96.97%
132	11788.44	97.36%	97.57%	97.56%

function of the pseudo random generator of Python 3.7 is used to simulate the data sets. The experiments are conducted for 10 separate instances for each size of the problem to remove noise and averaged out results are presented. We are interested in the quality of the solution provided by CSA in terms of the gap with the Lagrangian dual and also the performance of the dual model and the subgradient optimization. The same control algorithm has been used to determine the parameters for CSP (see Section 4.4.1) with 9 alternative start and end times for each home for pre and post-cool control period.

The results are reported in two batches. We begin with the solutions in terms of the load factor of the HVAC systems of the homes with comparison between different policies explained in Section (4.4.1). From Table (4.6), we observe that similar to the real-world instances, the load factor from the CSP model is significantly better than
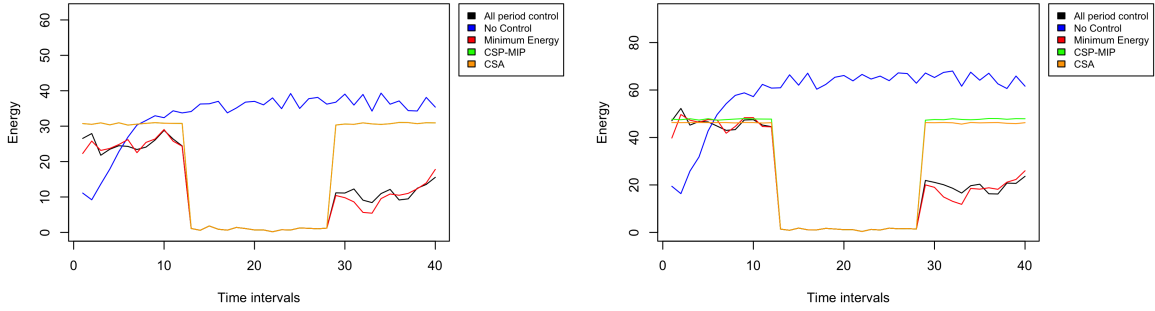
(a) Load profile of 75 homes for different policies

(b) Load profile of 132 homes for different policies

Figure 4.1: Load profile of multiple homes for control and scheduling policies

the others expect for the instance without direct load control. This is because under this policy homes consume significantly more energy during the CPP period and thus paying a severe monetary penalty but maintaining a flatter load profile. Included are the individual load factors from the pre and post-cool models for the CSP model. Load factors of the other models are reported as the overall during control period. The pre-cool load factor is lower because of the forced low consumption of energy during the peak period. This is enforced by the application of TSEP. Load profile for 1500 and 3000 homes are presented in Figure (4.2) as an example. These prove that CSP can help in flattening the demand curve even for a large number of homes. We observe that if no control is applied, the homes consume significantly more energy during the pre, post, and the expensive peak period.

Table (4.7) present the total cost in dollars for the different policies averaged over the 10 instances for each size. All the scheduling strategies significantly reduce the cost compared with the no control policy which portray the benefit of using the DLC during the CPP period. The scheduling policy from CSP provides a balance

between cost and load factor as shown in Figure (4.4) for 1500 and 3000 homes. The option witout control has a high load factor but a also a high average cost. Alongside in Table (4.8) the average solution times for Algorithm (1) are presented with Figure (4.3) showing the increase with instance size. Included in the same table are the average gap with the Lagrangian dual estimated with the subgradient optimization algorithm. The bounds are reasonably small for problems of all the sizes tested as an evidence to the solution quality. In addition, gap with the linear relaxation of CSP is also reported. This demonstrates that the Lagrangian dual provides a tighter bound and thus a better performance metric.

Table 4.6: Comparison of load factor

| Number of | All period | No control | Minimum energy | CSP | | |
Homes				overall	pre-cool	post-cool
500	0.47	0.89	0.43	0.61	0.45	0.9972
1000	0.47	0.88	0.44	0.60	0.44	0.9984
1500	0.49	0.86	0.42	0.62	0.45	0.9991
2000	0.46	0.89	0.44	0.61	0.45	0.9992
2500	0.45	0.85	0.45	0.62	0.45	0.9945
3000	0.46	0.86	0.43	0.61	0.45	0.9961

Table 4.7: Comparison of total cost

| Number of | No control | | vs. No control | |
Homes	total cost	CSP	Minimum energy	All control
500	46696	95.27%	96.45%	96.42%
1000	94159	95.37%	96.55%	96.51%
1500	141854	96.33%	96.52%	96.48%
2000	189115	96.38%	96.58%	96.54%
2500	235142	96.31%	96.50%	96.47%
3000	282095	94.45%	96.64%	96.61%
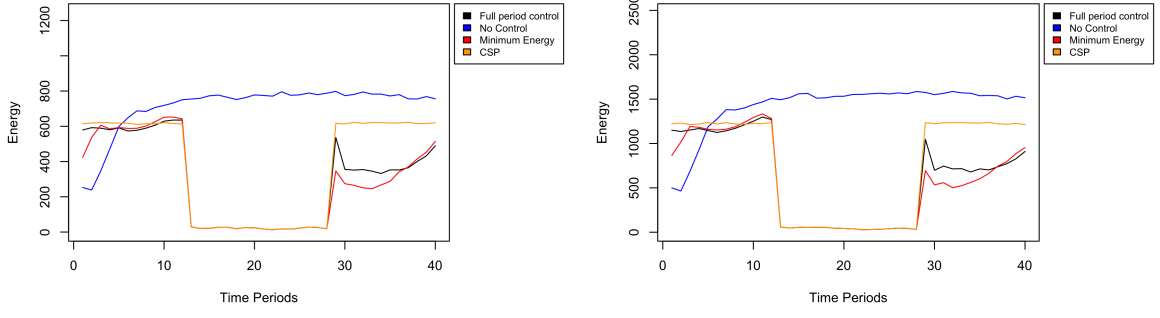
(a) Load profile of 1500 homes for different policies

(b) Load profile of 3000 homes for different policies

Figure 4.2: Load profile of multiple homes with simulated data
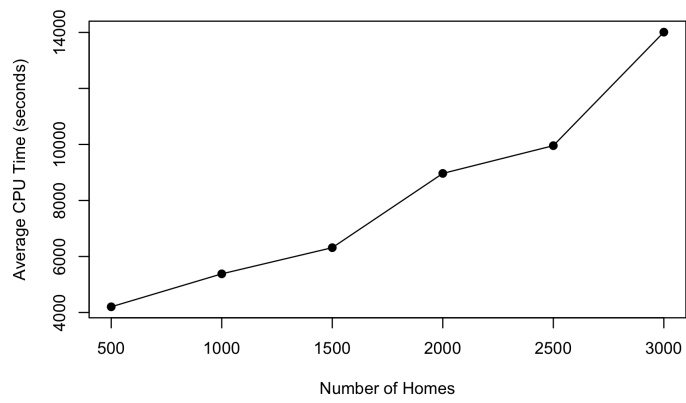
Figure 4.3: Average CPU times in seconds for CSA
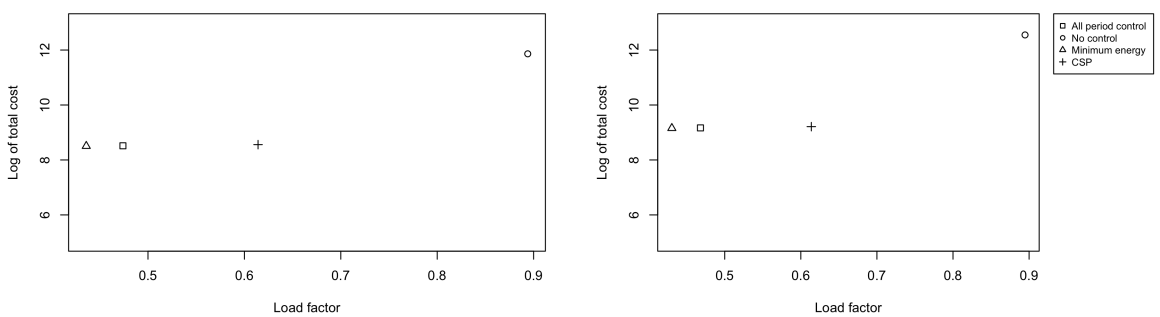
(a) Load profile of 1500 homes for different policies

(b) Load profile of 3000 homes for different policies

Figure 4.4: Load profile of multiple homes with simulated data

Table 4.8: Performance and solution quality

Number of Homes	CPU Time	Avg. Lagrangian dual gap	Avg. linear relaxation gap
500	4640.43	0.69%	1.16%
1000	5868.12	0.74%	2.06%
1500	6813.72	0.73%	2.28%
2000	9465.84	0.73%	3.19%
2500	10424.60	0.77%	1.22%
3000	14527.31	0.78%	2.14%

4.4.3 Computation of Lagrangian dual

The quality of the above presented solutions are compared with the upper bound provided by the approximation of the Lagrangian dual estimated by subgradient descent optimization. We already demonstrate the effectiveness of such a bound in Table (4.3). Analyses of the run times are required because we want to determine the bound within a reasonable amount of time corresponding to solving CSP by CSA. In Figure 4.5, we show the CPU times in seconds for the subgradient optimization algorithm to solve for $z_{D-}(\lambda^*), z_{D+}(\lambda^*)$ with the same parameters as given in Section 4.4.1. A maximum number of iteration limit of 1600 is provided to compare between problems of different size. As in the previous results, 10 different instances of each problem size is solved and the average metrics are reported. The method is able to provide results close to the final solution within 800 iterations as shown in Figure (4.6).

We proceed to offer some remarks as a summarization of the results. First, the bi-section search algorithm, CSA, offers significant improvement in solution time over general purpose solvers such as CPLEX and BARON in solving CSP. In addition,
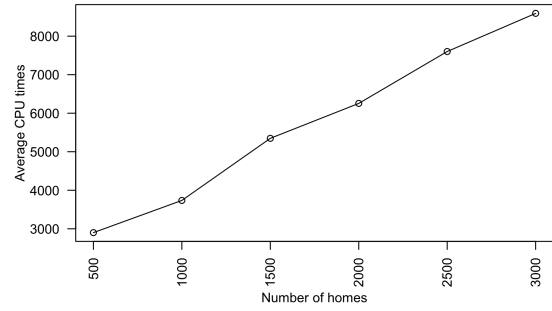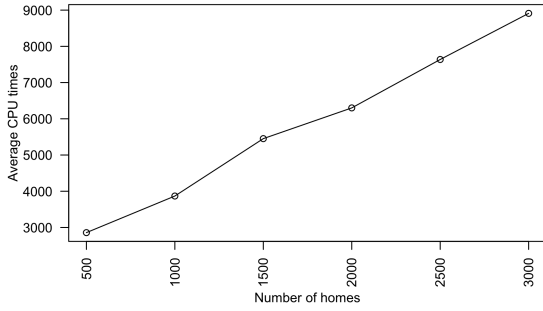
(a) Load profile of 1500 homes for different policies

(b) Load profile of 3000 homes for different policies

Figure 4.5: Load profile of multiple homes with simulated data
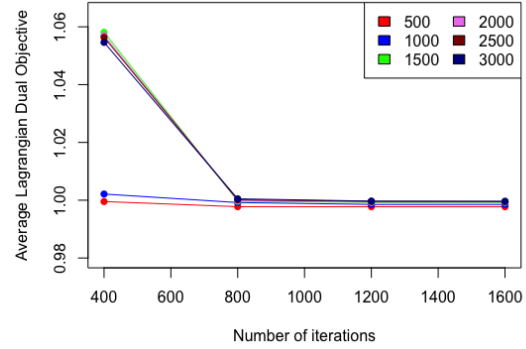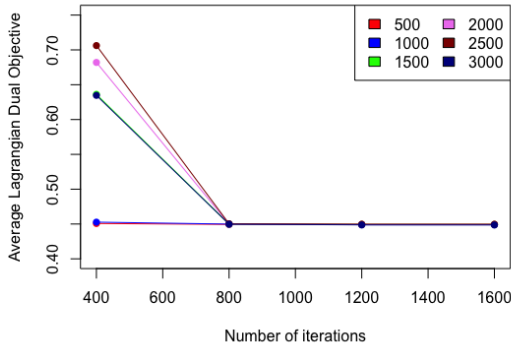
(a) *pre*-CSP(λ) objective through iterations

(b) *pre*-CSP(λ) objective through iterations

Figure 4.6: Lagrangian dual through iterations

the results also prove the efficacy of the optimization model, CSP, in flattening the demand curve. This is certainly desirable for planners when allocating resources. To observe the increase in solution time for larger instances we use simulated data generated systemically using a pseudo random number generator. CSA terminates within a reasonable amount of time for all instances with the largest containing 3000 homes and the solution leads to a flat demand curve. Next, we provide a Lagrangian dual approach to gauge the quality of the results. In case of small real-world instances, we demonstrate that the upper bound provided by such an approach is close to the

bound provided by CPLEX. In instances with simulated data, Lagrangian dual upper bound is tighter than the one provided by the linear relaxation (Table 4.8).

4.5 Conclusions

In this chapter, we propose an optimization model that aims to maximize the load factor to avoid shifting the peak as a consequence of deploying TSEP to multiple homes simultaneously. The problem is modelled as a MILFP problem with the binary decision variables indicating the start and end time of TSEP for each home. The MILFP model for the control scheduling problem (CSP) is computationally intractable for large instances even with state-of-the-art solvers. We transform the CSP from a MILFP to a MILP model. CSP-MIP can be solved to optimality with proven and efficient methods such as branch-and-cut and even with general purpose solvers (CPLEX) for medium sized- instances. Numerical studies with real world data demonstrate evidence of the efficacy of such an optimization model over other control strategies. However, application of the CSP on large sized instances results in excessive computational times. Hence, we present a bi-section search algorithm, namely the control scheduling algorithm (CSA), to solve the CSP with thousands of binary variables in reasonable CPU time. The methodology is able to reduce the solution time significantly when compared with off the shelf solvers such as CPLEX and BARON. n addition, a Lagrangian relaxation approach is developed to obtain an upper bound of CSP to evaluate the solution quality of CSA. Extensive numerical experiments are performed to investigate the performance of the model

and the algorithm for large-scale instances involving at most 3000 homes. The results show that a good quality solution can be achieved with the proposed bi-section search method. However, the solution times increase significantly with the increase in problem size. Though the scheduling does not have to be real-time for the application described here but efficient custom techniques (e.g. branch-and-cut) can be developed to solve the IP sub-problems at each iteration of the algorithm in future research.

CHAPTER 5

ESTIMATING UTILITIES OF PRICE-RESPONSIVE ELECTRICITY CONSUMERS

5.1 Introduction

In order to successfully implement DR programs, utility companies are interested in understanding the behavior of *price-responsive* consumers. However, formulating mathematical models of this behavior is difficult because of the general lack of data. Demand response programs have only started being implemented in the last few years by different utility companies. It is evident that designing consumer-friendly demand response program is very much dependent on understanding their price-responsive behavior. The same program may not reap similar benefits for every type of consumer because their reaction to change in electricity rates are different.

This chapter aims to develop quantitative models for estimating consumer's price-responsiveness using real-world data in the field demonstration project mentioned previously and the notion of utility theory. Utility functions have been widely used by researchers to model human's decision making with multiple factors (e.g., cost, convenience, risk). In electricity consumption, utility functions define the satisfaction level obtained from consuming electricity and from using various home appliances at

one's convenient times. We mention some of the pertinent literature in Section (2.2).

In this chapter, we propose a novel bi-level optimization model to estimate the utility coefficients representing price-responsiveness of electricity consumers under dynamic pricing in a DR program. The upper level objective is to minimize the estimation error between the measured data and the optimum consumption while the lower level has the objective of maximizing the consumer utility. The model uses hourly consumption data collected from smart-meters installed in the homes of residential consumers who are subjected to a dynamic pricing scheme. In subsequent sections, we will show that the utility coefficients calculated by the bi-level program model can be used to identify price-responsive consumers for DR programs. In addition, the model is developed at a per consumer level whereas most previous works have concentrated on a cluster of consumers. We propose a trust-region algorithm for solving the bi-level utility estimation model, based on a computationally efficient reformulation. We show a mathematical property of the optimal solution, which can be used as a cut in the algorithm. Extensive computational experiments demonstrate the validity and efficiency of the proposed TRA, especially showing substantial reduction on computational time when cuts are added to the re-formulation of the bi-level program.

5.2 Models and formulations

We propose a bi-level optimization model to estimate the utility coefficients for electricity consumers representing their level of price-responsiveness under a DR

program. The lower-level problem describes the electricity consumption behavior of the consumers given a DR pricing scheme where peak hour electricity rate is significantly higher than off peak hours. The optimization problem assumes users wish to maximize their total utility, i.e., the utility of consuming electricity minus the disutility of electricity cost. Such utility maximization is through optimal decision on curtailing usage on flexible load during peak hour. The upper-level problem aims to estimate the utility coefficients of users representing their varying degree of price-responsiveness, which will ultimately help utility firms to better understand their customers in designing a DR program. Such estimation is done through minimizing the least-square errors between users' estimated electricity usage (as a result of the lower-level problem) and their actual usage in a real-world DR pilot program.

5.2.1   Lower Level Problem

Consider a residential electricity consumer/household, who is subject to a time-of-use type of demand response pricing with a total of T (e.g., $T = 24$) time intervals. Let $\overline{\mathbf{T}}$ represent the set of time intervals during the peak period and $\underline{\mathbf{T}}$ represent that during the off peak period, where $\overline{\mathbf{T}} \bigcup \underline{\mathbf{T}} = \{1, \ 2, \ , \cdots, T\}$ and $\overline{\mathbf{T}} \bigcap \underline{\mathbf{T}} = \varnothing$. Under the time-of-use DR pricing, on any DR event day d, the electricity rate $\rho_d = (\rho_d^1, ..., \rho_d^T)^T$ must satisfy $\rho_d^{t_1} > \rho_d^{t_2}$ for $t_1 \in \overline{\mathbf{T}}$ and $t_2 \in \underline{\mathbf{T}}$. In response to such a DR pricing, a consumer can decide to curtail his/her electricity usage by the amount of $z_d^t = (z_d^1, ..., z_d^T)^T$ on any DR event day d ($d \in \{1, \cdots, D\}$). With such curtailment, the total electricity usage at time t is $f_d^t - z_d^t$, where f_d^t is the normal usage for the consumer at time t. In our computational experiment, f_d^t is estimated through a

random forest based prediction method using historical data collected by Advanced Metering Infrastructure (AMI). Note that z_d^t in our model can be positive or negative. A negative curtailment means the consumer increases their consumption likely during off-peak time $t \in \underline{\mathbf{T}}$ to compensate for the reduced usage during peak time $t \in \overline{\mathbf{T}}$. The consumption utility function is defined by $U(x)$ where x is the consumption. We consider the following quadratic utility:

$$U(x) = u_1 x - \frac{u_2}{2} x^2, \tag{5.1}$$

which satisfies the following properties:

$$\frac{\partial U(x)}{\partial u_1} > 0, \tag{5.2a}$$

$$\frac{\partial U(x)}{\partial u_2} < 0. \tag{5.2b}$$

Equation (5.2a) implies larger u_1 yields higher utility $U(x)$, due to higher level of electricity consumption. Similarly, no consumption translates to no gained utility, i.e., $U(0) = 0$, $\forall u_1 > 0$. The monotonicity expressed in (5.2a) allows us to rank customers based on parameter u_1. Equation (5.2b) helps to understand the price-responsive behavior of electricity consumers, which will be used in the sensitivity analysis.

Therefore, the lower-level problem for day d can be formulated as:

$$LL_d : \max_z \quad \sum_{t=1}^{T} U(f_d^t - z_d^t) - \rho_d^t(f_d^t - z_d^t), \qquad \forall d \qquad (5.3\text{a})$$

$$\text{s.t.} \quad \sum_{t=1}^{T} z_d^t \leq \sum_{t=1}^{T} f_d^t, \qquad \forall d \qquad (5.3\text{b})$$

$$\sum_{t=1}^{T} z_d^t \geq 0, \qquad \forall d \qquad (5.3\text{c})$$

$$z_d^t \text{ is unrestricted}, \qquad \forall t, \ \forall d \qquad (5.3\text{d})$$

The objective function (5.3a) maximizes the total utility over all time periods t minus the price paid for estimated consumption. Constraint (5.3b) imposes an upper bound on the total curtailed load over the time periods as the total predicted load, while constraint(5.3c) disallow an increase of total consumption over the entire planning period.

5.2.2 The Bi-Level Problem

The upper-level problem determines the utility function parameters u_1 and u_2, representing the price-responsiveness of electricity consumers. This is done by minimizing the gap between the estimated consumption $f_d^t - z_d^t$ and the actual consumption C_d^t, where C_d^t can be obtained from the AMI data for the local utility's system and $f_d^t - z_d^t$ is calculated from the lower-level problem. Note that without loss of generality, the notion of estimating the utility function parameters can be extended to utility functions other than the quadratic function (5.1) presented herein.

Let $\overline{U}_1$ and $\overline{U}_2$ be the upper bound of decision variables u_1 and u_2 in the upper

level, respectively, then the bi-level utility estimation (BLUE) problem is formulated as follows: Let $\overline{U}_1$ and $\overline{U}_2$ be the upper bound of decision variables u_1 and u_2 in the upper level, respectively, then the bi-level utility estimation (BLUE) problem formulated as follows:

$$BLUE : \min_{u_j} \quad \sum_{d=1}^{D}\sum_{t=1}^{T}(f_d^t - z_d^t - C_d^t)^2 \tag{5.4a}$$

$$\text{s.t.} \quad u_j \geq 0 \qquad\qquad \forall j = 1,\ 2 \tag{5.4b}$$

$$u_j \leq \overline{U}_j \qquad\qquad \forall j = 1,\ 2 \tag{5.4c}$$

$$\min_{z} \quad \sum_{t=1}^{T} u_1 z_d^t + \frac{u_2}{2} z_d^{t\,2} - \rho_d^t z_d^t, \qquad\qquad \forall d \tag{5.4d}$$

$$\text{s.t.} \quad \sum_{t=1}^{T} z_d^t \leq \sum_{t=1}^{T} f_d^t, \qquad\qquad \forall d \tag{5.4e}$$

$$\sum_{t=1}^{T} z_d^t \geq 0, \qquad\qquad \forall d \tag{5.4f}$$

$$z_d^t \text{ is unrestricted}, \qquad\qquad \forall t,\ \forall d \tag{5.4g}$$

For the above BLUE problem, the following theorem is established.

Theorem 5. *Let* $z^* = (z_d^1, ..., z_d^T)^T$ *be the optimum curtailed load for a given consumer for problem* LL_d, *then* z^* *must satisfy equation (5.5),*

$$-u_1 T + u_2 \sum_{t=1}^{T} f_d^t - u_2 \sum_{t=1}^{T} z^{*t}_d \leq 0, \qquad \forall d, \tag{5.5}$$

where f_d^t *is the predicted consumption at time t on day d.*

Proof: Without loss of generality, consider the lower-level problem LL_d for a single day i.e. $d = 1$ and hence omit d in the rest of this proof. Let $y_t = f^t - z^t$, the

lower-level problem can be rewritten as follows:

$$min \quad \sum_{t=1}^{T} (\rho^t - u_1)y_t + \frac{1}{2}u_2{y_t}^2 \tag{5.6}$$

$$\text{s.t.} \quad \sum_{t=1}^{T} y_t \geq 0,$$

$$\sum_{t=1}^{T} y_t \leq \sum_{t=1}^{T} f^t.$$

Letting λ_1 and λ_2 be the lagrangian multipliers for the two constraints in (5.6), respectively, yields the following Krush-Kuhn-Tucker (KKT) conditions at optimality:

$$\sum_{t=1}^{T} y_t^* \geq 0 \tag{5.7a}$$

$$\sum_{t=1}^{T} y_t^* \leq \sum_{t=1}^{T} f^t \tag{5.7b}$$

$$(\rho^t - u_1) + u_2 y_t^* - \lambda_1^* + \lambda_2^* = 0, \quad \forall t \tag{5.7c}$$

$$\lambda_1^* \sum_{t=1}^{T} y_t^* = 0 \tag{5.7d}$$

$$\lambda_2^* \sum_{t=1}^{T} (y_t^* - f^t) = 0 \tag{5.7e}$$

Adding all T (5.7c) equations, one obtains

$$Tu_1 - u_2 \sum_{t=1}^{T} y_t^* = \sum_{t=1}^{T} \rho^t + T(\lambda_2^* - \lambda_1^*). \tag{5.8}$$

Note that by adding equations (5.7d) and (5.7e), we have

$$\lambda_2^* - \lambda_1^* \geq 0. \tag{5.9}$$

Therefore, using (5.9) and (5.8), we obtain $Tu_1^* - u_2^* \sum_{t=1}^{T} y_t^* \geq 0$, or, $-Tu_1 + u_2 \sum_{t=1}^{T} f^t - u_2 \sum_{t=1}^{T} z^{*t} \leq 0$. $\blacksquare$

Given the above theorem, we add (5.5), as a cut, to the lower-level model LL_d

and obtain a more efficient lower-level model. We denote this lower-level model as

problem LLC_d. In Section 5.4.2, we will compare the performances the LL_d and LLC_d models when integrated in the BLUE problem.

5.3   Trust region algorithm to solve BLUE

Hansen et al. (1992) and Vicente et al. (1996) have shown that bi-level programming problems are strongly NP-hard even if all functions in the objectives and constraints are line. For nonlinear bi-level optimization problems, branch-and-bound methods (e.g., (?, ?)and (?, ?)) and penalty/descent methods (e.g., (Savard & Gauvin, 1994)) have been developed in the literature. The trust-region algorithm we develop for the BLUE problem is based on the work by Colson et al. (Colson et al., 2005) and falls into the branch-and-bound category.

Generally, in a trust-region method, an approximation of the original problem is solved at each iteration within some pre-defined region. If the solution gives an improvement with respect to the original objective, then the procedure continues with the region being expanded. On the other hand, an inferior objective function value leads to shrinking the region and computing based on a revised subproblem. Similar to Colson et al. (2005), our trust-region algorithm uses the first-order linear approximation to replace the original quadratic objective function in the upper level of the BLUE problem and uses a quadratic approximation to replace the original third-order polynomial objective function in the lower level. To do so, we obtain an initial feasible solution by solving the following single-level optimization problem, which in essence is the BLUE problem without the lower level objective function

(5.4d).

$$\min\{(5.4a)|(5.4b) - (5.4c), (5.3b) - (5.3d)\} \tag{5.10}$$

Suppose (u_1^*, u_2^*) is the optimal solution to (5.10). We then substitute (u_1^*, u_2^*) in the LL_d problem (5.3) and let $z^* = (z_1^{1*}, \cdots, z_1^{T*}, \cdots, z_d^{1*}, \cdots, z_d^{T*})$ be the stacked optimal solutions for solving $d = 1, \cdots, D$ independent lower-level problems (5.3). We then use (u_1^*, u_2^*, z^*) as the initial solution to obtain linear and quadratic approximation for the upper-level and lower-level objective function respectively for solving the BLUE problem.

In the BLUE problem (5.4), denote the upper-level objective function (5.4a) as $F(u, z)$ and the lower level objective function (5.4d) as $f(u, z)$. We develop the linear approximation of $F(u, z)$, $F_L(u, z)$, and the quadratic approximation of $f(u, z)$, $f_Q(u, z)$, using a solution (u^k, z^k) at iteration k. Let $\mathbf{e} = (1, \cdots, 1)^T$ and $\mathbf{I}$ be the

identity matrix, then we have the following:

$$
\begin{aligned}
F_L &= F(u^k, z^k) + \nabla_u F(u^k, z^k)(u - u^k) + \nabla_z F(u^k, z^k)(z - z^k) \\
&= \sum_{t,\,d} (-2f_d^t + 2z^{k^t}_d + 2C_d^t)(z_d^t - z^{k^t}_d)
\end{aligned}
$$

$$
\begin{aligned}
f_Q &= f(u^k, z^k) + \nabla_u f(u^k, z^k)(u - u^k) + \nabla_z f(u^k, z^k)(z - z^k) \\
&\quad + \tfrac{1}{2}(z - z^k)^T \nabla^2_{z,z} f(u^k, z^k)(z - z^k) \\
&\quad + \tfrac{1}{2}(u - u^k)^T \nabla^2_{u,u} f(u^k, z^k)(u - u^k) + (u - u^k)\nabla^2_{u,z} f(u^k, z^k)(z - z^k) \\
&= (z_d - z^{k^t}_d)^T(-\rho_d + \bar{u}_1 - u_2^k(f_d - z^{k^t}_d))\mathbf{e} - (f_d - z^{k^t}_d)^T(u_1 - u_1^k) \\
&\quad + (u_2 - u_2^k)^T \tfrac{1}{2}(f_d - z^{k^t}_d)^2 + \tfrac{1}{2}(z_d - z^{k^t}_d)^T \mathbf{e} u_2^k(z_d - z^{k^t}_d)
\end{aligned}
$$

The upper and lower-level objective function of the BLUE problem are then replaced by their linear and quadratic approximations respectively. As commonly done for the bi-level program, the KKT optimality conditions can be used to replace the lower-level problem and this converts the bi-level to the single-level problem. The latter however contains non-convex complimentary slackness constraints from the KKT conditions, which can then be replaced with big-M type of binary constraints. This ultimately converts the original bi-level problem to a single-level mixed integer program. Below is the mixed-integer program $SLMIP$ (5.11) approximation to the BLUE problem (5.4), where $\lambda_d = (\lambda_d^1, \lambda_d^2)$ are the Lagrangian multipliers for the lower level constraints (5.4e) and (5.4f), respectively.

$$SLMIP : \underset{u,\ z,\ \lambda,\ v}{\text{minimize}} \quad \sum_{d=1}^{D}\sum_{t=1}^{T}(-2f_d^t + 2\bar{z}_d^t + 2C_d^t)(z_d^t - \bar{z}_d^t)^T$$

s.t. $\quad u_j \geq 0, \ \forall j = 1, 2$ $\hspace{4cm}$ (5.11a)

$\quad u_j \leq \overline{U}_j, \ \forall j = 1, 2$ $\hspace{4cm}$ (5.11b)

$$\sum_{t=1}^{T} z_d^t \leq \sum_{t=1}^{T} f_d^t, \ \forall d \hspace{3cm} \text{(5.11c)}$$

$$\sum_{t=1}^{T} z_d^t \geq 0, \ \forall d \hspace{3.5cm} \text{(5.11d)}$$

$\quad \lambda_d^i >= 0, \ \forall i = 1, 2, \ \forall d$ $\hspace{3cm}$ (5.11e)

$\quad u_1 + (\bar{z}_d^t - f_d^t)u_2 + \bar{u}_2 z_d^t + \lambda_d^1 - \lambda_d^2$ $\hspace{2cm}$ (5.11f)

$\quad = -(-\rho_d^t + \bar{u}_1 - \bar{u}_2(f_d^t - \bar{z}_d^t)) + \bar{u}_1$

$\quad + (\bar{z}_d^t - f_d^t)\bar{u}_2 + \bar{u}_2 \bar{z}_d^t, \ \forall t, \ d$

$\quad \lambda_d^i <= Mv_d^i, \ \forall i = 1, 2, \ \forall d$ $\hspace{3cm}$ (5.11g)

$$\sum_{t=1}^{T} z_d^t - \sum_{t=1}^{T} f_d^t <= M(1 - v_d^1), \hspace{1.5cm} \forall d \hspace{1cm} \text{(5.11h)}$$

$$-\sum_{t=1}^{T} z_d^t <= M(1 - v_d^2), \ \forall d \hspace{2cm} \text{(5.11i)}$$

$\quad v_d^i \in \{0, 1\}, \ \forall i = 1, 2, \ \forall d.$ $\hspace{3cm}$ (5.11j)

In the above formulation (5.11), constraints (5.11a) and 5.11b) are the primal feasibility constraints while constraints (5.11c) and (5.11d) are the lower-level primal feasibility constraints. The dual feasibility of the lower-level problem is stated in (5.11e) and (5.11f). Constraints (5.11g) through (5.11j) are used to linearize the

complimentary slackness conditions of the lower-level problem, where v_d^i for $i = 1, 2$ are the binary variables for the big-M type of constraints. To integrate the trust region method, at each iteration, the trust-region constraints are added to the SLMIP problem (5.11):

$$u_j - u^k{}_j <= \delta_k, \ \forall j = 1, \ 2, \tag{5.12}$$

where δ_k is the radius of the trust region at iteration k. Table 5.1) displays the settings for all key parameters associated with the trust-region method for solving the BLUE problem. The outline of the trust-region method for the BLUE problem is presented in Algorithm 2.

Table 5.1: Trust-region algorithm parameters

Parameter	Notation
Trust-region radius	δ_k
Ratio of achieved vs. predicted reduction	r_k
Objective function improvement limit	η_1, η_2
Trust-region radius change	γ_1, γ_2
Maximum Iterations	K_{max}
Minimum radius	δ_{min}
Maximum unsuccessful iterations	Q_{max}

5.4   Computational results

We evaluate the proposed model and algorithms with computational experiments. We use the AMI data obtained from the real-world field demonstration demand response (DR) project. Since January 2016 the community has rolled out a dynamic pricing where electricity rate during peak hour (\$11.65 per kWh) is more than hundred folds of that (\$0.065 per kWh) during the off-peak hour. 3,000 participating homes were

Algorithm 2

1: **procedure** TRUST-REGION BASED UTILITY ESTIMATION ALGORITHM
2:     Initialize the problem by solving SL to get u_1^0 and u_2^0
3:     Solve d independent LL_d problems with u_1^0 and u_2^0
        from Step 1 to get z^0
4:     Initial feasible solution is $[u_1^0, u_2^0, z^0]$ from Step 1 & 2,
5:     Set iteration counter k, q to 0
        At each iteration k do,
6:     Solve $SLMIP$ along with (5.12)using z^k
        and $\bar{u} = u^k$ to get u^m and z^m
7:     Using u^m solve d independent LL_d to get z^*
8:     Calculate and update ratio of achieved vs. predicted reduction
       $r_k = \frac{F(u^k,\ z^k) - F(u^m,\ z^*)}{F_L(u^k,\ z^k) - F_L(u^m,\ z^m)}$
9:     **if** $r_k < \eta_1$ **then**   ▷ Check for improvement in solution
10:         $[u^{k+1}, z^{k+1}] = [u^k, z^k]$
11:         Decrease the radius δ_k in (5.12) by γ_1
12:         $q \leftarrow q + 1$
13:     **else**
14:         $[u^{k+1}, z^{k+1}] = [u^m, z^*]$
15:         **if** $r_k \geq \eta_2$ **then**
16:             Increase the radius δ_k in (5.12) by γ_2
17:         **end if**
18:     **end if**
19:     **if** $\left\| u^m - u^k \right\|_1 \leq \epsilon$ ▷ Check stopping criteria
        OR, $r_k \approx 1$
        OR, $k = K_{max}$
        OR, $\delta_k = \delta_{min}$
        OR, $q = Q_{max}$ **then**
20:         Stop,
21:         $[u^*, z^*] \leftarrow [u^m, z^*]$ as the solution of BL
22:     **else**
23:         $k \leftarrow k + 1$
24:         Go to Step 6
25:     **end if**
26: **end procedure**

---

equipped with smart appliances and AMI. The latter provides a rare real data set to study price-responsiveness of electricity consumers in a DR program and their utility attitude towards economic cost and life convenience.

This section consists of four parts. First, we evaluate the performance of various

algorithms for predicting households' energy consumption using several weather related factors and show that random forest provides the best performance. In the second and third subsections, we validate the trust-region algorithm and show its computational efficiency for the BLUE problem, respectively. Finally, we perform sensitivity analysis to draw managerial insights from our proposed model and algorithms.

5.4.1 Prediction of consumer load

We collected the AMI data with hourly resolution both for the year of 2015 and the year of 2016, before and after the implementation of the dynamic pricing. The 2015 data allows us to make predictions of households' energy consumption without DR pricing, i.e., f_d^t as in models LL_d and BLUE, while the 2016 data allows us to study consumers price-responsiveness when compared to their energy consumption before the DR pricing.

The electricity consumption inside a residence is dependent on multiple factors, some of which are general (e.g. weather factors such as temperature, dew-point) while some of which are specific (e.g. area, insulation level, and activity schedule). In this paper, due to confidentiality concerns, we are not provided with household specific information. Thus, in predicting energy consumption we only use temperature, humidity, dew point and day of the week as predictors. Figure 5.1 indicates there exists linear relationship between the electricity consumption in Summer 2015 (in the *y-axis*) and the temperature, humidity, and dew point. Also included in Figure 5.1 is the violin plots of energy consumption by day of the week. The widths of the violin

plots suggest that on Mondays and Thursdays the average hourly energy consumed by the given household is more distributed within the 1.5 - 2.5 kWh range, while on Wednesdays the hourly energy consumption for this household is more distributed within 0.5 - 1 kWh. This suggests to include the day of the week as a factor in the prediction model.

The above four factors: temperature, humidity, dew points and day of the week, are used to train four different models including linear regression (LR), decision trees (DT) (Breiman (2017)), random forests (RF) (Breiman (2001)), and k-Nearest Neighbor (k-NN) (Altman (1992)) to make predictions during the *off peak-periods* for Summer 2016. Because the DR pricing during that summer lasted for 3 hours, we have used the data for a 5-hour period (i.e., $T = 5$ in the $BLUE$ model), including 1-hour prior to and post the peak period and the 3-hour peak period. Note that households can choose to change their consumption behavior during the peak period as well as during the hours before and after according to their desired comfort level.

Figure 5.2 compares the results of the LR, DT, RF and kNN models for a single household and for 5 hours of a single DR pricing day. It suggests that the RF model outperforms all other three models as it provides the closet match with the actual consumption data. This is further confirmed by a different comparison analysis in Table 5.2, where the mean absolute percentage error (MAPE) is recorded for 80 homes during a period of 10 days (thus a total of 50 hours). Collectively, Figure 5.2 and Table 5.2 suggest to use the RF model to estimate the load f_d^t in the BLUE problem (5.4), and thereby the SLMIP problem (5.11). Therefore, we drop other prediction
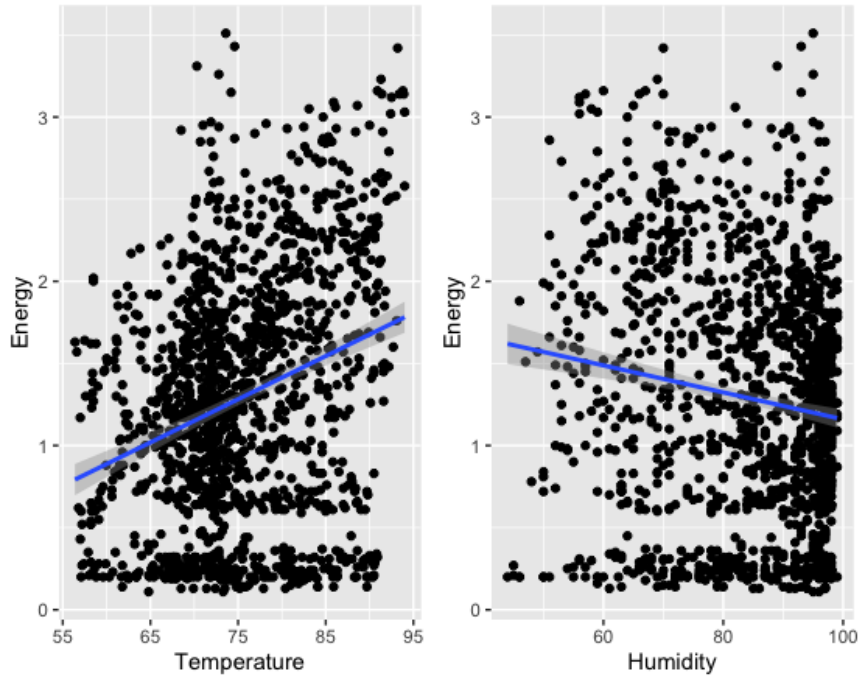
(a) Consumption vs. temperature & humidity
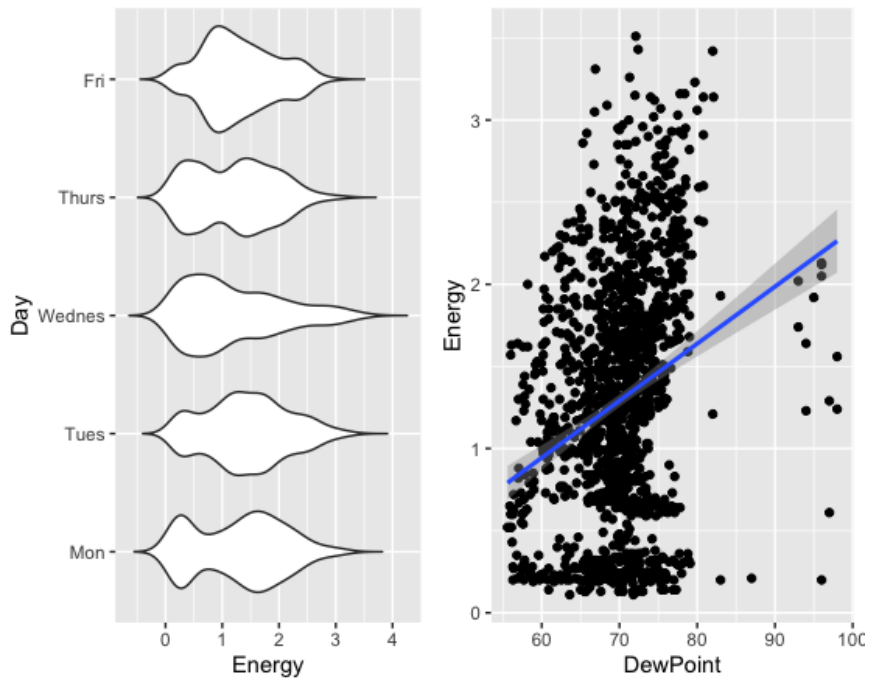
(b) Consumption vs. dew point & day of the week

Figure 5.1: Electricity consumption vs. four variables
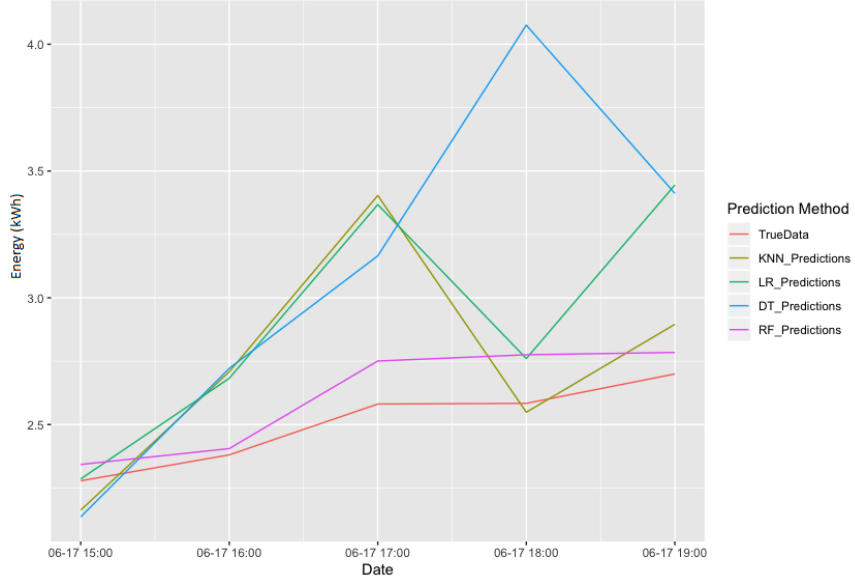
Figure 5.2: True vs. predicted consumptions for four models

models in our further modeling and analysis in the current paper.

Table 5.2: MAPE for four models

$Model$	$Accuracy\ (MAPE)$
Random Forest	93.7
Linear Regression	79.1
Decision Trees	84.3
k-NN	86.7

Table 5.3: Trust-region algorithm parameters values

Parameters	Values	Parameters	Values
δ_k	10	γ_2	1.4
η_1	0.01	K_{max}	10000
η_2	0.9	δ_{min}	1e-6
γ_1	0.6	Q_{max}	5

5.4.2 Validation and Performance of the Trust-Region Algorithm (TRA)

In this section we present computational results to validate the proposed trust-region

algorithm (TRA) from two aspects. First, we ran extensive preliminary tests to

ensure the TRA obtains accurate solutions for the bi-level model. To do so, suppose the optimal solution obtained by TRA is (u^*, z^*). We substitute both u_j^* for $j = 1, 2$ into the lower-level problem LL_d for day d, and solve the resulting single-level problem $LL_d(u^*)$, whose solution we denote as z_s^*. If $z^* = z_s^*$, then we conclude the bi-level problem (BLUE) is accurately solved by the TRA. This indeed is confirmed for all instances we tested. For example, for Home 5 on one day in Summer 2016 when total planning horizon is set to be $T = 5$, the TRA returns the optimal solution to the bi-level problem as $u_1^* = 10, u_2^* = 9.7374$ and $z^* = (0, 0.6368, 0.4145, 0.4454, -0.1146)$. Using u_1^* and u_2^*, we then obtained the single-level solution for the lower-level as $z_s^* = (0, 0.6332, 0.4115, 0.4418, -0.1141)$. This shows the TRA is able to solve the BLUE problem correctly. Second, we also evaluate the convergence of the proposed TRA. Figure 5.3 plots the upper level objective values against iterations of the TRA when run for an instance with $D = 10$ days (i.e., the BLUE model will have 10 independent lower level problems) for three homes. The figure suggests the proposed TRA converges after around 10-15 iterations.

To assess the computational efficiency of the TRA, we study the computational times for 20 homes. Table 5.4 presents the CPU times in seconds for the TRA to solve two alternative models: $BLUE$ and $BLUE - LLC_d$. The table indicates that, for all but one home (home #16), adding the cut (5.5) has substantially reduced the CPU time by a percentage ranging from 13% to 87%. On average the reduced CPU time is approximately 12.5 minutes or 750 seconds and the average percentage of reduction in CPU time is 53.6%.
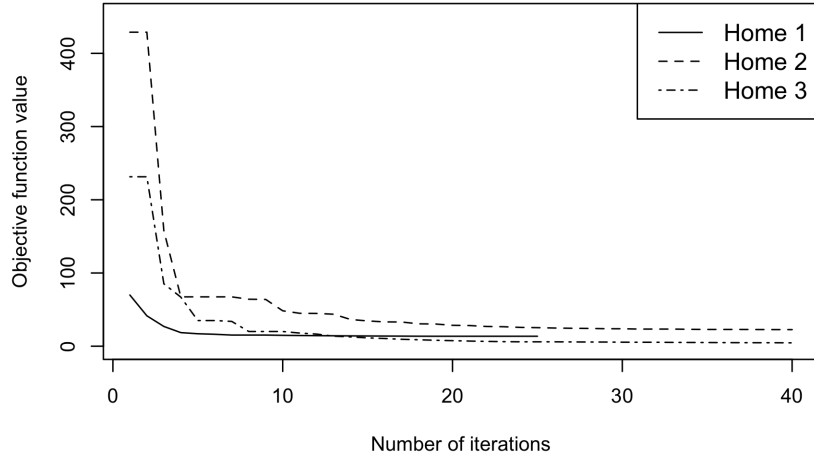
Figure 5.3: Convergence of TRA

We note that for three out of the 20 homes, different upper level objective functions were obtained by solving the $BLUE - LLC_d$, compared to solving the original $BLUE$ model. For example, for Home 7, $F^* = 46.2492$, $F_c^* = 54.3069$, where F^* and F_c^* are the optimal upper level objective values for models $BLUE$ and $BLUE - LLC_d$, respectively. In another instance for Home 13, $F^* = 5.8540$, $F_c^* = 21.4073$.

Investigating the discrepancies of these values has found that for these were caused by the TRA getting stuck in the local optimum. Hence, the trust region radius δ_k were increased from 10 to 100, which ultimately help the algorithm to reach the optimal solution. Figure 5.4 demonstrates the results for Home 4. The sub-plot (a) indicates that the upper-level decision variables are stuck in the local optima for the problem $BLUE - LLC_d$ while the sub-plot (b) suggests the algorithm successfully got out of the trap after increasing δ_k.

Table 5.4: Computational time of the TRA

Home #	BLUE	BLUE−LLC$_d$	difference (in %))	Home #	BLUE	BLUE−LLC$_d$	difference (in %)
1	66.61	45.59	32	11	566.41	144.29	75
2	329.76	73.73	78	12	4462.2	571.34	87
3	548.14	173.65	68	13	1096.9	624.52	43
4	142.95	32.43	77	14	4583.4	1443.5	69
5	234.77	83.07	65	15	1393.8	566.73	59
6	675.12	223.19	67	16	635.75	753.19	-18
7	1923.1	1113.14	42	17	5190.3	1663.23	68
8	38.93	23.57	39	18	107.21	55.98	48
9	52.06	31.1	40	19	148.06	128.89	13
10	619.26	184.61	70	20	218.35	106.55	51
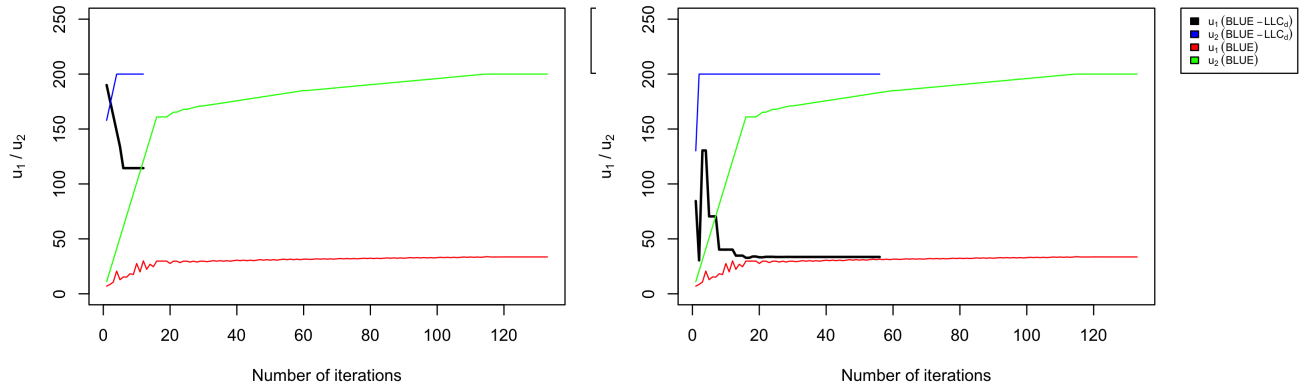
(a) TRA performance with $\delta_k = 10$ (b) TRA performance with $\delta_k = 100$

Figure 5.4: Effects of Increasing δ_k

5.4.3 Sensitivity Analysis

In this section, we correlate the two utility parameters under study with the commonly used economics measure, i.e., price elasticity; the latter captures the demand elasticity of a product or a service in response to the change in price. This allows us to conclude that parameter u_2 estimated by the BLUE model is a valid surrogate for measuring electricity consumers' price-responsiveness. Thereby, we argue that the quantitative

approach with the BLUE model in this paper can be very helpful for utility firms to understand their consumer's behavior and enable them to better design a customized DR pricing scheme.

In the energy market and economics literature, benefit function is used to quantitatively describe electricity consumers' economical gain through consuming electricity under a certain pricing policy. We adopt the following quadratic benefit function in Schweppe et al. (2013):

$$B(x) = B(x_0) + \rho_0(x - x_0)(1 + \frac{x - x_0}{2x_0\beta}), \tag{5.13}$$

where x is the energy consumed, β is the price elasticity of demand, and ρ_0 and x_0 are the nominal pricing and consumption of electricity, respectively. In our case, they represent the off-peak electricity price and individual homes' consumption during off-peak period.

On the other hand, the net benefit for an electricity consumer is commonly calculated as the benefit gained less the electricity cost, i.e.,

$$NB(x) = B(x) - \rho x, \tag{5.14}$$

where $B(x)$ is the benefit function of consuming x amount of electricity, and ρ is the proper electricity price during the study period. In our case, it is the peak pricing as we focus our analysis on consumer's price-responsiveness behavior during DR event hours.

Substituting (5.13) into the net benefit function (5.14), one obtains:

$$NB(x) = B(x_0) + \rho_0(x - x_0)(1 + \frac{x - x_0}{2x_0\beta}) - \rho x.$$

We assume electricity consumers wishes to maximize their net benefit, therefore $\frac{\partial NB(x)}{\partial x}|_{x^*} = 0$ at optimality can be used to characterize their behavior in response to changing price. Thus, the price elasticity for energy consumption satisfies the following:

$$\beta = \frac{(x - x_0)\rho_0}{(\rho - \rho_0)x_0}. \tag{5.15}$$

As expected, the price elasticity β calculated by (5.15) is negative during the DR event window as the consumption is less than the nominal consumption $(x < x_0)$ in response to prince spike during peak $(\rho > \rho_0)$.

Table 5.5: Homes sorted by price-responsiveness

Home #	$\frac{1}{N}\sum_{d,\, t} \frac{z_d^t}{f_d^t}$ (%)	β	u_2	Home #	$\frac{1}{N}\sum_{d,\, t} \frac{z_d^t}{f_d^t}$ (%)	β	u_2
4	95.74	-0.09311	332.5	7	15.56	-0.01365	25.01
13	62.17	-0.05572	116.95	3	13.74	-0.0115	18.16
16	35.81	-0.04683	79.41	20	12.14	-0.01035	16.61
19	23.35	-0.04183	72.60	9	10.97	-0.00925	16.25
12	21.02	-0.04045	69.68	10	10.49	-0.00909	13.91
15	19.41	-0.03148	69.10	17	9.04	-0.00078	13.23
14	18.43	-0.02856	40.31	2	8.34	-0.00057	12.62
11	17.58	-0.02141	30.85	8	7.33	-0.00051	12.53
18	17.36	-0.01845	26.87	6	5.53	-0.00028	11.05
1	15.61	-0.01732	25.79	5	4.46	-0.00009	6.97

In our computational experiment, x is the total consumption during the DR event hours, i.e., $\sum_t f_d^t - z_d^t$ for $t \in \overline{\mathbf{T}}$ peak periods and x_0 is the predicted load assuming

off-peak pricing, i.e., f_t^d. Therefore, the price elasticity is empirically calculated by

$$\beta = \frac{\sum_{d\in D,\ t\in\overline{\mathbf{T}}} \frac{-z_d^t \rho_0}{f_t^d(\bar{\rho}-\rho_0)}}{D\overline{T}},$$

where $\overline{T}$ is the number of time intervals in the peak period on a given day, and $\bar{\rho}$ and ρ_0 are the electricity prices during the peak and off-peak periods, respectively.

Table 5.5 displays the price elasticity β, the utility parameter u_2 obtained from the TRA as well as the average percentage of curtailed load (APCL) relative to the baseline consumption without DR pricing. All three measures were calculated for 20 homes and were arranged in the table in descending order with respect to the APCL. Several observations can be made. First, all homes show some level of price-responsiveness with their β's being negative. Second, as expected, among all homes, the trend for the APCL is consistent with that for the price elasticity. A home with higher APCL represents a home with higher price-responsiveness, thus higher value for u_2. Second, our estimated values of u_2 also correlates well with the widely used price elasticity value β. This validates the utility function defined in the paper and calculated by the proposed TRA. This positive correlation is further confirmed by Figure 5.5 where u_2 is plotted against the absolute value of β.

Finally, we randomly chose two representative homes: Home 13 as one with high level of price-responsiveness and Home 8 with low level of price-responsiveness. For both homes, we vary the price differential between the peak and off-peak periods and see how the two homes respond these varying differentials. In particular, let O_p be the original differential between rho_0 and $\bar{\rho}$. We ran experiments with six other
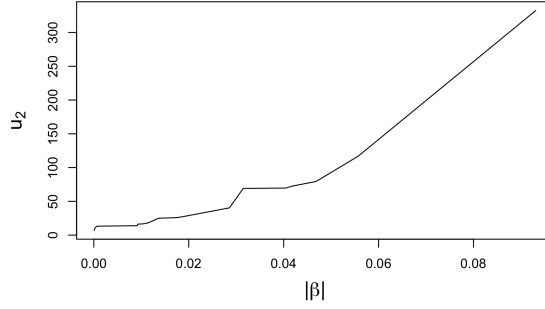
Figure 5.5: u_2 vs. $|\beta|$

alternative differentials being: $0.25O_p$, $0.5O_p$, $0.75O_p$, $1.25O_p$, $1.5O_p$ and $2O_p$.

The plots in Figure 5.6 are based on results for 3 DR event days each with 5 event hours, where the y-axis is the percentage of curtailed load relative to the baseline consumption without DR pricing, i.e., $\frac{z}{f}\%$, and x-axis is the time interval for a total of 15 DR event hours. From Figure 5.6a, we observe that for the more price-responsive Home 13, the curtailed load percentage changes significantly for all time intervals ranging from 25% to 70% under the baseline price differential O_p. It can be seen that as the price becomes more similar (e.g., under $0.5O_p$ and $0.25O_p$) in sub-figure 5.6a, the percentage of curtailed load decreases. The opposite behavior is observed as the difference is widened (e.g., under $1.5O_p$ and $2O_p$). On the other hand, the bottom sub-figure 5.6b shows that for a low price-responsive Home 8, the curtailed load percentage does not vary significantly with either the increase or decrease in the price differential. This shows the BLUE model effectively captures the price-responsiveness of the electricity consumers and is sensitive to the price differentials while preserving the inherent behavior of the consumers.
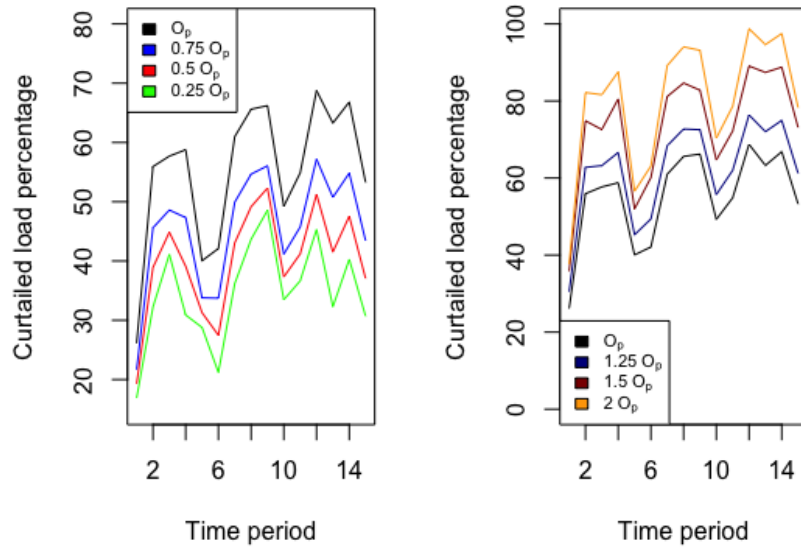
(a) Curtailed load percentage over 3-day period (Home 13)
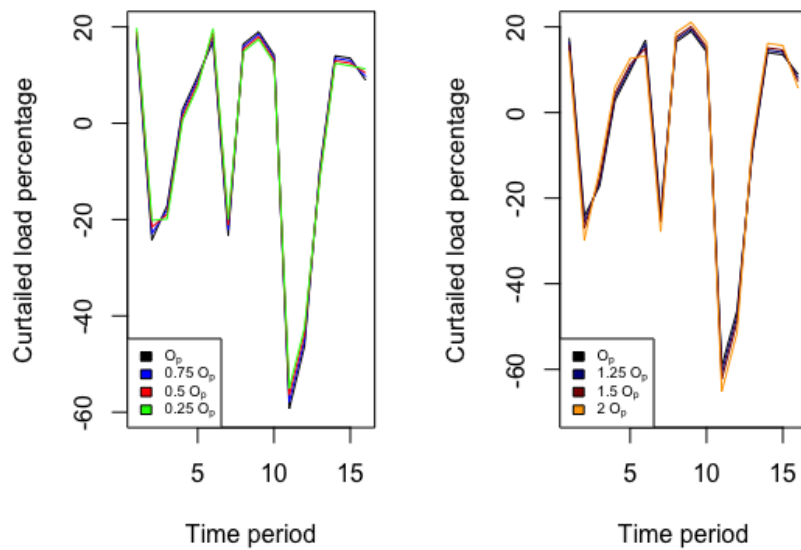
(b) Curtailed load percentage over 3-day period (Home 8)

Figure 5.6: Curtailed load percentage under various price differentials

5.5   Conclusions

In this chapter, we propose to use utility function to describe electricity consumers'
behavior and assumes consumers are to maximize their associated utility function
in their electricity consumption decision making when faced with dynamic pricing
in a demand response program. Thereby, we consider the problem of estimating
coefficients in the utility function. Particularly, we consider a quadratic function with
two coefficients: one representing the utility of consuming electricity and the other
representing consumer's price-responsiveness under dynamic pricing in a DR program.
The problem is formulated as bi-level utility estimation model (BLUE) where the
upper level program, in some form of inverse optimization, estimates the utility
coefficients by minimizing the error between the estimated and true consumption. The
lower level problem describes consumers' energy consumption behavior by maximizing
their total utility through properly curtailing electricity consumption during *peak-
pricing* periods, depending on their own utility towards consumption and price.

We approach the original BLUE by using a linear approximation for the upper
level problem and a quadratic approximation for the lower level problem. The
resulting of the bi-level program is then reformulated as an equivalent mixed integer
program (MIP). We then develop a trust region algorithm to solve the MIP model
efficiently, especially after introducing cuts into the re-formulation of the bi-level
problem model. The model is applied on the data collected from a real-world field
demonstration DR project. First, a random forest model is developed using the

hourly consumption AMI data to predict household's electricity consumption with weather factors and day-of-week as independent variables. Second. the proposed trust region algorithm (TRA) is validated by comparing its solution with a single-level optimization problem at optimality. Third, we show that adding a cut to the MIP model reduces the computational time by 53% on average over 20 BLUE instances. Finally, we conduct sensitivity analysis to show a strong positive correlation between the utility coefficients from the proposed BLUE model and the widely used price elasticity in economics when applied to our AMI data. This validates the use of the proposed model by utility companies to help them understand consumer's price-responsiveness under dynamic pricing.

Several future research directions can be considered. First, other forms of utility function can be considered in a study more focused on energy economics. For example, a logarithmic function has been used in the literature for modeling risk aversion of the consumer. Second, the lower-level problems for different days can be integrated into a single problem by considering energy consumption dynamics from day to day. Finally, it is interesting to develop a machine learning algorithm to predict electricity consumer's price-responsiveness without employ optimization models.

CHAPTER 6

AN EMPIRICAL STUDY OF SMART ENERGY TECHNOLOGIES IN RESIDENTIAL HOUSEHOLDS

6.1   Introduction

Alongside the development of connected technologies such as programmable thermostats and AMI to improve communication between the electricity utility and the consumer, regular household appliances such as HVAC systems and heat pump water heaters have also become more efficient. A municipal utility company in collaboration with a major appliance manufacturer initiated a demand response field study in midwestern U.S. in 2016. One of the objectives of the pilot program was to analyze the effect of efficient smart technologies in energy savings under a dynamic pricing model. The program included a high cost peak period. Consumers would be informed beforehand about the period. Essentially, the critical peak pricing (CPP) rate would discourage consumers to use electricity during forecasted peak period and lead to reducing stress on the resources. Smart technologies were installed in multiple homes in different levels to assess their role in energy saving and coincident load reduction for the pilot DR program.

This work aims to explore the cyber-physical interactions in residential power

distribution by employing the advanced technologies in a DR environment. These include wifi-enabled programmable smart thermostats, high-efficiency and connected water heaters, battery storage systems, improved weatherproofing and advanced metering infrastructure (AMI) to monitor the load real-time. We organize the chapter as follows. The design of the experimental study is presented first followed by the data collection methods. This is followed by the methodology to compare the coincident load of a household which includes the day-matching algorithm for the paired t-test. Finally, we run the models on the real-world data to assess the advantages of the installed smart electric technologies and the dynamic pricing rate. We finish the chapter with some concluding remarks.

6.2 Models

6.2.1 Experiment design and data collection

Residential consumption is mainly driven by convenience resulting in conincident demand that require ancillary services at power plants to satisfy the peak. Successful deployment of smart electric technologies with the suitable DR pricing program presents significant opportunities toward a sustainable power system. The study aims to test this notion on small set of residential consumers. The authors collaborated with a midwestern utility company along with the independent system operator (ISO) of the region. 330 households volunteered to be part of the program. The advanced technologies in question are smart thermostats (wi-fi connected and ability to operate remotely), advanced heat pump water heaters, ultra-efficiency home envelope, and

residential battery systems. Alongside these, a critical peak pricing (CPP) rate was also enforced on the customers. In such a program, consumers are charged a higher rate during the system peak period. This period is determined through forecasting methods by the utility companies or other agencies and are conveyed to the customer ahead of time. Such pricing encourages the customers to reduce consumption during the peak period and thus improving the system load factor. A two-fold objective was established; the first to determine the change in annual consumption because of the installation of smart electric technologies and the second to observe the effect of CPP rate with supporting technologies on the peak load.

The participant consumers were clustered into different groups according to the type of appliance installed. This will enable us to determine the effect on consumption of different technologies. The five groups formed were as follows:

- Ultra - *installed with every technology including new enveloping*

- Advanced Gas - *installed with every technology excluding new enveloping, battery, and having gas HVAC system*

- Advanced Electric - *installed with every technology excluding new enveloping, battery, and having electric HVAC system*

- Basic Gas - *only programmable thermostats, AMI, HPWH installed, and having gas HVAC system*

- Basic Electric -*only programmable thermostats, AMI, HPWH installed, and having electric HVAC system*

In addition, Table (6.1) identifies the distribution of the participants among the groups defined above according to type of HVAC system. The dual over here means having both electric and gas HVAC system.

Table 6.1: Home sub groups

Ultra	Advanced			Basic		
electric only	dual	electric	gas	dual	electric	gas
50	2	95	18	2	37	126

Weather factors play an important role in residential electric load. Any comparison of consumption between two periods should incorporate the attributes of weather. As such, we develop our comparison models centered around such factors. The smart technologies were installed in the homes in the year 2016-2017 along with the deployment of the CPP rate structure. We aim to compare the load during this year with the benchmark year 2014-15 when no such technology was introduced. However, AMI was present in both the cases to collect data at selected intervals.

Two major sets of data were requested for the analysis. The first set contains the AMI data with one-hour resolution in 2016-2017, as well as the AMI data in the benchmark year 2014-2015. In addition, we request weather data containing temperature, humidity, pressure, dew point and wind speed from the local weather station for the same time periods. The time interval in the weather data we have obtained is usually inconsistent; therefore, we have rounded the time to nearest 5-minute intervals. A sample AMI data for a prticular home is provided in Table (6.2) while Table (6.3) includes a snippet of the weather data.

Table 6.2: Sample AMI data

Home ID	Time	kWh
001	8/1/2016 0:00	1.512
001	8/1/2016 1:00	1.341
001	8/1/2016 2:00	1.266
001	8/1/2016 3:00	0.609
001	8/1/2016 4:00	0.561
001	8/1/2016 5:00	0.444
001	8/1/2016 6:00	0.579
001	8/1/2016 7:00	0.591
001	8/1/2016 8:00	0.339
001	8/1/2016 9:00	0.966
001	8/1/2016 10:00	1.803
001	8/1/2016 11:00	2.121

Table 6.3: Sample weather data

Date	Time	Temperature (°F)	Humidity (%)	Pressure (in)	Dew point (°F)	Wind speed (mph)
8/1/2016	0:00	74.8	95	30.00	73.3	0
8/1/2016	1:00	74.8	95	29.99	73.3	0
8/1/2016	2:00	74.7	95	29.99	73.2	0
8/1/2016	3:00	74.6	95	29.99	73.1	0
8/1/2016	4:00	74.6	96	29.99	73.4	0
8/1/2016	5:00	74.5	95	29.99	73.0	0
8/1/2016	6:00	74.4	96	29.99	73.2	0
8/1/2016	7:00	74.3	95	29.99	72.8	0
8/1/2016	8:00	74.1	95	29.99	72.6	0
8/1/2016	9:00	74.0	96	29.99	72.8	0
8/1/2016	10:00	73.9	96	29.99	72.7	0
8/1/2016	11:00	73.9	96	29.99	72.7	0

6.2.2 Concident load analysis

Coincident load is defined as the consumption of a particular household during the peak period. The metric gives a measure of the contribution of each home towards the peak. We want to examine the effects of the technologies and the CPP rate on the conincident load for each category of homes.

Comparison of the coincident loads of each home in each category can be accomplised through statistical testing such as *paired t-test*. But to conduct such a test we need to identify a corresponding load for each home to compare with. Since this load has to represent the consumption without the technologies, it has to be on a day before their installation or implementation. Thus we need to identify a day with similar weather conditions from a year before i.e. the profile for different weather attributes (e.g. temperature, humidity) for the paired days has to be close.

We denote the day to be matched i.e. the peak day as the target day and for each such day, we will have multiple candidate days to pair up. The final matched day is selected from the candidate days according to some metric. Recall in Table 6.3, five weather conditions are recorded in the original weather data. Among the five, we select three main factors, i.e., temperature, humidity and wind speed, in creating matching pairs between two days. We present the justification of such selection in Section(ref). The mean squared error between the target day and a candidate matching day is calculated over the weighted sum of squared error of each weather component for each time interval $t \in T$ where T is the index set of time intervals. In addition, the three factors have different scales, therefore we apply the feature scaling method (see eq. (6.1)) to normalize their values into the range of $[0, 1]$ so that they can be properly integrated into one index to represent the difference between target and matching days' weather. Note that in equation (6.1), x and x' represent the original and scaled values, respectively, while a and b represent the lower and upper

limits of the scaled interval, respectively. In our particular case, $a = 0$ and $b = 1$.

$$x' = a + \frac{(x - x_{min})(b - a)}{x_{max} - x_{min}} \tag{6.1}$$

The algorithm to determine a matched day for a target day d is given below. $i \in I$ is the set of weather components and $f_{d,t}^i$ is the value of component i for day d and for time interval $t \in T$. w_i is the corresponding weight of factor i.

Algorithm 3

1: **procedure** DAY MATCHING ALGORITHM
2:    Set a target day d
3:    **for** each candidate day d' in a ϵ day window ($\pm\epsilon$ days) in the previous year **do**
4:       **for** each time interval t **do**
5:          Calculate the weighted sum of squared error of weather factors:
             $SE_{d',t} = \sum_i w_i (f_{d,t}^i - f_{d',t}^i)^2$
6:       **end for**
7:       Calculate the mean squared error $MSE'_d = \frac{\sum_t SE_{d,t}}{|T|}$
8:    **end for**
9:    Choose a set of candidate matching days D'' such as,
       $MSE_{d''} \leq \underset{d'}{\mathrm{Minimum}}\{MSE'_d + \rho\} \quad \forall d'' \in D''$
10:   Select the matching day d^* from D'' such as,
       $d^* = \underset{d''}{\mathrm{argmin}}\{|Temp_{d''-1} - Temp_{d-1}|\}$
       where $Temp_d$ is the average temperature of day d
11:   Return d^* as the matched day for target day d
12: **end procedure**

We select a set of candidate matching days in Step 9 of Algorithm (3) because analysis of the weather data revealed that previous day weather has a significant effect on the weather of a particular day. Of all the weather factors analysed, temperature is the most co-related one to electric consumption and hence only temperature is taken as a selection criterion in Step 10 (see Section). The values of the paramters w_i, ρ, and ϵ varies according to application.

Once we have a matched day d^* for the target day d, we can find the coincident load for each such day. There are multiple system peak days reported in a month. Here, system peak refers to a day on which the demand is significantly hihjer than a regular day. These days are usually forecasted beforehand. There are multiple system peak days reported in a year and even in a single month especially during extreme seasons. Each such peak day constitutes a target day in Algorithm(3). Coincident load for each such target day and the corresponding matched day is determined. A paired t-test is then conducted on this set to determine the statistical significance of the difference. The tests are conducted by each home subgroup. This will reveal whether installation of the new technologies and implementation of the CPP rate have motivated the customers to change their consumption pattern. We present the numerical results of the empirical study in the following section.

6.3 Computational results

To conduct the analysis for coincident load comparison, we select the system peak days reported by the utility company for the period of September 2016 to September 2017. Each such target day is matched with day in 2015 having similar weather conditions using Algorithm (3). In Figure (6.1), we observe the temperature and humdity profile of target day 8/9/2016 and paired/matched day 7/29/2015. The profiles of the two days follow the same pattern and hence make them suitable for conducting the paired t-test. The difference between the coincident load on each of target and paired date is used to conduct the paired t-test.
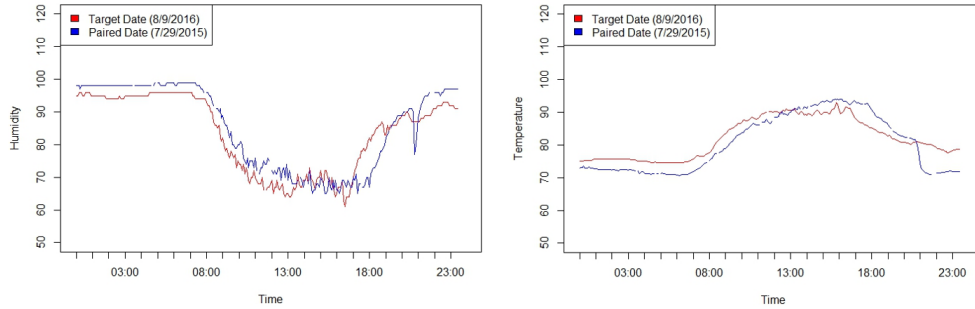
Figure 6.1: Comparison of weather conditions between two matching days

An illustration of such an analysis is provided in Table (6.4) with column 'Target Dates' indicating the peak days of 2016. 'Average load reduction' reports the coincident load reduction averaged over each category of home for each event day. A p-value less than or equal to 0.05 indicates this average coincident load reduction from 2015 to 2016-2017 is statistically significant. There are 48 event days from September 2016 to August 2017, among them one for ultra group, five for advanced electric group, 18 for advanced gas group, 10 for basic electric group and two for basic gas group have yielded p-value larger than 0.05 in the paired t-tests. This shows that the statistical power for analyzing advanced gas and basic electric groups are not great due to smaller sample sizes for the two groups.

We pool all winter event days and all summer event days separately and study the coincident load reductions broken down by seasons and by home types in order to better understand the impacts of technologies on coincident load reduction. Peak days which yield an insignificant p-value have been removed from this seasonal analysis. Table (6.5) shows the average, min, and max coincident load reductions over event days in winter (and summer) seasons during the study period. Several observations can be made from Table (6.5). First, all electric homes show consistent performances

Table 6.4: Sample AMI data

Target Dates	Home Type	p-value	Average load reduction
9/7/2016	Ultra	$1.83e{-}13$	2.407
9/8/2016		$1.33e{-}13$	2.557
9/9/2016		$1.45e{-}12$	2.359
9/7/2016	Advanced Electric	$2.74e{-}11$	1.665
9/8/2016		$3.37e{-}24$	2.782
9/9/2016		$4.21e{-}21$	2.380
9/7/2016	Advanced Gas	$2.77e{-}06$	2.771
9/8/2016		$2.68e{-}06$	2.755
9/9/2016		$6.86e{-}06$	2.393
9/7/2016	Basic Gas	0.0030	1.092
9/8/2016		0.0012	1.447
9/9/2016		0.0007	1.513

between two seasons. Ultra and advanced electric homes post similar higher reduction of 2.3-2.5 kW per event day per home, while basic electric homes post approximately 1.4-1.5 kW per event day per home. Second, on a given event day, all SET residential sites collectively are estimated to reduce peak load by 667 kW in the summer and 542 kW in the winter. Third, the coincident load reduction for basic electric is consistently less than ultra and advanced electric homes due to the non-existence of the residential battery system in these homes. Finally, as expected, gas-heated homes show approximately 50% less coincident reduction during winter than summer (e.g., for basic gas homes, 1.8 kW coincident load reduction per event day per home in the summer compared to 0.9 kW reduction in the winter).

We also attempt to examine the variability for coincident load reduction for three groups with sufficient sample size, i.e., ultra, advanced electric and basic gas, for summer and winter respectively. Particularly, we choose 8/22/2017, the system peak day August 2017, to represent summer and 2/3/2017, the system peak day in

Table 6.5: Summary of coincident load comparison in seasons

Seasonal coincident load reduction per home per event day			
Summer	Average	Minimum	Maximum
Ultra	2.5	1.4	3.2
Adavanced electric	2.3	0.5	3.0
Advanced gas	2.8	1.0	3.4
Basic electric	1.4	1.0	2.1
Basic gas	1.8	1.1	2.6
Winter	Average	Minimum	Maximum
Ultra	2.5	0.1	6.0
Adavanced electric	2.4	0.01	4.4
Advanced gas	1.2	0.4	1.9
Basic electric	1.5	0.7	2.6
Basic gas	0.9	0.4	1.6
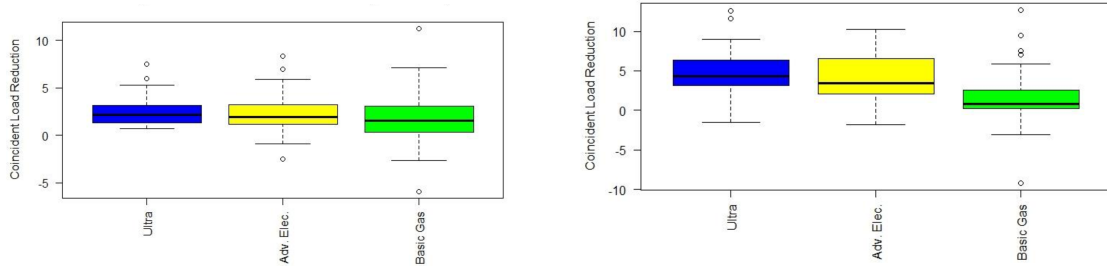
(a) Coincident load comparison for summer (b) Coincident load comparison for summer

Figure 6.2: Coincident load comparison by seasons and by home type

February 2017, to represent winter. Figures 8 and 9 use home-level coincident load reduction data on these two representative summer and winter days as sample to draw box plots for the three groups. Collectively, these two figures indicate that: 1) ultra homes have smallest variations on coincident load reduction among the three; and 2) the median coincident load reductions for three groups are similar in the summer while basic gas homes clearly have lower reduction in the winter when compared to the other two groups. Note that certain homes have demonstrated a negative reduction between the two years, however, they all lie in the bottom 25th quartile of

each subgroup.

We offer some remarks to summarize the results presented above. Our analysis revealed that ultra homes which in addition to the advanced technologies had efficient weatherproofing installed had the highest energy savings among all the categories. The savings translate to approximately 4300 kWh per home when extrapolated to a whole year. Homes with electric HVAC systems have shown greater savings in terms of coincident load. Ultra and advanced electric homes post similar higher reduction of 2.3-2.5 kW per event day per home in both seasons of summer and winter, while basic electric homes post approximately 1.4-1.5 kW per event day per home. The better performance of such homes can be attributed to the battery systems. Ultra homes show the most consistent behavior throughout the two seasons. The homes together are estimated to reduce 667 kWh and 542 kWh of energy from the peak. It is noteworthy that homes having gas powered HVACs significantly shave less load from the peak.

6.4 Conclusions

In this chapter, we aim to evaluate the effectiveness of residential smart electric technologies toward energy savings and coincident load reduction. A real-world study is conducted in collaboration with a municipal utility company. Though a number of empirical studies have been conducted in analyzing a direct load control architecture yet there have been no standard models developed to assess the effectiveness of these smart technologies. The model is utilized here to assess the change in coincident load

of each such home. Homes with electric HVAC systems reduce more coincident load consistently than gas powered ones. In addition, homes having battery systems for energy storage perform better than the ones without such facilities. The method estimates that the technologies along with the dynamic pricing can significantly reduce the peak load.

CHAPTER 7

CONCLUSIONS AND FUTURE RESEARCH

7.1 Conclusions

Smart grid is one of the pillars to facilitate our transition to a more sustainable and environment-friendly power system for the future. Demand side management as a service is becoming popular as it provides numerous opportunities for utility companies towards better management. Among many DSM schemes, demand response promotes market-based power distribution and transaction between utility companies and electricity end consumers. In this dissertation, we focus on investigating possibilities of load management for residential consumers. The motivation stemmed from participation in a field demonstration demand response project with a midwestern utility company.

The dissertation is composed of three parts. First, we present a direct load control framework for residential consumers which concentrates on energy savings for each customer and overall efficiency for the system. From the real-world study, we could observe that reaction to dynamic pricing is different for different consumers. Finally, based on the AMI data collected from the same field demonstration project, we develop an empirical model and a statistical approach for assessing the effects of

various smart energy technologies on energy savings and coincident load reduction. Next we detail the main conclusions for these three parts of the dissertation.

The first part of the dissertation is dedicated to developing a control algorithm to determine the optimal set-point of a residential HVAC system. Predictive controllers such as MPC require a reference temperature to determine the optimal control actions. Smart thermostats have enabled remote control of HVAC systems via the internet by changing the temperature set-point. We begin with presenting a dynamic programming model to estimate the optimal temperature set-points of a smart thermostat during a peak period (TSEP). A CPP rate is applied during the peak which makes consumption of less energy desirable. Hence the objective of the model is to consume the least energy during the peak. Adjoining to the peak period we consider two equal-length control periods as the pre-cool and post-cool, respectively. TSEP allows the inside temperature to hover between the maximum and minimum preferred temperature during these periods. Inside temperature is one parameter that represent users' experienced comfort. The pre-cool period is utilized to lower the inside temperature to the lowest preferred point consuming the least energy as possible while the temperature is let to rise during the peak period to the highest preferred point. This results in minimal consumption during the peak. The temperature is again lowered to the preferred point during post-cool. Application of the model to multiple homes along with analyses of different start and end times reveal the relationship between the insulation quality of such buildings and the control duration.

Application of control actions which takes into account only individual household

and not the total system may not achieve system-wide load leveling, which is the ultimate goal of a DR program. For example, application of control actions aimed towards individuals' electricity cost minimization by discouraging consumption during the forecasted peak period may result in shifting the peak. This is certainly undesirable for the utility companies. We study a control scheduling problem, a mixed integer linear fractional programing problem, to optimally deploy the TSEP controllers to a pool of homes in a staggered fashion so as to maximize load leveling or load factor for the distribution network. The control duration or the start and end times of TSEP for individual households are are decision variables and maximizing system load leveling or load factor is the objective. A bi-section search algorithm is developed to solve the CSP efficiently. The algorithm is capable of providing better results when compared to CPLEX and BARON for an instance of 132 homes and 9 alternative start and end times using much less computational time. When the number of binary variables is in the order of tens of thousands and general purpose solver cannot solve the CSP in reasonable CPU time, the propose a Lagrangian relaxation approach to estimate the bound thus optimality gap of the solution by the bi-section search algorithm. Computational results show that the optimization model is capable of flattening the load profile of multiple homes during the pre- and post-cool period. During these periods the control algorithm TSEP is applied to all the homes with different start and end times ensuring minimal consumption during the middle peak period.

In conducting the field demonstration of smart technologies and demand response programs, we observed that success of a load management program depends very

much on the price-responsiveness of the participants. Electricity as a commodity is very much related to the comfort of it's consumers and hence holds a different value for each individual. We assume that utility functions can describe the satisfaction gained from electricity consumption with the coefficients providing a quantitative assessment of their price-responsiveness. A bi-level optimization model is presented to estimate the utility coefficient values by utilizing consumption data from AMI in the field demonstration project. The upper level objective is to minimize the estimation error between the measured data and the optimum consumption while the lower level has the objective of maximizing the consumer utility. We propose a trust-region algorithm to solve the non-linear bi-level utility estimation (BLUE) model with a reformulated upper and lower level objective function which is computationally efficient. A mathematical property of the optimal solution is exploited to develop a cut which significantly improves the computational time. Numerical experiments with real world data are conducted to validate the proposed models. In addition, we show the strong positive correlation between the widely used price elasticity property and the utility coefficients from the bi-level model.

The last chapter of this dissertation focuses on the development of empirical models to study the effect of smart technologies on electricity consumption. Smart technologies in this context refer to smart thermostats, and heat pump water heater systems, and residential battery systems. The models developed here have been utilized in the aforementioned demand response pilot study. We present a model to estimate the change of coincident load of residential consumers with the installation

of efficient appliances. The method utilizes a day matching algorithm to with similar weather conditions to form paired samples for comparing energy consumption and coincident load reduction before and after installing smart energy technologies. The consumption data from the paired up days are subject to a paired t-test to evaluate the statistical significance of the changes in terms of both energy consumption and coincident load reduction. The results reveal that insulation plays an important role in energy savings along with battery systems.

Although the dissertation has reached its goals, there were some limitations which can be investigated on for further research directions. The first limitation is that all the models presented in Chapters 3,4, and 5 depend on the performance of a learning model, TSEP on the structure of the prediction model while BLUE on the predictions. Several learning models including k-NN, decision trees, and naïve Bayes are studied and ultimately the random forest model is selected due to its high prediction accuracy. With the development of sophisticated deep learning and reinforcement learning models, the proposed methodologies can be further improved. In addition, the prediction accuracy can be increased by using building information as predictors in models to predict the HVAC energy consumption.

Secondly, a fundamental assumption of the work or direct load control is that people would surrender control of their appliances in exchange for incentives. However, this may not be realistic in practice. HVAC control is directly related to the comfort level of consumers inside their homes. This reason has resulted in the unpopularity of DLC programs around the country. While at the same time demand response pricing

schemes such as TOU and CPP are being adapted successfully by utilities around the country. This shows that consumers are welcoming dynamic pricing but still prefer to have control over their electricity usage. DLC methodologies proposed in literature usually have energy usage minimization as their objective and many of them focus on customer comfort. As electricity demand increases it would be necessary to use resources judiciously. Thus, development of efficient DLC methodologies are necessary because as electric energy demand increases DLC programs offer an efficient way to use the resources effectively.

The third limitation is concerned with Chapter 5, where we assign a quantitative metric for human behavior, which may not be applicable directly in the real world. In our work, we stress that the proposed models are validated using the dataset available from the field demonstration project. We feel the need to stress that the proposed models are validated using the dataset we had in our resource. This data albeit being from a real-world study represents a very small fraction of electricity consumers. Hence, many of the underlying assumptions pertail to this specific dataset. The models should be adapted appropriately when extended to a new instance.

7.2 Future research

There are several future research to be extended from the current dissertation. First, the proposed temperature set-point estimation problem (TSEP) presented in Chapter (3) has to be applied sequentially but separately in three adjoining control periods. During summer days the inside of the household is cooled down in the first period

while during the next period which is the forecasted peak, the temperature is let to rise within the highest preferred point. The last period is then used to cool back down the building to the preferred point. The model can be extended to include the three periods so that the set-points for the total direct load control period can be determined jointly. This will significantly reduce the computational burden.

Second, a dynamic model can be developed and be solved with real-time data. As new data comes in, the learning model for predicting energy consumption would be re-trained and the TSEP would be solved for the remaining periods. Such dynamic learning model is anticipated to produce better prediction results. In particular, at any time interval t, only first n set-points would be applied and at interval $t + n - 1$, the TSEP would be solved with $t + n$ as the start time. Doing so, the TSEP is more adaptive in nature as it takes advantage of real-time data through the increased accuracy of the learning models.

Third, the solution for the TSEP are the set-point temperatures over time for a programmable thermostat. Using these as reference temperature points, one can design a model predictive controller which will output the signal actions of a HVAC system. MPCs can have an explicit objective function to optimize. In designing such a controller, a system level objective function can be given to have a holistic controller design. This results in an integrated controller design, which using TSEP, can output control actions which benefits the whole system as well as the individual consumers. (Avci et al., 2013b) have used a similar strategy however, their reference temperature determination algorithm takes into account a price-range and their MPC objective

function is in perspective of a single household.

Fourth, in Chapter 4, we have formulated the CSP as an MILFP problem to schedule the start and end times of TSEP for multiple homes. The objective function aims to flatten the demand curve during the pre- and post-cool periods so that the forecasted peak is not shifted. The problem is solved using a bi-section search algorithm within moderate amount of computational time for large-scale instances. The properties of the model can be studied further to develop an efficient branch-and-cut algorithms to solve the MILFP or the reformulated MILP model. This can further reduce the computational times enabling the application of TSEP over a large number of homes

Finally, in our bi-level utility estimation problem (BLUE) in Chapter 5, the lower-level problem is being solved for N instances which equals to the number of days in the dataset. Currently, these N lower-level problems are solve independently. An immediate extension is to integrate these multiple lower-level problems into one single problem. It is expected such integration will increase computational efficiency as well as increase energy efficiency by controlling interactions between multiple peak days.

REFERENCES

Aghaei, J., & Alizadeh, M.-I. (2013). Critical peak pricing with load control demand response program in unit commitment problem. *IET Generation, Transmission & Distribution*, *7*(7), 681–690.

Albadi, M. H., & El-Saadany, E. F. (2007). Demand response in electricity markets: An overview. In *2007 ieee power engineering society general meeting* (pp. 1–5).

Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, *46*(3), 175–185.

Armas, J. M., & Suryanarayanan, S. (2009). A heuristic technique for scheduling a customer-driven residential distributed energy resource installation. In *Intelligent system applications to power systems, 2009. isap'09. 15th international conference on* (pp. 1–7).

Ashok, S., & Banerjee, R. (2003). Optimal operation of industrial cogeneration for load management. *IEEE Transactions on power systems*, *18*(2), 931–937.

Avci, M., Erkoc, M., Rahmani, A., & Asfour, S. (2013a). Model predictive hvac load control in buildings using real-time electricity pricing. *Energy and Buildings*, *60*(Supplement C), 199 - 209. Retrieved from `http://www.sciencedirect.com/science/article/pii/S037877881300025X` doi: https://doi.org/10.1016/j.enbuild.2013.01.008

Avci, M., Erkoc, M., Rahmani, A., & Asfour, S. (2013b). Model predictive hvac load control in buildings using real-time electricity pricing. *Energy and Buildings*, *60*, 199–209.

Bazaraa, M. S., Sherali, H. D., & Shetty, C. M. (2013). *Nonlinear programming: theory and algorithms.* John Wiley & Sons.

Bigerna, S., Bollino, C. A., & Micheli, S. (2016). Socio-economic acceptability for smart grid development–a comprehensive review. *Journal of Cleaner Production*, *131*, 399–409.

Bompard, E., Ma, Y., Napoli, R., Abrate, G., & Ragazzi, E. (2007). The impacts of price responsiveness on strategic equilibrium in competitive electricity markets. *International Journal of Electrical Power & Energy Systems*, *29*(5), 397–407.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

Breiman, L. (2017). *Classification and regression trees.* Routledge.

Calderaro, V., Hadjicostis, C. N., Piccolo, A., & Siano, P. (2011). Failure identification in smart grids based on petri net modeling. *IEEE Transactions on Industrial Electronics*, *58*(10), 4613–4623.

Cappers, P., Goldman, C., & Kathan, D. (2010). Demand response in us electricity

markets: Empirical evidence. *Energy*, *35*(4), 1526–1535.

Charnes, A., & Cooper, W. W. (1962). Programming with linear fractional functionals. *Naval Research logistics quarterly*, *9*(3-4), 181–186.

Chen, C.-R., Lan, M.-J., Huang, C.-C., Hong, Y.-Y., & Low, S. H. (2013). Demand response optimization for smart home scheduling using genetic algorithm. In *Systems, man, and cybernetics (smc), 2013 ieee international conference on* (pp. 1461–1465).

Chen, H., Li, Y., Louie, R. H., & Vucetic, B. (2014). Autonomous demand side management based on energy consumption scheduling and instantaneous load billing: An aggregative game approach. *IEEE transactions on Smart Grid*, *5*(4), 1744–1754.

Chen, J., Lee, F. N., Breipohl, A. M., & Adapa, R. (1995). Scheduling direct load control to minimize system operation cost. *IEEE Transactions on Power Systems*, *10*(4), 1994–2001.

Chen, J., Lian, Z., Tan, L., Zhu, W., & Zhang, W. (2011). Modeling and experimental research on ground-source heat pump in operation by neural network. In *Computer distributed control and intelligent environmental monitoring (cdciem), 2011 international conference on* (pp. 459–462).

Colson, B., Marcotte, P., & Savard, G. (2005). A trust-region method for nonlinear bilevel programming: algorithm and computational experience. *Computational Optimization and Applications*, *30*(3), 211–227.

Conejo, A. J., Morales, J. M., & Baringo, L. (2010). Real-time demand response model. *IEEE Transactions on Smart Grid*, *1*(3), 236–242.

Cui, H., & Peng, X. (2015). Short-term city electric load forecasting with considering temperature effects: an improved arimax model. *Mathematical Problems in Engineering*, *2015*.

Dong, B., O'Neill, Z., Luo, D., & Trevor, B. (2013). Development and calibration of a reduced-order energy performance model for a mixed-use building. In *Proceedings of building simulation.*

Dorini, G., Pinson, P., & Madsen, H. (2013). Chance-constrained optimization of demand response to price signals. *IEEE Transactions on Smart Grid*, *4*(4), 2072–2080.

Dusparic, I., Harris, C., Marinescu, A., Cahill, V., & Clarke, S. (2013). Multi-agent residential demand response based on load forecasting. In *Technologies for sustainability (sustech), 2013 1st ieee conference on* (pp. 90–96).

Ericson, T. (2011). Households' self-selection of dynamic electricity tariffs. *Applied energy*, *88*(7), 2541–2547.

Espinosa, J. R. (1987). Implementation and integration of air conditioner, cycling at southern california edison. *IEEE transactions on power systems*, *2*(3), 792–798.

Gelazanskas, L., & Gamage, K. A. (2014). Demand side management in smart grid: A review and proposals for future direction. *Sustainable Cities and Society*, *11*, 22–30.

Ghazvini, M. A. F., Soares, J., Horta, N., Neves, R., Castro, R., & Vale, Z. (2015). A multi-objective model for scheduling of short-term incentive-based demand

response programs offered by electricity retailers. *Applied Energy*, *151*, 102–118.

Glover, F. (1975). Improved linear integer programming formulations of nonlinear integer problems. *Management Science*, *22*(4), 455–460.

Goffin, J.-L. (1977). On convergence rates of subgradient optimization methods. *Mathematical programming*, *13*(1), 329–347.

Graditi, G., Di Silvestre, M. L., Gallea, R., & Sanseverino, E. R. (2015). Heuristic-based shiftable loads optimal management in smart micro-grids. *IEEE Transactions on Industrial Informatics*, *11*(1), 271–280.

Ha, D. L., Ploix, S., Zamai, E., & Jacomino, M. (2006). A home automation system to improve household energy control. *IFAC Proceedings Volumes*, *39*(3), 15–20.

Hale, E. T., Bird, L. A., Padmanabhan, R., & Volpi, C. M. (2018). *Potential roles for demand response in high-growth electric systems with increasing shares of renewable generation* (Tech. Rep.). National Renewable Energy Lab.(NREL), Golden, CO (United States).

Hansen, P., Jaumard, B., & Savard, G. (1992). New branch-and-bound rules for linear bilevel programming. *SIAM Journal on scientific and Statistical Computing*, *13*(5), 1194–1217.

Higham, D. J., & Higham, N. J. (2016). *Matlab guide*. SIAM.

Hsu, Y.-Y., & Su, C.-C. (1991). Dispatch of direct load control using dynamic programming. *IEEE Transactions on Power Systems*, *6*(3), 1056–1061.

Huang, K.-Y., Chin, H.-C., & Huang, Y.-C. (2004). A model reference adaptive control strategy for interruptible load management. *IEEE Transactions on Power Systems*, *19*(1), 683–689.

Huang, K.-Y., & Huang, Y.-C. (2004). Integrating direct load control with interruptible load management to provide instantaneous reserves for ancillary services. *IEEE Transactions on power systems*, *19*(3), 1626–1634.

Huh, J.-H., & Brandemuehl, M. J. (2008). Optimization of air-conditioning system operating strategies for hot and humid climates. *Energy and Buildings*, *40*(7), 1202–1213.

Iqbal, Z., Javaid, N., Khan, M. R., Ahmed, I., Khan, Z. A., & Qasim, U. (2016). Cost and load reduction using heuristic algorithms in smart grid. In *Advanced information networking and applications workshops (waina), 2016 30th international conference on* (pp. 24–30).

Jalali, M. M., & Kazemi, A. (2015). Demand side management in a smart grid with multiple electricity suppliers. *Energy*, *81*, 766–776.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.

Jia, L., & Tong, L. (2012). Optimal pricing for residential demand response: A stochastic optimization approach. In *Communication, control, and computing (allerton), 2012 50th annual allerton conference on* (pp. 1879–1884).

Kara, E. C., Berges, M., Krogh, B., & Kar, S. (2012). Using smart devices for system-level management and control in the smart grid: A reinforcement learning framework. In *Smart grid communications (smartgridcomm), 2012 ieee third international conference on* (pp. 85–90).

Katipamula, S., & Lu, N. (2006). Evaluation of residential hvac control strategies for demand response programs. *ASHRAE transactions*, *112*(1).

Kinhekar, N., Padhy, N. P., Li, F., & Gupta, H. O. (2016). Utility oriented demand side management using smart ac and micro dc grid cooperative. *IEEE Transactions on Power Systems*, *31*(2), 1151–1160.

Kirschen, D. S., Strbac, G., Cumperayot, P., & de Paiva Mendes, D. (2000). Factoring the elasticity of demand in electricity prices. *IEEE Transactions on Power Systems*, *15*(2), 612–617.

Kusiak, A., & Li, M. (2010). Cooling output optimization of an air handling unit. *Applied Energy*, *87*(3), 901 - 909. Retrieved from `http://www.sciencedirect .com/science/article/pii/S0306261909002499` doi: https://doi.org/10 .1016/j.apenergy.2009.06.010

Kusiak, A., & Xu, G. (2012). Modeling and optimization of hvac systems using a dynamic neural network. *Energy*, *42*(1), 241–250.

Laurent, J.-C., Desaulniers, G., Malhamé, R. P., & Soumis, F. (1995). A column generation method for optimal load management via control of electric water heaters. *IEEE Transactions on Power Systems*, *10*(3), 1389–1400.

Li, N., Chen, L., & Low, S. H. (2011). Optimal demand response based on utility maximization in power networks. In *Power and energy society general meeting, 2011 ieee* (pp. 1–8).

Lixing, D., Jinhu, L., Xuemei, L., & Lanlan, L. (2010, May). Support vector regression and ant colony optimization for hvac cooling load prediction. In *2010 international symposium on computer, communication, control and automation (3ca)* (Vol. 1, p. 537-541). doi: 10.1109/3CA.2010.5533861

Logenthiran, T., Srinivasan, D., & Shun, T. Z. (2012). Demand side management in smart grid using heuristic optimization. *IEEE transactions on smart grid*, *3*(3), 1244–1252.

Ma, J., Qin, S. J., Li, B., & Salsbury, T. (2011, Jan). Economic model predictive control for building energy systems. In *Isgt 2011* (p. 1-6). doi: 10.1109/ISGT .2011.5759140

Ma, Y., Matuško, J., & Borrelli, F. (2015). Stochastic model predictive control for building hvac systems: Complexity and conservatism. *IEEE Transactions on Control Systems Technology*, *23*(1), 101–116.

Manual, C. U. (1987). Ibm ilog cplex optimization studio. *Version*, *12*, 1987–2018.

Mayne, D. Q., Rawlings, J. B., Rao, C. V., & Scokaert, P. O. (2000). Constrained model predictive control: Stability and optimality. *Automatica*, *36*(6), 789–814.

Meng, F.-L., & Zeng, X.-J. (2013). A stackelberg game-theoretic approach to optimal real-time pricing for the smart grid. *Soft Computing*, *17*(12), 2365–2380.

Mohsenian-Rad, A.-H., & Leon-Garcia, A. (2010). Optimal residential load control with price prediction in real-time electricity pricing environments. *IEEE transactions on Smart Grid*, *1*(2), 120–133.

Muratori, M., Marano, V., Sioshansi, R., & Rizzoni, G. (2012). Energy consumption of residential hvac systems: A simple physically-based model. In *Power and energy society general meeting, 2012 ieee* (pp. 1–8).

Muratori, M., Roberts, M. C., Sioshansi, R., Marano, V., & Rizzoni, G. (2013). A highly resolved modeling technique to simulate residential power demand. *Applied Energy*, *107*(Supplement C), 465 - 473. Retrieved from `http://www.sciencedirect.com/science/article/pii/S030626191300175X` doi: https://doi.org/10.1016/j.apenergy.2013.02.057

Muratori, M., Schuelke-Leech, B.-A., & Rizzoni, G. (2014). Role of residential demand response in modern electricity markets. *Renewable and Sustainable Energy Reviews*, *33*, 546–553.

O'Neill, D., Levorato, M., Goldsmith, A., & Mitra, U. (2010). Residential demand response using reinforcement learning. In *Smart grid communications (smartgridcomm), 2010 first ieee international conference on* (pp. 409–414).

Pedrasa, M. A. A., Spooner, T. D., & MacGill, I. F. (2009). Scheduling of demand side resources using binary particle swarm optimization. *IEEE Transactions on Power Systems*, *24*(3), 1173–1181.

Ramanathan, B., & Vittal, V. (2008). A framework for evaluation of advanced direct load control with minimum disruption. *IEEE Transactions on Power Systems*, *23*(4), 1681–1688.

Ratliff, L. J., Dong, R., Ohlsson, H., & Sastry, S. S. (2014). Incentive design and utility learning via energy disaggregation. *IFAC Proceedings Volumes*, *47*(3), 3158–3163.

Ruelens, F., Claessens, B. J., Vandael, S., De Schutter, B., Babuška, R., & Belmans, R. (2017). Residential demand response of thermostatically controlled loads using batch reinforcement learning. *IEEE Transactions on Smart Grid*.

Saez-Gallego, J., Morales, J. M., Zugno, M., & Madsen, H. (2016). A data-driven bidding model for a cluster of price-responsive consumers of electricity. *IEEE Transactions on Power Systems*, *31*(6), 5001–5011.

Safdarian, A., Fotuhi-Firuzabad, M., & Lehtonen, M. (2014). A distributed algorithm for managing residential demand response in smart grids. *IEEE Transactions on Industrial Informatics*, *10*(4), 2385–2393.

Sahinidis, N. V. (2017). BARON 17.8.9: Global Optimization of Mixed-Integer Nonlinear Programs, *User's Manual* [Computer software manual].

Samadi, P., Mohsenian-Rad, A.-H., Schober, R., Wong, V. W., & Jatskevich, J. (2010). Optimal real-time pricing algorithm based on utility maximization for smart grid. In *Smart grid communications (smartgridcomm), 2010 first ieee international conference on* (pp. 415–420).

Savard, G., & Gauvin, J. (1994). The steepest descent direction for the nonlinear bilevel programming problem. *Operations Research Letters*, *15*(5), 265–272.

Schweppe, F. C., Caramanis, M. C., Tabors, R. D., & Bohn, R. E. (2013). *Spot pricing of electricity*. Springer Science & Business Media.

Setlhaolo, D., Xia, X., & Zhang, J. (2014). Optimal scheduling of household appliances for demand response. *Electric Power Systems Research*, *116*, 24–28.

Shen, B., Ghatikar, G., Ni, C. C., Dudley, J., Martin, P., & Wikler, G. (2012). *Addressing energy demand through demand response. international experiences and practices* (Tech. Rep.). Ernest Orlando Lawrence Berkeley National

Laboratory, Berkeley, CA (United . . . .

Stoll, B., Buechler, E., & Hale, E. (2017). The value of demand response in florida. *The Electricity Journal*, *30*(9), 57–64.

Strbac, G. (2008). Demand side management: Benefits and challenges. *Energy policy*, *36*(12), 4419–4426.

Strickler, G. F., & Noell, S. (1988). Residential air conditioner cycling-a case study. *IEEE transactions on power systems*, *3*(1), 207–212.

Tang, F., Kusiak, A., & Wei, X. (2014). Modeling and short-term prediction of hvac system with a clustering algorithm. *Energy and Buildings*, *82*, 310–321.

Tashtoush, B., Molhim, M., & Al-Rousan, M. (2005). Dynamic model of an hvac system for control analysis. *Energy*, *30*(10), 1729 - 1745. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0360544204004761` doi: https://doi.org/10.1016/j.energy.2004.10.004

Tawarmalani, M., & Sahinidis, N. V. (2004). Global optimization of mixed-integer nonlinear programs: A theoretical and computational study. *Mathematical programming*, *99*(3), 563–591.

Tawarmalani, M., & Sahinidis, N. V. (2013). *Convexification and global optimization in continuous and mixed-integer nonlinear programming: theory, algorithms, software, and applications* (Vol. 65). Springer Science & Business Media.

Thimmapuram, P. R., & Kim, J. (2013). Consumers' price elasticity of demand modeling with economic effects on electricity markets using an agent-based model. *IEEE Transactions on Smart Grid*, *4*(1), 390–397.

Tsui, K. M., & Chan, S.-C. (2012). Demand response optimization for smart home scheduling under real-time pricing. *IEEE Transactions on Smart Grid*, *3*(4), 1812–1821.

U. s. energy information administration, annual energy outlook. (2020).

Venkatesan, N., Solanki, J., & Solanki, S. K. (2012). Residential demand response model and impact on voltage profile and losses of an electric distribution network. *Applied energy*, *96*, 84–91.

Vicente, L., Savard, G., & Judice, J. (1996). Discrete linear bilevel programming problem. *Journal of optimization theory and applications*, *89*(3), 597–614.

Virk, G. S., & Loveday, D. L. (1994). Model-based control for hvac applications. *IEEE, PISCATAWAY, NJ,(USA).*, *3*, 1861–1866.

Wei, D.-c., & Chen, N. (1995). Air conditioner direct load control by multi-pass dynamic programming. *IEEE Transactions on Power Systems*, *10*(1), 307–313.

Wemhoff, A., & Frank, M. (2010). Predictions of energy savings in hvac systems by lumped models. *Energy and Buildings*, *42*(10), 1807–1814.

Wen, Z., O'Neill, D., & Maei, H. (2015). Optimal demand response using device-based reinforcement learning. *IEEE Transactions on Smart Grid*, *6*(5), 2312–2324.

Wolsey, L. A. (1998). *Integer programming* (Vol. 52). John Wiley & Sons.

Wolsey, L. A., & Nemhauser, G. L. (1999). *Integer and combinatorial optimization* (Vol. 55). John Wiley & Sons.

Xu, X., Wang, S., & Huang, G. (2010). Robust mpc for temperature control of air-conditioning systems concerning on constraints and multitype

uncertainties. *Building Services Engineering Research and Technology*, *31*(1), 39-55. Retrieved from `https://doi.org/10.1177/0143624409352420` doi: 10.1177/0143624409352420

Xuemei, L., Lixing, D., Yuyuan, D., & Lanlan, L. (2010, May). Hybrid support vector machine and arima model in building cooling prediction. In *2010 international symposium on computer, communication, control and automation (3ca)* (Vol. 1, p. 533-536). doi: 10.1109/3CA.2010.5533864

Yang, H.-T., & Huang, K.-Y. (1999). Direct load control using fuzzy dynamic programming. *IEE Proceedings-Generation, Transmission and Distribution*, *146*(3), 294–300.

Yao, L., Chang, W.-C., & Yen, R.-L. (2005). An iterative deepening genetic algorithm for scheduling of direct load control. *IEEE Transactions on Power Systems*, *20*(3), 1414–1421.

Yi, P., Dong, X., Iwayemi, A., Zhou, C., & Li, S. (2013). Real-time opportunistic scheduling for residential demand response. *IEEE Transactions on smart grid*, *4*(1), 227–234.

Yin, R., Xu, P., Piette, M. A., & Kiliccote, S. (2010). Study on auto-dr and pre-cooling of commercial buildings with thermal mass in california. *Energy and Buildings*, *42*(7), 967–975.

You, F., Castro, P. M., & Grossmann, I. E. (2009). Dinkelbach's algorithm as an efficient method to solve a class of minlp models for large-scale cyclic scheduling problems. *Computers & Chemical Engineering*, *33*(11), 1879–1889.

Yue, D., Guillén-Gosálbez, G., & You, F. (2013). Global optimization of large-scale mixed-integer linear fractional programming problems: a reformulation-linearization method and process scheduling applications. *AIChE Journal*, *59*(11), 4255–4272.

Zlatanović, I., Gligorević, K., Ivanović, S., & Rudonja, N. (2011). Energy-saving estimation model for hypermarket hvac systems applications. *Energy and Buildings*, *43*(12), 3353–3359.

Zugno, M., Morales, J. M., Pinson, P., & Madsen, H. (2013). A bilevel model for electricity retailers' participation in a demand response market environment. *Energy Economics*, *36*, 182–197.

CURRICULUM VITA

Arnab Roy

2241 Arthur Ford Court #4, Louisville, KY 40217

(732)-861-7823   arnab.roy@louisville.edu

SUMMARY

PhD candidate in Industrial Engineering with interest in applying operations research methods in data science

EDUCATION

Ph.D. Candidate
August 2015-Present
  Industrial Engineering
  University of Louisville
  Current GPA: 3.952/4.00

M. Tech.
  Chemical Engineering
July 2012-June 2014
  Indian Institute of Technology, Roorkee, India
  GPA: 3.7/4.00

B. Tech.
  Chemical Engineering
July 2008-June 2012
  Jadavpur University, Kolkata, India
  GPA: 3.4/4.00

EXPERIENCE

Data Science – Operations Research Intern, Bayer Crop Science, May 2019 – August 2019

Determine optimal allocations for experiments in a research laboratory

- Develop a mixed integer programming model (Java and IBM DOC) to determine optimal allocations for over 100000 experiments for a biotechnology laboratory

- Implement the workflow for the large-scale optimization using cloud-based tools such as AWS and Domino
- Model is able to increase the space utilization of the lab by 23% over manual allocations

Research Assistant, Industrial Engineering Department, University of Louisville

Estimation of utility function of residential consumers under dynamic pricing of electricity

August 2017 – Present

- Estimate the consumption utility of price-responsive electricity consumers using real-world data through a bi-level optimization model
- Hourly consumption data of individual household is used to develop a custom random forest model (Python) whose structural information is then used by the lower-level optimization problem to optimize consumption for future peak-consumption days
- Implement a trust-region algorithm (Python with CPLEX 12.8) to solve the NP-hard problem

Mapping Automated Pavement Data to Windshield Visual Survey Data

June 2018 – Present

- Develop mapping process from the automated data into the older legacy format using machine learning methods
- Used R, SQL, Tableau to clean and combine millions of rows and hundreds of variables from the automated data, identified the significant factors for the visual survey data
- The modified AdaBoost (Python) method can predict pavement data such as severity and extent of cracks with approximately 90% accuracy
- Implement the developed algorithm in GUI form using Java

Dynamic Programming model for load control in Residential Households

January 2017 – July 2017

- Determine the optimal set points for smart thermostats during a demand response event using a Dynamic Programming (DP) model
- Collected, cleaned, filtered and combined big datasets of individual household consumption data at 30-second resolution using R, Python, Windows Powershell, AWS Tools
- Develop ARIMA models to predict the inside temperature of households and HVAC energy consumption with 92% accuracy; use these models as state-space equations in the DP
- The algorithm saved around 26% electricity consumption during peak-consumption period in a real-world scenario

A multi-criterion partner selection problem for virtual manufacturing enterprises under uncertainty

September 2016 – January 2017

- Proposed a new method of using the grey system theory to account for uncertainties in a project's start time, completion time, transportation time, as well as total cost
- Developed the concepts of delivery time agreement index and cost agreement index
- A chaotic particle swarm optimization (CPSO) algorithm (MATLAB) is developed. Extensive computational results show that the proposed CPSO method outperforms the standard PSO in providing quality solutions more reliably

An Empirical Study on a Demand Response Program with Battery Storage Systems
December 2015-September 2017

- Install and help in collect, structure, and store consumption data from smart devices such as smart thermostats, heat pump water-heaters, and battery for 3000 residential homes for a municipal utility company
- Develop objective comparisons to evaluate the performance of the smart devices and demand-response pricing
- Reduced average consumption of household by 20% by proposing control actions during demand response events to the utility company using custom algorithms for inferential statistics (R)
- Designed a GUI in Python for customers to monitor their electricity consumption behavior

Junior Research Fellow, Sailesh J. Mehta School of Management, Indian Institute of Technology, Bombay, September 2014-April 2015

Design and development of integrated forward-reverse supply chain under risk and uncertainty

- A closed loop supply chain network under risk and uncertainty is modeled through a system dynamics framework with product interchange and recuperation policies
- Risk measures such as, variability index; downside risk and conditional value at risk are integrated with the proposed model

COMPUTER SKILLS

- **Computer Languages:** C/C++, Java, Python, R
- **Data Analysis:** R, Python, SAS, SQL, Tableau, AWS
- **Scientific Applications:** MATLAB, CPLEX Concert Technology, GUROBI
- **Modeling and Simulation:** IBM DOC, ARENA 10.0, Minitab 17, Simio, AutoCAD, ANSYS Fluent

PUBLICATIONS

- Bin Huang, Lihui Bai, Arnab Roy, Ningning Ma, A multi-criterion partner selection problem for virtual manufacturing enterprises under uncertainty, In International Journal of Production Economics, 2017,ISSN 0925-5273, https://doi.org/10.1016/j.ijpe.2017.08.024.

- Roy Arnab, Bai Lihui, Control of HVAC System in Residential Buildings using Model Predictive Control: Plant Model Selection, *Proceedings of the 2017 Industrial and Systems Engineering Conference*
- Roy A, Srivastava P, Sinha S, Risk and reliability assessment in chemical process industries using Bayesian methods, Reviews in Chemical Engineering, ISSN (Online) 2191-0235, ISSN (Print) 0167-8299, DOI: 10.1515/revce-2013-0043, August 2014.
- Arnab Roy, Prashant Srivastava, Shishir Sinha, Dynamic Failure Assessment of a Storage Unit: A Case Study, Process Safety and Environmental Protection, Available online 30 September 2014, ISSN 0957-5820, http://dx.doi.org/10.1016/j.psep.2014.09.004.

PRESENTATIONS

- "A Bilevel Optimization Model for Estimating Utilities of Price-responsive Electricity Consumers" INFORMS Annual Meeting 2018, Phoenix, AZ
- "Dynamic Programming model for load control in Residential Households" INFORMS Annual Meeting 2017, Houston, TX
- "An Empirical Study on a Demand Response Program with Battery Storage Systems"
  INFORMS Annual Meeting 2016, Nashville, TN