A Data-Driven Strategy to Enable Efficient Participation of Diverse Social Classes in Smart Electric Grids

by

Mingyue He

A Thesis Presented in Partial Fulfillment
of the Requirement for the Degree
Master of Science

Approved November 2019 by the
Graduate Supervisory Committee:

Mojdeh Khorsand, Chair
Vijay Vittal
Anamitra Pal

ARIZONA STATE UNIVERSITY

December 2019

ABSTRACT

The grand transition of electric grids from conventional fossil fuel resources to intermittent bulk renewable resources and distributed energy resources (DERs) has initiated a paradigm shift in power system operation. Distributed energy resources (i.e. rooftop solar photovoltaic, battery storage, electric vehicles, and demand response), communication infrastructures, and smart measurement devices provide the opportunity for electric utility customers to play an active role in power system operation and even benefit financially from this opportunity. However, new operational challenges have been introduced due to the intrinsic characteristics of DERs such as intermittency of renewable resources, distributed nature of these resources, variety of DERs technologies and human-in-the-loop effect. Demand response (DR) is one of DERs and is highly influenced by human-in-the-loop effect. A data-driven based analysis is implemented to analyze and reveal the customers price responsiveness, and human-in-the-loop effect. The results confirm the critical impact of demographic characteristics of customers on their interaction with smart grid and their quality of service (QoS). The proposed framework is also applicable to other types of DERs. A chance-constraint based second-order-cone programming AC optimal power flow (SOCP-ACOPF) is utilized to dispatch DERs in distribution grid with knowing customers price responsiveness and energy output distribution. The simulation shows that the reliability of distribution gird can be improved by using chance-constraint.

ACKNOWLEDGMENTS

My deepest gratitude goes first and foremost to my supervisor, Dr. Mojdeh Khorsand, for her constant encouragement and guidance on this thesis research. I sincerely thank her for her consistent illuminating guidance and patience that walked me through all the stages of my master studies.

Second, I would like to express my sincere thanks to my committee members, Dr. Vijay Vittal and Dr. Anamitra Pal for giving me instructive advice and their direct and indirect help to me.

Special thank goes to Zahra Soltani who has put considerable time and efforts on helping me with second-order-cone programming ACOPF.

This work was supported by the Salt River Project (SRP).

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

NOMENCLATURE

Parameters

ρ_t^{DA}	The day − ahead wholesale electricity price at time period
ρ_t^{D}	The day − ahead DERs electricity price at time period
$B_{i,j}$	Susceptance of line
$C_{i,j}^{line}$	Capacity limit of line
$C_{i,n,t}^{pv1_P}$	Active power capacity limit of PV type 1
$C_{i,n,t}^{pv2_P}$	Active power capacity limit of PV type 2
$C_{i,n,t}^{pv3_P}$	Active power capacity limit of PV type 3
$C_{i,n,t}^{pv3_S}$	Apparent power capacity limit of PV type 3
$C_{i,t}^{S,H}$	Maximum capacity limit of the battery storage
$C_{i,t}^{S,L}$	Minimum capacity limit of the battery storage
$D_{n,t}^{P}$	Active power demand at bus n
$D_{n,t}^{Q}$	Reactive power demand at bus n
DR_m	Coefficient of maximum DR capacity
$F_{f,n}^{H}$	FACTS device maximum output
$F_{f,n}^{L}$	FACTS device minimum output
$G_{g,n}^{H}$	Generator maximum output limit

$G^L_{g,n}$	Generator minimum output limit
$G_{i,j}$	Conductance of line
$I^{p\%}_t$	Initial active power proportion of total DERs to total load
$I^{q\%}_t$	Initial reactive power proportion of total DERs to total load
N	Total number of buses
$R^H_{i,t}$	The maximum ramping rate of storage
$R^L_{i,t}$	Theminimum ramping rate of storage

Sets

BUS	Set of bus number
DR	Set of index and bus number of DR
F	Set of index and bus number of FACTS devices
GN	Set of index and bus number of generators
L	Set of from bus number and to bus number
$PV1$	Set of index and bus number of PV type 1
$PV2$	Set of index and bus number of PV type 2
$PV3$	Set of index and bus number of PV type 3
SG	Set of index and bus number of battery storage
T	Set of schedule period

Decision Variables

$\theta_{i,j,t}$	Bus angle difference
$I_{i,j,t}$	Auxiliary variables
$P_{g,n,t}^{grid}$	Scheduled active power purchasing from grid for substation
$P_{i,j,t}^{line}$	Scheduled active power line flow for line
$P_{i,n,t}^{S}$	Scheduled active power injecting to grid from batter storage
$P_{i,n,t}^{DR}$	Scheduled active power from DR
$P_{i,n,t}^{PV1}$	Scheduled active power from PV type 1
$P_{i,n,t}^{PV2}$	Scheduled active power from PV type 2
$P_{i,n,t}^{PV3}$	Scheduled active power from PV type 3
$Q_{f,n,t}^{facts}$	Scheduled reactive power from FACTS device
$Q_{g,n,t}^{grid}$	Scheduled reactive power purchasing from grid for substation
$Q_{i,j,t}^{line}$	Scheduled reactive power line flow for line
$Q_{i,n,t}^{DR}$	Scheduled reactive power from DR
$Q_{i,n,t}^{PV1}$	Scheduled reactive power from PV type 1
$Q_{i,n,t}^{PV3}$	Scheduled reactive power from PV type 3
$R_{i,j,t}$	Auxiliary variables
$SE_{i,n,t}$	Scheduled active power stored in batter storage
$U_{n,t}$	Auxiliary variables of voltage at bus n
$V_{i,t}$	Bus voltage magnitude at bus n

Chapter 1

INTRODUCTION

1.1 Background

With the increasing penetration of DERs in distribution systems, more and more customers are transitioning from the role of mere consumers into "prosumers" (producers and consumers, in one). However, this active involvement of consumers along with the ever-increasing penetration level of DERs also causes new challenges in power systems. These challenges are mainly caused due to the essential nature of DERs, namely, increased uncertainty and variability, which makes scheduling of DERs a sophisticated task [1] [2]. Successful transition from the conventional power system to this new paradigm requires an accurate planning and efficient management of these ressources.

Advanced measurement devices, which are enriched with communication technologies such as Advanced Metering Infrastructure (AMI) provides a high volume of reliable data such as daily household consumption data [3]. This information, along with other data such as weather forecast and social composition of customers, enables leveraging data analytical methods to enhance smart grid operation. These data-driven based analyses can be leveraged to estimate the participation of prosumers in grid services and assess their quality of service (QoS).

The ratio of peak electric load to average load is generally increasing [4]. The increasing peak load results in problems standing by more fast generators needed to prepare for the huge demand in a very short period of time. Starting more generators will not only cost a lot, but also be polluted since these generators operate in an inefficient manner. In order to handle the peak load condition issue, utilities are seeking approaches to improve capacity utilization. Instead of adapting generators to match the big change of the load demand, the demand itself can be more flexible to follow the requirement on the operation of the smart grid [5]. Demand response (DR) is one method that utilities use to shift peak customer's load and improve the operation of the smart grid. It has existed within a utility's arsenal as a means to manage the power system for decades. For one type of the DR programs, utility provides incentives and benefits to customers so that customers can reduce their load during on-peak periods and shift it to off-peak periods. The most common DR pricing structure adopted by utilities is the time of use (TOU), which is defined as different electricity price for different period of the day [6]. Currently, more and more residential household customers are participating in TOU DR programs in the distribution system. However, with the increasing number of the customers participating in DR programs, this brings system operational issues into the distribution grid since the customers' consumption behavior becomes different in a day due to the difference in the energy prices in distinct periods of the day. Florida Power and Light's (FPL) tested a DR program focused on two-way communication

based programmable thermostats where FPL could control the thermostat but that could also be overridden by the customer. The ex-post analysis by FPL showed that customers who enrolled in the program had increased consumption of 12 percent compared with both earlier usage and that of non-enrollees [7]. This issue points to a fundamental need to consider prosumers' interactions with the electric grids and the uncertainty of their responses.

1.2 Social Impact on Power System

Recent research from social science fields has revealed that energy is not merely a technological and economic phenomenon, but is rather deeply embedded within social and geopolitical landscapes [8] [9] [10]. For example, a general daily schedule of the customers, their social composition, and other demographic information may affect their QoS and participation in DR programs. Different DR programs have existed within a utility's arsenal as a means to manage the power system for years. One key hurdle that, to this day, has not been overcome is the human-in-the-loop challenge. The societal dependency of such method and lack of adequate incentives to unleash this potential, the societal desire to participate, are main reasons DR programs are underutilized. As a result, to fulfill the vision of the smart grid on a sustainable scale, socially-aware engineering solutions are needed.

1.3 Methodology and Data

This thesis aims to apply and extend concepts of the social value of energy to design a methodology to better operate DERs in distribution gird. This thesis explores this multifaceted problem in two stages. The first stage leverages advanced machine learning algorithms and available social and historical consumption data to evaluate customers' price responsiveness and consumption behavior during both the on-peak and off-peak periods. Various machine learning algorithms are examined and compared in order to find the most suitable approach. Artificial neural network (ANN) has a very strong ability to fit highly nonlinear functions [11]. K-nearest Neighbors (KNN), which is a nonparametric algorithm, requires a decent data structure. Kernel ridge regression (KRR) is a powerful regression algorithm that can have pretty good performance in terms of the linear function. The second stage shows the advantages of using the proposed method of the first stage via a chance-constraint based AC optimal power flow (ACOPF) scheduling tool. This model considers both technical and economic constraints in the distribution grid. There are various methods for implementing chance-constraint in the ACOPF problem including scenario approach, sample average approximation, robust optimization based method. In this thesis, a Monte Carlo Simulation based method is used to apply chance-constraint in the ACOPF model accounting for the uncertainty and variability nature of DERs.

In this thesis, firstly customers are classified based on customers' demographic

information. Papers show that groups of individuals with similar demographic characteristics exhibit similar behavior towards electric energy [12] [13]. In order to accurately estimate and evaluate the participation of prosumers in the smart grid, some papers already addressed demographic data in their work as a boosting strategy [12] [13]. Reference [12] utilizes demographic data from Mosaic, which is Experian's system for geodemographic classification of households, and divides customers into three subsets in order to improve the accuracy of the model. Reference [13] uses demographic data as indicators showing the similarity of customers' behavior. Secondly, several machine learning (ML) algorithms are leveraged to obtain each group's baseload during on-peak periods. The customers' baseload is defined as the answer to the question of "How much they would consume if they do not participate in DR". In [4], the authors apply a demographic data-based decision tree method and k-mean to cluster the customers, and then they propose three baseline calculation approaches which are regression-based, similar day-based and morning adjustment. One consumption baseline load prediction method focuses on finding the control group, which is defined as a group of selected customers whose daily consumption will be treated as a baseline for other [14]. Reference [15] studies the impact of less accurate customers' consumption baseline load on customers and utility benefits. Reference [16] [17] [18] [19] apply deep neural network to predict the customers' consumption without any price program involved. In this thesis, three regression-based ML algorithms are ap-

plied to attain customers' baseload using the consumption data points in off-peak periods. Thirdly, we calculate the consumption reductions and evaluate the customers' responsiveness from the same category to different DR programs. Finally, chance-constraint based SOCP-ACOPF is used to improve the reliability of the distribution grid by searching for the best proportion of DERs.

In order to show the effectiveness of the proposed data-driven strategy, electricity consumption, generation, weather, and demographic data of prosumers in a local electric utility in Phoenix area are used. Four different datasets from the local utility including different TOU plans are utilized in this thesis.

1.4 Literature Review

Load forecasting is an important research area in the power system. A precise short-term load forecasting in the residential level places a significant role in system operation. In [20], the author proposes a load decomposition-based forecasting method in order to have a better forecasting result by summing several disaggregated sub-load predictions. Unlike the load decomposition-based method, a load forecasting method in [21], which is based on measurement data from smart meters, is applying clustering methods to identify the groups of customers who have similar behaviors and load consumption patterns. Since machine learning has achieved lots of success in prediction, the application of machine learning in the power system has become an irresistible trend for many researchers. The approaches in [16] [17] [18] [19] explore the customers' behavior learning using deep

neural network (DNN), which recently has become a highly prevalent research field in many areas. Almost all of the paper studying in machine learning application area are applying recurrent neural network (RNN), which is a special type of DNN and have strong capability dealing with time-series data, and capturing the hidden factors behind the huge amount of measurement data.

Demand response is one of DERs. Before analyzing the effect of any type of demand response program, researchers should understand the calculation of consumption baseline, which is defined as the energy consumption that customers will consume, if customers did not participate in the DR program at that time. However, the customers' consumption baseline is tricky in some sense because no one can answer the question such as "what if the customers didn't participate in the demand response program, what would be their energy consumption?". In [22], the authors apply a demographic data-based decision tree method and k-mean to cluster the customers, and then they propose three baseline calculation approaches which are regression-based, similar day-based and morning adjustment. The regression-based method in [22] is piecewise linear regression. However, all the methods and framework mentioned in load forecasting [20] [21] [16] [17] [18] [19] can be directly exploited as customers' consumption baseline calculation routines. The similar day-based method selects the days that have similar temperature as event, and use the consumption in those days to calculate baseline. The morning adjustment method use the actual metered load in the morning

to adjust the consumption baseline for that day. There is a consumption baseline method mainly focusing on finding the control group in [14], which is defined as a group of selected customers whose daily consumption is treated as the baseline for others. Reference [15] studies the impact of social welfare loss and error analysis for customers' consumption baseline. A comprehensive discussion of DR consumption baseline calculation has been exhibited in [23].

The participation of DERs in the power system introduces the risk of dispatching residential DERs because of its characteristic of high uncertainty and variability, which is mainly caused by the human-in-the-loop effect. A more accurate DERs output prediction method can be really influential when integrating DERs in the electric market. Multiple price elasticity-based methods have been developed in [24] [25] [26] [27] in order to capture the customers' price responsiveness under different DR programs. Reference [25] utilizes five decreasing curve functions and seven sensitivity categories to represent customers' price responsiveness. Reference [26] considers rational customer model maximizing welfare with utility functions and self and cross-time elasticity. Reference [27] builds a mathematical price elasticity model to simulate customers' responsiveness. In [28], the authors construct an implementation framework using long short-term memory (LSTM), which is recurrent neural network, to predict customers' response behavior during different DR programs' events. In [13], the authors construct possibility distribution to represent the price responsiveness of the EV owners considering electricity

price and demographic information. Monte Carlo method in [29] is applied for simulating the EV owners' charging and discharging behavior. There is a satellite image based solar generation prediction method using SVM in [30] with comparatively high accuracy. Reference [31] advances a hybrid algorithm utilizing an optimization model to provide initial parameters of artificial neural network which is trained later by shuffled frog leaping algorithm. Reference [32] applies deep neural network framework to solar energy prediction problems and gains a satisfactory performance. An advanced diffusion-based kernel density estimator equipped with a fuzzy inference system and adaptation function is utilized to achieve high quality and accuracy of wind power generation in [33]. A wind power forecasting approach is proposed in [34] in which the author obtains the stochastic distribution of locally wind energy by applying neural network-based classification method utilizing global forecast system information. Reference [35] models the load demand, wind speed, and generator location are modeled according to Gaussian, Weibull, and discrete uniform distribution functions, respectively. For solar irradiance modeling, they cluster the data into two states which are high-irradiance state at daytime and low-irradiance state at night, and then they utilize a kernel smoothing technique to estimate the pdf of the solar irradiance of two states. Reference [36] uses k-mean to cluster daily PV output data and implements Epanechnikov kernel estimating technique to estimate the probability distribution of PV output. They also propose a way to approximate optimal solution of house-

hold energy management system using deep learning, SVM and ANN. Reference [37] proposes a deep learning-based short-term wind power forecasting algorithm by combining autoencoders, the back-propagation algorithm, and the genetic algorithm.

Since the human-in-the-loop aspect of emerging distributed smart grid technologies and energy prediction error, the uncertainty of DERs is necessary to be highly addressed in the power system operation scheme. In [38], the authors handle the uncertainty by applying stochastic dominance constraints, which is widely used in economics and finance due to its relation to models of risk-averse preference. References [39] [40] [41] [42] employ conditional value-at-risk (CVaR) in their optimization model to evaluate and control the risk of profit and cost variability of dispatching virtual power plant (VPP) because of its uncertainty and variability nature. In [43] [44], the authors come up with various special risk indices to assess the risk of DERs when integrating them into the power system. A fuzzy chance constrained programming approach to day-ahead scheduling VPP considering the uncertainty of DERs and its risk of dispatching is proposed in [45]. In this paper, the author associates uncertain factors in VPP with fuzzy parameters and fuzzy chance constraints, and also provide different confidence levels for the economy and risk sensitively of VPP. Reference [46] utilizes expected value and zero-mean normally distributed random variable with a standard deviation to represent the output and uncertainty of DERs. A scenario-based risk analysis

of VPP is introduced in [47], which accounts for the temporal correlation of solar and wind generation. Unlike [47], the authors in [48] directly generate scenarios based on price and load and applies stochastic programming to risk analysis approach with approximation. Reference [48] presents a robust optimization-based daily and weekly scheduling of VPP. The authors in [48] treats uncertainty of DERs as uncertainty control parameters embedded in their robust optimization model. Reference [49] combines chance-constraint with second-order-cone ACOPF. Reference [50] formulates an optimization model as an estimator to predict a group of DERs injection for the next time period. Their objective function is minimizing the square error between estimated values and true values, which subjects to a DERs controller's biased term.

1.5 Organization of this thesis

In chapter 2, the data, machine learning algorithms and analysis approach description are presented. In chapter 3, the results of the local utility project are exhibited. Simulation results, using chance-constraint based SOCP-ACOPF, are shown in chapter 4.

Chapter 2

DATA DESCRIPTION AND APPROACH

In the chapter, the description of local utility datasets and the explanation of the analyzing approach will be presented. The results of the local utility work will be shown in the next chapter.

2.1 Data Description

Four different residential customers' consumption datasets from the local utility are used for the purpose of this project. The first, second and third consumption datasets are used to analyze customers' behavior and price responsiveness in the local utility time of use (TOU) price plan. It is pertinent to note that three different TOU plans are mainly analyzed in this project, which are E21, E23 and E26. The detailed information of those three TOU plans is exhibited in the following section. The fourth dataset is used to analyze the customers' behavior during the emergency DR events in the local utility system. In the following subsections, these four consumption datasets are explained in detail.

2.1.1 First and Second Round Data

For the first and second round customers' consumption datasets, customers are randomly pooled from the local utility data system. The first two consump-

tion data-sets contain the customers' 15-mins consumption values from May 2017 to April 2018. We also received the customer's corresponding anonymized demographic information, which is obtained from a survey study conducted by local utility. The demographic information includes the following factors.

- Household annual income
- The number of occupants in the house
- Family composition (married, if they have Children)
- Age
- House ownership.

There are totally 3750 customers in these two datasets. There are 1776 customers who participated in E21 and 1974 customers who participated in E26. There is no missing data point in those 3750 customers' consumption data.

2.1.2 Third Round Data

The third consumption dataset only includes low-income customers in the local utility system. The resolution of the third consumption data is one hour. The consumption values are collected from May 2017 to April 2018. We also received the customer's corresponding anonymized demographic information obtained from a survey study conducted by local utility, which includes the following factors.

- Household annual income

- Number of occupants in the house
- Family composition
- Age
- Home ownership
- Livable space
- Rate start date
- Household appliance information
- Number of rooms
- Built year

There are totally 18609 customers in this dataset who are all low-income customers in local utility system. There are 8946 low-income customers who participated in E21 and 9663 low-income customers who participated in E26. There is no missing data point in those 18609 customers' consumption data.

2.1.3 Fourth Round Data

The fourth dataset includes the customers of the local utility system who participated in emergency DR events (i.e. conservation periods based on nest thermostats) in 2018. There are totally nine DR events that happened in 2018. One of them happened in Jun 2018. Three of them happened in Jul 2018. The rest of them

happened in Aug 2018. The duration of each event is from two to three hours. The information of the above emergency DR events is shown in Table 2.1.

Table 2.1: Emergency Demand Response Event Time Information

Event #	**Event Date**	**Thermostat Platform**	**Firm Peak Load Hour**	**Temp. at Peak (F)**	**Event Hours**
1	6/22/2018	Nest	17:00	109	5PM - 7PM
2	7/24/2018	Nest	18:00	115	4PM - 6PM
3	7/25/2018	Nest	17:00	115	4PM - 6PM
4	7/30/2018	Nest	17:00	108	3PM- 6PM
5	8/2/2018	Nest	17:00	108	3PM- 6PM
6	8/6/2018	Nest	17:00	113	3PM- 6PM
7	8/13/2018	Nest	18:00	100	4PM - 6PM
8	8/14/2018	Nest	17:00	102	5PM - 7PM
9	8/30/2018	Nest	18:00	102	4PM - 6PM

The resolution of this dataset is one hour. We also received the customer's corresponding anonymized demographic information. The demographic factors are shown below.

- Household annual income
- Number of occupants in the house
- Family composition

- Age
- Home ownership

There are totally 5180 customers in this fourth dataset. There is no missing data point in those 5180 customers' consumption data. There are 1301 customers in E23. The E23 customers' behavior during the DR events are mainly analyzed in this project.

2.1.4 Data Summary

The data information is summarized in the Table 2.2.

2.1.5 Local Utility Demand Response Programs

For the first, second and third datasets, we mainly focus on E21 and E26 of which information has been shown in Table 2.3. In the local utility system, summer is defined as the May, June, September, and October billing cycles. Summer Peak is defined as the July and August billing cycles.

For E21, the on-peak hours are from 3 p.m. to 6 p.m., Monday through Friday, Mountain Standard Time, excluding the holidays. All other hours are off-peak. The on-peak and off-peak prices in summer peak are 0.3588 $/kWh and 0.0864 $/kWh, respectively. Furthermore, the on-peak and off-peak price in summer are 0.3033 $/kWh and 0.0840 $/kWh, respectively [6].

For E26, the on-peak hours are from 1 p.m. to 8 p.m., Monday through Friday

Table 2.2: Summary of Data Information

Data	Resolution	Number of Customers	Time Period	Demographic Factors	Rate	Data Pulling
First Round Data	15-min	1322	May 2017 to April 2018	Income Level, Number of Occupants, If having Children	E21 and E26	Randomly Pulling
Second Round Data	15-min	2428	May 2017 to April 2018	Income Level, Number of Occupants, If having Children	E21 and E26	Randomly Pulling
Third Round Data	1-hour	18609	May 2017 to April 2018	Livable Space, Number of Occupants, Start Date	E21 and E26	All Low-Income Customers
Fourth Round Data	1-hour	5180	May 2018 to Sep 2018	Income Level, Number of Occupants	All Rate in local utility	All customers participating Emergency Demand Response

from May 1 through October 31, Mountain Standard Time, excluding the holidays. On-peak hours from November 1 through April 30 consist of those hours from 5 a.m. to 9 a.m. and from 5 p.m. to 9 p.m., Monday through Friday, Mountain Standard Time, excluding the holidays. All other hours are off-peak. The on-peak and off-peak prices in summer peak are 0.2226 $/kWh and 0.0741 $/kWh, respectively. The on-peak and off-peak price in summer are 0.1957 $/kWh and 0.0738 $/kWh,

Table 2.3: Local Utility DR Plans for E21 and E26

TOU DR Programs	**On-Peak Hours**	**Duration**	**On-Peak Price**	**Off-Peak Price**
E21	3 p.m. 6 p.m.	3 h	$0.3588 (Jul Aug) $0.3033 (May Jun Sep Oct)	$0.0864 (Jul Aug) $0. 0840 (May Jun Sep Oct)
E26	1 p.m. 8 p.m.	7 h	$0.2226 (Jul Aug) $0.1957 (May Jun Sep Oct)	$0.0741 (Jul Aug) $0.0738 (May Jun Sep Oct)

respectively [6].

The following holidays are off-peak days:

- New Year's Day (observed)
- Memorial Day (observed)
- Independence Day (observed)
- Labor Day
- Thanksgiving Day
- Christmas Day (observed).

For the fourth dataset, E23 customers are mainly considered. The E23 is the

basic rate in the local utility residential customers level. The E23 price information is shown below [6].

- Summer: 1st 700kWh: 0.1102 $/kWh; 701-2000 kWh: 0.1121 $/kWh; All Add'l kWh: 0.1226 $/kWh
- Summer Peak: 1st 700kWh: 0.1168 $/kWh; 701-2000 kWh: 0.1180 $/kWh; All Add'l kWh: 0.1331 $/kWh

For the rest of the DR plans, since the number of customers in these DR rate plans is very small, we excluded them in this data-driven analysis.

2.2 Approach

A data-driven based approach has been proposed in this project. This approach is implemented to analyze all of the four datasets. The flowchart of the proposed approach is shown in Fig. 2.1. First, several different categories are formed using customers' demographic data. Besides, three regression-based ML algorithms are applied to attain customers' baseload using their consumption data in off-peak periods. After that, the DR reduction values are calculated using the predicted customers' baseload and actual consumption during the on-peak periods. Finally, customers' behavior during on-peak periods and their price responsiveness are evaluated from the same category to different DR programs

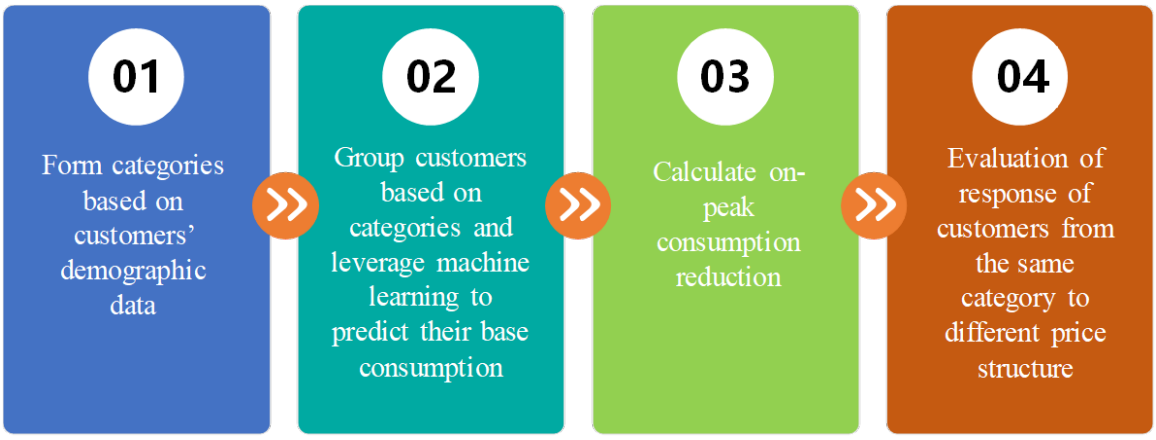

Figure 2.1: Approach Flowchart

2.2.1 Machine Learning

Three machine learning algorithms are implemented to predict customers' baseload during the on-peak periods. The following well-established regression-based customers' baseload calculation methods, which captures the relationship between the consumption values vector for the next time period Y and previous consumption values matrix X, are used in this paper.

- Artificial Neural Network (ANN)
- K Nearest Neighbors (KNN)
- Ridge Regression (RR)

2.2.2 Artificial Neural Network (ANN)

ANN, which is inspired by biological neural networks that constitute animal brains, is a prevalent research topic in many areas recently. The general structure

of ANN includes input, hidden and output layer.

For the input to hidden units we have:

$$a_h^t = \sum_{h'=1}^{I} w_{h'h} b_{h'}^{t-1} \tag{2.1}$$

$$b_h^t = \theta_h a(_h^t) \tag{2.2}$$

where $w_{h'h}$ are the weights between neuron h' in layer $t-1$ and neuron h in layer t. The $b_{h'}^{t-1}$ are the inputs of neuron h' from the last layer $t-1$. The I is the number of neurons in the layer $t-1$. The a_h^t is sum of $w_{h'h} b_{h'}^{t-1}$ from the last layer. The θ is activation function for the neuron h. The b_h^t are the output of neuron h in layer t. It will be the input to the layer $t+1$.

For the output unit we have:

$$a_k^t = \sum_{h=1}^{H} w_{hk} b_h^t \tag{2.3}$$

where the w_{hk} are weights and b_h^t are inputs from the last hidden layer. The H is the number of neurons in the last layer. The a_k^t will be the output of the ANN model, if the output layer is the t-th layer and has k neurons.

Common one-layer ANN structure and ANN neuron mathematical representation have been shown in Fig. 2.2 and Fig. 2.3.

Due to its structure, ANN can have a higher ability to fit into highly nonlinear functions. The higher number of hidden layers in ANN results in improved capa-

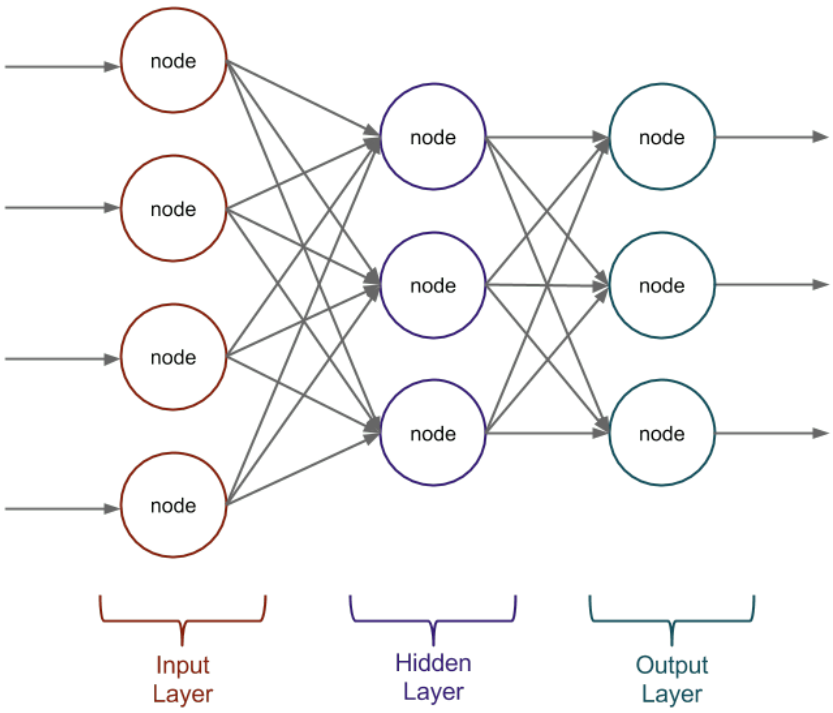

Figure 2.2: One-layer ANN Structure

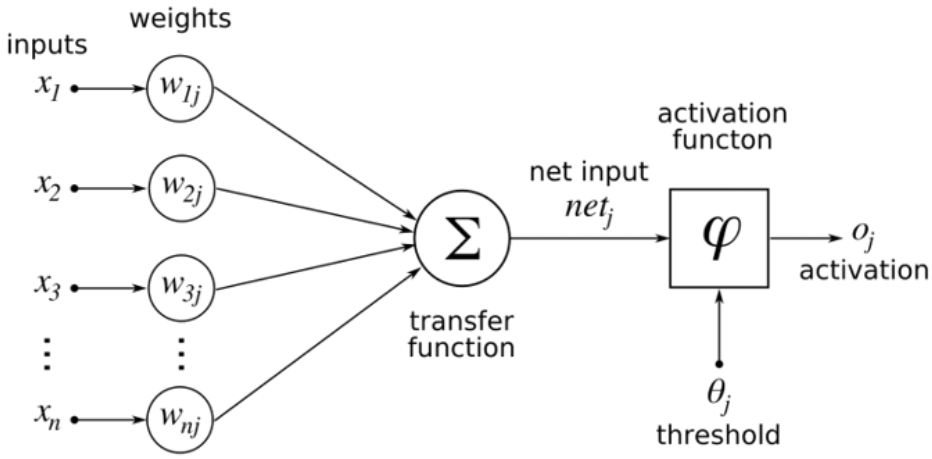

Figure 2.3: ANN Neuron Mathematical Representation

bility of the model to extract more abstracted features from the input. However, the drawbacks of ANN are its less interpretability and obtaining local optimal so-

lutions.

2.2.3 K-Nearest Neighbors (KNN)

KNN is a non-parametric ML algorithm used for classification and regression problems. For the regression problem, KNN first calculates the distance for each sample. The distance can be calculated by different distance functions such as Euclidean, Manhattan and Minkowski function. To forecast the response to a new sample, the trained model uses k samples that are closest to the new data point.

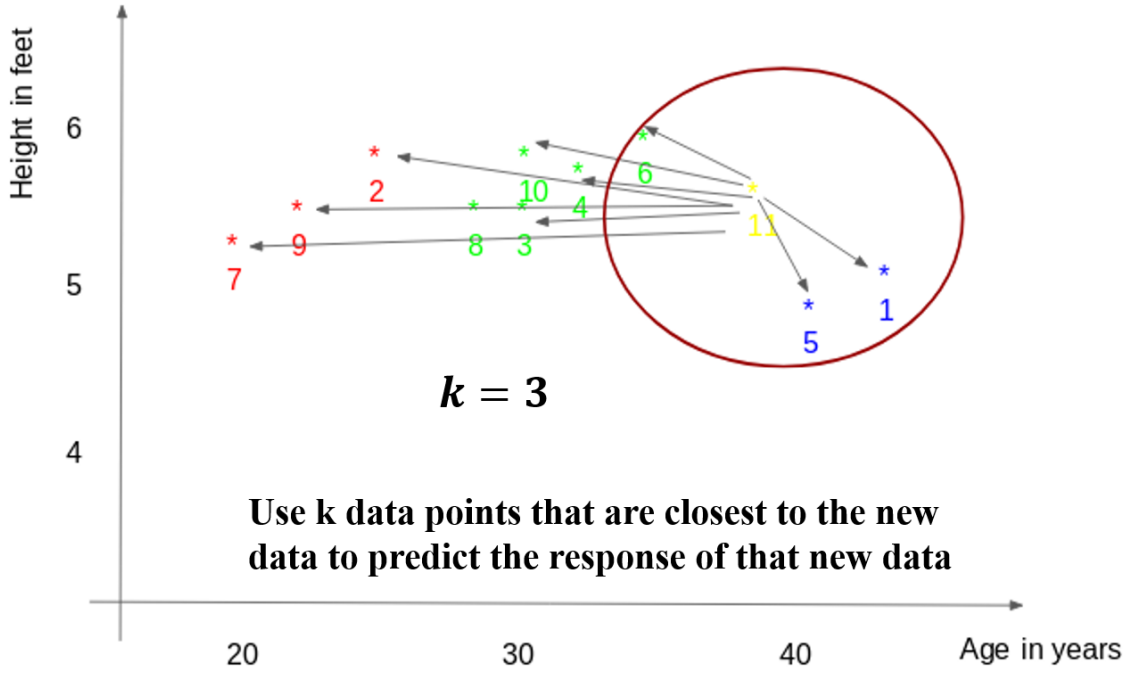

Figure 2.4: A Schema of KNN Algorithm

Fig. 2.4 shows a schema of the KNN algorithm. In Fig. 2.4, we need to predict the response of the yellow point. The KNN algorithm will find the k points that are closest to the yellow point and use their values to predict the response of the

yellow point.

2.2.4 Ridge Regression (RR)

RR is a commonly used algorithm that deals with linear or near-linear data. The loss function of the ridge regression is the combination of the linear least squares function and regularization given by the l2-norm (also called Euclidean norm). The objective function can be mathematically be written as below,

$$\min \sum_{n=1}^{N} (Y_n - W^T X_n)^2 + \lambda W^T W \tag{2.4}$$

where the N is the number of samples in the training matrix. The Y_n is the response of each sample. The W^T is the weight for each independent variable X_n. The λ is the hyper-parameter used in the regularization term. Note that the RR is strict convex, which guarantees a global optimal solution.

2.2.5 Baseload Prediction

The customers' baseload is defined as the answer to the question of "How much they would consume if they do not participate in DR".

For the purpose of this study, customers' consumption usage and their DR on-peak load reduction are analyzed for groups of customers instead of each individual household. The off-peak period consumption values and ML algorithms are used to build the customers' baseload model for the on-peak periods.

Since the resolution of the first and second datasets is 15-mins and resolution of the third and fourth datasets is one hour, we make different assumptions before building customers' baseload model.

For the first and second datasets, due to their 15-mins resolution, if we use more than prior 8 hours consumption values, the accuracy of the baseload model is high enough to predict customers' baseload. Besides, the accuracy only slightly changes if we increase the number of prior hours consumption values that are used for training the ML algorithms. Thus, we utilize consumption values from 8 previous hours as features in the training matrix and assume that the group's consumption values are only correlated to the previous 8 hours of group's consumption values. The presentation of the first and second data baseload models can be presented using (2.5),

$$P(Y|x_1, x_2, ..., x_{32}) \tag{2.5}$$

where Y is the consumption value of the group for the next time interval. $x_1, x_2, ..., x_{32}$ are the prior 8 hours historical consumption values with 15-min resolution. All of the values in (2.5) are the consumption values during the off-peak period.

For the third and fourth datasets, the accuracy of the baseload model changes a lot, if we use a different number for previous hours consumption. The relationship between model accuracy and the number of prior hours consumption is tested

using a group of customers' data whose demographic information are all 1-2 occupants in the house, livable space less 1500 sqft and start date of becoming local utility customers after 2016. The start date is the date that the customer becomes local utility's customer. For the third and fourth data baseload estimation, we finally use the prior 6 hours of consumption values and assume that the group's consumption values for the next time period are only correlated to customers' consumption values of the previous prior 6 hours. The prediction error of the baseload model for different training structures is shown in Fig. 2.5. Root-mean-square error (note equation (2.7) in the next page) is used as a measure of prediction error. Fig. 2.5 shows that the most accurate baseload prediction is obtained when the off-peak consumption values of the prior 6 hours are used for training the ML algorithms.

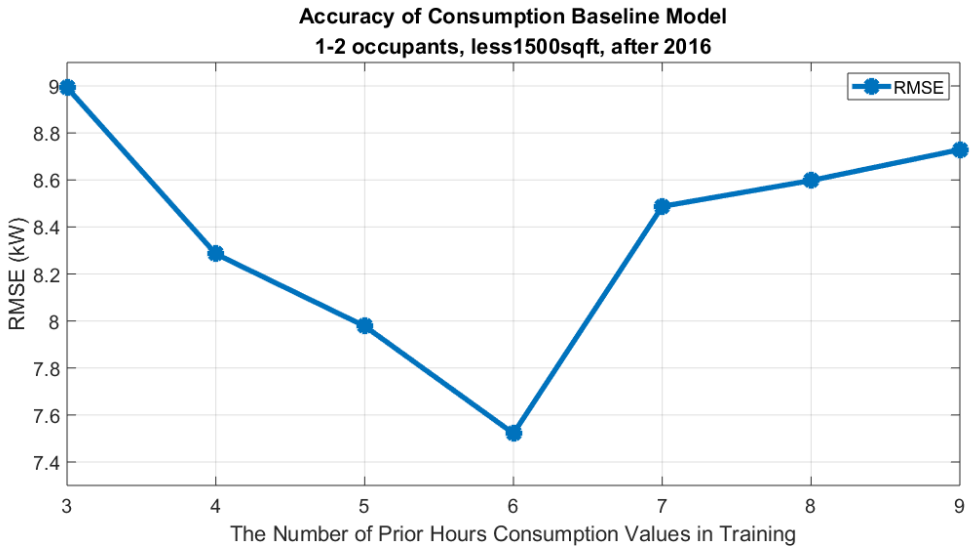

Figure 2.5: Trained Model Accuracy Testing for Third and Fourth Dataset

The presentation of the third and fourth data baseload model can be shown in

(2.6),

$$P(Y|x_1, x_2, ..., x_6) \tag{2.6}$$

where Y is the consumption value of the group for the next time interval. $x_1, x_2, ..., x_6$ are the prior 6 hours historical consumption values with one-hour resolution. All of the values in (6) are the consumption values during the off-peak period.

2.2.6 Training

Scikit-learn is used to implement the ANN, KNN and RR algorithms, which is a machine learning library embedded in Python. A three-hidden-layers ANN structure is built in this project and the hyper-parameter of KNN is set to be 3. The coefficient of determination (R^2) and root-mean-square error ($RMSE$) are used to evaluate the performance of the trained model. These metrics are presented in (2.7)-(2.8),

$$R^2 = 1 - \frac{\sum_i (y_i - y_i^{predict})^2}{\sum_i (y_i - y_{mean})^2} \tag{2.7}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{T} (y_i^{predict} - y_i)^2}{T}} \tag{2.8}$$

where y_i is the actual value for the test sample i. The $y_i^{predict}$ is the predicted value for the test sample i. The y_{mean} is the average value for the T test samples.

Note that the closer R^2 value is to 1, the more accurate the baseload model is. The smaller $RMSE$ represents a higher accuracy of prediction.

2.2.7 Consumption Baseline Accuracy: First and Second Datasets

As mentioned before, the customer categories are formed based on customers' demographic information. For each class of customers, e.g. medium income with 1-2 occupants and no child, groups of 20-100 customers are formed. Different ML algorithms are trained for each group to predict their baseload load. Table 2.4 lists the accuracy of the trained model for the first and second datasets where the resolution is 15-mins. The prior 8 hours consumption values are used to train the model. The results suggest that the model accuracy increases when larger sample groups are used. The more people in the training model, the more accurate the trained model is. The models trained by ANN and RR have higher accuracy compared with the trained model using KNN. However, ANN needs more training time. Thus, in most of the cases, RR is utilized to get customers' baseload.

In order to exhibit the accuracy of the trained model, the actual values versus predicted values are shown in Fig. 2.6. The test samples are randomly picked from the test data. The evaluation metric values are shown in the figures. Fig. 2.6 confirms that the trained model is able to predict the baseload consumption with very high accuracy.

As can be seen, some of the predicted values exactly match the true values. Furthermore, Fig. 2.7 presents an exemplary result of consumption baseline load

Table 2.4: Accuracy of the Trained Model for Different Size of Customers' Group: Medium Income, 1-2 occupants, no Child

Metrics	**Number of Customers**	**ANN**	**KNN**	**RR**
R2	20	0.900	0.880	0.908
RMSE (kW)	20	6.240	6.852	5.986
R2	30	0.943	0.932	0.948
RMSE (kW)	30	7.259	7.900	6.883
R2	50	0.959	0.944	0.962
RMSE (kW)	50	10.001	11.703	9.644
R2	100	0.982	0.974	0.983
RMSE (kW)	100	14.328	17.002	13.961

calculation.

Note that Fig. 2.7 shows the off-peak and on-peak periods in order to provide a comparison to actual and predicted summed consumption values on Jul. 3rd. Due to the highly accurate baseload load model, in the following study, the customers' baseloads predicted by these trained models are considered as their true baseload.

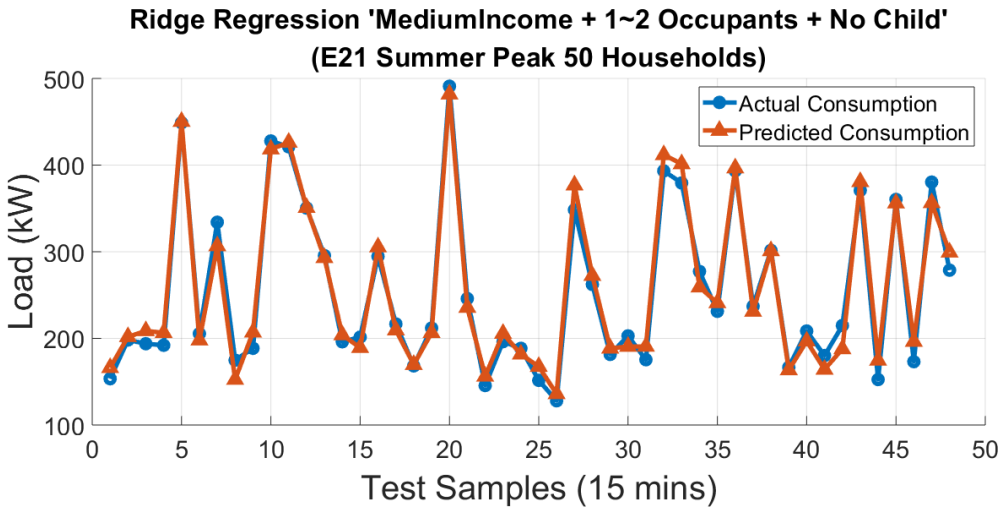

Figure 2.6: An Example of Trained Model Accuracy for First and Second Round Dataset

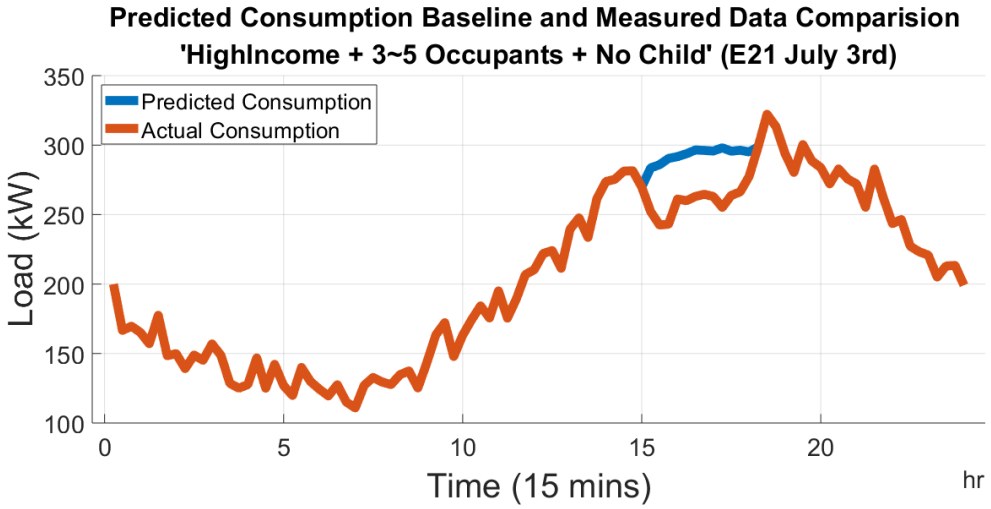

Figure 2.7: An Exemplary Result of Baseload Load Calculation: Aggreated Consumption of 60 Customers

2.2.8 Consumption Baseline Accuracy: Third and Fourth Datasets

Similarly, the customer categories are formed based on customers' demographic information using third and fourth datasets. Table 2.5 presents the accuracy of the trained model for the third and fourth datasets, the resolution of which is one hour. The same effect can be seen in Table 2.5. The model accuracy increases when larger sample groups are used. The models trained by ANN and RR have higher accuracy compared with the trained model using KNN. For the analysis of the third and fourth dataset, RR is used to predict customers' baseload model.

Due to the difference of resolution, the accuracy of the trained model and an exemplary result of baseload calculation are shown in Fig. 2.8 and Fig. 2.9. The test samples are randomly picked from the test data. Furthermore, the evaluation metric values are shown in Fig. 2.8.

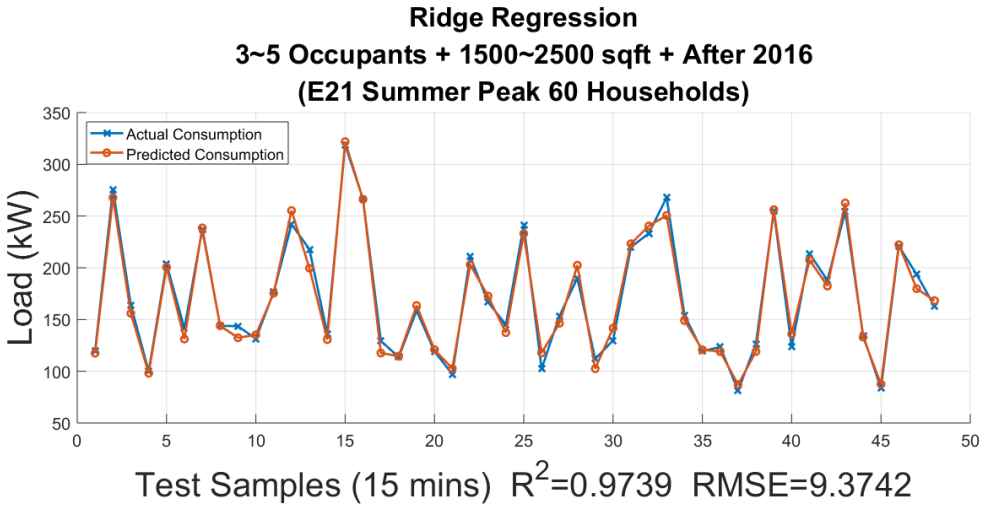

Figure 2.8: An Example of Trained Model Accuracy for Third and Fourth Round Dataset

Table 2.5: Accuracy of the Trained Model for Different Size of Customers' Group: 1-2 occupants, less1500sqft, after 2016

Metrics	**Number of Customers**	**ANN**	**KNN**	**RR**
R2	20	0.935	0.914	0.940
RMSE (kW)	20	4.220	4.871	4.074
R2	30	0.943	0.943	0.942
RMSE (kW)	30	5.601	5.591	5.657
R2	50	0.955	0.940	0.955
RMSE (kW)	50	8.349	9.644	8.354
R2	100	0.974	0.970	0.973
RMSE (kW)	100	13.124	14.028	13.174

As can be seen, the baseload models for the third and fourth datasets are highly accurate. Some of the predicted values exactly match the true values.

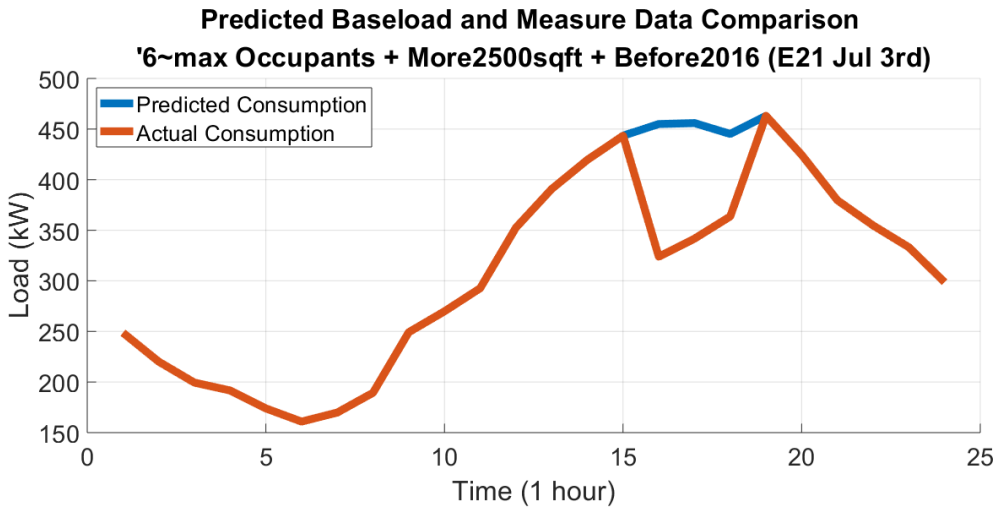

Figure 2.9: An Exemplary Result of Baseload Calculation: Aggreated Consumption of 60 Customers

Note that Fig. 2.9 shows the off-peak and on-peak periods in order to provide a comparison to actual and predicted summed consumption values on Jul. 3rd. Due to the highly accurate baseload model, in the following analysis of the third and fourth datasets, the customers' baseloads predicted by these trained models are considered as their true baseload.

The datasets provided by SRP is shown in this chapter. The machine learning algorithms, including ANN, KNN, and KRR, are utilized to predict customers' baseload during the on-peak period. The trained baseload model is highly accurate. The customers' consumption baseloads predicted by the trained baseload model in chapter 3 are considered as customers' true baseload.

Chapter 3

PROJECT RESULTS

The first and second datasets analysis will be shown in this chapter first. The third dataset analysis and corresponding one demographic factor analysis will be presented next. Finally, the analysis of the fourth dataset, which is emergency demand response data, will be exhibited.

3.1 First and Second Datasets Analysis

3.1.1 Customers Responsiveness for the First and Second Datasets

For the first and second dataset analysis, a decision-tree-like approach is applied to categorize customers based on their demographic information such as income level, number of occupants, and if they have Children. However, the customers are randomly picked from the local utility data pool and the demographic data information of those customers is not evenly distributed. For some of the demographic categories, there are only a few customers, which is not enough to conduct the data-driven analysis. Due to this issue, we mainly focus on the following seven categories shown in Table 3.1 for this study.

The analysis of the first and second datasets for the E21 has been shown in Fig. 3.1, 3.2 and 3.3. The blue bar shows the average predicted consumption during the

Table 3.1: Analyzed Demographic Categories for the First Round Dataset

Categories	Income Level	Number of Occupants	Child
1	High-Income	1-2 people	no child
2	High-Income	3-5 people	no child
3	High-Income	3-5 people	have children
4	Medium-Income	1-2 people	no child
5	Medium-Income	3-5 people	no child
6	Medium-Income	3-5 people	have children
7	Low-Income	1-2 people	no child

on-peak periods. The orange bars represent the average load reduction during the on-peak periods. The red percentage values are the percentage of load reduction for each category. For each demographic category, 60 customers are aggregated and analyzed as a group. Table 2.4 confirms that an accurate prediction can be obtained with this number of customers in a group. The detailed baseload model accuracy can be seen in the Appendix.

As can be seen, the category "High-income, 3 to 5 occupants, having child" has the highest energy consumption and load reduction during on-peak. It is revealed that if having children may have a great impact on high-income customer groups. In contrast, the least percentage reduction is observed for low-income customers.

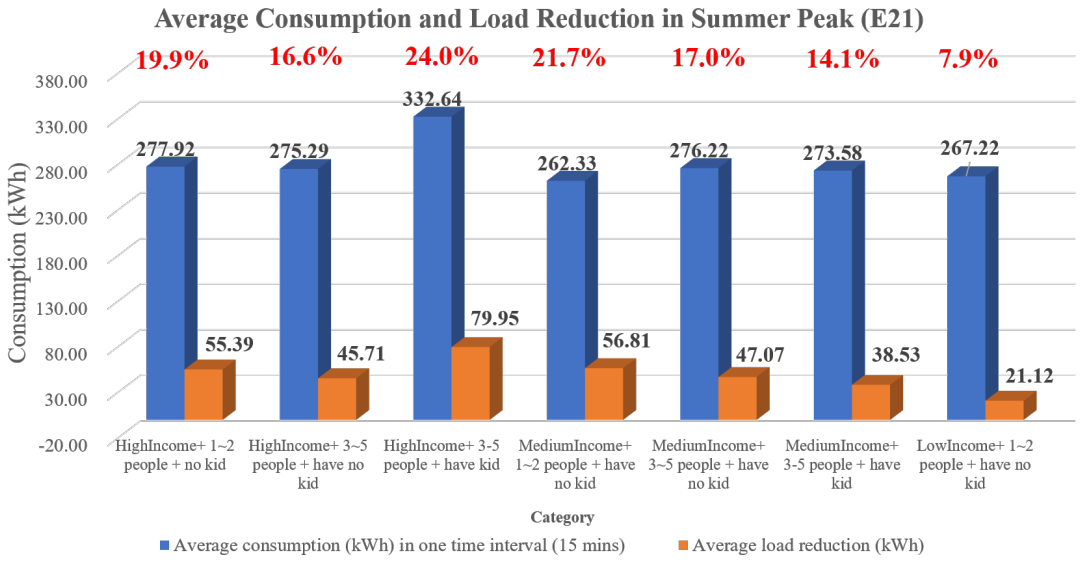

Figure 3.1: First Round Dataset Analysis for E21 Summer Peak, 60 Households in Each Group

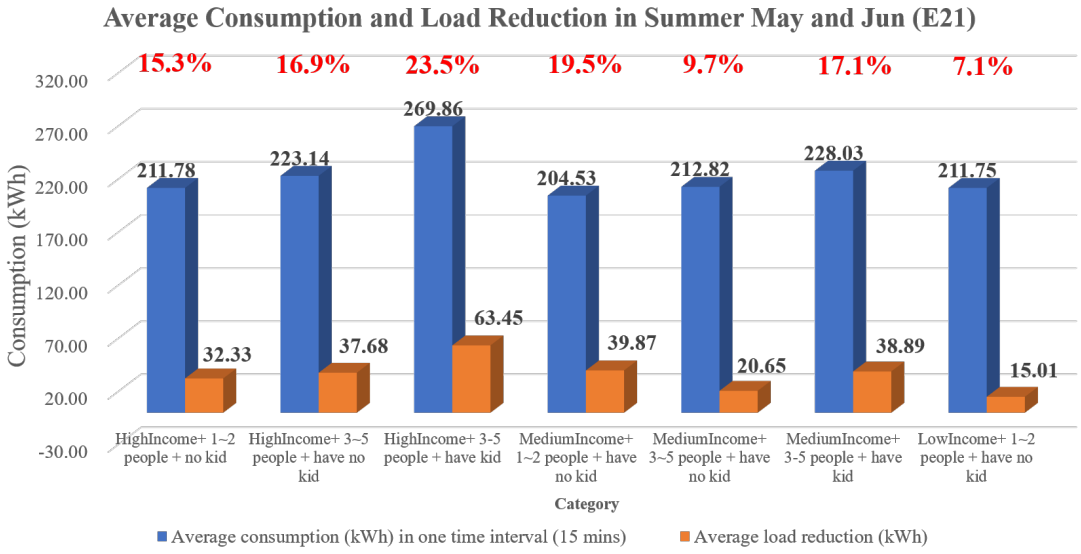

Figure 3.2: First Round Dataset Analysis for E21 Summer (May and Jun), 60 Households in Each Group

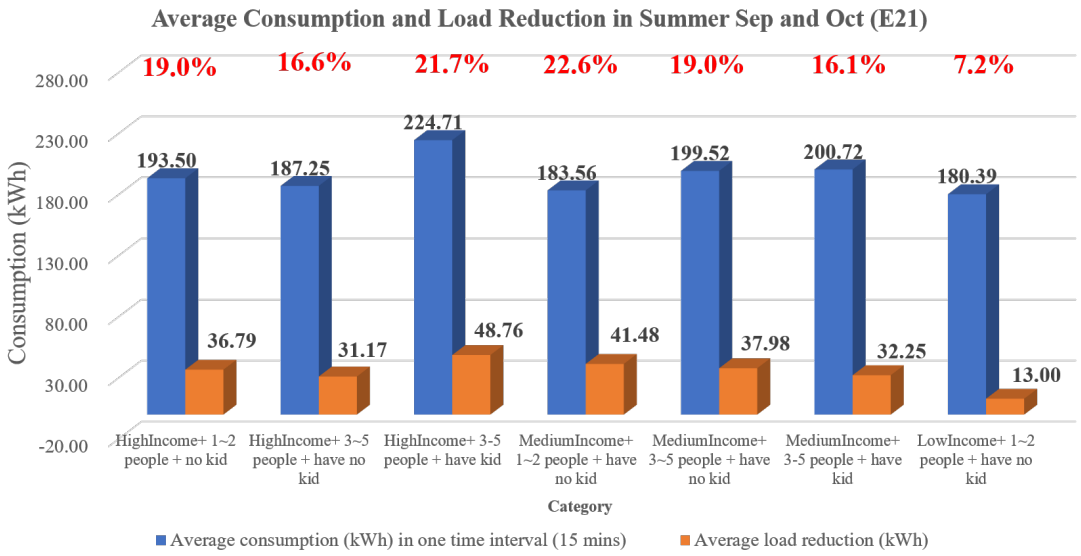

Figure 3.3: First Round Dataset Analysis for E21 Summer (Sep and Oct), 60 Households in Each Group

The lower percentage of consumption reduction can have multiple reasons. For example, this group of customers may not have programmable thermostats, or maybe pre-occupied with different other life challenges. To further understand this group, a more detail analysis of this group is conducted using the second dataset. From this analysis, we can conclude that the income level is greatly related to the customers' behavior during the on-peak periods. Comparing Fig. 3.1, 3.2, and 3.3, the load reduction value is higher during the Summer Peak periods, which is consistent with the designed incentives, i.e. higher on-peak pricing.

The analysis of the first and second datasets for the E26 DR program is also shown in Fig. 3.4, 3.5 and 3.6.

As can be seen from Fig. 3.4, 3.5, and 3.6, the average load reduction values

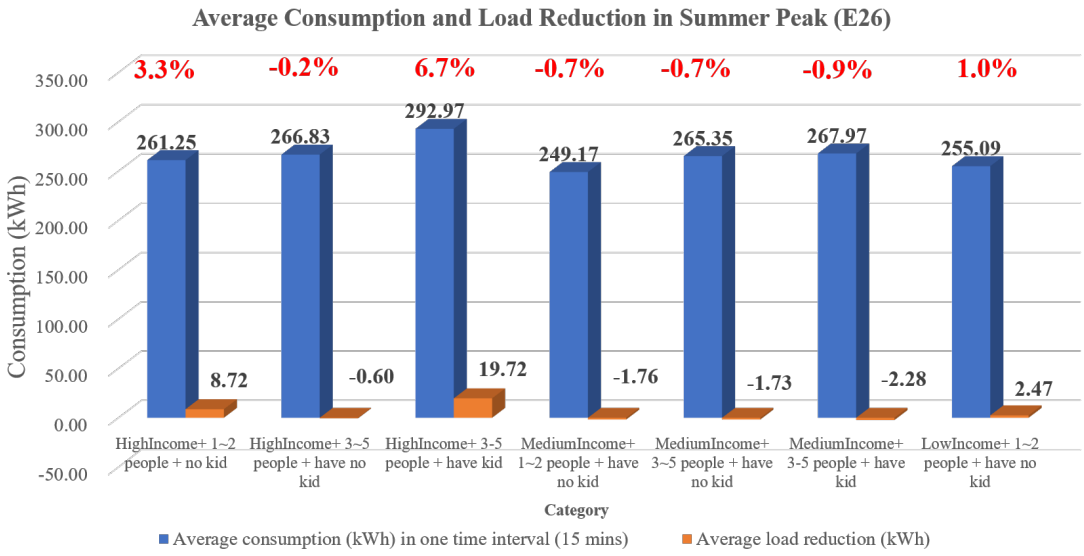

Figure 3.4: First Round Dataset Analysis for E26 Summer Peak, 60 Households in Each Group

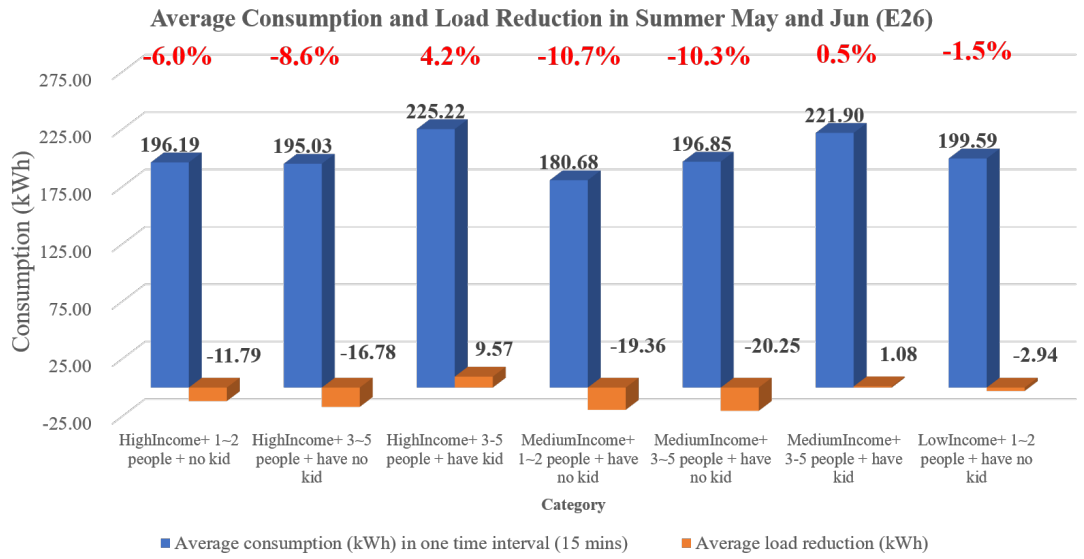

Figure 3.5: First Round Dataset Analysis for E26 Summer (May and Jun), 60 Households in Each Group

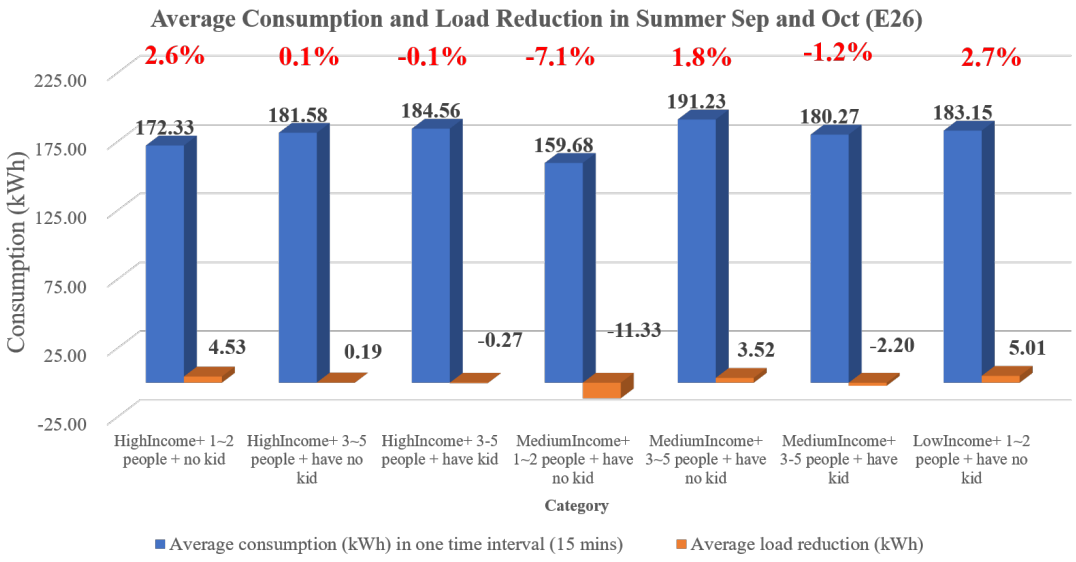

Figure 3.6: First Round Dataset Analysis for E26 Summer (Sep and Oct), 60 Households in Each Group

of E26 customers during on-peak periods are very small compared with the reductions in E21. We even find negative values. However, the variance of load reduction for E26 is large, and the average representation of results is inconclusive. We further present the result of E26 and E21 in the next section and give the explanation of why the average load reduction of E26 is small or even negative. The detailed analysis results in this section are exhibited in Appendix.

3.1.2 Detailed Analysis for the First and Second Datasets

Two exemplary plots of average consumption comparison for E21 and E26 have been shown in Fig. 3.7 and Fig. 3.8. The blue curve shows the predicted customers' baseload during the on-peak periods. The red curve represents the actual energy

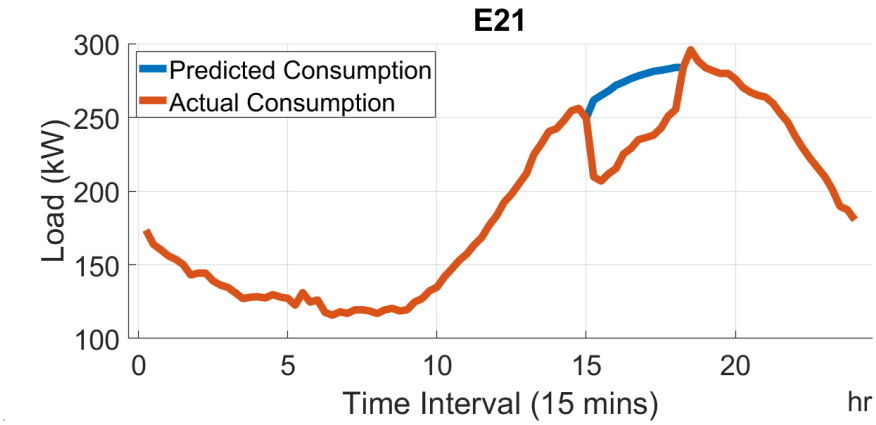

Figure 3.7: An Example of Average Consumption Comparison for "High-Income, 3-5 Occupants, No Child" Category During On-Peak Period in July and August E21

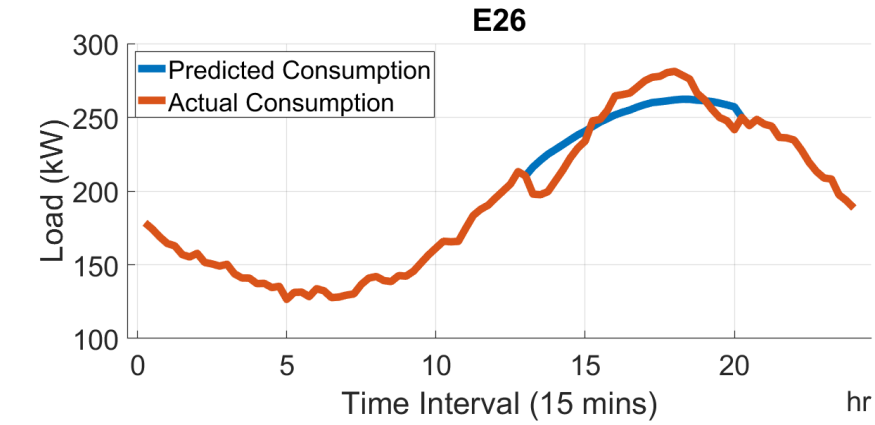

Figure 3.8: An Example of Average Consumption Comparison for "High-Income, 3-5 Occupants, No Child" Category During On-Peak Period in July and August E26

consumption. As can be seen in Fig. 3.7, there is a clear load reduction and load shifting effect happening during the on-peak periods. However, in Fig. 3.8, due to the long on-peak periods and smaller price gap, the customers who participate

in the E26 tend to reduce their loads at the beginning of the on-peak periods. After that, when the time is close to 6 p.m., people may come back home and need to consume more electricity, such as cooking, which leads to increased electricity consumption and less responsiveness to on-peak pricing. However, the load reduction effect can be seen again when the time is close to the end of the on-peak periods. We can see the actual load consumption is almost changing around the predicted consumption.

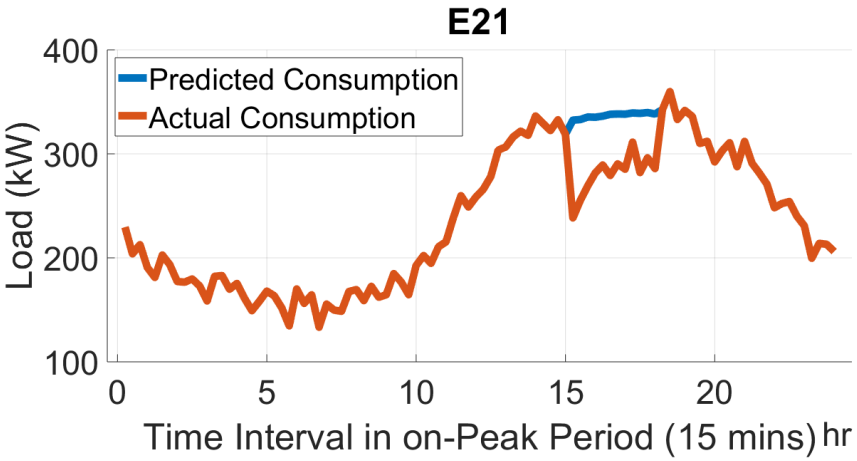

Figure 3.9: An Example of Daily Consumption Comparison in July 3rd for "High-Income, 3-5 Occupants, Have Child" Category During On-Peak Period in July and August E21

Two exemplary plots of daily consumption comparison E21 and E26 are shown in Fig. 3.9 and Fig. 3.10. Fig. 3.9 and Fig. 3.10 show an example of daily actual and predicted load consumption curve. Although the red curve is not smooth like the average load curve, we can still clearly observe the load reduction and load shifting

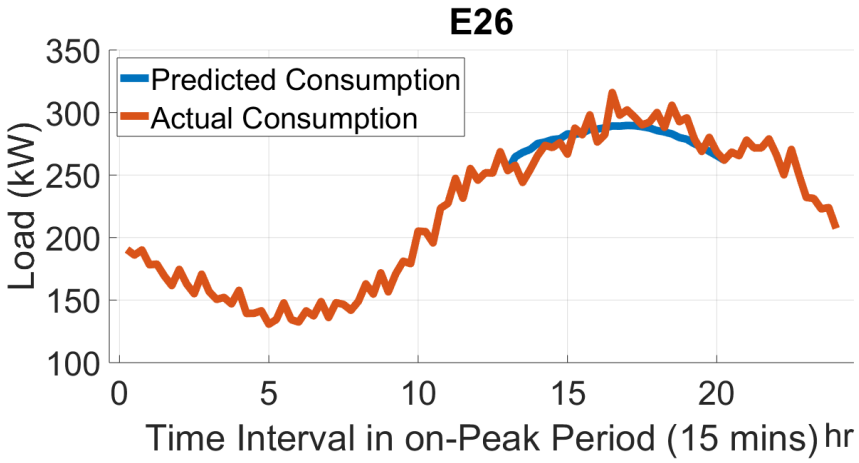

Figure 3.10: An Example of Daily Consumption Comparison in July 3rd for "High-Income, 3-5 Occupants, Have Child" Category During On-Peak Period in July and August E26

effect in the E21 plot. Moreover, the behavior of the customers who participate in E26 is further shown in this figure. The predicted baseload values are really close to the actual consumption values, which means that the customers in E26 provide less average load reduction during the on-peak periods.

Two histogram examples of load reductions for E21 and E26 customers are illustrated in Fig. 3.11 and Fig. 3.12. The load reduction values during the on-peak periods almost follow a Gaussian distribution. Most of load reduction values in E21 are positive. However, almost half of the load reduction values in E26 are negative, and the center of them is around 0. The variance of load reduction in E26 is slightly higher than the variance of load reduction in E21, which is mainly due to the long on-peak periods.

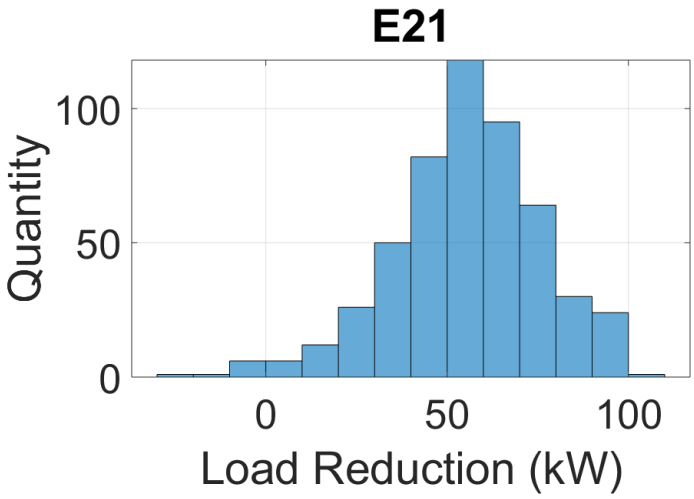

Figure 3.11: An Example of Load Reduction Histogram for "High-Income, 1-2 Occupants, No Child" Category During On-Peak Period in July and August E21

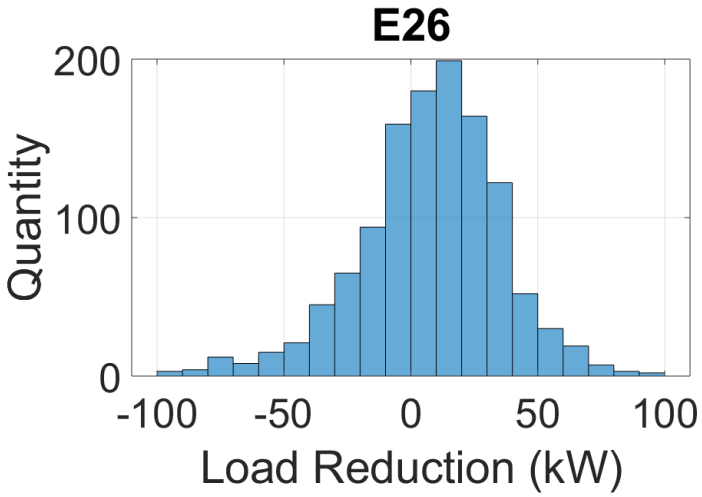

Figure 3.12: An Example of Load Reduction Histogram for "High-Income, 1-2 Occupants, No Child" Category During On-Peak Period in July and August E26

Two scatter plot examples for the E21 and E26 comparison during the on-peak period has been shown in Fig. 3.13 and Fig. 3.14.The two scatter plots show the changing of the uncertainty of customers' behavior as time changing. For the cus-

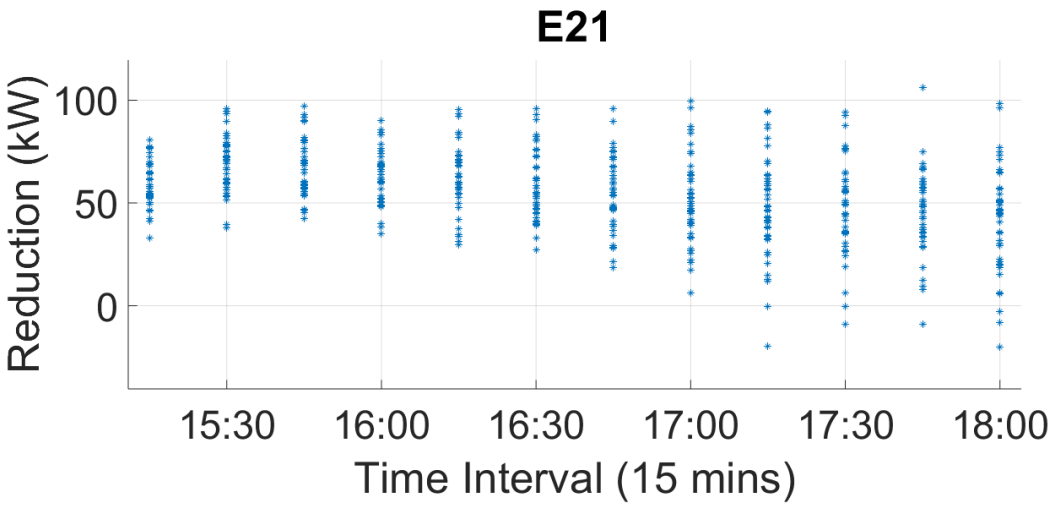

Figure 3.13: An Example of Scatter Plot for the E21 Comparison for "High-Income, 1-2 Occupants, No Child" Category During On-Peak Period in July and August E21

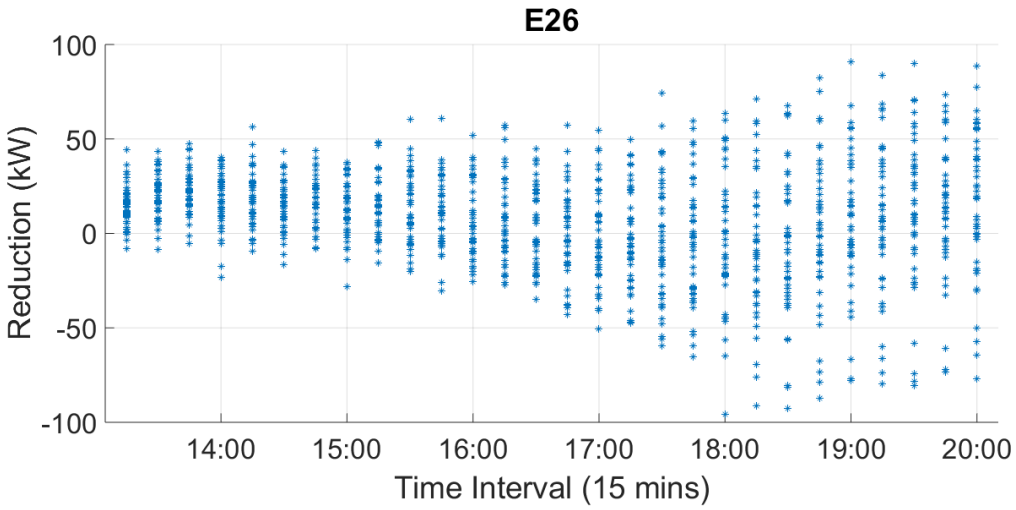

Figure 3.14: An Example of Scatter Plot for the E26 Comparison for "High-Income, 1-2 Occupants, No Child" Category During On-Peak Period in July and August E26

tomers in E21, almost all of the load reduction points are positive, and the uncertainty of customers' behavior is larger towards the end of the on-peak period. The behavior uncertainty of customers in E26 becomes more random after 6 p.m. in Fig. 3.14. This again shows the fact that after around 6 p.m. people come back home, and the uncertainty of behavior of customers is increased due to their essential electricity usage, e.g. cooking appliances. These two figures present the impact of on-peak electric price and duration on customers' behavior and the changing of customers' price responsiveness during the on-peak periods.

3.2 Third Datasets Analysis

3.2.1 Customers Responsiveness for the Third Dataset

The third dataset contains only low-income customers in the local utility. For analyzing the dataset of low-income customers, we focus on three demographic factors, including the number of occupants and two new factors: livable space and start date. The start date refers to when the customer participated in the local utility program. The customers whose start date is before 2016 are considered old customers in this study. The customers whose start date is after 2016 are treated as new customers in this study. The old customers are assumed to have more knowledge and experience about their current DR plans than new customers. The same approach is used to analyze low-income customers' consumption during the on-peak periods.

There are 60 households in each group in order to keep a highly accurate baseload model. There are enough customers to train a highly accurate baseload model, except the following three categories for E26 low-income customers.

- 6-max Occupants, Less 1500 sqft, After 2016
- 1 2 Occupants, More 2500 sqft, After 2016
- 6 max Occupants, More 2500 sqft, After 2016

We have not included these three categories in this analysis.

There are a few categories in which there are less than 60 customers. However, the accuracy of the baseload model is still relatively high with less than 60 customers. We included these categories in our analysis. The detailed low-income customers' analysis results have been shown in the Appendix. In order to exhibit better results visualization, several plots are shown below. In these figures, the impact of the number of occupants and livable space on the customers' behavior during the on-peak periods is revealed while the start date is fixed.

Fig. 3.15 and Fig. 3.16 show the same results with two different arrangements in order to facilitate visual analysis of the impact of the number of occupants and livable space on low-income customers' consumption reduction during on-peak periods when the start date is fixed as 'after 2016'. As can be seen, the customers' percentage of load reduction values are increasing as the number of occupants or the livable space increases. The new local utility customers (start date is 'after

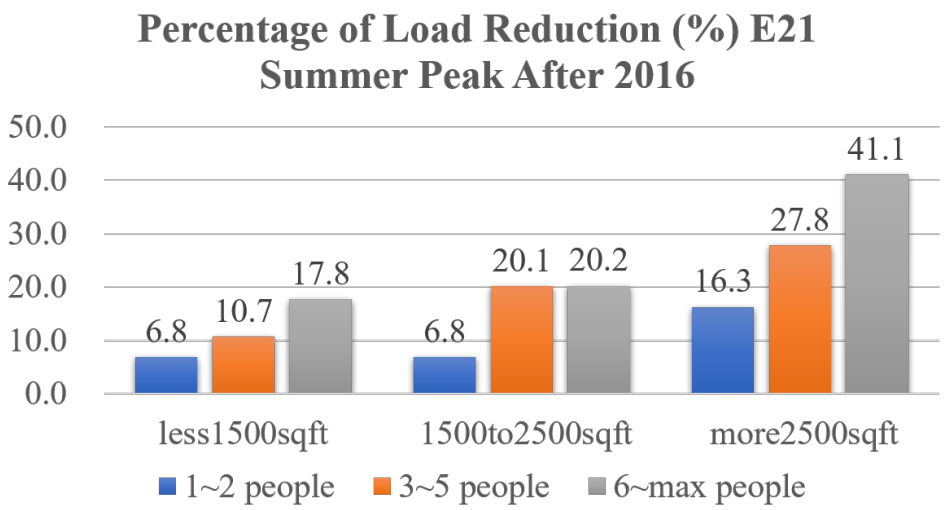

Figure 3.15: An Example of Results Visualization for Low-Income Customers in Summer Peak E21: Fixed Start Date (After 2016), Compare Impact of Number of Occupants

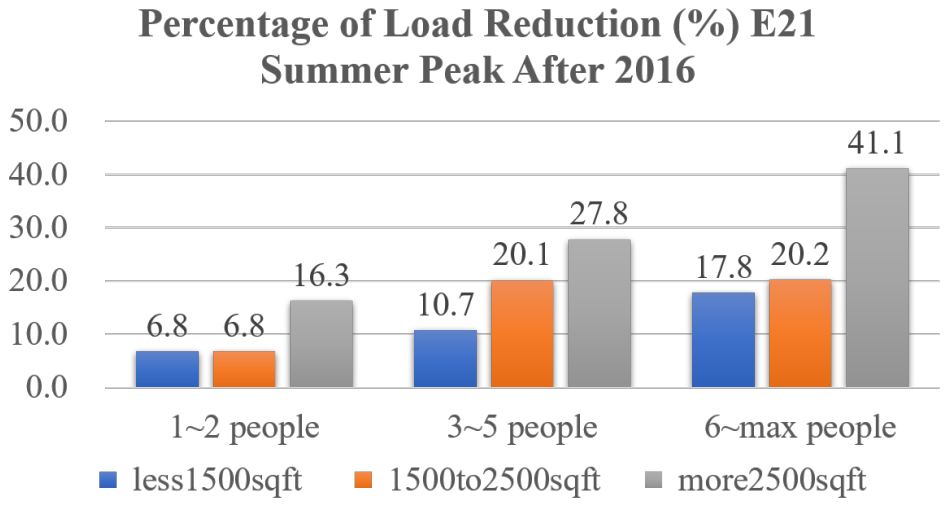

Figure 3.16: An Example of Results Visualization for Low-Income Customers in Summer Peak E21: Fixed Start Date (After 2016), Compare Livable Space

2016') tend to reduce more load during the on-peak periods, if they have more livable space or there are more people living in the house.

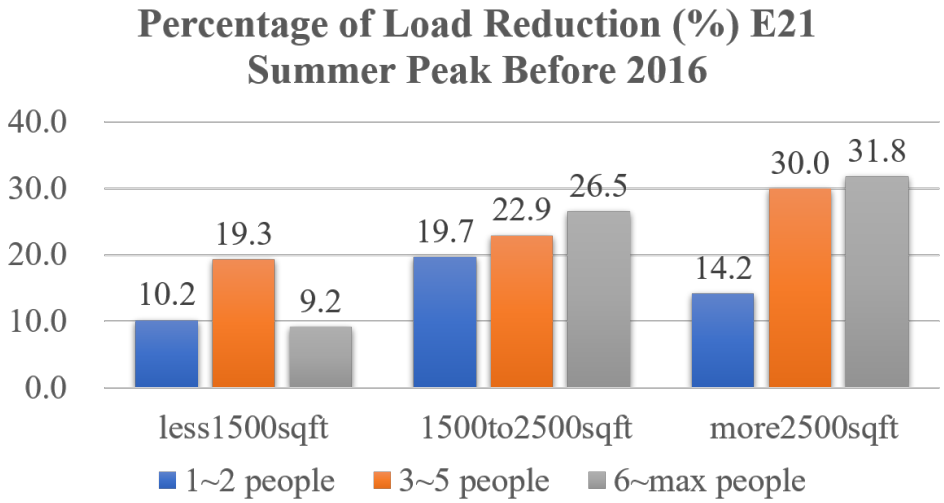

Figure 3.17: An Example of Results Visualization for Low-Income Customers in Summer Peak E21: Fixed Start Date (Before 2016), Compare Impact of Number of Occupants

Fig. 3.17 and Fig. 3.18 exhibit the impact of the number of occupants and livable space on the old low-income local utility customers' behavior. As can be seen, the old low-income local utility customers' behavior almost follows the same trend during the on-peak periods, except the customers in 1-2 occupant, "1500 2500 sqft, before 2016" category. In Fig. 3.17 and Fig. 3.18, the behavior of old low-income local utility customers has been shown. The number of occupants and livable space play an important role in residential customers' behavior during the on-peak periods.

Fig. 3.19 shows the percentage of load reduction comparison for start date (af-

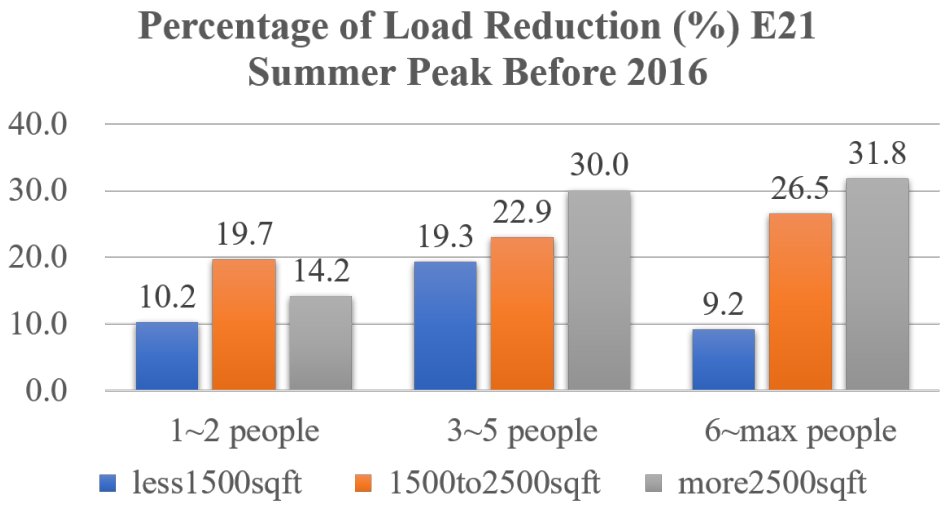

Figure 3.18: An Example of Results Visualization for Low-Income Customers in Summer Peak E21: Fixed Start Date (Before 2016), Compare Livable Space

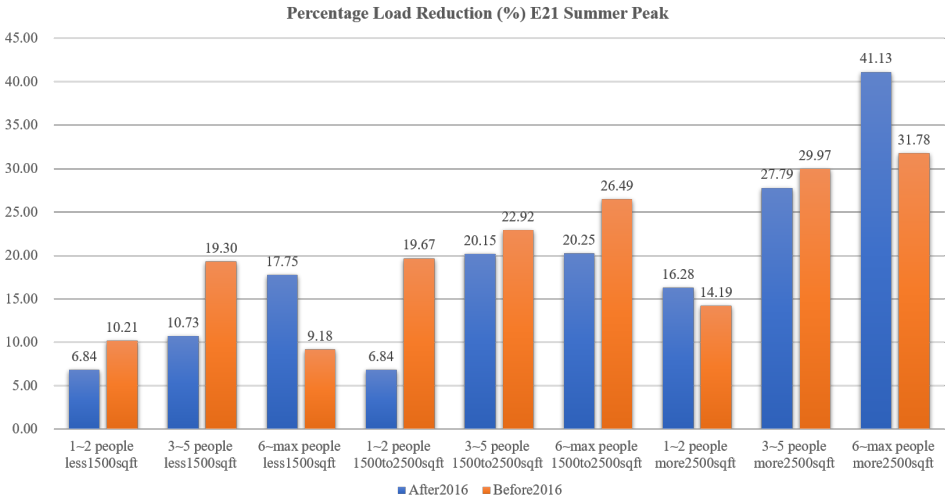

Figure 3.19: An Example of Results Visualization for Low-Income Customers in Summer Peak E21: Impact of Start Date (Old and New Customers)

ter 2016 and before 2016) for E21. From Fig. 3.19, in most of the cases, the old customers provide more percentage of load reduction, except three customers cat-

egories, which are “6 max people, more 2500 sqft”, “1 2 people, more 2500 sqft”, and “6 max, less 1500 sqft”. However, note that there are not enough customers in “6 max people, more 2500 sqft” and “1 2 people, more 2500 sqft” categories. Only 22 customers are in “6 max people, more 2500 sqft” category, and 55 customers in the “6 max people, more 2500 sqft” category. One reason why the load reduction values don’t follow the same trend (the new customers in these two categories provide more percentage of load reduction than old customers) is the data bias issue. However, there is one special category which is the “6 max, less 1500 sqft” for E21. There are enough customers in this category, but the new customers in this category are willing to provide more percentage of load reduction compared with the old customers.

3.2.2 Detailed Analysis for the Third dataset

The variance of load reduction for E26 low-income customers is large, the average representation of results is inconclusive. Thus, we present the detailed result of E26 and E21 low-income customers in the following sections.

Fig. 3.20, Fig. 3.21, Fig. 3.22 and Fig. 3.23 present the average customers’ baseload and average customers actual energy consumption. The blue curve shows the predicted customers’ baseload during the on-peak periods. The red curve represents the actual energy consumption. The four figures show the impact of demographic factors, electric price, and duration on customers’ behavior during the on-peak periods. In Fig. 3.20, the load reduction of customers in “1500 2500 sqft, 1-

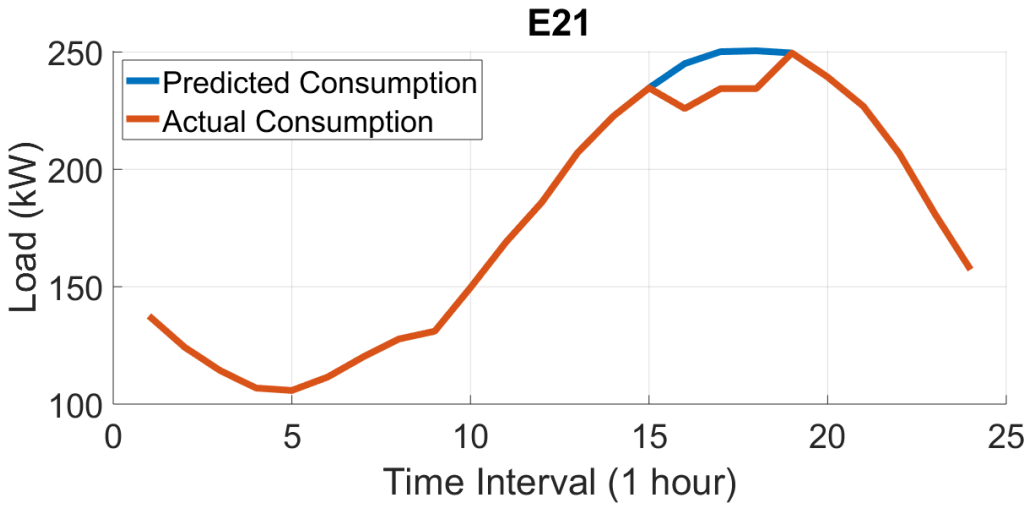

Figure 3.20: An Example of Average Consumption Comparison for "1500 2500 sqft, 1-2 Occupants, After 2016" Category During On-Peak Period in July and August E21

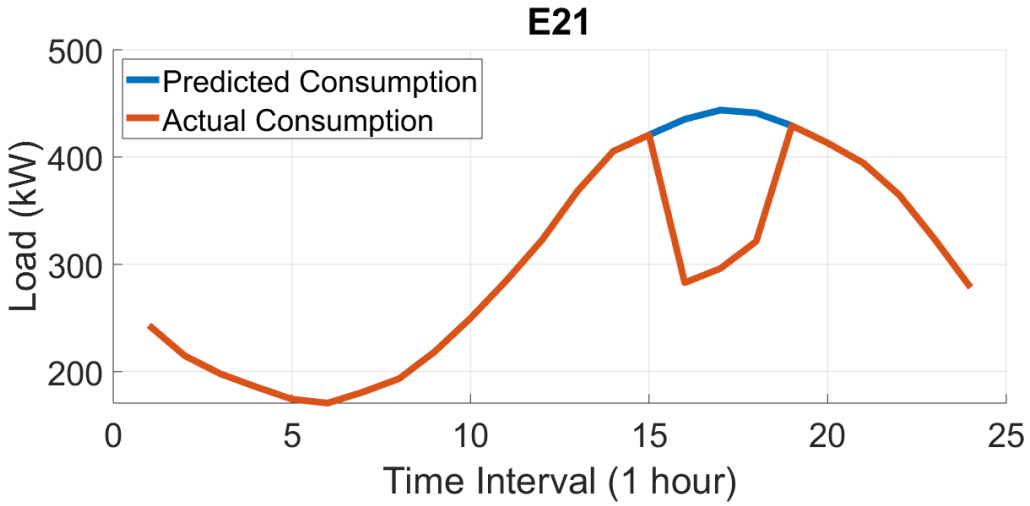

Figure 3.21: An Example of Average Consumption Comparison for "More Than 2500 sqft, More Than 6 Occupants, Before 2016" Category During On-Peak Period in July and August E21

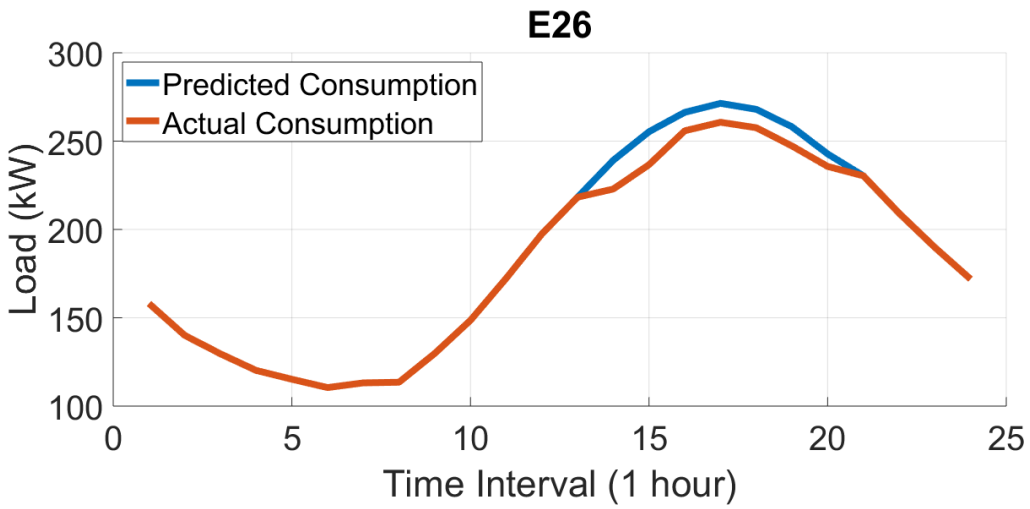

Figure 3.22: An Example of Average Consumption Comparison for "1500 2500 sqft, 1-2 Occupants, After 2016" Category During On-Peak Period in July and August E26

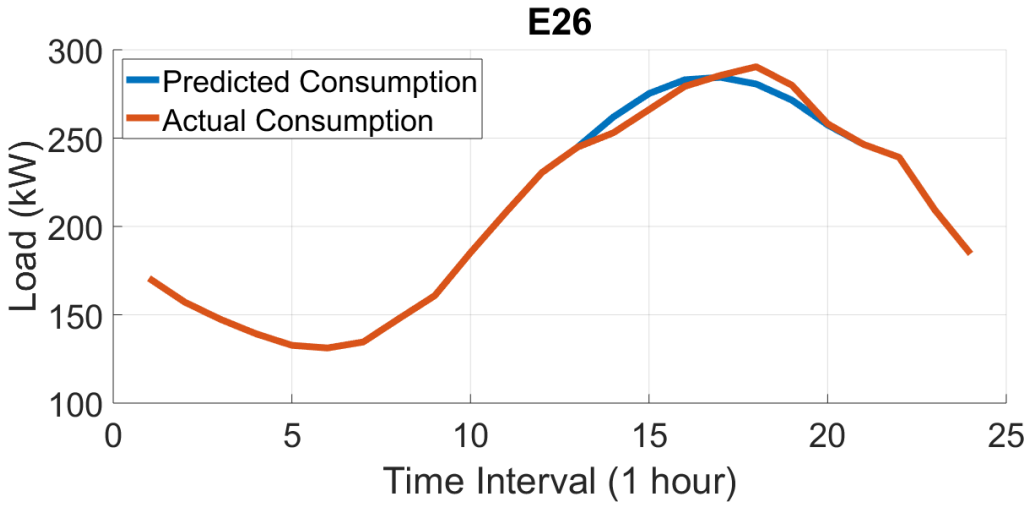

Figure 3.23: An Example of Average Consumption Comparison for "1500 2500 sqft, 3-5 Occupants, Before 2016" Category During On-Peak Period in July and August E26

2 Occupants, After 2016" category is smaller than the load reduction of customers in "More Than 2500 sqft, More Than 6 Occupants, Before 2016" category. In Fig. 3.22, the load reduction effect can be seen all over the on-peak periods for the E26 program. However, Fig. 3.23 shows that customers may increase energy consumption level during the on-peak periods. As can be seen, the intersection point of the baseload curve and actual consumption curve is very close to the time when the people usually come back home after work. The customers' consumption is greater than the predicted baseload. This is the reason why there is negative average load reduction for some of the categories. These customers consume less energy at the beginning of the on-peak periods, but when they go home, the energy consumption rise and exceed the predicted baseload. It leads to a negative calculated load reduction value.

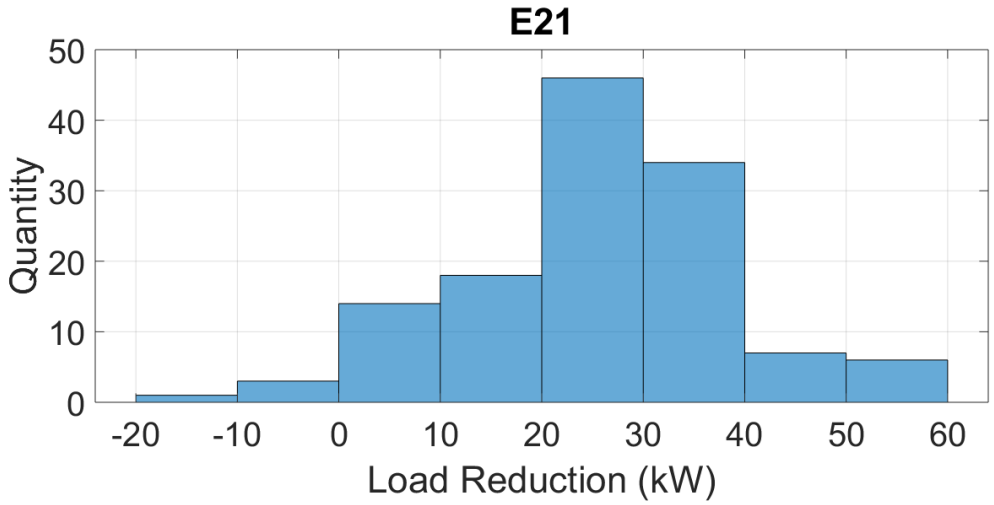

Figure 3.24: An Example of Load Reduction Histogram for "Less 1500 sqft, 3-5 Occupants, After 2016" Category During On-Peak Period in July and August E21

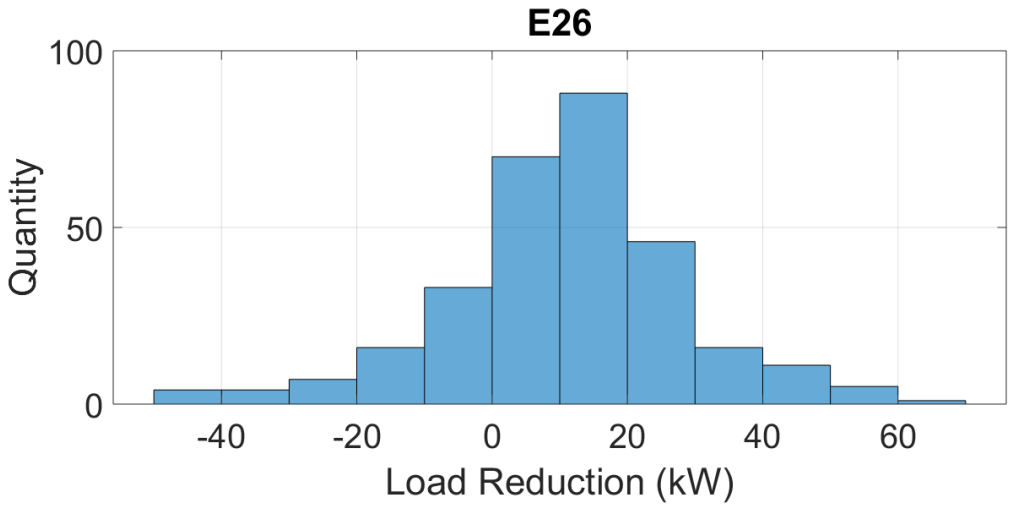

Figure 3.25: An Example of Load Reduction Histogram for "Less 1500 sqft, 3-5 Occupants, After 2016" Category During On-Peak Period in July and August E26

Fig. 3.24 and Fig. 3.25 show the histogram of on-peak reduction values for E21 and E26 low-income customers. As can be seen, the reduction values almost follow a Gaussian distribution. In Fig. 3.24, the mean of reduction values is positive and close to 30 kW. However, the mean of reduction values in Fig. 3.25 is close to 0 kW. We can observe several negative values in Fig. 3.24, but lots of negative values in Fig. 3.25. This shows the reason why the most of percentage values of average load reduction of E26 customers in Appendix table are small or even negative, while most of the percentage values of average load reduction of E21 customers are large and non-negative. The two figures also illustrate the impact of on-peak electric price and duration on the customer's behavior. Higher on-peak electric price and smaller duration lead to a higher customers' price responsiveness during the on-peak periods.

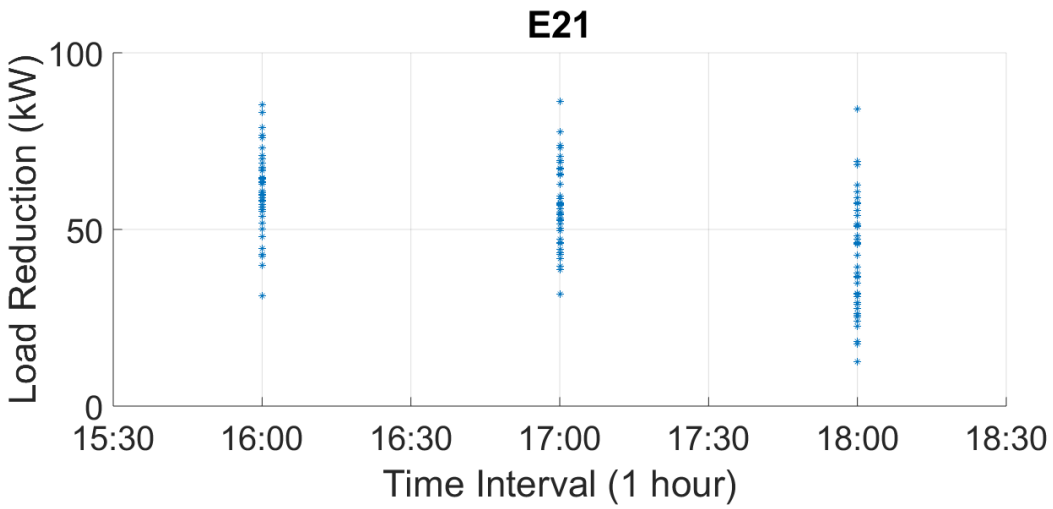

Figure 3.26: An Example of Scatter Plot for the E21 Comparison for "1500 2500 sqft, 3-5 Occupants, After 2016" Category During On-Peak Period in July and August E21

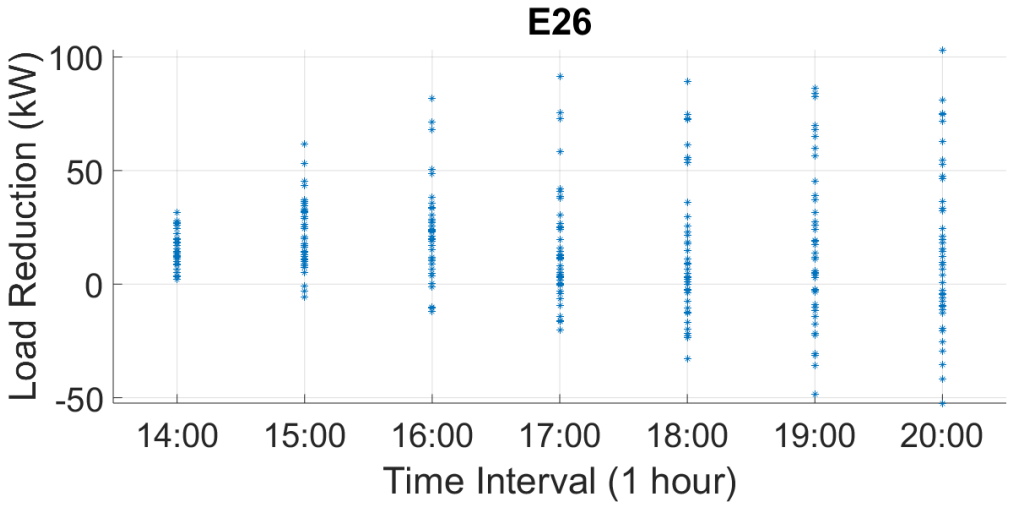

Figure 3.27: An Example of Scatter Plot for the E21 Comparison for "1500 2500 sqft, 3-5 Occupants, After 2016" Category During On-Peak Period in July and August E26

Two exemplary scatter plots are shown in Fig. 3.26 and Fig. 3.27, which exhibits the details of customers' price responsiveness for each the on-peak intervals. Many large positive load reduction values are observed in Fig. 3.26. The variance is almost unchanged. However, as can be seen, the variance of load reduction values increases over time in Fig. 3.27, which means that the customers' behavior becomes more random when it gets close to the end of the on-peak periods. A large load reduction variance is observed after 5 pm. It is time for people to come home from work, and people have their own schedule after coming back home. For example, they may need to use more energy for a series of behavior such as cooking and watching TV. This results in more negative load reduction values that are observed, and the increased uncertainty of customers' behavior.

3.2.3 One Demographic Factor Analysis for the Third Dataset

In this third dataset analysis, one demographic factor analysis is used to reveal which demographic factors have the largest impact on low-income customers' behavior during on-peak periods. The same approach is used for one demographic factor analysis. For each demographic factor analysis, all of the customers who have the same demographic factors are used. This analysis exhibits the impact of the household's livable space, number of occupants and start date on low-income customers' price responsiveness during the on-peak periods. Percentage of average load reduction values are considered in this analysis. The detailed analysis results are shown in the Appendix.

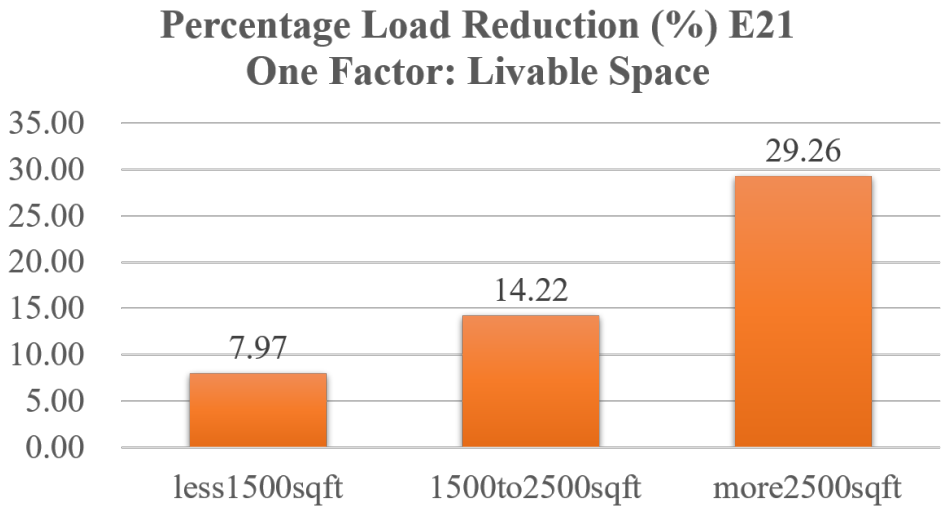

Figure 3.28: Percentage of Load Reduction During On-Peak, An Example of One Demographic Factor Analysis for the E21: Livable Space

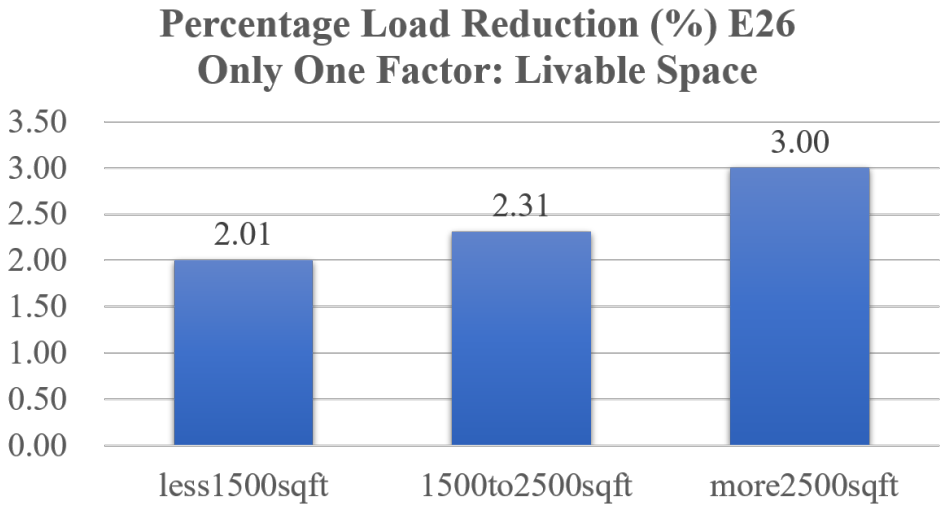

Figure 3.29: Percentage of Load Reduction During On-Peak, An Example of One Demographic Factor Analysis for the E26: Livable Space

Fig. 3.28 and Fig. 3.29 show the impact of the household's livable space. As can be seen, the household's livable space plays an important role in low-income

customers' responsiveness. The larger the house is, the higher the percentage of load reduction the customers tend to provide during the on-peak periods for both E21 and E26 low-income customers. This is expected as people with smaller living space may have a less flexible load (e.g. only one air conditioning unit versus two for larger living spaces) to reduce their consumption. E21 low-income customers provide more load reduction comparing with E26 low-income customers. In Fig. 3.28, the E21 low-income customers who live in a house having more than 2500 sqft provide more than twice the load reduction as much as the customers who live in a house having 1500 to 2500 sqft. The E21 low-income customers who live in a house having 1500 to 2500 sqft provide almost twice the load reduction as much as the customers who live in a house having less than 1500 sqft. In Fig. 3.29, only a small difference between each class can be observed. The percentage of load reduction values in Fig. 3.29 are small compared with the values in Fig. 3.28.

Fig. 3.30 and Fig. 3.31 exhibit the impact of the number of occupants. For E21 low-income customers, the more occupants living in the house, the more load reduction they tend to provide during the on-peak periods. The households that have 3-5 occupants reduce twice as much as the households that have 1-2 occupants. The households having more than 6 occupants provide the largest percentage of load reduction. However, for E26 low-income customers, the households that have 3 5 occupants provide the least load reduction, while the households that more than 6 occupants provide the most load reduction. As can be seen, the

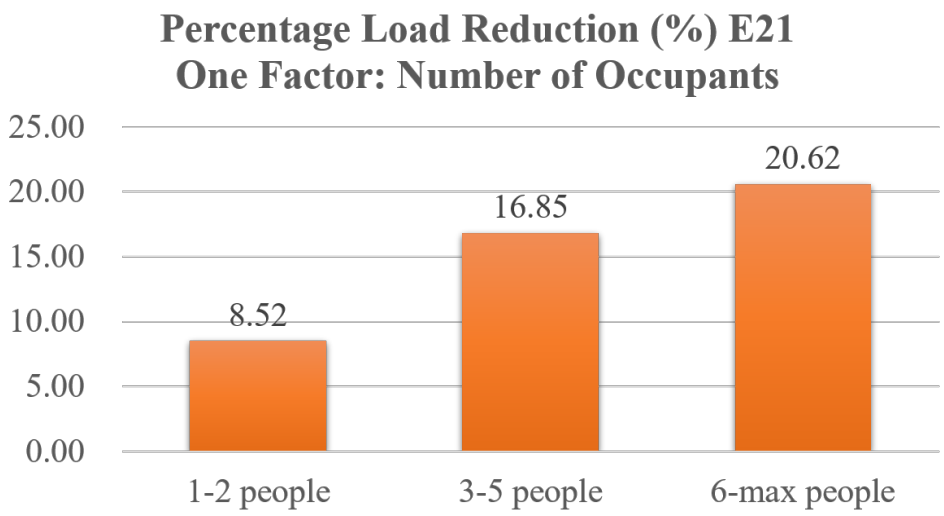

Figure 3.30: Percentage of Load Reduction During On-Peak, An Example of One Demographic Factor Analysis for the E21: Number of Occupants

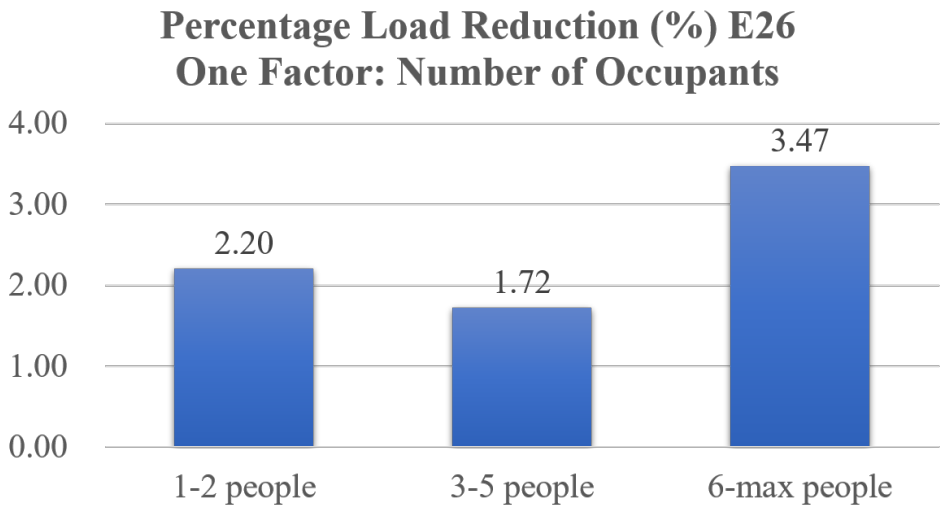

Figure 3.31: Percentage of Load Reduction During On-Peak, An Example of One Demographic Factor Analysis for the E26: Number of Occupants

E26 low-income customers percentage of load reduction values are much smaller than E21 low-income customers' percentage of load reduction values.

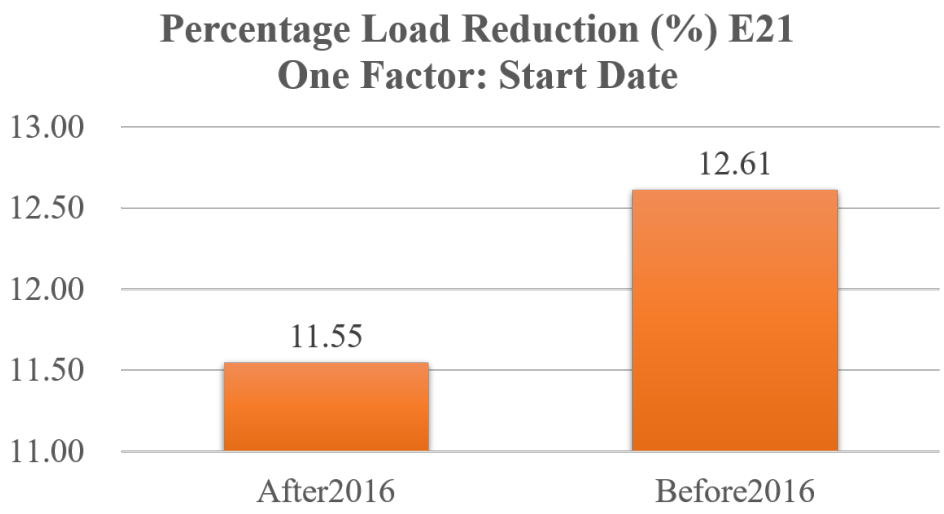

Figure 3.32: Percentage of Load Reduction During On-Peak, An Example of One Demographic Factor Analysis for the E21: Start Date

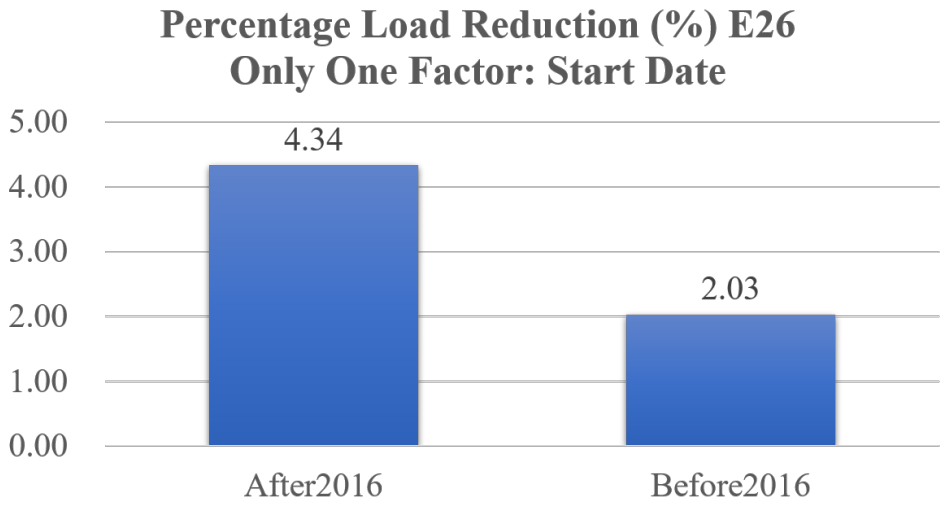

Figure 3.33: Percentage of Load Reduction During On-Peak, An Example of One Demographic Factor Analysis for the E26: Start Date

Another analysis compares the load reduction effect between the new customers (start date before 2016) and the old customers (start date after 2016) in Fig.

3.32 and Fig. 3.33. As can be seen, for E21 low-income customers, the old customers tend to provide more load reduction and new customers reduce less during the on-peak period. This result shows that the customers start to learn their price plan policy and how to adjust their consumption in response to the on-peak pricing after they continue to be local utility customers for more than a certain period of time. These customers get more knowledge and experience over time. This knowledge and experience lead customers how to make a rational choice in order to save some money for their electric bill. However, for E26 low-income customers, this start date has the opposite effect. The new customers tend to provide more load reduction and old customers reduce less during the on-peak period. It is shown that more experience of their TOU plan makes customers pay less attention to the on-peak load reduction. The main reason why this happens is that the on-peak price of E21 is much larger than the on-peak price of E26, while the on-peak duration of E26 is much longer than the on-peak duration of E21.

3.3 Fourth Dataset Analysis

3.3.1 Customers Responsiveness for the Fourth Dataset

The Fourth dataset contains all local utility customers who participated in the emergency demand response (EDR) program with different rate policies in 2018. For each customer in this dataset, there is a smart nest thermostat installed. SPR communicates each DR event through the customers' thermostat. Nest thermostat

allows local utility to control the thermostat temperature, however, the customers have the right to re-adjust their device at any time. Thermostat settings during an event may seem extreme and may adjust frequently. These settings are designed to maximize the pre-cooling or conservation impacts. DR events typically occur between noon and 9 p.m. on weekdays only (holidays excluded). Customers receive local utility bill credit for participating in this program.

In this dataset, there are totally 9 DR events in 2018. The event time has been presented in Table 1. In this analysis, we mainly focus on E23 customers. Note that E23 is the flat rate for residential customers in local utility. The same strategy in the third dataset analysis is used to predict customers' baseload. However, due to the characteristic of EDR, utility needs to notify customers one hour in advance. This issue impacts customers electricity consumption within one hour before the beginning of the DR event. Thus, we predict customers' baseload from one hour before the beginning of the DR event.

3.3.2 All E23 Customers Analysis for the Fourth Dataset

An analysis is conducted for all E23 customers in the fourth dataset.

Fig. 3.34 shows an example of baseload prediction for the fourth dataset for only E23 customers. The blue curve is the predicted baseload. The red curve is the actual customers consumption. The two red marked points are the start and end points of the DR event. The DR event time in Fig. 3.34 is 17:00 to 19:00 at Jun/22/2018. As can be seen, customers reduce their load dramatically after the

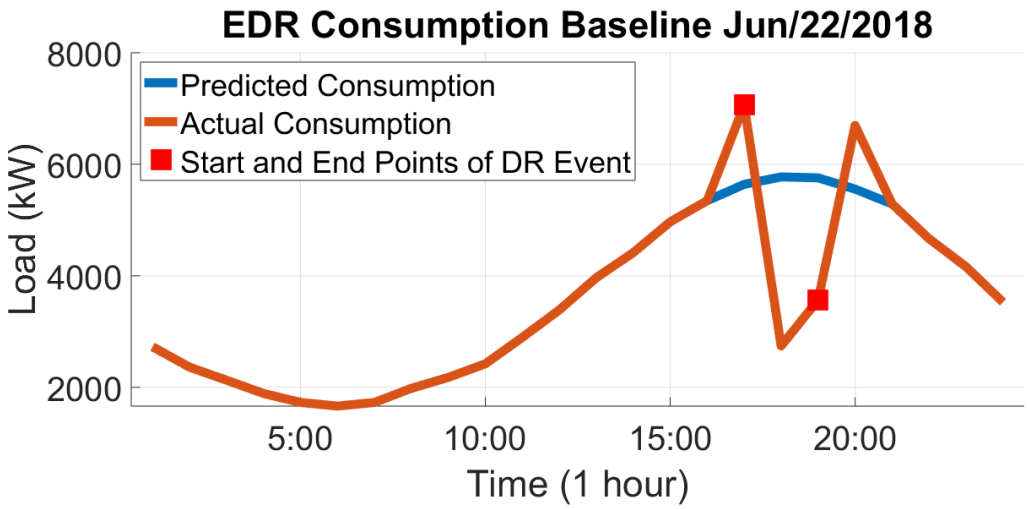

Figure 3.34: An Example of EDR Consumption Baseline and Actual Consumption Comparison for All E23 Customers at Jun/22/2018 (17:00 - 19:00)

start of the DR event. The customers consumption remains at a very low level during the DR event. local utility adjusts the nest thermostat temperature setting to precool the customers houses before the start of DR event and cool their customers house after the end of DR event, two spikes can be seen in Fig. 3.34. This shows that the baseload model is accurate enough to detect the abnormal period before and after the DR event.

Fig. 3.35 and Fig. 3.36 present scatter plots of load reduction values during the DR events by grouping all E23 customers in the fourth dataset. Fig. 3.35 and Fig. 3.36 exhibit the customers' load reduction and behavior during conservation periods from two different aspects. Fig. 3.35 shows consumption reductions during various time periods of conservation events. Fig. 3.36 shows load reductions for first, second, and third intervals of all conservation events despite varying start

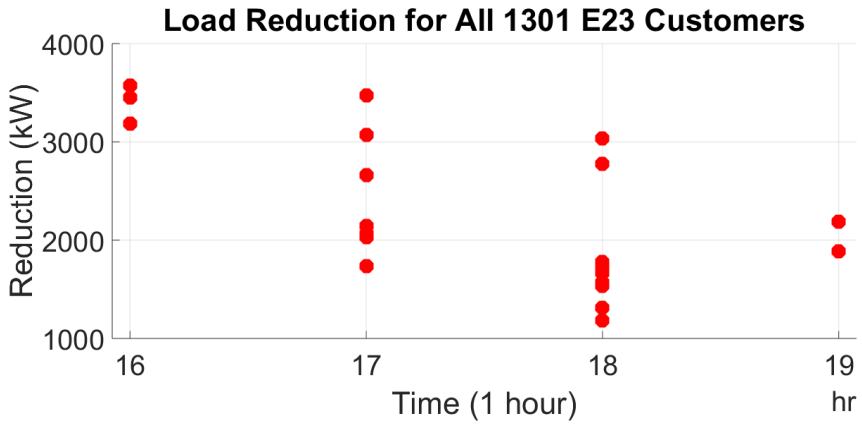

Figure 3.35: An Example of All E23 Customers Load Reduction Scatter Plot for Each Hour During the DR event at Jun/22/2018 (17:00 - 19:00)

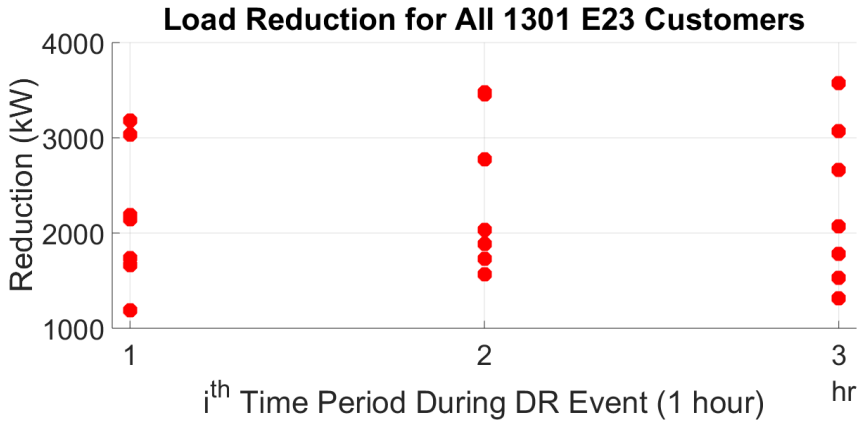

Figure 3.36: An Example of All E23 Customers Load Reduction Scatter Plot for i^{th} Hour During the DR event at Jun/22/2018 (17:00 - 19:00)

and end times of each event. The detailed analysis results are shown in the Appendix.

3.3.3 *One Demographic Factor Analysis for the Fourth Dataset*

One demographic factor analysis has been done in the fourth dataset to reveal the customers behavior considering different demographic factors. The same approach is used for one demographic factor analysis. For each demographic factor analysis, all of the customers who have the same demographic factors are used. This analysis shows the impact of the household's income level and the number of occupants on customers' price responsiveness during DR events. The percentage of the average load reduction values is considered in this analysis.

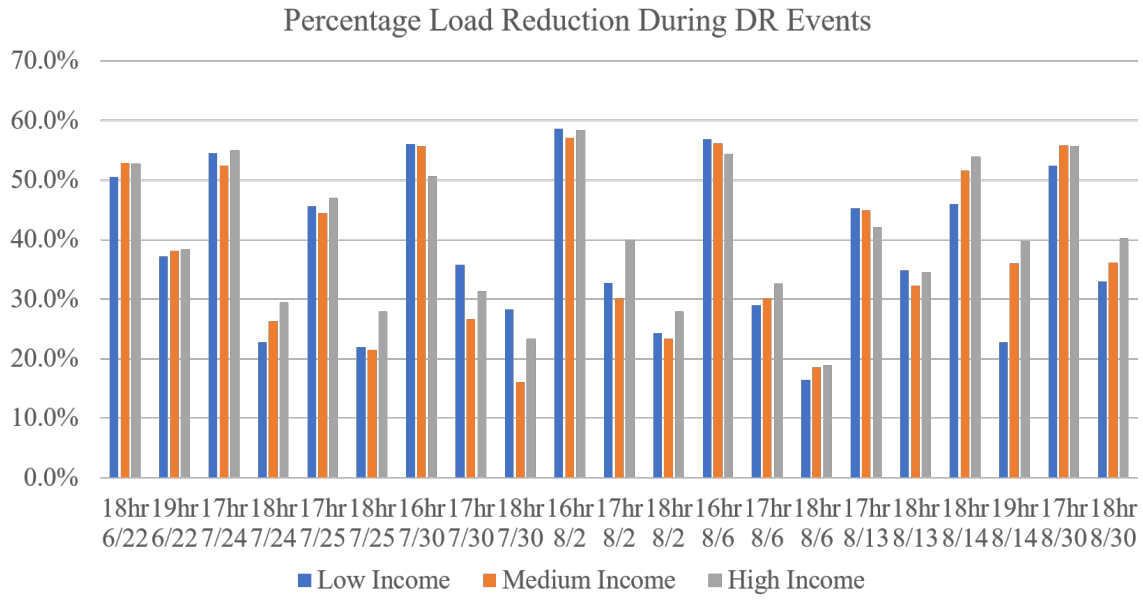

Figure 3.37: One Demographic Factor Analysis, Percentage Load Reduction Result: Income Level

Fig. 3.37 and Fig. 3.38 show all of the percentage of the load reduction results for one demographic factor analysis in the fourth dataset. More detailed results

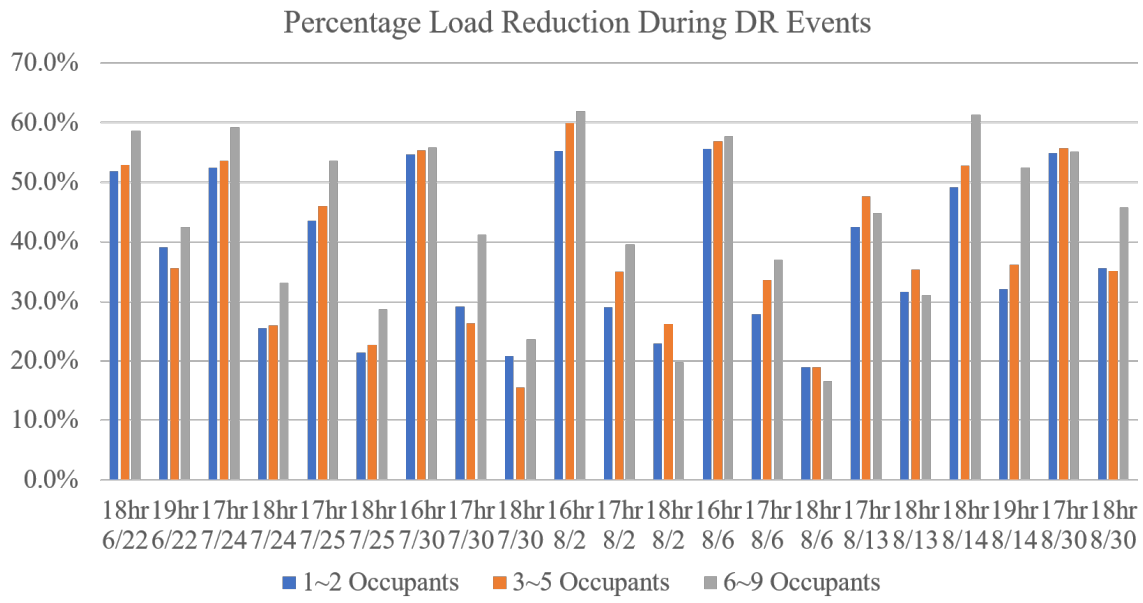

Figure 3.38: One Demographic Factor Analysis, Percentage Load Reduction Result: Number of Occupants

are shown in the Appendix.

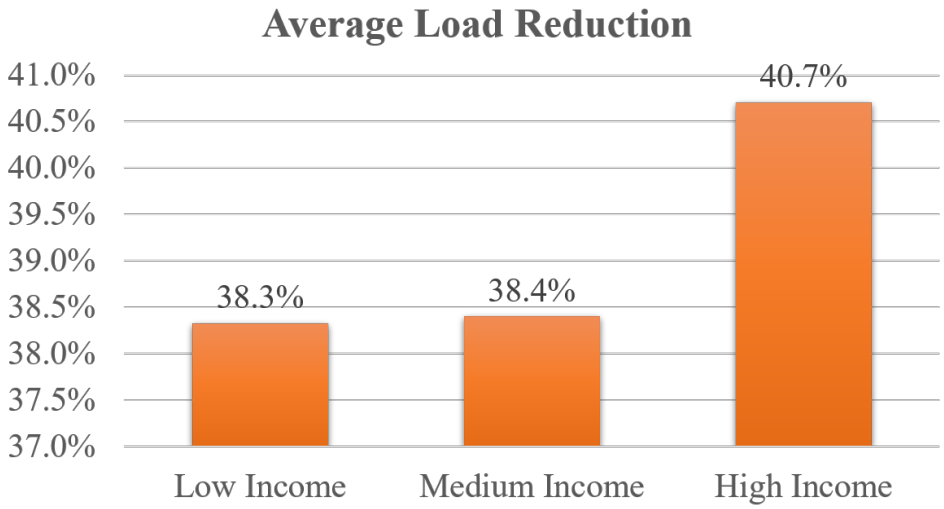

Figure 3.39: One Demographic Factor Analysis, Average Percentage Load Reduction Result: Income Level

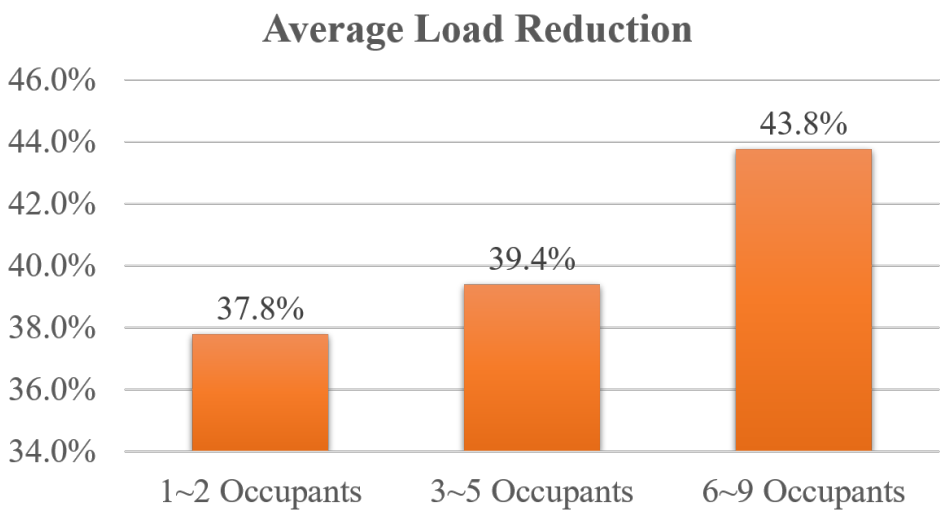

Figure 3.40: One Demographic Factor Analysis, Average Percentage Load Reduction Result: Number of Occupants

Fig. 3.39 and Fig. 3.40 exhibit the average percentage of load reduction results for each income level and the number of occupants. The average percentage of load reduction is computed by calculating the average load reduction values for each demographic factor group overall 9 EDR periods. In Fig. 3.39, low-income and medium-income customers perform similarly during the DR events. Moreover, high-income customers tend to provide more load reduction during the DR events comparing with the other two groups. As can be seen in Fig. 3.40, the average percentage of load reduction values are increasing as the number of occupants living in the house is increasing.

3.4 Summary

The customers' behavior for different price plans and different demographic factors during the on-peak period is presented in this chapter.

The results of the first and second dataset analyses show that the demographic factors play an important role in customers' behavior, which is reflected by how much load reduction the customers in each category will provide. The impact of income level and if having child has been shown in the first round dataset. The high-income and medium-income level customers have relatively high load reduction. One reason is that they may have more interruptable loads comparing with low-income customers. If having child places an important role in high-income families. Although it leads to higher daily electricity consumption, it also results in a larger load reduction during the on-peak period.

The impact of the number of occupants, livable space and start date has been exhibited in the third round analysis. For E21 customers, the more people that live in the house, the more load reductions the household customers will tend to provide. For both E21 and E26 customers, if the house is larger, the customers will have a higher capability to reduce more load during the on-peak period. It's also revealed that the start date of becoming customers is related to customers' behavior during the on-peak period. For E21 customers, the more experience and knowledge the customers have, the more rational choice the customers will make. However, for E26 low-income customers, this start date has the opposite effect.

The new customers tend to provide more load reduction and old customers reduce less during the on-peak period. It is shown that more experience of their TOU plan makes customers pay less attention to the on-peak load reduction.

For the fourth dataset analysis (emergency DR), high-income customers tend to provide more load reduction during the DR events comparing with the other two groups. Low-income and medium-income customers perform similarly during the DR events. The more people that live in the house, the more load reductions the household customers will tend to provide.

The analysis results in this chapter will be further utilized in the distribution grid operation optimization model. The customers will be divided based on their demographic information. The price responsiveness of each demographic category will be represented by different Gaussian distributions. Monte Carlo Simulation (MCS) will be utilized to sample scenarios, which are used for chance-constraint based SOCP-ACOPF, from these Gaussian distributions.

The next chapter will illustrate the chance-constraint based SOCP-ACOPF optimization model.

Chapter 4

CHANCE-CONSTRAINT BASED SOCP-ACOPF

The chance-constraint is utilized to handle the optimization model that has high uncertainty in its input data. In the distribution grid, the penetration of DERs is increasing dramatically nowadays. Due to the uncertainty of DERs, the operation of the distribution system is facing huge challenges. In the distribution grid, if the operator chooses to use and rely on more energy coming from DERs, the distribution system will face higher risk due to the characteristic of DERs, namely uncertainty and variability. However, chance-constraint can be powerful to overcome uncertainty issues. For example, a violation threshold can be set to guarantee that the solution of the model can have an acceptable risk level. By implementing a chance-constraint, the operation of the distribution grid can be improved.

4.1 Uncertainty of Customers' Price Responsiveness

From chapter 3, the customers' behavior during the on-peak period for different demographic categories is analyzed. The customers' price responsiveness can be represented as different Gaussian distributions.

An exemplary figure of the customers' price responsiveness for different time periods, demographic categories and price plans is exhibited in Fig. 4.1. The five different Gaussian distributions are the probability density function (PDF) of the

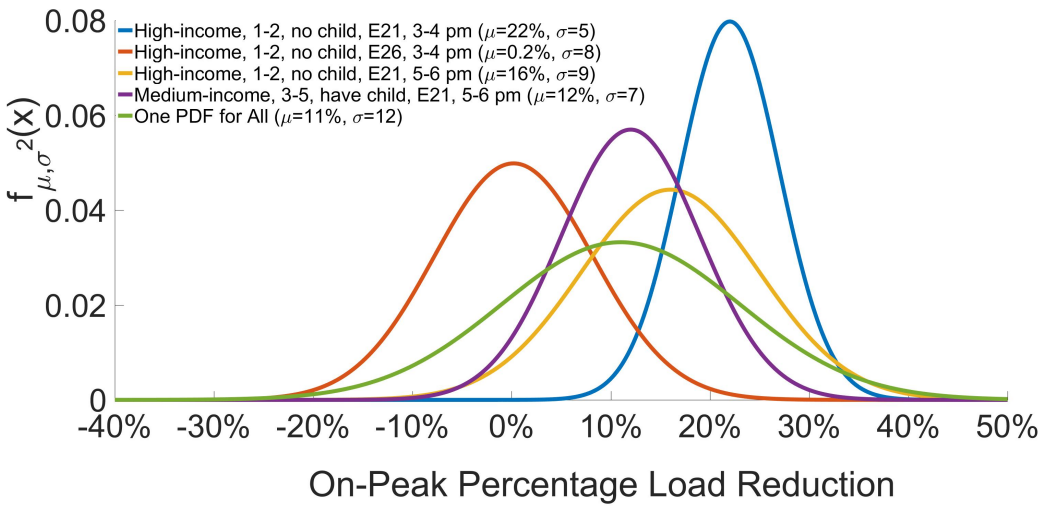

Figure 4.1: Customers' Price Responsiveness for Different Time Periods, Demographic Categories and Price Plans

customer's price responsiveness. As can be seen from Fig. 4.1, for different electricity price plans, different time periods, different demographic categories, the customer's electricity usage behavior is different. If we don't consider these differences, but only use one Gaussian distribution to represent the price responsiveness of all customers at different time periods and different price plans, the green curve in Fig. 4.1 is what we can obtain. Comparing with the other Gaussian distributions in Fig. 4.1, the green curve has higher variance, which means higher uncertainty. This shows that analyzing customers' behavior based on time periods, demographic categories and price plans can reduce the uncertainty of DR, which will then improve the operation of the power system. The utility and operator can also benefit from these analyzed customers' price responsiveness distribu-

tions. For example, they can call the customer groups, which have less uncertainty of customers' behavior, to reduce consumption. They can target customer groups that have larger uncertainty of customers' behavior and provide less load reduction in order to find the reasons why these happen.

The Monte Carlo Simulation (MCS) is utilized to sample the scenarios from these Gaussian distributions. However, due to the time limit, fake distributions are used in this chapter. In future work, the obtained customers' price responsiveness distributions will be used for MCS scenarios sampling.

The following sections will illustrate chance-constraint SOCP-ACOPF.

4.2 Flowchart of Chance-Constraint Based Model

A flowchart of implementing chance-constraint in the optimization model is shown in Fig. 4.2. This algorithm borrows the idea of contingency analysis in the power system.

Firstly, in the flowchart, second-order-cone-programming (SOCP) ACOPF is solved by using Cplex in order to obtain the dispatch of the distribution system. The uncertainty and variability of DERs are not considered in the first step. The initial proportion of total DERs to total load $I_t^{p}\%$ and $I_t^{q}\%$ are defined as the dispatch of the active and reactive power proportion of DERs without any uncertainty and variability. They are calculated by the total dispatch of active and reactive power of DERs divided by the total load in the system. Secondly, the Monte Carlo Simulation (MCS) is used to generate scenarios. In this study, due to the time limit,

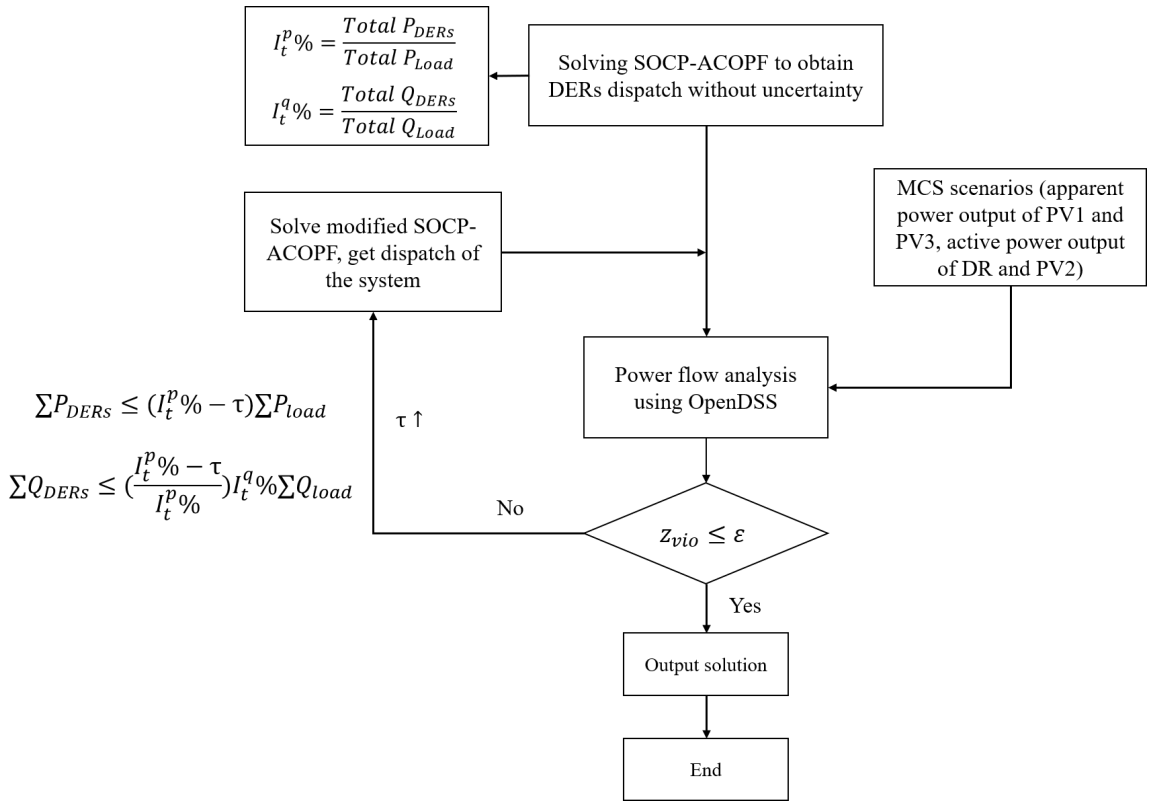

Figure 4.2: A Flowchart of Chance-constraint model

fake distributions are utilized for MCS scenarios sampling. The outputs of six DR buses are formed by six different distributions. For solar panel generation distributions, two different weather conditions are considered, which are sunny and cloudy days. OpenDSS is a commercial software for simulating the distribution grid. Power flow analyses are conducted by using MCS scenarios and OpenDSS. After solving the power flow, the bus voltage magnitudes, angles, and energy injection from substation are obtained for each MCS scenario. Thirdly, the algorithm analyzes the result of each scenario, if the scenario violates the threshold. The percentage of the number of violated scenarios z_{vio} is calculated by the number of

violated scenarios divided by the total number of scenarios. Fourthly, if the percentage of the number of violated scenarios z_{vio} is greater than the violation possibility ε, the algorithm adjusts the proportion of DERs and retest the result using MCS scenarios. The violation possibility ε here is defined as either the possibility of the scenario violating the threshold or the risk of using energy coming from DERs. The threshold can be voltage deviation, power compensated from transmission grid and if violating line capacity. If z_{vio} is greater than ε, the coefficient τ increases. The coefficient τ is defined as the reduction of the initial active power proportion of total DERs to total load $I_t^{p\%}$. The coefficient τ is initialized to be 0, and the increased τ is put into modified SOCP-ACOPF. The active power proportion of total DERs to total load in the model is decreased by subtracting $I_t^{p\%}$ by τ. Since if the active power proportion of total DERs to total load is decreased, the reactive power proportion of total DERs to total load will also need to decrease. Note that $I_t^{q\%}$ may not be equal to $I_t^{p\%}$. In this chapter, due to the reactive proportion of DERs should be reduced in order to follow the decrease of the active proportion of DERs, the reactive proportion of DERs is decreased by $(\frac{I_t^{p\%}-\tau}{I_t^{p\%}})I_t^{q\%}$. The modified model is solved with less proportion of DERs and get dispatch of the system. The new dispatch result and MCS scenarios are put in OpenDSS for solving power flow analysis again. When the percentage of the number of violated scenarios z_{vio} is smaller than violation possibility ε, then the algorithm stops.

4.2.1 Second Order Cone Programming (SOCP) ACOPF

For the ACOPF model, the global optimal solution is hard to obtain due to the non-linear and non-convexity of AC power flow equations. In this section, a SOCP-ACOPF model is introduced to guarantee the global optimal solution of the ACOPF model.

The power flow equations are shown in (4.1)-(4.2), if $Y_{i,k} = G_{i,k} - jB_{i,k}$

$$P_i = \sum_{k=1}^{N} |V_i||V_k|(G_{i,k}\cos\theta_{i,k} - B_{i,k}\sin\theta_{i,k}) \tag{4.1}$$

$$Q_i = \sum_{k=1}^{N} |V_i||V_k|(G_{i,k}\sin\theta_{i,k} + B_{i,k}\cos\theta_{i,k}) \tag{4.2}$$

If there are only two buses, the power flow equations would be presented in (4.3)-(4.4) from bus i to bus j for time period t.

$$P^{line}_{i,j,t} = G_{i,j}V^2_{i,t} + G_{i,j}V_{i,t}V_{j,t}\cos\theta_{i,j,t} - B_{i,j}V_{i,t}V_{j,t}\sin\theta_{i,j,t} \tag{4.3}$$

$$Q^{line}_{i,j,t} = B_{i,j}V^2_{i,t} + B_{i,j}V_{i,t}V_{j,t}\cos\theta_{i,j,t} + G_{i,j}V_{i,t}V_{j,t}\sin\theta_{i,j,t} \tag{4.4}$$

If we introduce auxiliary values (4.5)-(4.7) [51],

$$U_{i,t} = \frac{V^2_{i,t}}{\sqrt{2}} \tag{4.5}$$

$$R_{i,j,t} = -V_{i,t}V_{j,t}\cos\theta_{i,j,t} \tag{4.6}$$

$$I_{i,j,t} = -V_{i,t}V_{j,t}\sin\theta_{i,j,t} \tag{4.7}$$

and substitute them into power flow equations (4.3)-(4.4). The equations would be

[51],

$$P_{i,j,t}^{line} = \sqrt{2}G_{i,j}U_{i,t} - G_{i,j}R_{i,j,t} + B_{i,j}I_{i,j,t} \tag{4.8}$$

$$Q_{i,j,t}^{line} = \sqrt{2}B_{i,j}U_{i,t} - B_{i,j}R_{i,j,t} - G_{i,j}I_{i,j,t} \tag{4.9}$$

For auxiliary values,

$$2U_{i,t}U_{j,t} \geq R_{i,j,t}^2 + I_{i,j,t}^2 \tag{4.10}$$

$$R_{i,j,t} = R_{j,i,t} \tag{4.11}$$

$$I_{i,j,t} = -I_{j,i,t} \tag{4.12}$$

After making ACOPF convex, the convex SOCP-ACOPF model is formulated. When modeling SOCP-ACOPF, DERs are considered in the optimization model. Due to the utility needs to ask customers how much load reduction they will provide, DR is modeled as the load reduction values that customers are willing to provide. It depends on residential customers' behavior. Three types of solar panels are considered. For PV type 1, the reactive power output can be controlled by its inverter. There is no control over the active power output. The output of the active power of PV type 1 is always the maximum available active power capacity. For PV type 2, it has no capability to provide reactive power support. The active power output can be controlled for PV type 2. For PV type 3, both active and reactive power can be controlled by its inverter. The charging rate, capacity limit, and power balance are considered in battery storage modeling.

The objective function of the SOCP-ACOPF model is minimizing the total sys-

tem cost.

Objective function

$$\min \sum_{t\in T}\sum_{(g,n)\in GN}\rho_t^{DA}P_{g,n,t}^{grid}+\sum_{t\in T}\sum_{(i,n)\in PV1}\rho_t^{D}P_{i,n,t}^{PV1}$$
$$+\sum_{t\in T}\sum_{(i,n)\in PV2}\rho_t^{D}P_{i,n,t}^{PV2}+\sum_{t\in T}\sum_{(i,n)\in PV3}\rho_t^{D}P_{i,n,t}^{PV3}$$
$$+\sum_{t\in T}\sum_{(i,n)\in DR}\rho_t^{D}P_{i,n,t}^{DR} \tag{4.13}$$

Constraints

$$P_{i,j,t}^{line}=\sqrt{2}G_{i,j}U_{n,t}-G_{i,j}R_{i,j,t}+B_{i,j}I_{i,j,t}\quad (i,j)\in L,t\in T,n\in BUS \tag{4.14}$$

$$Q_{i,j,t}^{line}=\sqrt{2}B_{i,j}U_{n,t}-B_{i,j}R_{i,j,t}-G_{i,j}I_{i,j,t}\quad (i,j)\in L,t\in T,n\in BUS \tag{4.15}$$

$$2U_{i,t}U_{j,t}\geq R_{i,j,t}^2+I_{i,j,t}^2\quad (i,j)\in L,t\in T \tag{4.16}$$

$$R_{i,j,t}=R_{i,j,t}\quad (i,j)\in L,t\in T \tag{4.17}$$

$$I_{i,j,t}=-I_{i,j,t}\quad (i,j)\in L,t\in T \tag{4.18}$$

$$-C_{i,j}^{line}\leq P_{i,j,t}^{line}\leq C_{i,j}^{line}\quad (i,j)\in L,t\in T \tag{4.19}$$

$$G_{g,n}^{L}\leq P_{g,n,t}^{grid}\leq G_{g,n}^{H}\quad (g,n)\in GN,t\in T \tag{4.20}$$

$$F_{f,n}^{L}\leq Q_{f,n,t}^{facts}\leq F_{f,n}^{H}\quad (f,n)\in F,t\in T \tag{4.21}$$

$$P_{i,n,t}^{DR}\leq DR_m D_{n,t}^{P}\quad (i,n)\in DR,t\in T \tag{4.22}$$

$$P_{i,n,t}^{PV1}=C_{i,n,t}^{pv1_P}\quad (i,n)\in PV1,t\in T \tag{4.23}$$

$$Q_{i,n,t}^{PV1}\leq\sqrt{(C_{i,n,t}^{pv1_S})^2-(P_{i,n,t}^{PV1})^2}\quad (i,n)\in PV1,t\in T \tag{4.24}$$

$$P_{i,n,t}^{PV2} \leq C_{i,n,t}^{pv2_P} \quad (i,n) \in PV2, t \in T \tag{4.25}$$

$$P_{i,n,t}^{PV3} \leq C_{i,n,t}^{pv3_P} \quad (i,n) \in PV3, t \in T \tag{4.26}$$

$$(P_{i,n,t}^{PV3})^2 + (Q_{i,n,t}^{PV3})^2 \leq (C_{i,n,t}^{pv3_S})^2 \quad (i,n) \in PV3, t \in T \tag{4.27}$$

$$SE_{i,n,t} + P_{i,n,t}^{S} = \alpha C_{i,n}^{S,H} \quad (i,n) \in SG, t = 1 \tag{4.28}$$

$$SE_{i,n,t} + P_{i,n,t}^{S} = SE_{i,n,t-1} \quad (i,n) \in SG, t \in T, t \neq 1 \tag{4.29}$$

$$R_{i,n}^{L} \leq P_{i,n,t}^{S} \leq R_{i,n}^{H} (i,n) \in SG, t \in T \tag{4.30}$$

$$C_{i,n}^{S,L} \leq SE_{i,n,t} \leq C_{i,n}^{S,H} \quad (i,n) \in SG, t \in T, t \neq 1 \tag{4.31}$$

$$\beta C_{i,n}^{S,H} \leq SE_{i,n,t} \leq C_{i,n}^{S,H} \quad (i,n) \in SG, t = 24 \tag{4.32}$$

$$\begin{aligned} &\sum_{(g,n)\in GN} P_{g,n,t}^{grid} + \sum_{(i,n)\in DR} P_{i,n,t}^{DR} + \sum_{(i,n)\in PV1} P_{i,n,t}^{PV1} + \sum_{(i,n)\in PV2} P_{i,n,t}^{PV2} \\ &+ \sum_{(i,n)\in PV3} P_{i,n,t}^{PV3} + \sum_{(i,n)\in SG} P_{i,n,t}^{P} = D_{n,t}^{P} + \sum_{j\in\phi_n} P_{n,j,t}^{line} \quad t \in T \end{aligned} \tag{4.33}$$

$$\begin{aligned} &\sum_{(g,n)\in GN} Q_{g,n,t}^{grid} + \sum_{(i,n)\in DR} Q_{i,n,t}^{DR} + \sum_{(i,n)\in PV1} Q_{i,n,t}^{PV1} + \sum_{(i,n)\in PV3} Q_{i,n,t}^{PV3} \\ &+ \sum_{(f,n)\in F} Q_{i,n,t}^{facts} = D_{n,t}^{Q} + \sum_{j\in\phi_n} Q_{n,j,t}^{line} \quad t \in T \end{aligned} \tag{4.34}$$

$$\frac{0.9^2}{\sqrt{2}} \leq U_{n,t} \leq \frac{1.1^2}{\sqrt{2}} \quad n \in BUS, t \in T, n \neq 1 \tag{4.35}$$

$$U_{n,t} = 1 \quad n = 1, t \in T \tag{4.36}$$

The constraints (4.14) and (4.15) are line flow from bus i to bus j for time period t, which are derived from SOCP convex strategy. (4.16)-(4.18) are cone constraints.

Line capacity limit constraint is given by (4.19) to guarantee that the line flow will not exceed the line limit. The energy injection of substation and FACTS devices are limited by (4.20) and (4.21) due to the capacity of substation and FACTS are not infinite. For DR modeling, if bus n can provide DR service, the maximum DR reduction is thirty percent of the load in bus n. The DR load reduction values follow Gaussian distributions. Since the utility needs to ask the customers participating in DR programs, how much load reduction they are willing to provide for a certain time period. In the SOCP-ACOPF, the dispatched DR reduction should be less or equal to $DR_m D^P_{n,t}$. The coefficient DR_m is the maximum percentage of the load reduction, which is set to 0.3. The $D^P_{n,t}$ is the total active power demand in bus n. The constraints of PV type 1 are shown in (4.23) and (4.24). Reactive power is controlled by inverter (4.23) and active power is equal to the maximum capacity (4.24). The active control of PV type 2 constraint is presented in (4.25). The active and reactive controls of PV type 3 are introduced in (4.26) and (4.27). The battery storage power balance constraints are (4.28) and (4.29). For the first period, we assume that there is stored energy in batteries, which is represented by the coefficient α in (4.28). For the rest of the time period t, the stored and injecting energy should be equal to the stored energy for the last time period $t-1$, which is shown in (4.29). The minimum and maximum charging rates are shown in (4.30), which aims to restrict the battery storage energy injecting to the grid. Due to stored energy can not exceed the capacity of the battery storage, the battery storage constraints are listed

in (4.31) and (4.32). We assume that the stored energy cannot be zero for the last time period, which is represented by the coefficient β in (4.32). The node active and reactive power balancing constraints of the system are exhibited in (4.33) and (4.34). In order to keep voltage magnitude within reasonable range, voltage limit constraints are shown in (4.35) and (4.36). Due to the substation is connected to bus 1, the voltage magnitude of bus 1 is equal to 1 p.u. for any period of time, which is represented in (4.36).

4.2.2 Modified Second Order Cone Programming (SOCP) ACOPF

If the percentage of the number of violated scenarios z_{vio} is greater than the violation possibility ε, the algorithm increases the τ, and then runs a modified SOCP-ACOPF model for obtaining the dispatch of the distribution grid with less proportion of DERs. In order to reduce the proportion of DERs in the SOCP-ACOPF model, several adjustments are needed to implement.

The first adjustment is (4.37),

$$P_{i,n,t}^{PV1} = \frac{I_t^{p\%} - \tau}{I_t^{p\%}} C_{i,n,t}^{pv1_P} \quad (i,n) \in PV1, t \in T \tag{4.37}$$

From constraint (4.23), the active power output of PV type 1 is equal to the maximum available active power capacity. The available active power capacity of PV type 1 should be shrunk, as the proportion of DERs is reduced by τ. If there is no $\frac{I_t^{p\%}-\tau}{I_t^{p\%}}$ term, $P_{i,n,t}^{PV1}$ will always be equal to $C_{i,n,t}^{pv1_P}$ which is a fixed value.

The second and third adjustments are adding another two constraints (4.38)

and (4.39),

$$\sum_{(i,n)\in DR} P_{i,n,t}^{DR} + \sum_{(i,n)\in PV1} P_{i,n,t}^{PV1} + \sum_{(i,n)\in PV2} P_{i,n,t}^{PV2} + \sum_{(i,n)\in PV3} P_{i,n,t}^{PV3} \leq (I_t^{p\%} - \tau) D_{n,t}^{P} \quad t \in T \tag{4.38}$$

$$\sum_{(i,n)\in DR} Q_{i,n,t}^{DR} + \sum_{(i,n)\in PV1} Q_{i,n,t}^{PV1} + \sum_{(i,n)\in PV3} Q_{i,n,t}^{PV3} \leq (\frac{I_t^{p\%} - \tau}{I_t^{p\%}}) I_t^{q\%} D_{n,t}^{Q} \quad t \in T \tag{4.39}$$

the constraint (4.38) is added in order to limit the total active power output of DERs. The constraint (4.39) is appended to modified SOCP-ACOPF for restricting the reactive power output of DERs, as the proportion of DERs to load is decreased. Since if the active power proportion of DERs is decreased, the reactive power proportion of DERs will also need to decrease. Note that $I_t^{q\%}$ is not equal to $I_t^{p\%}$. The reactive proportion of DERs should be reduced to follow the decrease of the active proportion of DERs. In this chapter, the reactive proportion of DERs is decreased by $(\frac{I_t^{p\%} - \tau}{I_t^{p\%}}) I_t^{q\%}$.

The modified SOCP-ACOPF model includes the same objective function (4.13). The constraints are (4.14)-(4.22), (4.37), (4.24)-(4.36) and (4.38)-(4.39). The modified SOCP-ACOPF model is exhibited in the following.

Objective function

$$\min \sum_{t\in T}\sum_{(g,n)\in GN}\rho_t^{DA}P_{g,n,t}^{grid}+\sum_{t\in T}\sum_{(i,n)\in PV1}\rho_t^{D}P_{i,n,t}^{PV1}$$
$$+\sum_{t\in T}\sum_{(i,n)\in PV2}\rho_t^{D}P_{i,n,t}^{PV2}+\sum_{t\in T}\sum_{(i,n)\in PV3}\rho_t^{D}P_{i,n,t}^{PV3}$$
$$+\sum_{t\in T}\sum_{(i,n)\in DR}\rho_t^{D}P_{i,n,t}^{DR}$$

Constraints

$$P_{i,j,t}^{line}=\sqrt{2}G_{i,j}U_{n,t}-G_{i,j}R_{i,j,t}+B_{i,j}I_{i,j,t}\quad (i,j)\in L,t\in T,n\in BUS$$

$$Q_{i,j,t}^{line}=\sqrt{2}B_{i,j}U_{n,t}-B_{i,j}R_{i,j,t}-G_{i,j}I_{i,j,t}\quad (i,j)\in L,t\in T,n\in BUS$$

$$2U_{i,t}U_{j,t}\geq R_{i,j,t}^2+I_{i,j,t}^2\quad (i,j)\in L,t\in T$$

$$R_{i,j,t}=R_{j,i,t}\quad (i,j)\in L,t\in T$$

$$I_{i,j,t}=-I_{j,i,t}\quad (i,j)\in L,t\in T$$

$$-C_{i,j}^{line}\leq P_{i,j,t}^{line}\leq C_{i,j}^{line}\quad (i,j)\in L,t\in T$$

$$G_{g,n}^{L}\leq P_{g,n,t}^{grid}\leq G_{g,n}^{H}\quad (g,n)\in GN,t\in T$$

$$F_{f,n}^{L}\leq Q_{f,n,t}^{facts}\leq F_{f,n}^{H}\quad (f,n)\in F,t\in T$$

$$P_{i,n,t}^{DR}\leq DR_m D_{n,t}^{P}\quad (i,n)\in DR,t\in T$$

$$Q_{i,n,t}^{PV1}\leq\sqrt{(C_{i,n,t}^{pv1_S})^2-(P_{i,n,t}^{PV1})^2}\quad (i,n)\in PV1,t\in T$$

$$P_{i,n,t}^{PV2}\leq C_{i,n,t}^{pv2_P}\quad (i,n)\in PV2,t\in T$$

$$P_{i,n,t}^{PV3} \leq C_{i,n,t}^{pv3_P} \quad (i,n) \in PV3, t \in T$$

$$(P_{i,n,t}^{PV3})^2 + (Q_{i,n,t}^{PV3})^2 \leq (C_{i,n,t}^{pv3_S})^2 \quad (i,n) \in PV3, t \in T$$

$$SE_{i,n,t} + P_{i,n,t}^{S} = \alpha C_{i,n}^{S,H} \quad (i,n) \in SG, t = 1$$

$$SE_{i,n,t} + P_{i,n,t}^{S} = SE_{i,n,t-1} \quad (i,n) \in SG, t \in T, t \neq 1$$

$$R_{i,n}^{L} \leq P_{i,n,t}^{S} \leq R_{i,n}^{H} (i,n) \in SG, t \in T$$

$$C_{i,n}^{S,L} \leq SE_{i,n,t} \leq C_{i,n}^{S,H} \quad (i,n) \in SG, t \in T, t \neq 1$$

$$\beta C_{i,n}^{S,H} \leq SE_{i,n,t} \leq C_{i,n}^{S,H} \quad (i,n) \in SG, t = 24$$

$$\sum_{(g,n)\in GN} P_{g,n,t}^{grid} + \sum_{(i,n)\in DR} P_{i,n,t}^{DR} + \sum_{(i,n)\in PV1} P_{i,n,t}^{PV1} + \sum_{(i,n)\in PV2} P_{i,n,t}^{PV2}$$

$$+ \sum_{(i,n)\in PV3} P_{i,n,t}^{PV3} + \sum_{(i,n)\in SG} P_{i,n,t}^{P} = D_{n,t}^{P} + \sum_{j\in\phi_n} P_{n,j,t}^{line} \quad t \in T$$

$$\sum_{(g,n)\in GN} Q_{g,n,t}^{grid} + \sum_{(i,n)\in DR} Q_{i,n,t}^{DR} + \sum_{(i,n)\in PV1} Q_{i,n,t}^{PV1} + \sum_{(i,n)\in PV3} Q_{i,n,t}^{PV3}$$

$$+ \sum_{(f,n)\in F} Q_{i,n,t}^{facts} = D_{n,t}^{Q} + \sum_{j\in\phi_n} Q_{n,j,t}^{line} \quad t \in T$$

$$\frac{0.9^2}{\sqrt{2}} \leq U_{n,t} \leq \frac{1.1^2}{\sqrt{2}} \quad n \in BUS, t \in T, n \neq 1$$

$$U_{n,t} = 1 \quad n = 1, t \in T$$

$$P_{i,n,t}^{PV1} = \frac{I_t^{p}\% - \tau}{I_t^{p}\%} C_{i,n,t}^{pv1_P} \quad (i,n) \in PV1, t \in T$$

$$\sum_{(i,n)\in DR} P_{i,n,t}^{DR} + \sum_{(i,n)\in PV1} P_{i,n,t}^{PV1} + \sum_{(i,n)\in PV2} P_{i,n,t}^{PV2} + \sum_{(i,n)\in PV3} P_{i,n,t}^{PV3}$$

$$\leq (I_t^{p}\% - \tau) D_{n,t}^{P} \quad t \in T$$

$$\sum_{(i,n)\in DR} Q^{DR}_{i,n,t} + \sum_{(i,n)\in PV1} Q^{PV1}_{i,n,t} + \sum_{(i,n)\in PV3} Q^{PV3}_{i,n,t} \leq (\frac{I^{p\%}_t - \tau}{I^{p\%}_t}) I^{q\%}_t D^Q_{n,t} \quad t \in T$$

This modified SOCP-ACOPF model is used in the flowchart Fig. 4.2 to obtain the new dispatch of the distribution grid with less proportion of DERs, if the percentage of the number of violated scenarios z_{vio} is greater than the violation possibility ε.

4.2.3 Scenarios Generation Using Monte Carlo Simulation (MCS)

MCS method is utilized in this study in order to analyze the reliability of the distribution grid under the uncertainty and variability of DERs. In the distribution system, there are six buses that can provide DR service. There are five PV type 1, five PV type 2 and five PV type 3. Because the inverter of PV type 1 can control the reactive power and PV type 3 can control both active and reactive power, it is not reasonable to sample active nor reactive power generation for PV type 1 and 3. In this case, apparent power is sampled for the generation of PV types 1 and 3. In each MCS scenario, the values include six active power output of DR, five apparent power output of PV type 1, five active power output of PV type 2 and five apparent power output of PV type 3. These values are drawn from different Gaussian distributions. Since the characteristics of residential customers' behavior towards DR buses are different from each other, six different Gaussian distributions are used to sample DR reduction values. Due to PV generation is highly related to weather conditions, two different Gaussian distributions are utilized to simulate the power

generation of PV type 1, 2 and 3 under sunny and cloudy weather. The Gaussian distribution for sunny weather has lower forecasting deviation and smaller variance. The Gaussian distribution for cloudy weather has higher forecasting deviation and larger variance. In future work, all of the Gaussian distributions will be obtained by using machine learning.

4.2.4 Case Study

IEEE 33 bus distribution system is used for the case study. The topology of the IEEE 33 bus is shown in Fig. 4.3.

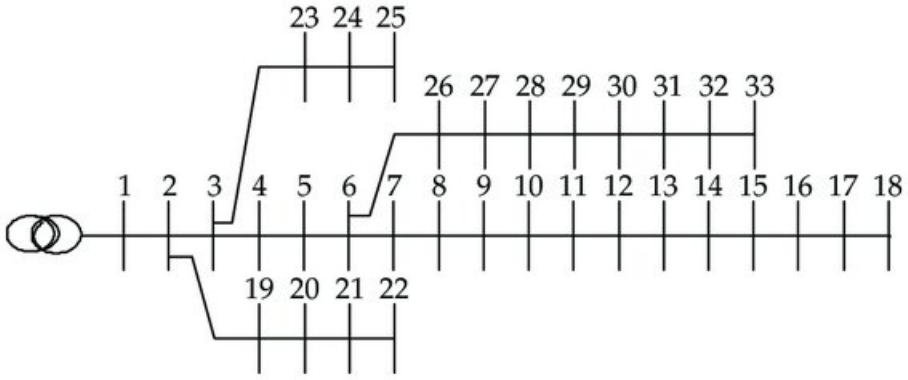

Figure 4.3: Topology of IEEE 33 Bus Distribution System

There are one substation and four feeders. All of four feeders are radial.

The DERs information is presented in Table 4.1. The table shows buses that are connected with DR, PV types 1, 2 and 3, and battery.

Table 4.1: DERs Information

	DR	PV1	PV2	PV3	Battery
Bus	7	2	9	2	2
Bus	8	6	15	18	6
Bus	24	12	21	22	12
Bus	25	18	25	25	29
Bus	30	29	33	33	
Bus	32				

The 13th time period from 12pm - 13pm is considered in this case study. For the 13th time period, the wholesale electricity price is higher than DERs price, which means that if there is no uncertainty in DERs generation, DERs will provide as much energy as possible in this time period. Moreover, the sun radiance is strong during the 13th time period, and the available capacity of PV types 1, 2 and 3 tends to be large.

4.2.5 Voltage Deviation

In this case study, the voltage deviation from the dispatched voltage value is considered. The maximum percentage of the voltage deviation from the dispatched voltage is defined in 4.40.

If there are N scenarios, the actual voltage $V_{i,j}^{a}$ is smaller than the dispatch volt-

age V_{Dis},

$$V_D^L = \max\{\frac{V_{Dis} - V_{i,j}^a}{V_{Dis}} \times 100\%, \quad i \in BUS, j \in \{1, 2, ..., N\}\} \tag{4.40}$$

The case study is simulated using Python. The SOCP-ACOPF and modified SOCP-ACOPF optimization are solved by Cplex, and power flow analysis is solved by OpenDSS. The result of the case study is presented in Fig. 4.4.

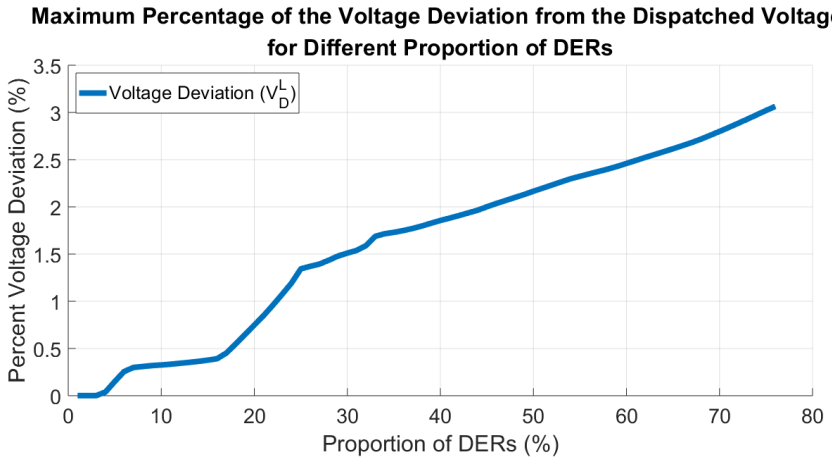

Figure 4.4: Voltage Deviation from the Expected Value for Different Proportion of DERs

In Fig. 4.4, The voltage deviation V_D^L is increasing as the proportion of DERs is increased. This shows that if we reduce the proportion of DERs, the lower bound of voltage (lowest voltage) will be close to the dispatched value. Comparing with the small proportion of DERs case, if we use more DERs, the lowest voltage value may far away from the desired value. Since the dispatched voltage values are the solutions coming from the SOCP-ACOPF model, greatly deviating from the dispatched values is not what operators expect. Due to the characteristic of one

substation distribution grid (voltage is going down from substation to the end), slightly increasing voltage will not harm the distribution grid, but improve the power quality of the system. Thus, the voltage deviation that actual voltage is smaller than desired voltage is considered in this study.

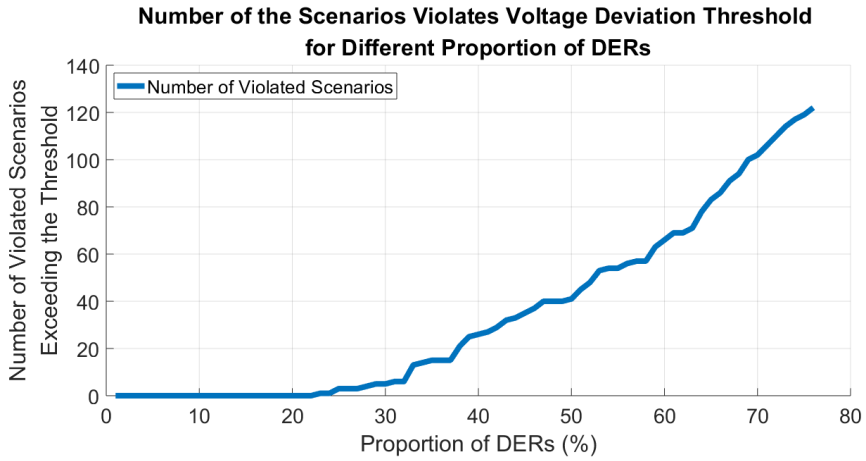

Figure 4.5: Number of Violated Scenarios: Voltage Deviation from the Expected Value for Different Proportion of DERs

Since voltage values are not desired to greatly deviate from the dispatched value, the voltage deviation from the dispatched voltage can be considered as one chance-constraint criterion. We can set the voltage deviation smaller than a specified threshold and run the algorithm iteration shown in Fig. 4.2. In each iteration, the algorithm counts how many scenarios violate the voltage deviation threshold. A threshold of voltage deviation is set to be 1%. The number of violated scenarios is recorded for each different proportion of DERs if the voltage deviation of the scenario is greater than 1%. The result is shown in Fig. 4.5.

As can be seen in Fig. 4.5, the number of violated scenarios is decreasing, as

we reduce the proportion of DERs. This means that the voltage will not be much lower than the dispatched voltage value if we reduce the proportion of DERs.

We can also set one probability violation ε to stop the algorithm and output the solution from modified SOCP-ACOPF, when the probability of violation is smaller than the probability violation ε. The obtained solution can guarantee the possibility of voltage deviation is smaller than the determined value. The probability violation ε can be considered as the risk that the operator prefers. If the probability violation ε is set to be 5%, the relationship between voltage deviation thresholds and dispatched DERs is shown in Fig. 4.6.

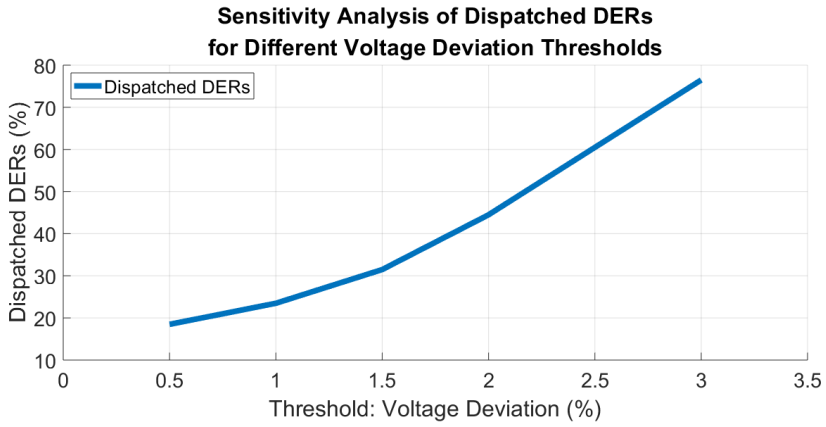

Figure 4.6: Sensitivity Analysis of Dispatched DERs for Different Voltage Deviation Threshold

The dispatched DERs is defined as the solution of modified SOCP-ACOPF satisfying the chance-constraint (4.41).

$$z_{vio} \leq \varepsilon \tag{4.41}$$

From Fig. 4.6, the larger the voltage deviation threshold is, the more DERs energy will be dispatched for a determined risk level value ε. As can be seen, the small change of voltage deviation threshold has a large impact on the proportion of dispatched DERs.

4.2.6 Energy Compensation from Transmission grid

In the case study, the energy that is compensated from the transmission grid is considered. The compensated energy is equal to the mismatch between the dispatched and actual energy from the substation. The compensated can also be defined as the reserves that the distribution grid needs in the real-time operation from the transmission. The dispatched active and reactive power from substation are denoted as P_{Dis}^{grid} and Q_{Dis}^{grid}. The actual active and reactive power from substation are denoted as P_{i}^{grid} and Q_{i}^{grid}.

If there are N scenarios, and the active and reactive power that are compensated from transmission grid P_{C}^{grid} and Q_{C}^{grid} can be

$$P_{C}^{grid} = \max\{P_{i}^{grid} - P_{Dis}^{grid}, \quad i \in \{1, 2, ..., N\}\} \tag{4.42}$$

$$Q_{C}^{grid} = \max\{Q_{i}^{grid} - Q_{Dis}^{grid}, \quad i \in \{1, 2, ..., N\} \tag{4.43}$$

The percentage of active and reactive power that are compensated from trans-

mission grid $P^{grid}_{C,\%}$ and $Q^{grid}_{C,\%}$ can be

$$P^{grid}_{C,\%} = \max\{\frac{P^{grid}_{i} - P^{grid}_{Dis}}{P^{grid}_{Dis}} \times 100\%, \quad i \in \{1, 2, ..., N\}\} \tag{4.44}$$

$$Q^{grid}_{C,\%} = \max\{\frac{Q^{grid}_{i} - Q^{grid}_{Dis}}{Q^{grid}_{Dis}} \times 100\%, \quad i \in \{1, 2, ..., N\}\} \tag{4.45}$$

The case study results are shown in the Fig. 4.7 and 4.8.

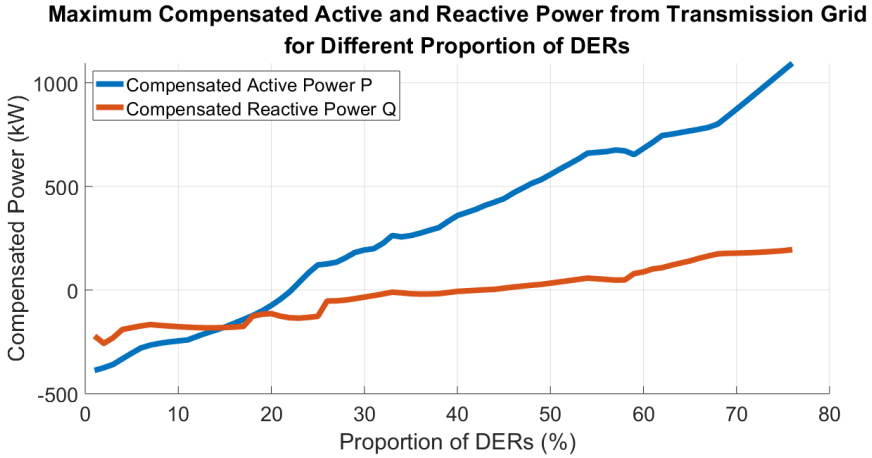

Figure 4.7: Active and Reactive Power Compensated from Transmission Grid for Different Proportion of DERs

The increasing proportion of DERs is corresponding to the increasing uncertainty and variability in the distribution grid. From Fig. 4.7 and 4.8, the active and reactive power compensated from the transmission grid are increasing, as the proportion of DERs is increased. This shows that due to the characteristic of DERs, namely uncertainty and variability, the compensated energy from the transmission grid is increasing if the uncertainty and variability in the distribution grid are increased. The compensated energy from the transmission grid can also be considered as the backup from the transmission grid. The higher the uncertainty in the

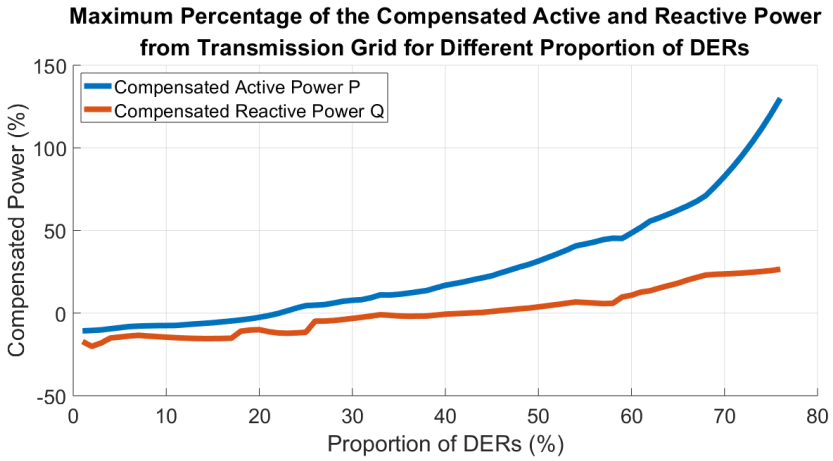

Figure 4.8: Percentage of Active and Reactive Power Compensated from Transmission Grid for Different Proportion of DERs

distribution grid, the more backup from the transmission grid the substation will tend to require.

Since the reserves from the transmission grid are sometimes limited, compensating lots of energy from the transmission grid may bring troubles to system operation and undermine the reliability of the system. Moreover, the real-time electricity price is much higher than the day ahead wholesale price, purchasing too much energy from the real-time market will seriously affect customers and the utility's benefit. Requiring a large amount of backup from the transmission grid should be avoided when dispatching the system. The compensated energy from the transmission grid constraint can be considered as one chance-constraint criterion. This case study applies a similar idea as Fig. 4.5. A threshold for compensated energy is set to be 20%. If the percentage of the compensated energy from the transmission grid in the scenario is greater than 20%, the scenario is defined as the violated

scenario. The algorithm records the number of scenarios violating the threshold for different proportions of DERs. The result is exhibit in Fig. 4.9.

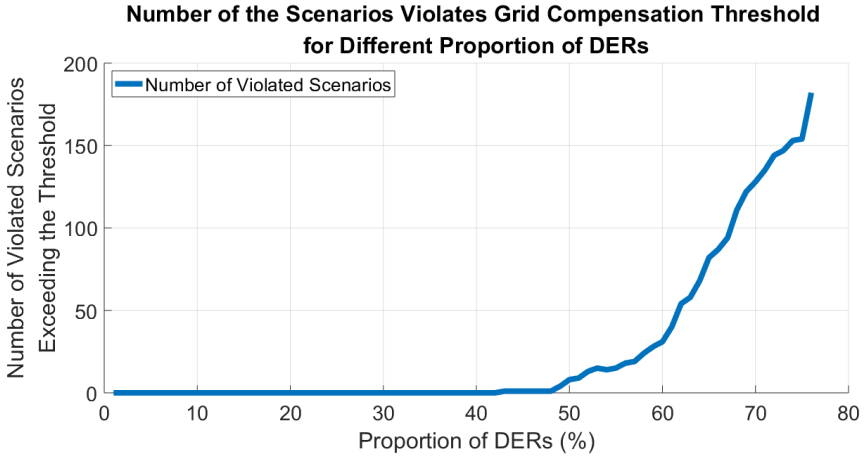

Figure 4.9: Number of Violated Scenarios: Energy Compensated from the Transmission Grid for Different Proportion of DERs

Fig. 4.9 shows that the number of scenarios that violate the compensated energy threshold is increasing, as the proportion of DERs is increased. It can be seen that the number of violated scenarios has increased dramatically after about 60% of proportion of DERs. Besides, the number of violated scenarios becomes zero, if the proportion of DERs is less than around 48%. The result indicates that dispatching more DERs in the distribution grid enlarges the uncertainty of DERs, which results in requiring more energy compensated from the transmission grid.

A probability violation ε can be set to stop the algorithm when the number of violated scenarios is lower than the determined value. The probability violation ε can be considered as the risk that the system will have when dispatching DERs. The algorithm will output the solution that can guarantee that the possibility of vi-

olating the compensated energy threshold is less than the determined value. If the probability violation ε is set to be 5%, the relationship between voltage deviation thresholds and dispatched DERs is shown in Fig. 4.10.

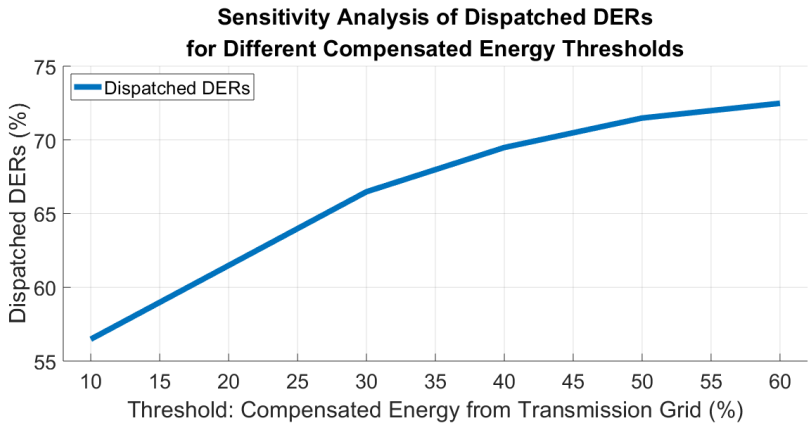

Figure 4.10: Sensitivity Analysis of Dispatched DERs for Different Compensated Energy Threshold

From Fig. 4.10, the larger the compensated energy threshold is, the more DERs energy will be dispatched. The results can also be translated as the more DERs will be dispatched, if there are enough reverses provided from the transmission grid for a determined risk level.

4.2.7 Sensitivity Analysis of Probability Violation

The last study is the sensitivity analysis of probability violation. If the voltage deviation threshold and the compensated energy threshold are set to be 1% and 20%, respectively, the result is shown in Fig. 4.11.

As can be seen, the dispatched DERs energy is increasing as the violation prob-

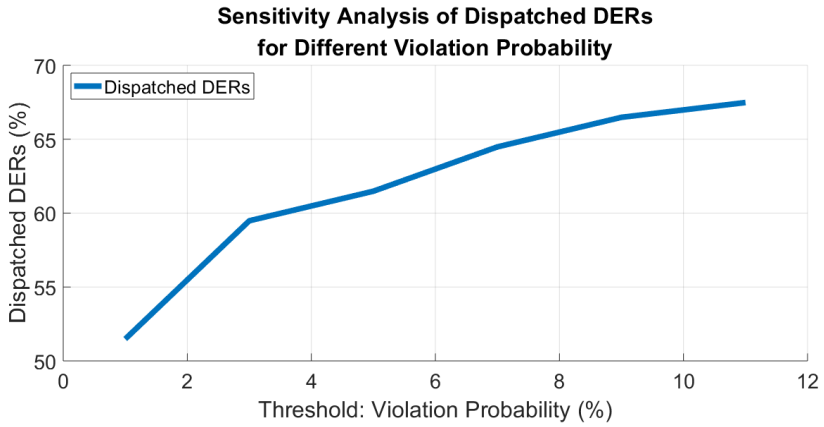

Figure 4.11: Sensitivity Analysis of Dispatched DERs for Different Violation Probability Threshold

ability increases. The small violation probability value refers to a conservative operation strategy. The large violation probability value means an optimistic operation strategy. Fig. 4.11 shows that the more conservative the operator is, the less DERs will be dispatched. When operating the distribution grid with DERs, conservative sometimes is related to higher system costs. The trade-off between system cost and risk can be further studied in the future.

4.2.8 Summary

Chapter 4 illustrates one chance-constraint based SOCP-ACOPF. The SOCP-ACOPF is a convex model, which means that the global optimal solution is guaranteed. Monte Carlo Simulation is used to sample the scenarios. The case study results show the impact of the determined threshold on the solution of dispatched DERs.

Chapter 5

FUTURE WORK

Machine learning and chance-constraint based SOCP-ACOPF model will be the main research direction for future work.

5.1 Machine Learning

The machine learning integrating with customers' demographic information can improve the representation of the uncertainty of DERs, which is already shown in chapter 3 and chapter 4. One powerful exemplary result figure is Fig. 4.1. Since demographic information can reduce the uncertainty of customers' price responsiveness distribution, the socially-aware aggregation of DERs for operational scheduling should be studied by implementing different machine learning algorithms.

Combining machine learning in the ACOPF model will be one of the important research direction in my future work. The accuracy of the trained model has a direct impact on the uncertainty of DERs in the chance-constraint based SOCP-ACOPF model. The higher the accuracy of the trained model, the less the uncertainty of DERs will have. The operation of the distribution grid can be greatly improved if the uncertainty of DERs is reduced.

Forecasting the generation of DERs is a regression problem. There are many

well-established machine learning algorithms for regression such as ridge regression, LASSO, support vector regression, k-nearest neighbors, artificial neural network and recurrent neural network.

Since different types of DERs have different characteristics, it is very important to find the most appropriate machine learning algorithm for each type of DERs data. Besides, after training the data using machine learning, providing the interpretation of why this machine learning algorithm performs best is another key research direction in my future work.

5.2 Chance-constraint

For the chance-constraint based model, there are three primary categories of existing methods.

- Scenarios approach
- Sample average approximation
- Robust optimization based methods

For the scenarios approach method, multiple scenarios are used to estimate the chance-constraint based model. Scenarios approach method searches the feasible region of all scenarios in order to find the optimal solution. Scenarios approach method can be implemented without any distribution assumption. It is a data-driven based approach. However, the number of scenarios and the complexity of scenarios are highly related to the solution. More scenarios and larger complexity

of the scenarios result in a more conservative solution. If there are N scenarios $\{\xi_i\}_{i=1}^{N}$, the scenarios approach method can be formulated in the following [52].

$$\min_{x} c^T x \tag{5.1}$$

$$s.t.\ f(x, \xi_i) \leq 0, ..., f(x, \xi_N) \leq 0 \tag{5.2}$$

In the distribution optimization model, due to the uncertainty and variability of DERs, scenarios approach method can be utilized to implement chance-constraint on ACOPF. However, the number of scenarios and the complexity of scenarios require careful study.

The sample average approximation method is similar to what is applied in chapter 4. It utilizes empirical distribution to approximate the true distribution of the targeted problem, which can be formulated in the following [52].

$$\min_{x} c^T x \tag{5.3}$$

$$s.t. \frac{1}{N} \sum_{i=1}^{N} [\mathbf{1}\{f(x, \xi) > 0\}] \leq \varepsilon \tag{5.4}$$

In sample average approximation, empirical distribution is used to estimate the possibility of violation. It can be formulated as a mixed integer programming us-

ing big-M.

$$\min_{x,z} c^T x \tag{5.5}$$

$$s.t. f(x, \xi_1) - M z_1 < 0 \tag{5.6}$$

...

$$f(x, \xi_N) - M z_N < 0 \tag{5.7}$$

$$\frac{1}{N} \sum_{i=1}^{N} z_i \leq \varepsilon \tag{5.8}$$

$$z_i \in \{0, 1\} \tag{5.9}$$

Sample average approximation methods can be utilized to implement chance-constraint for ACOPF in distribution gird. It is a data-driven method. Scenarios are needed to generate. However, Big-M is generally weak formulation. This shortcoming requires more research study.

Robust optimization related methods are searching for an optimal solution that is feasible under a predefined uncertainty set. If the predefined uncertainty set is defined to be $\mathcal{U}$, uncertainties are represented by $\xi \in \mathcal{U}$, the chance-constraint programming and its robust counterpart can be formulated in the following [52].

$$\min_{x} c^T x \tag{5.10}$$

$$s.t. \ \mathbf{P}_{\xi}(x_0^i + \xi^T x^i \leq 0, i = 1, 2, ...) \geq 1 - \epsilon \tag{5.11}$$

$$\min_{x} c^T x \tag{5.12}$$

$$s.t. \ x_0^i + \xi^T x^i \leq 0 \quad \forall \xi \in \mathcal{U}, i = 1, 2, .. \tag{5.13}$$

Defining the uncertainty set is the most important part of using robust optimization. A properly defined uncertainty set makes problems tractable, and can also make the solution neither too conservative nor too risky. However, the robust optimization always needs huge computation.

In the distribution grid, the uncertainty of DERs should be properly defined in the uncertainty set, and the robust optimization should be appropriately formulated in the SOCP-ACOPF model.

In a distribution system, due to the characteristic of DERs, some of the constraints can be formulated as chance-constraints. For example, the voltage magnitude constraint can be converted into the possibility of voltage magnitude V_i violating the system criteria should be less than determined violation possibility ε.

$$\mathbf{P}(V^L \leq V_i \leq V^H, i \in BUS) \leq 1 - \varepsilon$$

V^L and V^H are the bus voltage limits.

The line capacity limit constraint can be converted into the chance-constraint in the following.

$$\mathbf{P}(-C_{i,j}^{line} \leq P_{i,j}^{line} \leq C_{i,j}^{line}, (i,j) \in L) \leq 1 - \varepsilon$$

The transmission grid repurchasing constraint, which restricts the energy that is compensated from transmission grid, can be formulated into chance-constraint.

$$\mathbf{P}(P_{act} - P_{dis} \leq \alpha P_{dis}) \leq 1 - \varepsilon$$

$$\mathbf{P}(Q_{act} - Q_{dis} \leq \beta Q_{dis}) \leq 1 - \varepsilon$$

The dispatched active power and reactive power purchasing from the transmission are denoted as P_{dis} and Q_{dis}. The actual active and reactive power consumption in the distribution grid are denoted as P_{act} and Q_{act}. Energy repurchasing threshold is denoted as αP_{dis} and βQ_{dis}. The difference between the actual and dispatched energy should be within determined threshold.

Since compensated energy from the transmission grid can be considered as the reserves from the transmission grid, the chance-constraint can be also formulated in the following.

$$\mathbf{P}(P_{grid} \leq P_{grid}^{sched} + R_{grid}^{sched}) \leq 1 - \varepsilon$$

P_{grid}^{sched} and R_{grid}^{sched} are the scheduled active power and scheduled available reserves from the transmission grid. P_{grid} is the actual active power injected from the transmission grid. The actual injected active power from the transmission grid should be smaller than the sum of the scheduled injected power and scheduled reserves.

Other chance-constraints can also be considered in future work.

REFERENCES

[1] G. Constante-Flores and M. Illindala, "Data-driven probabilistic power flow analysis for a distribution system with renewable energy sources using monte carlo simulation," *IEEE Transactions on Industrial Application*, vol. 55, no. 1, pp. 174–181, 2019.

[2] D. O. T. Wang and H. Kamath, "Dynamic control and optimization of distributed energy resources in a microgrid," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 2884–2894, 2015.

[3] M. Macedo, J. Galo, L. Almeida, and A. Lima, "Demand side management using artificial neural networks in a smart grid environment," *Renewable and Sustainable Energy Reviews*, vol. 41, pp. 128–133, 2015.

[4] Z. Yi, C. Weiwei, X. Rui, and B. Jason, "A Cluster-Based Method for Calculating Baselines for Residential Loads," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2368–2377, 2016.

[5] M. Muratori and G. Rizzoni, "Residential demand response: Dynamic energy management and time-varying electricity pricing," *IEEE Transactions on Power Systems*, vol. 31, no. 2, pp. 1108–1117, 2016.

[6] SRP, *"SRPs Standard Electric Price Plans Effective with the May 2019 Billing Cycle"*, 2019, Available:http://https://www.srpnet.com/prices/pdfx/rateboo k.pdf.

[7] J. S. Lopes and P. Agnew, *"Fpl residential thermostat load control pilot project evaluation"*, 2010, Available:http://aceee.org/files/proceedings/2010/data/pap ers/1953.pdf.

[8] C. Miller, A. Iles, and C. Jones, "The social dimensions of energy transitions," *Science as Culture*, vol. 22, no. 2, pp. 135–148, 2013.

[9] S. T. Moe, R. Mariann, and B. Thomas, "A travelers guide to smart grids and the social sciences," *Energy Research Social Science*, vol. 9, pp. 1–8, 2015.

[10] K. Zimmerer, "New geographies of energy: introduction to the special issue," *Annals of the Association of American Geographers*, vol. 101, no. 4, pp. 705–711, 2011.

[11] P. R. P. S. C. Cecati, J. Kolbusz and B. Wilamowski, "A novel rbf training algorithm for short-term electric load forecasting and comparative studies," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 10, pp. 6519–6529, 2015.

[12] M. A. S. D. Percy and A. Berry, "Residential demand forecasting with solar-battery systems: A survey-less approach," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 4, pp. 1499–1507, Oct. 2018.

[13] S. K. N. Y. Soltani and G. B. Giannakis, "Real-time load elasticity tracking and pricing for electric vehicle charging," *IEEE Transactions on Smart Grid*, vol. 6, no. 3, pp. 1303–1313, May 2015.

[14] P. C. L. Hatton and E. Matzner-Lber, "Statistical estimation of the residential baseline," *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 1752–1759, May 2016.

[15] A. A. S. Mohajeryami, M. Doostan and P. Schwarz, "Error analysis of customer baseline load (cbl) calculation methods for residential customers," *IEEE Transactions on Industry Applications*, vol. 53, no. 1, pp. 5–14, Jan.-Feb. 2017.

[16] D. J. H. F. L. W. Kong, Z. Y. Dong and Y. Xu, "Short-term residential load forecasting based on resident behaviour learning," *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 1087–1088, Jan. 2018.

[17] M. X. H. Shi and R. Li, "Deep learning for household load forecastinga novel pooling deep rnn," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 5271–5280, Sept. 2018.

[18] Y. J. D. J. H. Y. X. W. Kong, Z. Y. Dong and Y. Zhang, "Short-term residential load forecasting based on lstm recurrent neural network," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 841–851, Jan. 2019.

[19] Y. J. R. Jiao, T. Zhang and H. He, "Short-term non-residential load forecasting based on multiple sequences lstm recurrent neural network," *IEEE Acces*, vol. 6, pp. 59 438–59 448, 2018.

[20] S. Y. K. Park and E. Hwang, "Hybrid load forecasting for mixed-use complex based on the characteristic load decomposition by pilot signals," *IEEE Access*, vol. 7, pp. 12 297–12 306, 2019.

[21] H. H. D. Y. W. F. L. Quilumba, W. Lee and R. L. Szabados, "Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities," *IEEE Transactions on Smart Grid*, vol. 6, no. 2, pp. 911–918, March 2015.

[22] R. X. Y. Zhang, W. Chen and J. Black, "A cluster-based method for calculating baselines for residential loads," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2368–2377, Sept. 2016.

[23] ENERNOC, *The Demand Response Baseline*, 2011, Available:https://library.cee1.org/sites/default/files/library/10774/CEE_EvalDRBaseline_2011.pdf.

[24] J. Q. S. Z. M. L. H. Yang, J. Zhang and Z. Y. Dong, "A practical pricing approach to smart grid demand response based on load classification," *IEEE Transactions on Smart Grid*, vol. 9, no. 1, pp. 179–190, Jan. 2018.

[25] L. W. Z. Zhao and G. Song, "Convergence of volatile power markets with price-based demand response," *IEEE Transactions on Power Systems*, vol. 29, no. 5, pp. 2107–2118, Sept. 2014.

[26] W. Y. R. Yu and S. Rahardja, "A statistical demand-price model with its application in optimal real-time price," *IEEE Transactions on Smart Grid*, vol. 3, no. 4, pp. 1734–1742, Dec. 2012.

[27] H. J. C. Douglas, C. Laurits, "Consistency of residential customer response in time-of-use electricity pricing experiments," *Journal of Econometrics*, vol. 26, pp. 179–203, 1984.

[28] Y. Q. B. L. D. Liu, Y. Sun and Y. Xu, "Analysis and accurate prediction of users response behavior in incentive-based demand response," *IEEE Access*, vol. 7, pp. 3170–3180, 2019.

[29] Z. S. Y. Zhou, Zhifei Li and W. Li, "The charging and discharging power prediction for electric vehicles," *IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society*, pp. 4014–4019, 2016.

[30] H. P. H. S. Jang, K. Y. Bae and D. K. Sung, "Solar power prediction based on satellite images and support vector machine," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 3, pp. 1255–1263, July 2016.

[31] T. X. W. A. Asrari and B. Ramos, "A hybrid algorithm for short-term solar power predictionsunshine state case study," *IEEE Transactions on Sustainable Energy*, vol. 8, no. 2, pp. 582–591, April 2017.

[32] Q. Liu and Q. Zhang, "Accuracy improvement of energy prediction for solar-energy-powered embedded systems," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 24, no. 6, pp. 2062–2074, June 2016.

[33] N. S. B. Khorramdel, C. Y. Chung and G. C. D. Price, "A fuzzy adaptive probabilistic wind power prediction framework using diffusion kernel density estimators," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 7109–7121, Nov. 2018.

[34] L. J. O. C. S. Ioakimidis and K. N. Genikomsakis, "Wind power forecasting in a residential location as part of the energy box management decision tool," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 4, pp. 2103–2111, Nov. 2014.

[35] G. E. Constante-Flores and M. Illindala, "Data-driven probabilistic power flow analysis for a distribution system with renewable energy sources using monte carlo simulation," *IEEE/IAS 53rd Industrial and Commercial Power Systems Technical Conference (ICPS)*, pp. 1–8, 2017.

[36] A. C. C. C. Keerthisinghe and G. Verbi, "Energy management of pv-storage systems: Policy approximations using machine learning," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 1, pp. 257–265, Jan. 2019.

[37] Z. Zhou, "Game-theoretical energy management for energy internet with big data-based renewable power forecasting," *IEEE Access*, vol. 5, pp. 5731–5746, 2017.

[38] M. S.-E.-E. M. Shabanzadeh and M. Haghifam, "Risk-based medium-term trading strategy for a virtual power plant with first-order stochastic dominance constraints," *IET Generation, Transmission Distribution*, vol. 11, no. 2, pp. 520–529, 2017.

[39] ——, "Optimal offering strategy of a virtual power plant: A stochastic bi-level approach," *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 794–806, March 2016.

[40] Y. Z. J. Qiu, K. Meng and Z. Y. Dong, "Optimal scheduling of distributed energy resources as a virtual power plant in a transactive energy framework," *IET Generation, Transmission Distribution*, vol. 11, no. 13, pp. 3417–3427, 2017.

[41] S. R. Dabbagh and M. K. Sheikh-El-Eslami, "Risk assessment of virtual power plants offering in energy and reserve markets," *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3572–3582, Sept. 2016.

[42] T. J. H. P. Z. Liang, Q. Alsafasfeh and W. Su, "Risk-constrained optimal energy management for virtual power plants considering correlated demand response," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 1577–1587, March 2019.

[43] Z. L. B. W. Y. Z. H. Jia, W. Qi and T. Xu, "Hierarchical risk assessment of transmission system considering the influence of active distribution network," *IEEE Transactions on Power Systems*, vol. 30, no. 2, pp. 1084–1093, March 2015.

[44] Y. Wang, "An adaptive importance sampling method for spinning reserve risk evaluation of generating systems incorporating virtual power plants," *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 5082–5091, Sept. 2018.

[45] ——, "Fuzzy day-ahead scheduling of virtual power plant with optimal confidence level," *IET Generation, Transmission Distribution*, vol. 10, no. 1, pp. 205–212, 2016.

[46] S. H. P. Malahat and S. Kazem, "Two-stage approach for optimal dispatch of distributed energy resources in distribution networks considering virtual power plant concept," *International Transactions on Electrical Energy System*, vol. 24, no. 1, pp. 43–63, 2014.

[47] X. W. B. L. G. Zhang, C. Jiang and H. Zhu, "Bidding strategy analysis of virtual power plant considering demand response and uncertainty of renewable energy," *IET Generation, Transmission Distribution*, vol. 11, no. 13, pp. 3268–3277, 2017.

[48] S. M.-K. S. Morteza and H. Mahmoud-Reza, "The design of a risk-hedging tool for virtual power plants via robust optimization approach," *Applied Energy*, vol. 115, pp. 766–777, 2015.

[49] P. P. L. Halilbai and S. Chatzivasileiadis, "Convex relaxations and approximations of chance-constrained ac-opf problems," *EEE Transactions on Power Systems*, vol. 34, no. 2, pp. 1459–1470, March . 2019.

[50] A. D. D.-G. H. Xu and P. W. Sauer, "Data-driven coordination of distributed energy resources for active power provision," *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 3047–3058, July 2019.

[51] R. A. Jabr, "Radial distribution load flow using conic programming," *IEEE Transactions on Power Systems*, vol. 21, no. 3, pp. 1458–1459, Aug. 2006.

[52] X. Geng and L. Xie, "*Data-driven Decision Making with Probabilistic Guarantees (Part 1): A Schematic Overview of Chance-constrained Optimization*", 2019, Available:https://www.researchgate.net/publication/332011109_Data-driven_Decision_Making_with_Probabilistic_Guarantees_Part_1_A_Schematic_Overview_of_Chance-constrained_Optimization.

APPENDIX A

DETAILED RESULTS FOR THE FOUR DATASETS ANALYSIS

A.1 FIRST AND SECOND DATASETS

Table A.1: E21 Customers Responsiveness for First and Second Datasets

E21	Summer Peak						Summer May and Jun						Summer Sep and Oct					
Category	NO. of customer in the group	Accuracy of baseline prediction model R2	Accuracy of baseline prediction model RMSE (kWh)	Avg. consumption (kWh) in one time interval (15 mins)	Avg. load reduction (kWh)	Total load reduction (kWh)	NO. of customer in the group	Accuracy of baseline prediction model R2	Accuracy of baseline prediction model RMSE (kWh)	Avg. consumption (kWh) in one time interval (15 mins)	Avg. load reduction (kWh)	Total load reduction (kWh)	NO. of customer in the group	Accuracy of baseline prediction model R2	Accuracy of baseline prediction model RMSE (kWh)	Avg. consumption (kWh) in one time interval (15 mins)	Avg. load reduction (kWh)	Total load reduction (kWh)
HighIncome + 1-2 people + no child	60	0.968	11.478	182.85	55.39	28731.81	60	0.975	10.622	144.30	32.33	17072.56	60	0.964	9.594	124.67	36.79	18543.84
HighIncome + 3-5 people + no child	60	0.970	11.111	176.96	45.71	23585.27	60	0.979	9.410	131.64	37.68	19894.86	60	0.971	9.832	114.88	31.17	15707.49
HighIncome + 3-5 people + have children	60	0.963	13.333	217.54	79.95	41253.90	60	0.975	11.854	170.65	63.45	33500.96	60	0.964	11.005	150.91	48.76	24577.46
MediumIncome + 1-2 people + no child	60	0.969	10.859	173.61	56.81	29313.27	60	0.977	9.814	127.93	39.87	21052.16	60	0.968	9.069	113.41	41.48	20903.90
MediumIncome + 3-5 people + no child	60	0.965	11.510	181.89	47.07	24286.88	60	0.979	9.788	135.69	20.65	10901.54	60	0.972	8.978	120.40	37.98	19139.73
MediumIncome + 3-5 people + have children	60	0.970	11.292	181.58	38.53	19882.29	60	0.981	10.171	138.76	38.89	20534.93	60	0.970	10.253	120.90	32.25	16251.53
LowIncome + 1-2 people + no child	60	0.968	11.134	175.46	21.12	10895.55	60	0.981	9.495	132.04	15.01	7923.87	60	0.974	9.098	110.47	13.00	6549.56

Table A.2: E26 Customers Responsiveness for First and Second Datasets

E26	Summer Peak						Summer May and Jun						Summer Sep and Oct					
Category	NO. of customer in the group	Accuracy of baseline prediction model R2	Accuracy of baseline prediction model RMSE (kWh)	Avg. consumption (kWh) in one time interval (15 mins)	Avg. load reduction (kWh)	Total load reduction (kWh)	NO. of customer in the group	Accuracy of baseline prediction model R2	Accuracy of baseline prediction model RMSE (kWh)	Avg. consumption (kWh) in one time interval (15 mins)	Avg. load reduction (kWh)	Total load reduction (kWh)	NO. of customer in the group	Accuracy of baseline prediction model R2	Accuracy of baseline prediction model RMSE (kWh)	Avg. consumption (kWh) in one time interval (15 mins)	Avg. load reduction (kWh)	Total load reduction (kWh)
HighIncome + 1-2 people + no child	60	0.969	10.910	180.21	8.72	10495.82	60	0.977	9.826	142.10	-11.79	-14521.90	60	0.962	9.517	119.19	4.53	5332.27
HighIncome + 3-5 people + no child	60	0.952	13.090	182.52	-0.60	-727.71	60	0.970	11.164	140.90	-16.78	-20674.75	60	0.963	9.758	122.80	0.19	220.42
HighIncome + 3-5 people + have children	60	0.967	11.714	186.25	19.72	23739.08	60	0.978	10.145	140.09	9.57	11789.17	60	0.966	9.981	119.32	-0.27	-318.92
MediumIncome + 1-2 people + no child	60	0.959	11.418	174.05	-1.76	-2114.39	60	0.973	10.372	132.46	-19.36	-23853.17	60	0.952	10.288	117.03	-11.33	-13327.30
MediumIncome + 3-5 people + no child	60	0.961	11.381	178.34	-1.73	-2085.16	60	0.974	10.757	139.07	-20.25	-24949.82	60	0.971	9.519	119.03	3.52	4136.94
MediumIncome + 3-5 people + have children	60	0.964	11.685	175.12	-2.28	-2745.37	60	0.976	10.252	135.81	1.08	1328.10	60	0.961	9.450	114.07	-2.20	-2584.42
LowIncome + 1-2 people + no child	60	0.963	10.802	175.24	2.47	2977.93	60	0.968	10.458	134.75	-2.94	-3623.17	60	0.965	9.462	115.30	5.01	5886.90

A.2 THRID DATASETS

Table A.3: E21 Customers Responsiveness for Third Datasets (Summer Peak)

NO of Occupants	Livable Space	Start Date	NO. of Customers	R2	RMSE	Avg. Load Reduction (kW)	Percentage Avg. Load Reduction (%)
1-2	less1500sqft	After2016	60	0.9774	7.2194	14.05	6.84
3-5	less1500sqft	After2016	60	0.9782	7.6320	25.40	10.73
6-max	less1500sqft	After2016	60	0.9747	8.9031	45.27	17.75
1-2	1500to2500sqft	After2016	60	0.9720	9.3695	17.00	6.84
3-5	1500to2500sqft	After2016	60	0.9771	8.7200	52.96	20.15
6-max	1500to2500sqft	After2016	60	0.9820	10.1913	61.93	20.25
1-2	more2500sqft	After2016	55	0.9563	12.7116	50.57	16.28
3-5	more2500sqft	After2016	60	0.9759	13.0941	118.60	27.79
6-max	more2500sqft	After2016	22	0.9383	6.6320	57.48	41.13
1-2	less1500sqft	Before2016	60	0.9727	8.6440	24.43	10.21
3-5	less1500sqft	Before2016	60	0.9667	9.3171	47.42	19.30
6-max	less1500sqft	Before2016	60	0.9703	9.9641	23.64	9.18
1-2	1500to2500sqft	Before2016	60	0.9643	10.0608	50.73	19.67
3-5	1500to2500sqft	Before2016	60	0.9739	9.8538	69.77	22.92
6-max	1500to2500sqft	Before2016	60	0.9741	10.9881	81.59	26.49
1-2	more2500sqft	Before2016	60	0.9733	12.9001	51.23	14.19
3-5	more2500sqft	Before2016	60	0.9766	13.3278	121.73	29.97
6-max	more2500sqft	Before2016	60	0.9800	14.6901	139.87	31.78

Table A.4: E21 Customers Responsiveness for Third Datasets (Summer Sep Oct)

NO of Occupants	Livable Space	Start Date	NO. of Customers	R2	RMSE	Avg. Load Reduction (kW)	Percentage Avg. Load Reduction (%)
1-2	less1500sqft	After2016	60	0.9628	8.3739	13.85	9.47
3-5	less1500sqft	After2016	60	0.9647	8.1317	25.29	14.47
6-max	less1500sqft	After2016	60	0.9620	10.4800	36.93	19.99
1-2	1500to2500sqft	After2016	60	0.9702	9.1318	21.15	11.46
3-5	1500to2500sqft	After2016	60	0.9661	10.3202	49.12	24.67
6-max	1500to2500sqft	After2016	60	0.9630	12.0356	49.35	22.71
1-2	more2500sqft	After2016	55	0.9719	11.1362	40.78	18.71
3-5	more2500sqft	After2016	60	0.9710	12.5142	78.93	25.94
6-max	more2500sqft	After2016	22	0.9155	7.6169	36.54	39.96
1-2	less1500sqft	Before2016	60	0.9492	10.3945	14.77	9.11
3-5	less1500sqft	Before2016	60	0.9649	9.0492	34.37	19.47
6-max	less1500sqft	Before2016	60	0.9620	10.9456	19.48	10.58
1-2	1500to2500sqft	Before2016	60	0.9669	9.1223	44.70	23.65
3-5	1500to2500sqft	Before2016	60	0.9673	11.2294	35.82	16.73
6-max	1500to2500sqft	Before2016	60	0.9718	10.4920	52.42	24.29
1-2	more2500sqft	Before2016	60	0.9611	11.8571	38.44	15.05
3-5	more2500sqft	Before2016	60	0.9759	11.1032	100.06	32.83
6-max	more2500sqft	Before2016	60	0.9637	15.7048	99.39	31.29

Table A.5: E21 Customers Responsiveness for Third Datasets (Summer May Jun)

NO of Occupants	Livable Space	Start Date	NO. of Customers	R2	RMSE	Avg. Load Reduction (kW)	Percentage Avg. Load Reduction (%)
1-2	less1500sqft	After2016	60	0.9742	8.6617	13.20	7.75
3-5	less1500sqft	After2016	60	0.9775	9.5824	27.75	13.74
6-max	less1500sqft	After2016	60	0.9809	9.0808	49.82	22.89
1-2	1500to2500sqft	After2016	60	0.9717	10.6510	22.06	10.13
3-5	1500to2500sqft	After2016	60	0.9732	9.9266	43.66	20.66
6-max	1500to2500sqft	After2016	60	0.9795	11.9005	52.60	20.69
1-2	more2500sqft	After2016	55	0.9713	12.5641	40.98	16.27
3-5	more2500sqft	After2016	60	0.9804	12.4024	93.99	26.78
6-max	more2500sqft	After2016	22	0.9528	6.3210	34.25	32.78
1-2	less1500sqft	Before2016	60	0.9710	9.7369	23.01	11.23
3-5	less1500sqft	Before2016	60	0.9766	9.3173	28.25	14.51
6-max	less1500sqft	Before2016	60	0.9702	10.4940	15.48	7.41
1-2	1500to2500sqft	Before2016	60	0.9771	8.6452	39.51	18.53
3-5	1500to2500sqft	Before2016	60	0.9789	9.8773	27.54	11.16
6-max	1500to2500sqft	Before2016	60	0.9754	10.9469	70.19	27.18
1-2	more2500sqft	Before2016	60	0.9796	13.5768	41.69	13.88
3-5	more2500sqft	Before2016	60	0.9843	11.7368	103.57	30.34
6-max	more2500sqft	Before2016	60	0.9810	14.2655	97.80	26.81

Table A.6: E26 Customers Responsiveness for Third Datasets (Summer Peak)

NO of Occupants	Livable Space	Start Date	NO. of Customers	R2	RMSE	Avg. Load Reduction (kW)	Percentage Avg. Load Reduction (%)
1-2	less1500sqft	After2016	60	0.9714	7.8878	10.41	4.96
3-5	less1500sqft	After2016	60	0.9712	9.7310	11.17	5.00
1-2	1500to2500sqft	After2016	60	0.9706	10.1515	12.07	4.69
3-5	1500to2500sqft	After2016	60	0.9817	9.1397	17.38	6.31
6-max	1500to2500sqft	After2016	23	0.9487	5.2721	11.46	10.47
3-5	more2500sqft	After2016	38	0.9686	10.1373	5.41	2.17
1-2	less1500sqft	Before2016	60	0.9734	8.0724	7.08	3.36
3-5	less1500sqft	Before2016	60	0.9753	8.6665	-7.49	-3.54
6-max	less1500sqft	Before2016	60	0.9787	9.2521	15.01	5.72
1-2	1500to2500sqft	Before2016	60	0.9618	10.0778	6.09	2.50
3-5	1500to2500sqft	Before2016	60	0.9728	9.7418	0.25	0.09
6-max	1500to2500sqft	Before2016	60	0.9749	10.8574	15.86	5.21
1-2	more2500sqft	Before2016	60	0.9749	11.3718	-4.81	-1.67
3-5	more2500sqft	Before2016	60	0.9687	15.2892	28.77	7.57
6-max	more2500sqft	Before2016	60	0.9687	15.4258	24.93	6.55

Table A.7: E26 Customers Responsiveness for Third Datasets (Summer Sep Oct)

NO of Occupants	Livable Space	Start Date	NO. of Customers	R2	RMSE	Avg. Load Reduction (kW)	Percentage Avg. Load Reduction (%)
1-2	less1500sqft	After2016	60	0.9541	7.8082	4.23	3.03
3-5	less1500sqft	After2016	60	0.9681	8.8777	11.84	7.55
1-2	1500to2500sqft	After2016	60	0.9628	9.9307	19.17	9.87
3-5	1500to2500sqft	After2016	60	0.9770	9.4999	9.17	4.78
6-max	1500to2500sqft	After2016	23	0.9389	5.4865	5.64	5.64
3-5	more2500sqft	After2016	38	0.9654	8.5809	2.17	1.27
1-2	less1500sqft	Before2016	60	0.9668	7.6298	5.36	3.79
3-5	less1500sqft	Before2016	60	0.9650	8.4682	0.73	0.50
6-max	less1500sqft	Before2016	60	0.9607	10.1398	6.43	3.58
1-2	1500to2500sqft	Before2016	60	0.9696	9.1644	24.94	13.25
3-5	1500to2500sqft	Before2016	60	0.9777	9.5047	2.49	1.30
6-max	1500to2500sqft	Before2016	60	0.9579	11.0347	2.32	1.15
1-2	more2500sqft	Before2016	60	0.9756	8.8255	2.96	1.50
3-5	more2500sqft	Before2016	60	0.9726	10.9621	27.44	10.31
6-max	more2500sqft	Before2016	60	0.9628	12.2618	15.63	5.91

Table A.8: E26 Customers Responsiveness for Third Datasets (Summer May Jun)

NO of Occupants	Livable Space	Start Date	NO. of Customers	R2	RMSE	Avg. Load Reduction (kW)	Percentage Avg. Load Reduction (%)
1-2	less1500sqft	After2016	60	0.9679	8.8958	2.18	1.31
3-5	less1500sqft	After2016	60	0.9715	9.7922	22.12	11.59
1-2	1500to2500sqft	After2016	60	0.9730	10.0020	12.65	5.78
3-5	1500to2500sqft	After2016	60	0.9806	10.2643	18.49	7.83
6-max	1500to2500sqft	After2016	23	0.9494	5.9965	9.30	10.27
3-5	more2500sqft	After2016	38	0.9631	10.7319	9.84	4.54
1-2	less1500sqft	Before2016	60	0.9783	8.0759	9.13	5.30
3-5	less1500sqft	Before2016	60	0.9679	9.0555	5.15	2.86
6-max	less1500sqft	Before2016	60	0.9707	10.6497	22.41	9.99
1-2	1500to2500sqft	Before2016	60	0.9814	8.5932	12.78	6.24
3-5	1500to2500sqft	Before2016	60	0.9746	10.3415	10.69	4.62
6-max	1500to2500sqft	Before2016	60	0.9765	10.3353	33.73	12.34
1-2	more2500sqft	Before2016	60	0.9752	12.1527	-16.80	-7.48
3-5	more2500sqft	Before2016	60	0.9791	13.0396	25.94	8.19
6-max	more2500sqft	Before2016	60	0.9769	15.2741	39.08	11.74

Table A.9: E21 and E26 One Demographic Factor Analysis for Third Dataset (Summer Peak)

Rate	Demographic factor	NO. of Customers	R2	RMSE (kW)	Avg. Load Reduction (kW)	Percentage Avg. Load Reduction (%)
E21	less1500sqft	5483	0.9955	311.0858	1592.75	7.97
	1500to2500sqft	2939	0.9961	188.3553	1867.76	14.22
	more2500sqft	524	0.9907	63.9577	1022.71	29.26
	After2016	3202	0.9939	204.7800	1447.95	11.55
	Before2016	5744	0.9950	367.3647	3036.02	12.61
	1-2 people	5774	0.9950	339.9138	1852.33	8.52
	3-5 people	2388	0.9914	204.0060	1826.33	16.85
	6-max people	784	0.9909	82.0036	837.64	20.62
E26	less1500sqft	4529	0.9892	345.9386	336.27	2.01
	1500to2500sqft	4047	0.9936	316.2021	413.49	2.31
	more2500sqft	1087	0.9904	147.2767	195.07	3.00
	After2016	666	0.9890	59.5779	118.86	4.34
	Before2016	8997	0.9922	754.0461	778.70	2.03
	1-2 people	4452	0.9964	240.7565	369.31	2.20
	3-5 people	3185	0.9948	235.4018	243.62	1.72
	6-max people	2026	0.9919	212.6305	355.21	3.47

Table A.10: E21 and E26 One Demographic Factor Analysis for Third Dataset (Summer Sep Oct)

Rate	Demographic factor	NO. of Customers	R2	RMSE (kW)	Avg. Load Reduction (kW)	Percentage Avg. Load Reduction (%)
E21	less1500sqft	5483	0.9921	361.7314	1289.90	9.24
	1500to2500sqft	2939	0.9951	196.5508	1416.51	15.32
	more2500sqft	524	0.9864	62.4531	631.51	26.33
	After2016	3202	0.9943	197.3997	1087.27	12.41
	Before2016	5744	0.9920	374.0446	2216.27	13.18
	1-2 people	5774	0.9955	314.5156	1399.11	9.29
	3-5 people	2388	0.9937	166.4658	1271.79	16.76
	6-max people	784	0.9840	92.2772	581.11	20.23
E26	less1500sqft	4529	0.9934	275.9834	778.48	6.50
	1500to2500sqft	4047	0.9915	325.0663	586.80	4.69
	more2500sqft	1087	0.9909	116.4776	142.67	3.20
	After2016	666	0.9888	55.6323	132.83	6.84
	Before2016	8997	0.9935	628.9174	1375.78	5.09
	1-2 people	4452	0.9939	268.5560	738.13	6.28
	3-5 people	3185	0.9920	242.2960	395.52	3.98
	6-max people	2026	0.9839	226.2338	289.61	4.04

Table A.11: E21 and E26 One Demographic Factor Analysis for Third Dataset (Summer May Jun)

Rate	Demographic factor	NO. of Customers	R2	RMSE (kW)	Avg. Load Reduction (kW)	Percentage Avg. Load Reduction (%)
E21	less1500sqft	5483	0.9953	361.8017	1512.03	9.19
	1500to2500sqft	2939	0.9946	210.5073	1587.05	14.60
	more2500sqft	524	0.9939	64.2697	750.70	26.49
	After2016	3202	0.9942	208.1004	1220.18	11.86
	Before2016	5744	0.9952	389.3419	2631.11	13.24
	1-2 people	5774	0.9971	291.0692	1659.45	9.31
	3-5 people	2388	0.9944	198.6030	1490.65	16.71
	6-max people	784	0.9890	97.0758	711.21	20.83
E26	less1500sqft	4529	0.9948	313.0059	928.97	6.50
	1500to2500sqft	4047	0.9966	264.8593	618.91	4.07
	more2500sqft	1087	0.9914	134.8799	378.83	6.46
	After2016	666	0.9932	57.1282	213.99	8.98
	Before2016	8997	0.9930	708.4868	1984.82	6.08
	1-2 people	4452	0.9915	378.9022	912.87	6.43
	3-5 people	3185	0.9952	246.9225	684.22	5.68
	6-max people	2026	0.9900	233.7112	638.65	7.23

A.3 FOURTH DATASET

Table A.12: All Customers Analysis for Fourth Data

Year	Month	Day	Start Time	End Time	Reduction (kW)			Percentage Reduction (%)		
2018	6	22	17	19	3025.46	2185.57	-	52.5%	38.0%	-
2018	7	24	16	18	3462.03	1709.89	-	53.1%	26.3%	-
2018	7	25	16	18	3073.48	1529.43	-	45.2%	22.9%	-
2018	7	30	15	18	3166.47	1718.66	1156.29	55.2%	29.1%	19.6%
2018	8	2	15	18	3443.10	2031.64	1586.98	57.8%	32.6%	25.1%
2018	8	6	15	18	3575.30	2078.12	1340.26	56.5%	31.5%	20.2%
2018	8	13	16	18	2148.12	1672.87	-	44.5%	33.0%	-
2018	8	14	17	19	2773.74	1881.96	-	51.1%	34.6%	-
2018	8	30	16	18	2668.76	1796.75	-	55.1%	35.9%	-
R2	0.992				Avg Consumption for Non-DR-Event Period (kW)			3571.54		
RMSE	118.79				Number of Customers			1301		

Table A.13: One Demographic Factor Analysis for Fourth Data: Low Income

Year	Month	Day	Start Time	End Time	Reduction (kW)			Percentage Reduction (%)		
2018	6	22	17	19	471.27	350.53	-	50.5%	37.2%	-
2018	7	24	16	18	566.98	235.08	-	54.5%	22.8%	-
2018	7	25	16	18	475.40	227.24	-	45.6%	22.0%	-
2018	7	30	15	18	492.20	327.62	261.26	56.0%	35.8%	28.2%
2018	8	2	15	18	536.33	310.24	231.97	58.7%	32.8%	24.3%
2018	8	6	15	18	540.18	284.19	161.95	56.9%	29.0%	16.4%
2018	8	13	16	18	351.87	285.94	-	45.3%	34.9%	-
2018	8	14	17	19	352.82	174.00	-	45.9%	22.8%	-
2018	8	30	16	18	376.53	244.74	-	52.4%	32.9%	-
R2	0.984				Avg Consumption for Non-DR-Event Period (kW)			561.21		
RMSE (kW)	27.22				Number of Customers			249		

Table A.14: One Demographic Factor Analysis for Fourth Data: Medium Income

Year	Month	Day	Start Time	End Time	Reduction (kW)			Percentage Reduction (%)		
2018	6	22	17	19	2125.08	1523.73	-	52.9%	38.1%	-
2018	7	24	16	18	2372.40	1180.46	-	52.4%	26.3%	-
2018	7	25	16	18	2100.53	993.36	-	44.5%	21.5%	-
2018	7	30	15	18	2227.42	1090.07	656.96	55.8%	26.6%	16.1%
2018	8	2	15	18	2375.84	1313.48	1029.89	57.1%	30.2%	23.4%
2018	8	6	15	18	2461.17	1372.84	839.86	56.1%	30.2%	18.5%
2018	8	13	16	18	1492.07	1121.28	-	44.9%	32.2%	-
2018	8	14	17	19	1978.54	1383.75	-	51.6%	36.0%	-
2018	8	30	16	18	1910.16	1275.82	-	55.8%	36.1%	-
R2	0.992				Avg Consumption for Non-DR-Event Period (kW)			2480.88		
RMSE (kW)	87.50				Number of Customers			916		

Table A.15: One Demographic Factor Analysis for Fourth Data: High Income

Year	Month	Day	Start Time	End Time	Reduction (kW)			Percentage Reduction (%)		
2018	6	22	17	19	434.63	311.85	-	52.8%	38.4%	-
2018	7	24	16	18	525.70	284.30	-	55.0%	29.5%	-
2018	7	25	16	18	475.11	278.48	-	47.0%	28.0%	-
2018	7	30	15	18	422.47	267.32	198.50	50.7%	31.3%	23.3%
2018	8	2	15	18	495.41	352.26	247.66	58.4%	40.0%	28.0%
2018	8	6	15	18	503.26	312.83	181.10	54.5%	32.6%	18.9%
2018	8	13	16	18	305.04	262.12	-	42.1%	34.5%	-
2018	8	14	17	19	452.90	342.99	-	53.9%	39.9%	-
2018	8	30	16	18	404.99	303.07	-	55.7%	40.3%	-
R2	0.981				Avg Consumption for Non-DR-Event Period (kW)			529.45		
RMSE (kW)	26.50				Number of Customers			136		

Table A.16: One Demographic Factor Analysis for Fourth Data: 1-2 Occupants

Year	Month	Day	Start Time	End Time	Reduction (kW)			Percentage Reduction (%)		
2018	6	22	17	19	1676.47	1270.42	-	51.8%	39.1%	-
2018	7	24	16	18	1876.68	912.64	-	52.4%	25.4%	-
2018	7	25	16	18	1619.47	777.53	-	43.5%	21.3%	-
2018	7	30	15	18	1737.88	951.51	681.80	54.6%	29.1%	20.9%
2018	8	2	15	18	1781.43	970.05	771.02	55.3%	29.0%	22.9%
2018	8	6	15	18	1887.83	975.01	662.03	55.6%	27.8%	18.9%
2018	8	13	16	18	1112.18	862.98	-	42.5%	31.6%	-
2018	8	14	17	19	1446.97	936.53	-	49.2%	32.0%	-
2018	8	30	16	18	1481.90	989.01	-	54.8%	35.6%	-
R2	0.991				Avg Consumption for Non-DR-Event Period (kW)			1980.53		
RMSE (kW)	70.72				Number of Customers			791		

Table A.17: One Demographic Factor Analysis for Fourth Data: 3-5 Occupants

Year	Month	Day	Start Time	End Time	Reduction (kW)			Percentage Reduction (%)		
2018	6	22	17	19	1156.15	770.04	-	52.9%	35.7%	-
2018	7	24	16	18	1349.15	652.60	-	53.6%	26.1%	-
2018	7	25	16	18	1208.20	606.35	-	46.1%	23.3%	-
2018	7	30	15	18	1188.23	581.82	344.50	55.4%	26.4%	15.6%
2018	8	2	15	18	1380.13	850.68	654.86	59.9%	35.1%	26.4%
2018	8	6	15	18	1417.34	885.16	513.13	56.9%	33.9%	19.5%
2018	8	13	16	18	908.90	717.90	-	47.7%	35.5%	-
2018	8	14	17	19	1115.45	770.27	-	52.8%	36.1%	-
2018	8	30	16	18	1015.09	660.42	-	55.7%	35.1%	-
R2	0.991				Avg Consumption for Non-DR-Event Period (kW)			1355.67		
RMSE (kW)	51.11				Number of Customers			435		

Table A.18: One Demographic Factor Analysis for Fourth Data: 6-9 Occupants

Year	Month	Day	Start Time	End Time	Reduction (kW)			Percentage Reduction (%)		
2018	6	22	17	19	151.17	108.38	-	58.7%	42.5%	-
2018	7	24	16	18	174.63	96.01	-	59.2%	33.1%	-
2018	7	25	16	18	168.51	87.63	-	53.6%	28.6%	-
2018	7	30	15	18	151.04	112.88	64.85	55.8%	41.2%	23.7%
2018	8	2	15	18	177.37	114.96	56.71	61.9%	39.5%	19.7%
2018	8	6	15	18	161.85	106.30	47.58	57.7%	36.9%	16.6%
2018	8	13	16	18	95.96	67.44	-	44.8%	31.0%	-
2018	8	14	17	19	178.46	157.03	-	61.3%	52.4%	-
2018	8	30	16	18	130.99	111.15	-	55.1%	45.7%	-
R2	0.961				Avg Consumption for Non-DR-Event Period (kW)			170.31		
RMSE (kW)	12.67				Number of Customers			44		