

UCLA

UCLA Electronic Theses and Dissertations

Title

Management of Demand Response Programs in the Electricity Industry

Permalink

<https://escholarship.org/uc/item/5s73d8h7>

Author

Rebeiz, Paul Pierre

Publication Date

2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Management of Demand Response Programs in the Electricity Industry

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Management

by

Paul Pierre Rebeiz

2016

ABSTRACT OF THE DISSERTATION

Management of Demand Response Programs in the Electricity Industry

by

Paul Pierre Rebeiz

Doctor of Philosophy in Management

University of California, Los Angeles, 2016

Professor Reza H. Ahmadi, Chair

Daily electricity load profile is characterized by peak hours which are periods in which electrical power is expected to be provided for a sustained period at a significantly higher than average supply level. As a result, satisfying the electricity demand throughout the day will entail utility companies to build additional plants that are only used during the highest peak hours of the year or to buy high-priced wholesale energy. Further, such costs will increase given the expected growth of electricity demand in the next decades. To avoid these additional costs and address the resulting supply-demand mismatch, utility companies have designed Demand Response Programs (DRP) which are programs that incentivize customers to shift their electricity demand from peak hours to off-peak hours. In this work, I study the problem of an electricity utility company that offers DRP to its commercial and industrial customers with the objective of reducing its electricity costs.

In Chapter 1, I give an overview of the electricity industry in the United States and describe the important role that DRP play in improving the electric grid reliability and reducing the costs of electricity generation for the utility companies. I then describe and formulate the problem of an electricity retailer that offers interruptible demand response programs, which are a type of DRP, to their commercial and industrial customers. These programs consist of the Base Interruptible Program (BIP) and the Agricultural and Pumping Interruptible Program (API). Using these contracts, enrolled customers agree to curtail their consumption by a pre-specified load when instructed and obtain in return financial payments from the utility company. The operational challenges of these programs are in their implementation and management due to the large number of interruption possibilities, the uncertainty in electricity demand and the limited number of interruptions the electricity retailer. To address these challenges, I propose and describe the solution adopted to solve the dynamic program. The approach I use consists of a certainty equivalence algorithm that had two components: an electric load forecasting model and the deterministic model of the dynamic program which I discuss in chapters 2 and 3 respectively.

In Chapter 2, I present an electric load forecasting model in the context of demand response for both the short and long term horizons. The short term model consists of predicting by nonparametric regression the hourly electricity demand at the start of a given day using the previous day load and same day temperature as the driving variables. The long term forecasting model consists of first predicting the peak load through multivariate and semiparametric regression taking into account the temperature variable and calendar effects. Then, I approximate the hourly load profile by nonparametric regression using the predicted peak load. Further, I construct the peak load distribution by temperature simulation and kernel density approximation. The proposed methodology had been used to forecast the short and long term electricity demand as well as the

probability distribution of the peak load for the area served by the Southern California Edison (SCE) electric utility company. The performance of the methodology is evaluated by comparing the forecasts results to the ones of the California Independent System Operator (CAISO) for the area served by SCE.

In Chapter 3, I study the problem of implementing these contracts by determining their execution policy using a certainty equivalence approach. A central component of the certainty equivalence algorithm is the deterministic problem in which the electricity demand is known. Given that this problem is NP-hard, we propose a heuristic that efficiently solves the deterministic problem and test its efficiency by determining the optimality gap with a lower bound. Using the developed electricity forecasting and demand simulation models we then solve the certainty equivalence model in order to devise near optimal strategies for executing such contracts and verify its effectiveness.

The dissertation of Paul Pierre Rebeiz is approved.

Rakesh K. Sarin

Guillaume Y. Roels

Sriram Dasu

Reza H. Ahmadi, Committee Chair

University of California, Los Angeles

2016

DEDICATION

For my mother, Arlette Rebeiz

and

my brother, Eric Rebeiz

and

in the memory of my dear father, Pierre Rebeiz

TABLE OF CONTENTS

1	Introduction	1
1.1	Electricity Industry Overview.....	1
1.2	Problem Formulation and Solution Approach.....	3
2	Load Forecasting Model	10
2.1	Introduction.....	10
2.2	Short Term Load Forecasting Model.....	14
2.2.1	Hourly Load Forecasting.....	14
2.2.2	Analysis of the residuals.....	18
2.2.3	Asymptotic Distribution.....	20
2.2.4	Benchmarking the model with an Artificial Neural Network.....	21
2.3	Long Term Load Forecasting Model.....	21
2.3.1	Peak Load Forecast.....	21
2.3.2	Long Term Hourly Load Profile Forecast.....	24
2.4	Temperature Modeling and Peak Load Probability Distribution.....	25
2.4.1	Temperature Modeling.....	25
2.4.2	Peak Load Probability Distribution.....	27
2.5	Conclusion.....	28

3 Certainty Equivalence Algorithm.....	48
3.1 Introduction.....	48
3.2 The Deterministic Demand Response Model.....	50
3.3 Aggregation and Lower Bound.....	51
3.4 DDRP Heuristic.....	53
3.4.1-Master Problem.....	53
3.4.2-Assignment Problem.....	55
3.5 Numerical Analysis.....	60
3.6 Certainty Equivalence Algorithm.....	61
3.7 Conclusion.....	62
References.....	65

LIST OF FIGURES

1.1	Electricity Supply Chain.....	8
1.2	Electricity Supply Curve for the Texas Wholesale Market.....	8
1.3	Solution Approach.....	9
2.1	Autocorrelation (left) and Partial Autocorrelation (right) of the residuals of in summer data.....	30
2.2	Autocorrelation (left) and Partial Autocorrelation (right) of the residuals of in sample summer data.....	31
2.3	Autocorrelation (left) and Partial Autocorrelation (right) of the residuals of in sample winter data.....	32
2.4	Forecasted Load versus Actual Load for Summer Season.....	33
2.5	Forecasted Load versus Actual Load for Winter Season.....	34
2.6	Illustration of a Neural Network with a Single Layer.....	35
2.7	Peak Load Time Series.....	36
2.8	Peak Load versus Peak Temperature for Summer Season.....	36
2.9	Peak Load versus Peak Temperature for Winter Season.....	37
2.10	Peak Temperatures from 2011 to 2013 for the counties of Los Angeles (Top) and Riverside (Bottom).....	38
2.11	Actual and Deterministic Temperature Plots of Los Angeles (Top) and Riverside Bottom).....	39
2.12	Serial Correlation and Sample Partial Autocorrelation Plots for Los Angeles (Left) and Riverside (Right).....	40
2.13	Regression Residuals and Their Serial Correlation for Los Angeles (Left) and Riverside (Right).....	41
2.13	Original Temperature time series for 2010 and four realistic simulations of Temperature Realizations.....	41
2.14	Peak Load Density.....	42

LIST OF TABLES

2.1	Smoothing Parameters (DL: Day Light Saving)	43
2.2	Short Term Forecasting Model Comparison.....	43
2.3	Multivariate Regression Results for Summer Season.....	44
2.4	OLS Regression Results for Winter Season.....	45
2.5	Covariance of hourly load and peak loads.....	46
2.6	Long Term Forecasting Model Comparison.....	46
2.7	Sine Model Parameters Estimates.....	46
2.8	Day Types.....	46
2.9	Monthly Probability Distribution of Day Types.....	47
2.10	Load Profile Approximation for Each Day Type.....	47
3.1	50 Days Horizon Results.....	64
3.2	100 Days Horizon Results.....	64
3.3	200 Days Horizon Results.....	64

ACKNOWLEDGEMENTS

I am extremely grateful for the opportunity to study at the UCLA Anderson School of Management. The school has provided me an excellent environment that facilitated my development as a researcher. My dissertation and progress throughout the doctoral program would not have been possible without the guidance and support of my committee members. I am very grateful for the support and assistance of my advisor, Reza Ahmadi. He has been a great mentor and has facilitated my development as a researcher.

I am also thankful for Sriram Dasu, Rakesh Sarin and Guillaume Roels for their valuable insights and comments on my research. I am also grateful to Felipe Caro, Kumar Rajaram, Charles Corbett and Chris Tang for their constant encouragement and support. I also would like to thank my colleagues in the doctoral program Sandeep Rath, Christian Blanco, Wei Zhang and Bobby Nyotta for their invaluable support.

I also thank my parents for their support. Their encouragements and prayers were a constant source of motivation for me. I am also extremely grateful to my brother Eric and his wife Emily who were always there for me during my doctoral studies.

VITA

2005	B.Eng. in Civil Engineering American University of Beirut, Beirut, Lebanon
2005-2007	Environmental Engineer, Dar Al Handasah, Beirut, Lebanon
2007-2009	M.S. in Operation Research University of Texas, Austin, Austin, Texas
2009-2010	Master in Management, Fuqua School of Business, Durham, North Carolina
2010-2011	Research Associate, Fuqua School of Business, Durham, North Carolina
2011-2016	Graduate Student Researcher, UCLA Anderson School of Management

CHAPTER 1

Introduction

1.1-Electricity Industry Overview

The electric grid in the United States consists of a network of transmission lines, substations and transformers that deliver electricity from the power plant to end users. Throughout the years the electric grid was improved with the advancements of technology through each decade. Today it consists of more than 9,000 units generating more than 1 million megawatts that are transmitted across 300,000 miles of transmission lines. Figure-1.1 shows the electricity supply chain. At the upstream, power plants produce electricity that is transmitted across distribution lines before reaching a substation that reduces the electricity voltage before transmitting it to distribution lines that supply end-customers. In order to move forward, a different kind of electricity grid is needed to handle the needs and increasing complexity of our electricity demand. Such a grid will need to harness the advances in computerized and digital technology. This digital technology will allow the grid to provide a two-way communication between the utility and its customers and to have sensing capabilities along the transmission lines. This new grid concept is referred nowadays by the Smart Grid. The Smart Grid represents an opportunity for the electricity industry to move into a new era of reliability, availability and efficiency that will improve our current economic and environmental status. The benefits associated with the Smart Grid will include a more efficient transmission of electricity, a quicker restoration of outages, a reduction in operating costs and peak demand and a larger integration of renewable energy systems. As mentioned above, the Smart Grid will enable an unprecedented level of customer participation and thus will increase customer awareness about the role they could play in enhancing the overall reliability and security of the grid. Technologies like smart meter and other mechanisms will allow customers to decide how

much electricity to use, when to use it and its cost. The main goal of this two-way communication is to reduce the electricity demand during periods when it is significantly higher than average supply level. The resulting reduction in the supply demand mismatch will mitigate the high costs incurred by a utility company and decrease the stress on the distribution lines. Figure-1.2 illustrates the supply curve of the Texas wholesale electricity market. We notice that after a certain electricity demand threshold, the electricity generation costs increase substantially as Peakers are invoked in order to supply the demand. Peakers consist of power plants that are on hold and supply electricity when demand exceeds a supplier's capacity. Since Peakers provide electricity in real time, the produced electricity is purchased at the spot price by the electricity provider which is much higher than the average electricity price rate. Demand Response Programs provide a mechanism to better address the mismatch between supply and demand and enable customers to be active participants in the overall electric network reliability by shifting their demand from peak to off peak hours. These programs are divided into categories: Time Based Programs and Incentive Based Programs. In Time Based Programs, the electricity price changes depending on the period of the day. This offers end users the opportunity to shift their consumption from periods with high electricity price rates to periods of lower electricity price rates. There is no penalty if an end user does not curtail her electricity consumption. The second type of programs are the Incentive-Based programs in which end customers agree to curtail or shift their load consumption to periods of lower demand. More detailed explanations of demand response programs can be found on the FERC website. In this work we focus on one type of Incentive based programs which are the Interruptible Curtailable load programs. We particularly focus on the interruptible contracts under which commercial and industrial consumers are offered payments to reduce their electricity consumption to a pre-specified level that they choose. We consider the problem of a producer and provider of electricity

that offers interruptible contracts to its industrial and commercial customers. The interruptible contracts that we consider are the ones offered by the Southern California Edison utility company. These programs consist of the Base Interruptible Program (BIP) and the Agricultural and Pumping Interruptible Program (API). Each program has multiple customers that are enrolled in. Customers are divided into groups. Each group consists of a set of customers enrolled in the same program that are close geographically to each other. Each enrolled customer can be interrupted at most once per day for duration of at most six hours. For the customer enrolled in the BIP program, there is a limit of 180 hours of total interruption per year while customers enrolled in the API program can be interrupted at most 25 times per year. The objective of the utility company is to minimize its electricity generation costs by deciding on a given day, in the case of an interruption which group to call as well as the timing and duration of the interruption for each of the interrupted groups. In the next sections, we formally state the problem, discuss its challenges and solution approach.

1.2-Problem Formulation and Solution Approach

We first formulate the corresponding dynamic program. To give a formal illustration of the program we define the following sets and indices:

Indices:

D : The set of days.

G : The set of groups.

G_1, G_2 : The set of BIP and API groups respectively (subsets of G).

T : The set of hours of the day during which the interruptible programs can be called. These are the hours from 9am to 9pm.

T' : The set of hours of the day representing the starting time of an interruption. These are the hours from 8am to 8pm.

H : The set of the duration of interruptions in hour.

d : The subscript for days.

g : The subscript for groups.

g_1, g_2 : The subscripts for the set of BIP and API groups.

t : The subscript for the set of time of the day.

t' : The subscript for the set of starting time for interruption.

h : The subscript for the set of duration of interruption.

Parameters:

$u_{d,t}$: Demand on day d at time t in Megawatts (MW).

l_g : Load impact of group g in Megawatts (MW).

(a, b) : Parameters for the hourly cost function for electricity generation.

$I(t, t', h)$: Indicator equal to 1 if a group called at hour t' for h consecutive hours reduces the demand at time t on day d by its corresponding load impact.

S_d : The information available on day d .

States:

$bh_{g_1,d}$: Number of BIP hours of interruption remaining for group g_1 on day d

$ac_{g_2,d}$: Number of API calls remaining for group g_2 on day d .

Control:

$$X_{d,g,t',h} = \begin{cases} 1 & \text{if group } g \text{ in interrupted from hour } t' \text{ for } h \text{ consecutive hours} \\ 0 & \text{otherwise} \end{cases}$$

Transition:

$$bh_{g_1,d+1} = \begin{cases} bh_{g_1,d} - h X_{d,g_1,t',h}, & \text{if } X_{d,g_1,t',h} = 1 \text{ and } bh_{g_1,d} > h \\ bh_{g_1,d}, & \text{otherwise} \end{cases}$$

$$ac_{g_2,d+1} = \begin{cases} ac_{g_2,d} - X_{d,g_2,t',h}, & \text{if } X_{d,g_2,t',h} = 1 \text{ and } ac_{g_2,d} > 0 \\ ac_{g_2,d}, & \text{otherwise} \end{cases}$$

In the set of indices above, the set T represents the hours of the day where a reduction in the electric

load can occur. These hours are the hours from 9am to 9pm. The set T' represents the starting time of an interruption which can take place from 8am to 8pm. The electric load at time $t \in T$ represents the total load demand from hours $t - 1$ to t . The indicator $I(t, t', h)$ is equal to 1 if an interruption at time t' for h consecutive hours reduces the demand at time t by the aggregate load impact of the interrupted groups during that period. For example, if $t'=3$ and $h=2$, then $I(4, 3, 2)$ and $I(5, 3, 2)$ are equal to 1. The set S_d represents the set of available information on day d which consists of the current and week ahead hourly load electricity demand. These quantities are forecasted by the methodology summarized in Chapter 2.

The state space consists of the remaining number of hours $bh_{g_1,d} \forall g_1 \in G_1$ and calls $ac_{g_2,d} \forall g_2 \in G_2$ for each of the groups enrolled in the BIP and API programs respectively. At each day, the utility company has to decide whether an interruption should occur as well as which groups to interrupt and the timing and duration of each interruption. The decision to interrupt is represented by the integer variable $X_{d,g,t',h}$. If an enrolled group in the BIP program is interrupted for h hours its state transitions from $bh_{g_1,d}$ on day d to $bh_{g_1,d} - h$ the next day. On the other hand, if a group from the API program is interrupted its state transitions from $ac_{g_2,d}$ on day d to $ac_{g_2,d} - 1$ the next day.

We define by C_d the hourly electricity generation costs on day d . C_d is expressed as follows:

$$C_d = \sum_{t \in T} a \left(u_{d,t} - \sum_{g \in G} \sum_{t' \in T'} \sum_{h \in H} I(t, t', h) l_g X_{d,g,t',h} \right)^2 + b \left(u_{dt} - \sum_{g \in G} \sum_{t' \in T'} \sum_{h \in H} I(t, t', h) l_g X_{d,g,t',h} \right) \quad (1)$$

The electricity generation costs are expressed by a quadratic function which is commonly used in the literature and captures the fact that the electricity generation cost increases at a higher rate than

the electricity demand. The objective function consists of minimizing the total hourly electricity generation costs. At each hour the electricity generation cost can be reduced by calling one or more groups from the BIP or API programs. Each called group g will reduce the hourly electricity load by an amount l_g .

Dynamic Programming Formulation:

Let $F_d(bh_{g_1,d}, ac_{g_2,d}) \forall g_1 \in G_1, g_2 \in G_2$ be the value function of the expected electricity generation costs on day d . The dynamic program is expressed by the following bellman equation.

$$F_d(bh_d, ac_d) = \min_{X_{d,t',h}} \{C_d + E(F_{d+1}(bh_{d+1}, ac_{d+1})|S_d)\} \quad (2)$$

Where the minimization is done over the following set of constraints:

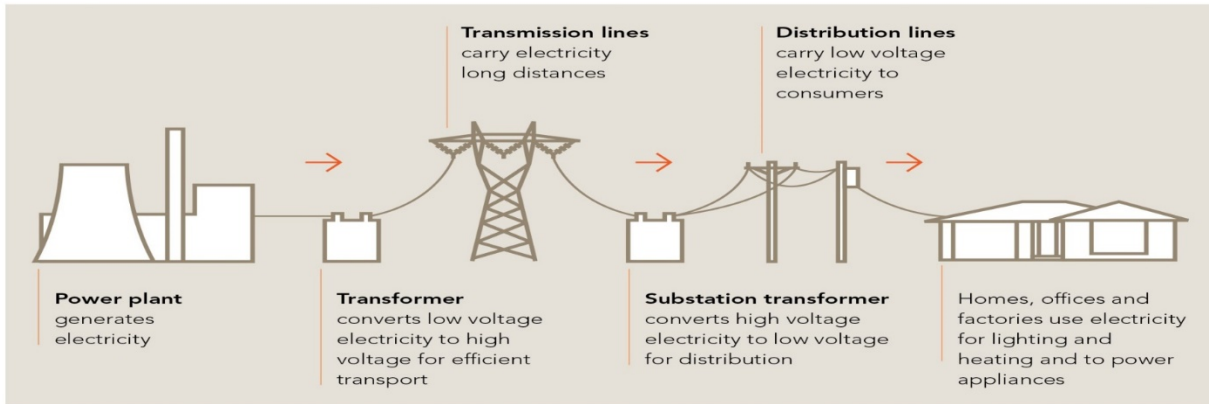
$$\sum_{t' \in T'} \sum_{h \in H} X_{d,g,t',h} \leq 1 \forall d \in D, g \in G \quad (3)$$

$$X_{d,g,t',h} \in \{0,1\} \forall d \in D, g \in G, t' \in T', h \in H \quad (4)$$

As seen from the above formulation, the challenges of the above problem stem from both the uncertainty in the load electricity demand and the large number of configurations through which groups can be interrupted. To address these challenges, we resort to approximate programming techniques and more specifically the certainty equivalence approach. To this end, we first develop an electricity forecasting model for the both the short and long term horizons. The forecasting model details are discussed in Chapter 2. The output of the forecasting model is then used in a deterministic model which we refer to as the Deterministic Demand Response Program (DDRP). The DDRP is used in the context of our certainty equivalence approach that we describe in Chapter 3. Figure 1.3 shows the solution approach. The first step consists of forecasting the electricity load

demand for both the short and long term horizons. In the short term, we forecast the week ahead hourly electricity demand while in the long term we determine number of day types across the months of the year. A day type is determined by the range in which the peak load of that day falls into and are discussed in Chapter 1. Each day type has a corresponding hourly load profile approximation. The output of the forecasting model consists of the hourly load profile for the short and long term forecasts. These load profiles are used as input in a Deterministic Demand Response Problem (DDRP) which represents the deterministic formulation of the problem's dynamic program. The DDRP solution consists of the interruption schedule of all groups across the horizon. We pick from the DDRP solution the interruption schedule of the current day and update the number of hours and calls of the groups enrolled in the BIP and API programs respectively. This procedure is repeated on a daily basis. In the next chapter we present our forecasting model and test its performance.

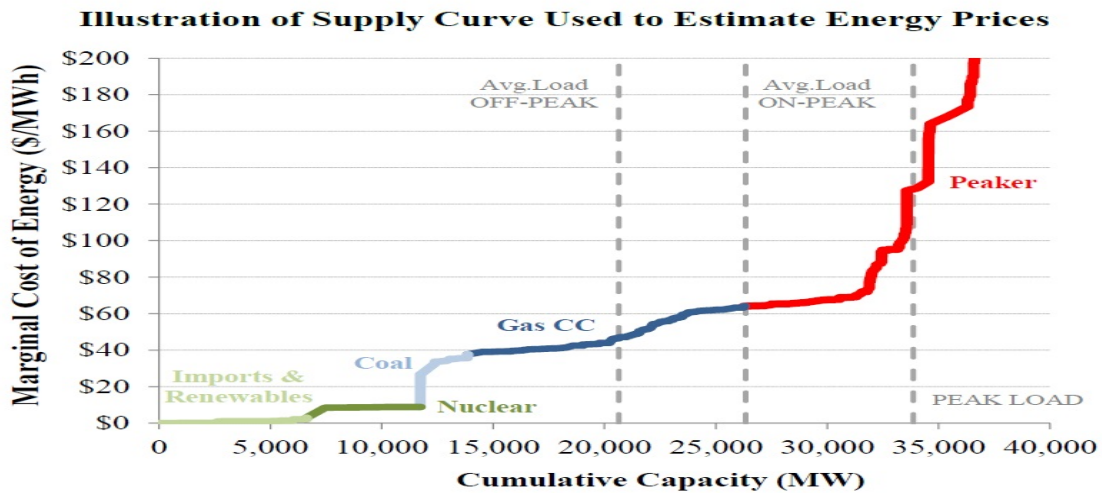
Figure-1.1: Electricity Supply Chain



TRANSPORT OF ELECTRICITY

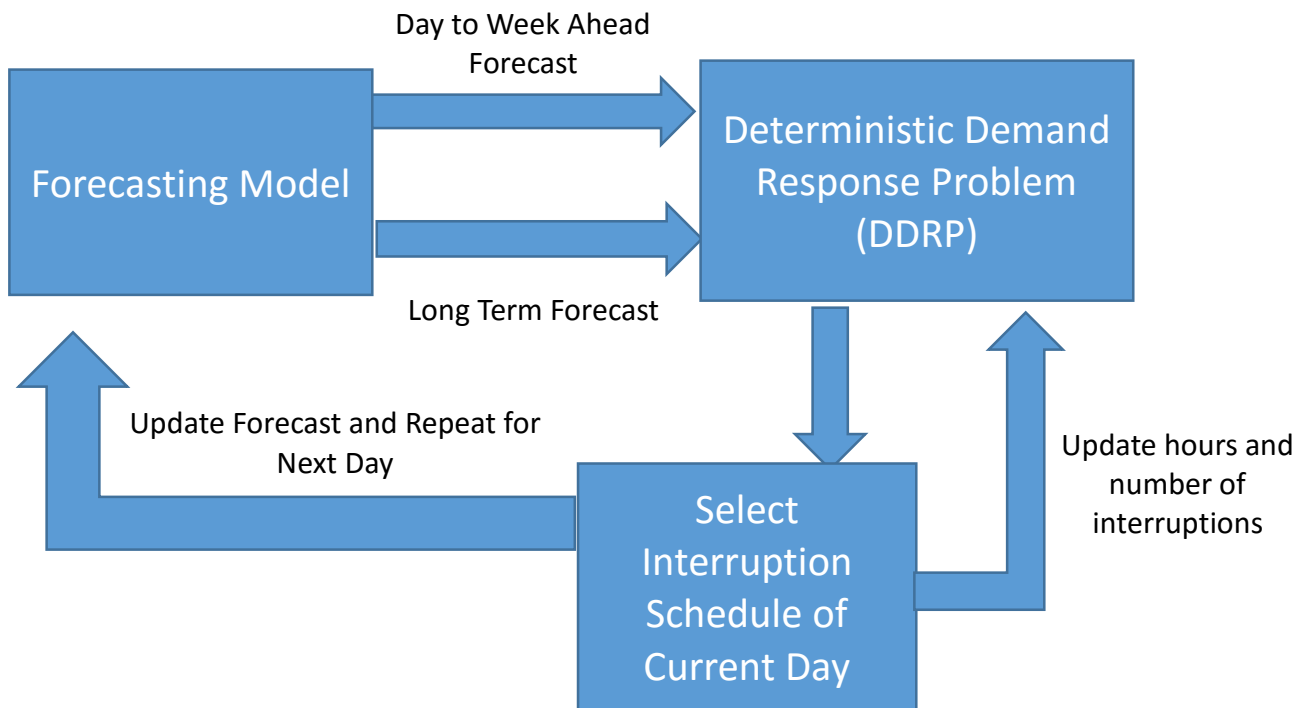
(Source: <http://eex.gov.au/>)

Figure-1.2: Electricity Supply Curve of the Texas Wholesale Market



(Source: Solar Energy Association)

Figure-1.3: Solution Approach



CHAPTER 2

Load Forecasting Model

2.1-Introduction

Load forecasting plays an important role in the operation of power systems. In the short term, the hourly load forecasts are typically determined for one day to one week ahead. Load forecasts are used by power companies to establish operational plans for power stations and their generation units as well as security studies including security analysis and load management. Another application of short term load forecasting is in optimizing the operational state of a power system in terms of load flow, power management and interchange scheduling. Long term electric load forecasting corresponds to a forecast horizon from several months to several years ahead. It is critical for scheduling the construction of new generation facilities and in the development of transmission and distribution systems. In contrast to short term load forecasting, point forecasts are not very useful as they cannot be used to account and hedge for the financial risk due demand. Instead, density forecasts are more helpful for long-term planning. Another difference between short-term and long-term demand forecasting is in their use of meteorological variables such as temperature and humidity which among the main determinants for demand. In the short term such weather information is available up to one week ahead but is unavailable for long-term forecasts. Therefore, a method to generate future temperatures is required.

In this paper, we present a combined model for short and long term electricity demand forecasting and apply them to the Southern California Edison (SCE) electric utility company in Los Angeles.

The forecasts are used for managing Demand Response Programs (DRP) adopted by most utility companies. DRP are contracts through which utility companies incentivizes customers to shift their electric demand load from peak hours to off peak hours. DRP are divided into commercial and residential programs. Interruptible programs are the most common commercial DRP. Under such programs, customers receive monthly payments and in return the utility company has the right to instruct them to reduce their electric load to a pre-specified level that each firm chooses when it enrolls. The program is run on a yearly basis and there is a limit on the number and duration of interruptions. For the short term, the goal is to determine the hourly load forecast during peak demand hours. The peak daily hours consist of the hours of the day between 9:00 am and 9:00 pm. The short term load forecast allows a company to determine the load profile during peak hours in order to decide whether to dispatch the program on that day or postpone interruption. However, to decide whether to postpone or dispatch the interruptible program, a long term load forecast should be determined given the limit that exists on the number and duration of the interruptions. Therefore, a utility company could decide to interrupt its customer on a different day based on the predictions of the long term forecast. To address the above goals, a combined model for short and long term predictions has to be developed. For the short-term, we propose a nonparametric regression model for the day-ahead forecast of hourly load during peak hours. For the long term, we first propose multivariate and nonparametric regressions for determining the expected peak load as well as nonparametric regression methods for approximating the load profile in the long profile during peak hours for the long run. Further given the absence of temperature data in the long term, we simulate year ahead daily peak temperature realizations. Since interruptible DR programs are used when its enrolled customers are operating, we develop short and long term load forecasting models for weekdays since those are the days when most enrolled customers operate.

There has been a significant body of research on load forecasting. Alfares and Nazeeruddin (2001) provide a comprehensive review of this area. Short term load forecasting methods can be classified into two broad categories: parametric method and artificial intelligence based methods. Parametric methods formulate a mathematical or statistical model of load by examining qualitative relationships between the load and load affecting factors. Some examples of the models used are: polynomial functions, ARMA models, Fourier series and multiple linear regression. The model parameters are estimated from historical data and the accuracy of the model is verified by analysis of forecast errors. The model developed is often data dependent and hence cannot be applied to another utility company. Artificial intelligence based methods use artificial neural networks as load models. The main advantage of using neural networks lies in their abilities to learn the dependencies from the historical data without selecting an appropriate model. The methodology selected for short term load forecasting in this paper is based on nonparametric regression. Our short term forecasting model is related to Olinda, Chen and Charytoniuk (1998) who develop a short term load forecast using nonparametric regression. In their model, the load forecast is a conditional expectation of load given the time, weather conditions and other explanatory variables. The forecast is calculated from historical data as a local average of observed past loads with the size of the local neighborhood and the specific weights defined by a multivariate product kernel. Another related work is in the short term load forecasting model of Vilar, Cao and Aneiros (2012) who present a nonparametric a semiparametric model for next day forecasting of electricity demand and price for the electricity market in Spain. Also, in the realm of short term load forecasting and nonparametric methods, Hyndman and Fan (2012) propose a semiparametric additive model for short term load prediction from one to seven days ahead. They further propose a forecast interval prediction by a modified bootstrapping method. In this paper we adopt a

nonparametric regression for the day-ahead load forecast in which we forecast the electric load demand during peak hours. The load affecting variables consist of the previous day hourly load and the current day hourly temperature. With respect to short term load forecasting applied to California, Weron and Nowicka-Zagrajek (2002) develop a day ahead forecast for the system wide load in California. To this end they fit an ARMA model after deseasonalizing the data and compare their forecasts to the ones obtained by CAISO.

Despite their importance for system planning, medium and long term load forecast have not received as much attention as short term load forecasting. McSharry et al (2005), propose a model that provides probabilistic forecast for both the magnitude and timing of peak load for lead times of one year for a province of the Netherlands. The daily load forecast is determined by a multivariate linear regression. Weather forecasts are determined using a simulation technique referred to as the method of surrogates (M.Small, C.K.Tse. 2002) which provides realistic realizations of different weather variables such as temperature, wind speed and luminosity by replicating both the distribution and autocorrelation of each variable and preserving the cross-correlation between those variables. Also, Hyndman and Fan (2010) provide a comprehensive methodology to forecast the density and weekly peak electricity demand up to ten years. The electric load demand is determined by a semi-parametric additive model where the half-hourly load is forecasted in function of the temperature, calendar effect as well as demographic and economic variables. Temperature simulation is done by a double season block bootstrap methodology where different temperature realizations are obtained by randomly selecting each block of temperature from a different year. Another notable work in long term load prediction is the model proposed by Morita, Kase, Tamur and Iwamoto (1996). In their paper an interval prediction model of peak annual electric demand is developed using a grey dynamic model.

Relative to these papers, our work is different in the following ways. First, we develop linear and semi parametric regressions in which the peak load is determined for a given peak temperature taking also into account calendar effects. Further, we approximate for a given peak load the long term load profile across the daily peak hours by nonparametric regression. In the long-term, the daily peak temperature is the only load affecting that we use for determining the peak load. Since temperature data is not available in the long term, we simulate temperature realizations by modeling the deterministic component of the temperature by a sine model and its stochastic component by modeling the residuals as an AR process. Also, to the best of our knowledge this work is the first to propose a combined short and long term forecasting model for the management of demand response programs. As previously mentioned, for the management of such contracts next day load forecasting is not enough as the utility company had a limited number of interruptions. Hence, long term forecast is also needed in order to decide whether it is optimal for a utility company to interrupt its customers during the same day or delay interruption.

The remainder of this chapter is organized as follows. In Section 2 we propose a day-ahead hourly load forecast. In Section 3 we determine the long term peak load and provide approximations for the hourly load profile. In Section 4 we develop a simulation model for daily peak temperatures and use the realizations to determine the probability distribution of peak loads. Finally, we summarize the main results in Section 5.

2.2-Short Term Load Forecasting Model

2.2.1-Hourly Load Forecasting

In this section, we develop a day-ahead electricity demand forecasting model based on nonparametric regression. Our model can be described by regressing the scalar dependent variable $y = L_{d,h}$ which represents the electric load on day d during peak hour h in Gigawatts (GW) on a

regressor vector $x = (L_{d-1,h}, T_{d,h})$, where $L_{d-1,h}$ represents the electric load on day $d - 1$ during peak hour h , and $T_{d,h}$ the hourly temperature in Fahrenheit during peak hour h on the current day.

The daily peak hours are from 9am to 9pm. The regression model is:

$$L_{d,h} = m(L_{d-1,h}, T_{d,h}) + \epsilon_{d,h}, i = 1, \dots, N, h = 1, \dots, H \quad (1)$$

Hence, we develop for each hour h a nonparametric regression equation where the electric demand load is determined given the previous day load and current day hourly temperature at hour h . Hours 9:00 am through 9:00pm are denoted by hours 1 through 13.

The load model in (1) can be described in terms of a multivariate probability density function $f(L_{d,h}, L_{d-1,h}, T_{d,h})$. A pdf estimator can be obtained from historical load and temperature measurements by means of nonparametric density estimation. This is achieved by estimating the entire density function directly from the sample. We compute the nonparametric expectation of next day's hourly load using a dataset of daily measurement of hourly loads and temperatures. The hourly electricity load measurements represent actual hourly electricity demand for the region served by the Southern California Edison (SCE) utility company. The electric load measurements are obtained from the California Independent System Operator (CAISO) website. On the other hand, the hourly temperatures are collected from the National Climatic Data Center. Given that the electricity load demand data is at the aggregate level, the hourly temperature is taken as the weighted average of the counties of Los Angeles and the combined counties of Orange and Riverside, with weights of 0.65 and 0.35 respectively. The temperature in Riverside is considered as the representative temperature of both Riverside and Orange counties. The weights correspond to ratio of each county's population over the total population of the three counties. The choice of these three counties was due to the fact that they are the most populated counties in the region served by SCE.

Conceptually, if there are n sets observations $o_i = (L_{i,h}, L'_{i,h}, T_{i,h})$. $i = 1, \dots, n$, where $L'_{i,h}$ is the previous day hourly load for observation i respectively, the estimated density function $\hat{f}(L_{d,h}, L_{d-1,h}, T_{d,h})$ can be calculated as follows:

$$\begin{aligned} & \hat{f}(L_{d,h}, L_{d-1,h}, T_{d,h}) \\ &= \frac{\sum_{i=1}^n \left\{ k\left(\frac{L_{d,h} - L_{i,h}}{b_L}\right) k\left(\frac{L_{d-1,h} - L'_{i,h}}{b_{L'}}\right) k\left(\frac{T_{d,h} - T_{i,h}}{b_T}\right) \right\}}{nb_L b_{L'} b_T} \quad (2) \end{aligned}$$

Where, $k(\cdot)$ is a Kernel functions and b_L , $b_{L'}$ and b_T are the bandwidths which represent smoothing parameters corresponding to the current day hourly load, the previous day hourly load and the current day hourly temperature respectively.

A Kernel function $k(\cdot)$ is a function that satisfies the following: $k(u) = k(-u)$, $\int |k(u)| du < \infty$, $\int uk(u) du = 0$, $\int k(u) du = 1$ and $k(u) \geq 0$.

Using the normal kernel, (2) will have the following form:

$$\hat{f}(L_{d,h}, L_{d-1,h}, T_{d,h}) = \frac{\sum_{i=1}^n \left\{ \exp - \left(\frac{(L_{d,h} - L_{i,h})^2}{b_L^2} + \frac{(L_{d-1,h} - L'_{i,h})^2}{b_{L'}^2} + \frac{(T_{d,h} - T_{i,h})^2}{b_T^2} \right) \right\}}{(2\pi)^{5/2} nb_L b_{L'} b_T} \quad (3)$$

Given the pdf estimator $\hat{f}(L_{d,h}, \mathbf{x})$, with $\mathbf{x} = (L_{d-1,h}, T_{d,h})$, the day-ahead peak load can be determined as the conditional expectation of the load $L_{d,h}$ given the regressor vector \mathbf{x} .

$$E(L|\mathbf{x}) = \frac{\int L \hat{f}(L, \mathbf{x}) dL}{\int \hat{f}(L, \mathbf{x}) dL} \quad (4)$$

Using the estimated pdf, the equation above can be calculated from the original historical data using the following equation:

$$\hat{E}(L_{d,h}|L_{d-1,h}, T_{d,h}) = \frac{\sum_{i=1}^n \left\{ L_{i,h} k \left(\frac{L_{d-1,h} - L'_{i,h}}{b_L} \right) k \left(\frac{T_{d,h} - T_{i,h}}{b_T} \right) \right\}}{\sum_{i=1}^n k \left(\frac{L_{d-1,h} - L'_{i,h}}{b_L} \right) k \left(\frac{T_{d,h} - T_{i,h}}{b_T} \right)} \quad (5)$$

As seen from the equation (5) above, the hourly load forecast $L_{d,h}$ is a local weighted average of observed past loads on the neighborhood of a given vector of observed previous day hourly loads and same day hourly temperature and hour of the day. The closer (farther) this vector is from the vector our forecast is computed against, the higher (lower) the weight on the observed load. Hence, provided the smoothing parameters $b = (b_L, b_{L'}, b_T)$ are available, the forecasted load can be directly calculated from past observations. The expected load conditional on a given temperature and previous hourly load is referred to as the Kernel Regression Estimator.

For the normal Kernel, the above equation takes the following form:

$$\hat{E}(L_{d,h}|L_{d-1,h}, T_{d,h}) = \frac{\sum_{i=1}^n \left\{ L_{i,h} \exp \left(-\frac{(L_{d-1,h} - L'_{i,h})^2}{2b_{L'}^2} - \frac{(T_{d,h} - T_{i,h})^2}{2b_T^2} \right) \right\}}{\sum_{i=1}^n \left\{ \exp \left(-\frac{(L_{d-1,h} - L'_{i,h})^2}{2b_{L'}^2} - \frac{(T_{d,h} - T_{i,h})^2}{2b_T^2} \right) \right\}} \quad (6)$$

The smoothing parameters can be selected by applying the cross validation technique. We denote by $\hat{L}_{-j}(x)$ the estimate of the load for day j based on the samples $x_i, (i = 1, \dots, n)$ without observation j . As a result, the optimal bandwidths (b^*, λ^*) are obtained by solving the following:

$$b^* = \operatorname{argmin} \left\{ \frac{1}{n} \sum_{i=1}^n [L_i - \hat{L}_{-j}(x_i)]^2 \right\} \quad (7)$$

Where, $\hat{L}_{-j}(x) = \frac{\sum_{i=1, i \neq j}^n L_i K\left(\frac{x-x_i}{b}\right)}{\sum_{i=1, i \neq j}^n K\left(\frac{x-x_i}{b}\right)}$, $K(\cdot)$ is the kernel product which consists of the product of kernel functions and x the vector of regressors. The optimal smoothing parameters are summarized in Table-1.1.

To account for seasonality, we separate our dataset into two datasets, one for the summer and one for the winter season. The summer season extends from the months of June until September while the winter season extends from October until May. For the winter season, the data set was further divided into two sets to account for months that fall during the daylight saving period and those that do not. The smoothing parameters were computed using the `npregbw` package using the R software. We test the prediction accuracy of our proposed model on two out of sample data sets, one for each of the winter and summer seasons for the year 2014. The accuracy of the forecasts was measured by computing the Mean Absolute Percentage Error (MAPE), which can be expressed as $MAPE = \frac{100\%}{N} * \sum_{i=1}^N \left| \frac{\hat{L}(T_i) - L(T_i)}{L(T_i)} \right|$, where N is the sample size. The MAPE results for both tests were measured to be 3.07% for the summer season and 2.56% for the winter season.

2.2.2-Analysis of the residuals

An analysis of the errors from the insample data shows evidence of serial correlation. The autocorrelation and partial correlation for the day ahead summer forecast are respectively shown in Figure-1.1. The autocorrelation plot reveals the evidence of exponential decay or a damped sine wave whereas the partial correlation plot has a large spike at lag 1 and can be considered as nonsignificant afterwards. The behavior suggest an AR(1) modeling for the errors. The same behavior is also detected for the winter dataset and are shown in Figure-2.2.

Given the previous two observations, the residual can be described by a simple autoregressive process of order 1 AR(1). As a result,

$$e_d(h + 1) = \phi e_d(h) \quad (8)$$

where $e_d(h) = L_{d,h} - \hat{L}_{d,h}$ is the forecasting residual on day d during peak hour h and $0 < \phi < 1$ is a parameter estimated from past values of e . If there are n past errors available then ϕ is the least squares solution to the system of $n - 1$ linear equations having the form (8). As a result, a load forecast made at time t for a time lead l can be calculated as:

$$\hat{L}_{d,h}(e) = \hat{L}_{d,h} + e_d(t)\phi^l \quad (9)$$

The resulting coefficients ϕ were equal to 0.88 and 0.87 for the summer and winter season respectively. The same observation was also found in the nonparametric regression model developed by Olinda et. al (1998). The second term in the above equation corrects the forecast for small time leads based on the recent error. As a result, we use the forecasting error on the same day for the hour between 9:00 am and 9:00 pm and consequently correct the forecasted update for the daily peak hours from 9:am to 9:00 pm. The resulting out of sample MAPE was reduced to 2.56% and 2.18% for the summer and winter seasons respectively. Figures 2.3 and 2.4 show a sample of the predicted days for both the summer and winter seasons respectively. From the plot forecasts we notice the difference in the hourly demand pattern which is consistent in for the summer and winter seasons. Another difference between the two figures is the timing of the peak load occurrence. For the summer season, the peak load occurs between the hours of 3 and 5pm whereas for the winter season it occurs later in the day between the hours of 6 and 9 pm.

In comparison to the CAISO forecasts, the day ahead summer season prediction for CAISO had a MAPE around 2.56% ,while the winter season day ahead forecasts had an MAPE of 1.85%.

Hence our suggested model performs relatively well in comparison to the day ahead prediction of CAISO. In fact, CAISO disposes of more disaggregated data which consists of inland and coastal electricity demand as well industrial electricity consumption. In contrast the data we used consists of aggregate electricity demand data. As a result, we expect our forecast to further improve if more disaggregated data was at our disposal.

2.2.3-Asymptotic Distribution

After determining non-parametrically the expected electric demand load during day d , at peak hour h , $L_{d,h}$, we determine the asymptotic distribution of our estimates in order to obtain a confidence interval. It can be shown that for values of hourly temperature $T_{d,h}$ and previous day load $L_{d-1,h}$, we have that

$$\begin{aligned} & \sqrt{Nh}(\hat{m}(T_{d,h}, L_{d-1,h}) - m(T_{d,h}, L_{d-1,h}) - b(T_{d,h}, L_{d-1,h})) \\ & \xrightarrow{d} N\left(0, \frac{\sigma_\epsilon^2}{f(T_{d,h}, L_{d-1,h})} \int K(z)^2 dz\right) \end{aligned} \quad (10)$$

Where $\hat{m}(T_{d,h}, L_{d-1,h}) = \hat{E}(L_{d,h} | T_{d,h}, L_{d-1,h})$.

If the bias $b(T_i)$ is ignored, the above limiting distribution yields the following 95% confidence interval for $m(T_i)$:

$$\hat{m}(T_{d,h}, L_{d-1,h}) \pm 1.96 \sqrt{\frac{1}{Nh_T h_L} \frac{\hat{\sigma}_\epsilon^2}{\hat{f}(T_{d,h}, L_{d-1,h})} \int K(z)^2 dz} \quad (11)$$

Where:

$$\hat{f}(T_i, L_i) = \frac{1}{Nh_T h_L} \sum_{j=1}^N k\left(\frac{T_j - T_i}{h_T}\right) k\left(\frac{L_j - L_i}{h_L}\right) \text{ and } \hat{\sigma}_\epsilon^2 = \frac{1}{N} \sum_{j=1}^N \left(L_j - \hat{m}(T_j, L_j)\right)^2 * \frac{K\left(\frac{T_j - T_i}{h_T}\right) K\left(\frac{L_j - L_i}{h_L}\right)}{\sum_{j=1}^N K\left(\frac{T_j - T_i}{h_T}\right) K\left(\frac{L_j - L_i}{h_L}\right)}$$

is the estimated error variance.

2.2.4-Benchmarking the model with an Artificial Neural Network

Next, we benchmark our proposed forecasting model with an artificial neural network (ANN). The artificial neural network is among the most used models in machine learning and is a flexible nonlinear regression model. An illustration of a single layer neural network is shown in Figure-2.5. The hidden units are linear combinations of the predictors that have been transformed by a sigmoidal function. The output is modeled by a linear combination of the hidden units. To this end, we fit for each hour of the day, a model averaged artificial neural network which is an aggregation of several neural networks where the day ahead electric load demand at a given hour is predicted given the previous day load and same day temperature (i.e. the same vector regressor). For each hour, the averaged artificial neural network models were fit using hourly electricity load and temperature measurements for the years 2009 through 2013 and were tested on the out of sample data of 2014. There are two tuning parameters: the number of layers and the weigh decay. The latter is a penalty on the regression coefficients in order to mitigate overfitting. For more details, we refer the reader to Khun and Johnson (2013).

2.3-Long Term Load Forecasting Model

In this section, we develop long term load forecasting models. The forecast is done in two steps. In the first step, we forecast the peak load then we approximate the load profile for each hour of the day from 9am to 9pm given the previously estimated load.

2.3.1- Peak Load Forecast

To estimate the relationship between peak demand and peak daily temperature we use a dataset of electric demand load from March 2009 to December 2013. The data consists of hourly loads for all the area covered by SCE. The first part of the long term forecasting analysis is to find a relation

between the peak load and peak temperature. Figure-2.6 shows a plot of daily peak load profile from January 2009 until December 2013.

The time series plot reveals a strong seasonality in the peak load where the peak load values in the summer are higher and more variable than the peak load values that occur during the winter season. Therefore, we divide the data into two sets, one for each season, and establish a relation between peak load and peak temperature. Figures 2.7 and 2.8 show the relation between the peak load and peak temperature for the summer and winter seasons. The peak temperature is the weighted peak temperature of the Los Angeles county and both counties of Orange and Riverside. We notice the peak load and peak temperature exhibits a linear relationship for the summer season and a nonlinear relationship for the winter season.

Motivated by the linear relation between the peak load and peak temperature, the most suitable model for peak load prediction is a multivariate regression. The best regression model was found as follows:

$$L_d = \alpha + \beta T_d + \mu_1 Jun_1 + \mu_2 Jul_2 + \mu_3 Aug_3 + \epsilon_d \quad (12)$$

Where, L_d and T_d are the peak load and peak temperature on day d , and M_1 , M_2 and M_3 binary variables representing the months of June, July and August respectively. The R^2 of the regression was equal to 0.87. The regression model was tested on the out of sample data and both generated a MAPE of 3.5%. The results of the regression are summarized in Table-2.3.

Given the nonlinear relation between the peak load and peak temperature for the winter season, we develop a semiparametric model for peak load prediction. The proposed model is as follows:

$$L_d = g(T_d) + \sum_{i=1}^7 \alpha_i M_i + \sum_{j=1}^4 \beta_j D_j + \epsilon_i, \quad (13)$$

Where, L_d is the peak load during day d , $m(T_d)$ an unknown function for the temperature effect on the load, $M_i (i = 1, \dots, 7)$ are dummy variables for the months of January through May and October and November respectively with the month of December considered as the reference month. Finally, $D_j (j = 1, \dots, 4)$ are dummy variables for the weekdays Monday through Thursday respectively with Friday taken as the reference weekday.

The above semiparametric model is a partially linear model such that the conditional mean is a composed of a usual regression function plus an unspecified linear component. Therefore:

$$E(L_d|T_d) = g(T_d) + \sum_{i=1}^7 \alpha_i E[M_i] + \sum_{j=1}^4 \beta_j E[D_j] \quad (14)$$

Following the methodology proposed by Robinson (1988), we find consistent estimators for $\alpha_i (i = 1, \dots, 7)$, $\beta_j (j = 1, \dots, 4)$ and $g(T_d)$. Given that M_i and D_j are dummy variable, $E[M_i]$ and $E[D_j]$ are computed as the proportion of months and days equal to M_i and D_j respectively.

Subtracting equations (13) and (14) yields:

$$L_d - E(L_d|T_d) = \sum_{i=1}^7 \alpha_i (M_i - E[M_i]) + \sum_{j=1}^4 \beta_j (D_j - E[D_j]) \quad (15)$$

Following the same methodology as is Section 2, we can express the expected peak load for a given peak temperature for the normal Kernel as:

$$E(L_d|T_d) = \frac{\sum_{i=1}^n \left\{ L_i \exp \left(-\frac{(T_d - T_i)^2}{2h_T^2} \right) \right\}}{\sum_{i=1}^n \exp \left(-\frac{(T_d - T_i)^2}{2h_T^2} \right)} \quad (16)$$

Where h_T is the smoothing parameter for the temperature variable. The optimal smoothing parameter is determined by cross validation and was determined to be equal to 0.98.

Replacing the conditional moment $E(L_d|T_d)$ by its nonparametric estimators, we can therefore estimate the coefficients α_i and β_j by OLS regression provided that $E(\epsilon_t|\mathcal{F}_{t-1}) = 0$, where \mathcal{F}_{t-1} has the interpretation of the information available at time $t - 1$. The OLS results are summarized in Table-2.4 below. From the table below, the signs of the coefficients seem intuitive, the calendar effect for each of the winter season months except for the month of May decreases the peak load. Further, the calendar effects of the weekdays increase the peak load relatively to Friday with Monday's effect being the smallest one probably due to the fact that Monday is the day the follows the weekend.

Since $g(T_d) = E(L_d|T_d) - \sum_{i=1}^7 \alpha_i E[M_i] + \sum_{j=1}^4 \beta_j E[D_j]$ it can consistently be estimated by

$$\hat{g}(T_i) = E(L_d|T_d) - \sum_{i=1}^7 \hat{\alpha}_i E[M_i] + \sum_{j=1}^4 \hat{\beta}_j E[D_j] \quad (17)$$

We test our method on an out of sample data set consisting of daily peak temperatures for the summer of 2011. The resulting MAPE was equal to 3.3%.

2.3.2- Long Term Hourly Load Profile Forecast

To estimate the long term load profile, we determine the expected hourly load for a given peak load. The rationale behind our analysis is from the strong correlation between the loads at each hour and the peak load. As shown in Table-2.5, this correlation is precisely very strong for the summer season while the correlation values for the winter season is still considerable but to a lesser extent than the ones for the summer season. Therefore, the peak load is a strong predictor for the hourly electric demand load.

$$L_h = m(L_p) + \epsilon_h, h = 1, \dots, H \quad (18)$$

In the above model, we estimate nonparametrically the hourly load L_h for a given estimated peak load L_p . Using the same approach as in Section 2, we can express hourly load for as:

$$E(L_h | L_p) = \frac{\sum_{i=1}^n \left\{ L_i \exp\left(-\frac{(L_p - L_{p,i})^2}{2h^2}\right) \right\}}{\sum_{i=1}^n \exp\left(-\frac{(L_p - L_{p,i})^2}{2h^2}\right)} \quad (19)$$

Where, L_i is the hourly load of observation i and $L_{p,i}$ is the peak load corresponding to observation i with h being the bandwidth estimated by cross validation. Similar to our approach in Section 2 the bandwidth values are estimated for each hour and across the summer and winter seasons. After estimating the peak load for the out of sample data of 2014, we estimate the hourly peak loads using equation (19). The MAPE for the winter and summer season where respectively equal to 4.6% and 3.9% respectively compared to the 3.2% and 6.6% seven days ahead MAPE forecast of the CAISO. Therefore, although inferior to the CAISO forecasts during the winter season, our forecasts are superior to those of the CAISO for the summer season. The discrepancy for the winter season is probably due to the low correlation of the between the hourly loads and peak loads in comparison to the summer season. The forecasting results are summarized in Table-2.6.

2.4-Temperature Modeling and Peak Load Probability Distribution

2.4.1-Temperature Modeling

Given values for temperatures, it is possible to forecast for one year ahead the expected peak load. However, values for peak daily temperature are unknown in the long run and therefore must be simulated using historical records. Therefore, using historical daily peak temperatures for the years 2011 through 2013, we simulate the daily peak temperatures by modeling it as a sum of a

deterministic component and a stochastic component with a seasonal AR model. Figure-2.9 shows the time series plot of the daily peak temperature for the time period from January 2011 to December 2013 for the counties of Los Angeles and Riverside region.

From the observed periodicities in the data, the deterministic or expected temperature component for both counties is modeled with a sum of sines model. The sine model has the following form:

$$model(D) = \alpha_1 \sin(\beta_1 * D + \delta_1) + \alpha_2 \sin(\beta_2 * D + \delta_2) \quad (20)$$

Where, D is the day number in the dataset. After computing the mean of the temperature time series, we subtract it from each point in the temperature dataset and fit the resulting data to the sine model. The estimates of the coefficients are summarized in Table-2.7.

We visualize the validity of the resulting model we plot the predicted daily peak temperatures with their actual observations. The predicted peak temperature on a given day d , $\hat{T}(d)$ is equal to

$$\hat{T}(d) = Model(d) + m \quad (21)$$

Where m is the mean peak temperature calculated from the data. The absolute error was computed to be equal to 5.3 and 6.8 degrees Fahrenheit for the counties of Los Angeles and Riverside respectively. The plots of the actual temperature and deterministic temperature component are shown in Figure-2.10.

To analyze the serial correlation of the residuals, we plot the serial and partial correlation of the residuals as shown in Figure-2.11. One of the features apparent in the below plot is that the residuals are serially correlated. This is expected as above average temperatures are likely to follow above average temperatures.

The autocorrelation function is decaying exponentially and the partial autocorrelation function has values close to 0 for lags beyond 2 days. This suggests that the residuals can be described as a simple autoregressive process of order 1, AR (2):

$$e(d + 1) = \alpha_1 e(d) + \alpha_2 e(d - 1) + \varepsilon \quad (22)$$

where $e(d) = \hat{T}(d) - T(d)$, $T(d)$ is the actual load during on day d and $\hat{T}(d)$ is the peak temperature predicted obtained from equation (15), $0 < \alpha_i < 1, (i = 1, 2)$ are parameters estimated from past values of e by least-squares solutions using equation (22) and ε are the residuals of the regression equation. The estimated value of α_1 and α_2 for Los Angeles and Riverside are respectively equal to (0.88, -0.26) and (0.94, -0.31).

As shown from Figure-2.12 the residuals ε from the correlation above are serially uncorrelated. Since the residuals are mostly uncorrelated, they can be modeled as independent draws from an appropriate distribution. A t-location-scale distribution can be shown to provide a good fit.

2.4.2-Peak Load Probability Distribution

Using the generated simulations, we can determine the probability distribution of the peak loads. To this end, we nonparametrically construct the density of the peak load by simulating 1000 temperature paths and estimating subsequently the peak loads using the multivariate and semiparametric regressions for the summer and winter seasons respectively. The kernel density estimator for a peak load L is:

$$\hat{f}(L) = \frac{1}{Nh} \sum_{i=1}^N k\left(\frac{L_i - L}{h}\right) \quad (23)$$

Where $k(\frac{L_i-L}{h})$ is the normal kernel and h is the smoothing parameter. For the normal Kernel, Scott's plug-in estimate is $h = N^{-1/(5)}s$, where s is the sample standard deviation of the simulated peak loads and N the size of the simulated peak loads sample. Using equation (23) we can therefore estimate the probability distribution of the peak load.

Figure-2.14 illustrates the resulting peak load probability distribution. We notice from the figure below how narrow the density distribution of the peak load is for load values less than 1.5 GW and the wider density for loads above 1.5 GW. This observation corroborates our data values where for the winter season most daily peak loads fall within the range of 1.2 and 1.5 GW whereas for the summer season they fall within a wider range of 1.5 and 2.2 GW. Given that temperature values beyond a week are not accurate, the demand at a time horizon exceeding a week is unknown and therefore we replace the hourly electricity load demand by an expected value. The expected value for the hourly load is computed for every month by day type. We define 9 day types depending on the range in which the peak load of a given day falls into as shown in Table-2.8. Using the simulated temperature paths we compute the peak load for each day and construct for each month the empirical distribution of the peak load from which we determine the probability distribution of day types. The probability distribution of day types is shown in Table-2.9. We then nonparametrically estimate the expected load profile per day type for the corresponding day load of that day type. The peak load consists of the mean of peak load range in which the day type falls into. The load profile approximations are shown in Table-2.10.

2.5-Conclusion

In this paper, we proposed a combined load forecasting model for demand response programs. We presented forecasting methods for both short and long term load forecasting. In the short term, we

first determine the day-ahead hourly electricity demand and then update our forecast at the start of a given day. We model the short term load forecasting by a nonparametric regression. In comparison to the forecasts of the CAISO, it turned out that our short term forecasting model performs quite well despite of the larger amount of information at the disposal of CAISO. For the long run, we first estimate the peak load and then estimate the long term hourly demand by nonparametric regression with the predicted peak load as regressor. Peak temperature and calendars effects are used as regressors to determine the peak load. For the summer season, a multivariate regression model was used whereas a semiparametric regression model was used for the winter season. Our results showed that for the summer season we outperform CAISO's long term load forecasting results. Finally, we propose a new simulation methodology to generate potential future temperature paths from which we determine the peak load density by kernel density estimation.

There exist some areas of possible future improvement. For the short term load regression our mode was calibrated using data from the period of 2009 up to 2013. It is reasonable to expect that economic conditions will not remain constant. As a result, it would be desirable to take into account such changes in both the short and long term load forecasting models. Another area for development is to increase the number of temperature sites. For Los Angeles' geographically concentrated population two temperature sites may be enough but for other regions more temperature sites might be needed.

Figure-2.1: Autocorrelation (top) and Partial Autocorrelation (bottom) of the Residuals of the Insample Summer Data

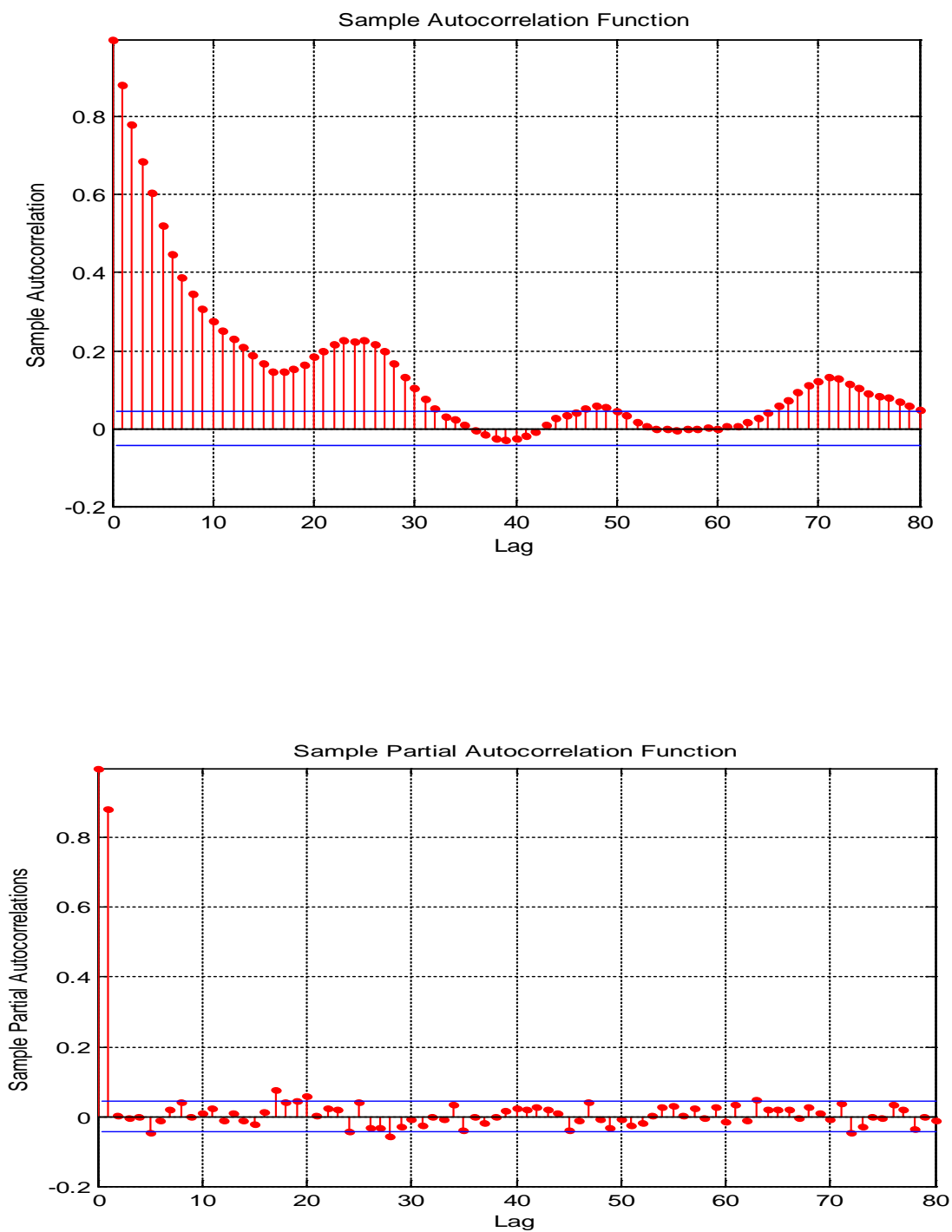


Figure-2.2: Autocorrelation (top) and Partial Autocorrelation (bottom) of the Residuals of the Insample Winter Data

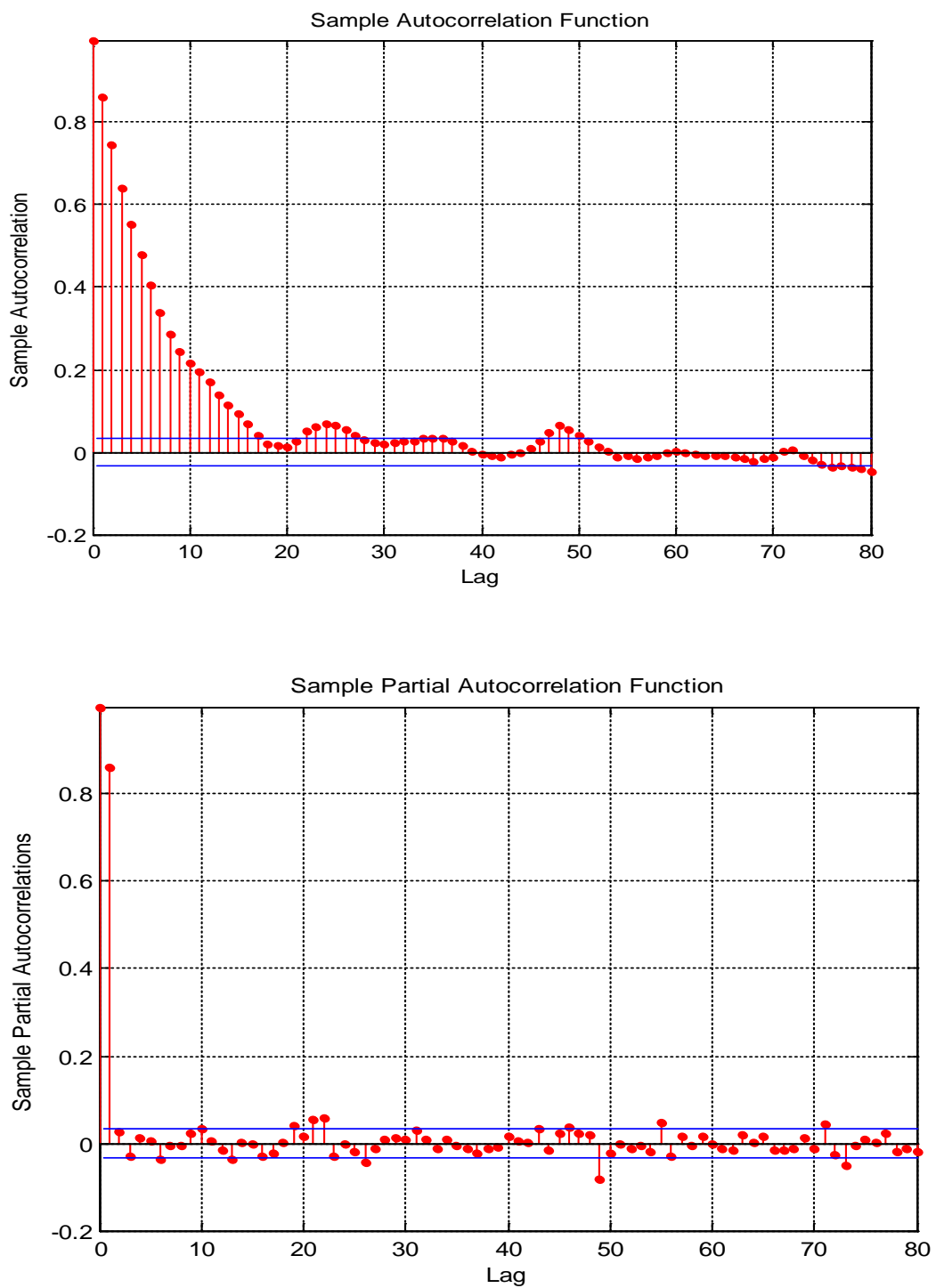
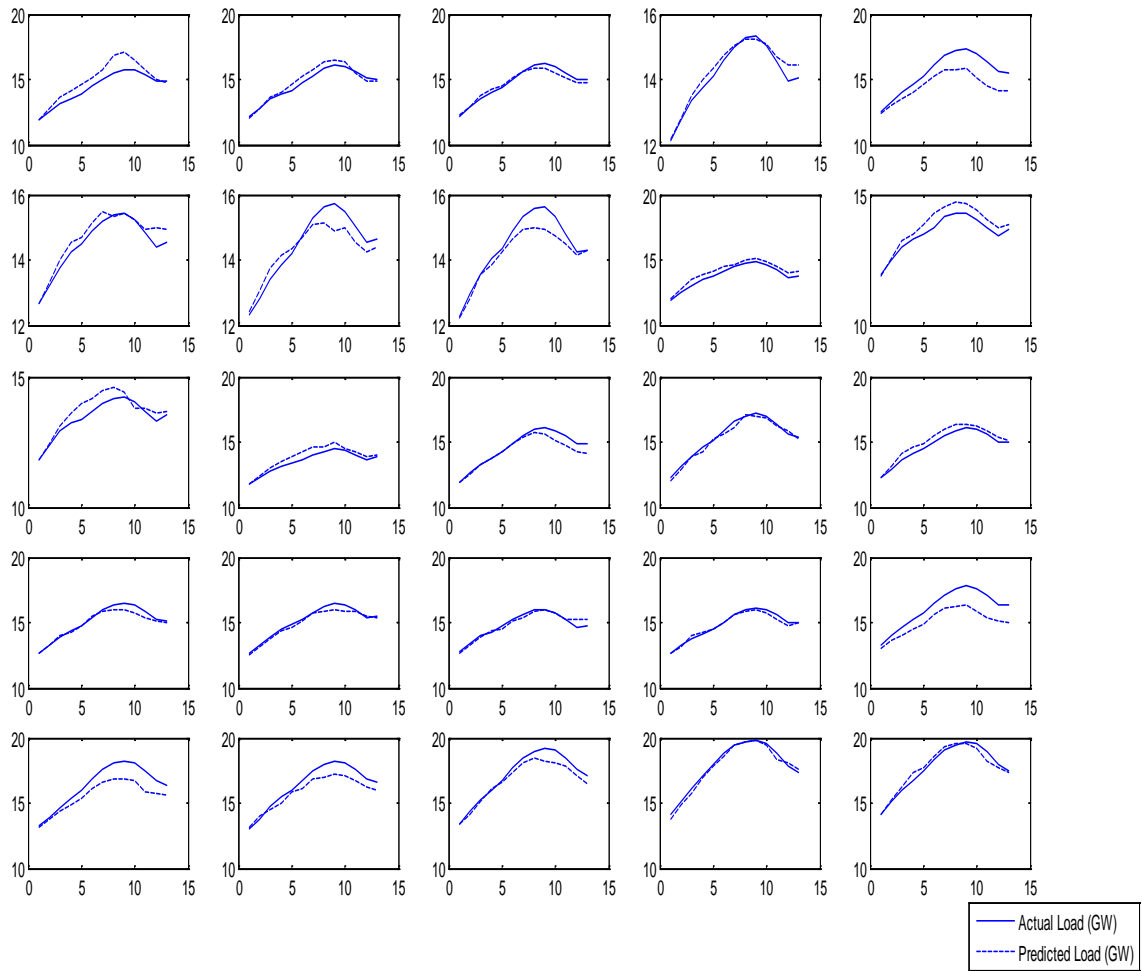


Figure-2.3: Forecasted Load versus Actual Load for Summer Season



The above figure represents a sample of the actual versus predicted hourly load profile from your out of sample data. The solid line represents the actual hourly load while the dotted lines are the forecasted ones. The y axis represents the electricity demand in Gigawatts (GW) and while the x axis represents the peak hours from 9:00 am (0) to 9:00 pm (13).

Figure-2.4: Forecasted Load versus Actual Load for Winter Season

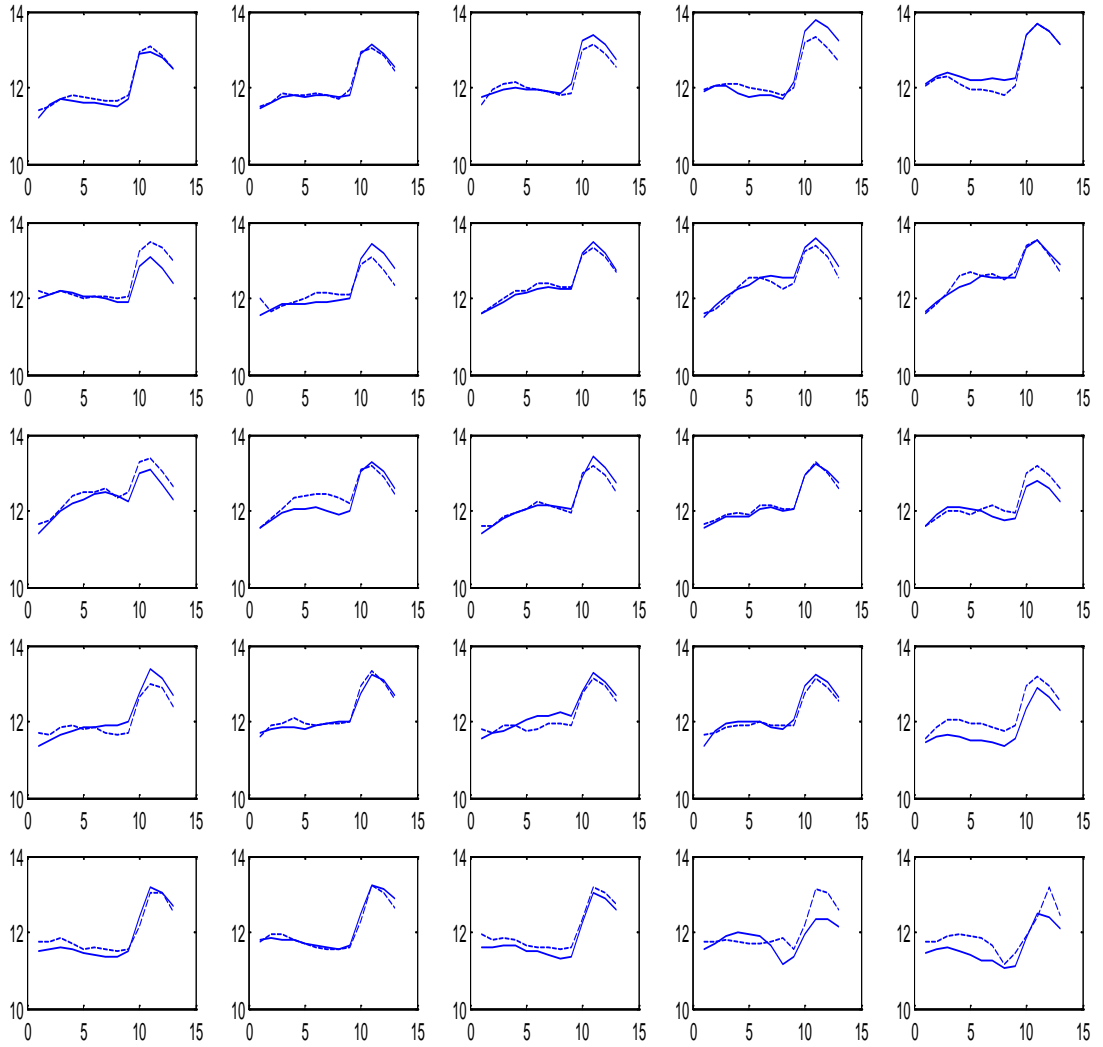


Figure-2.5: Illustration of a Neural Network with a Single Layer

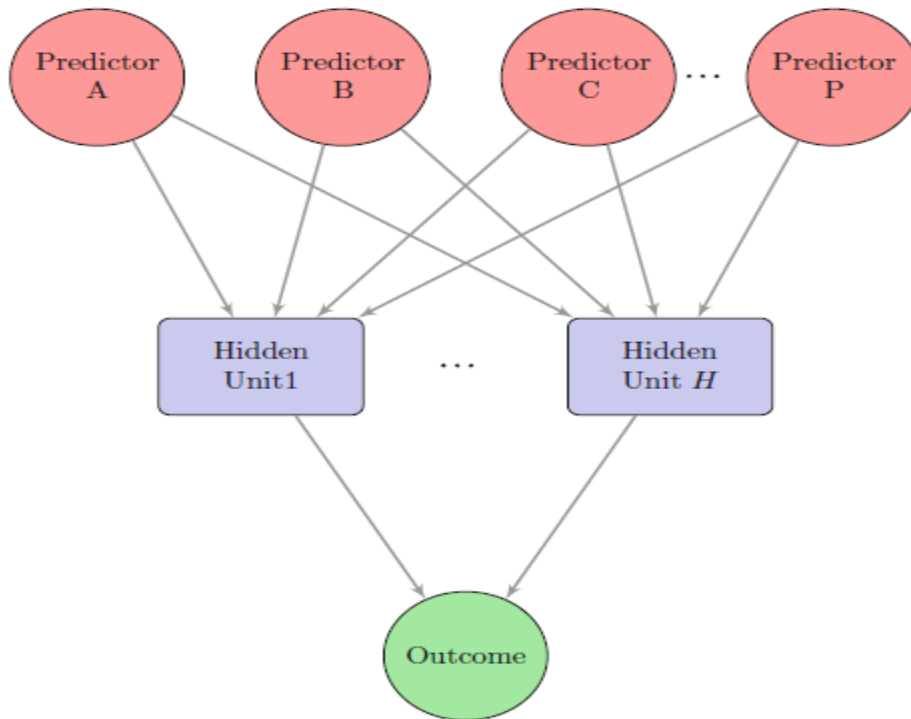


Figure-2.6: Peak Load Time Series

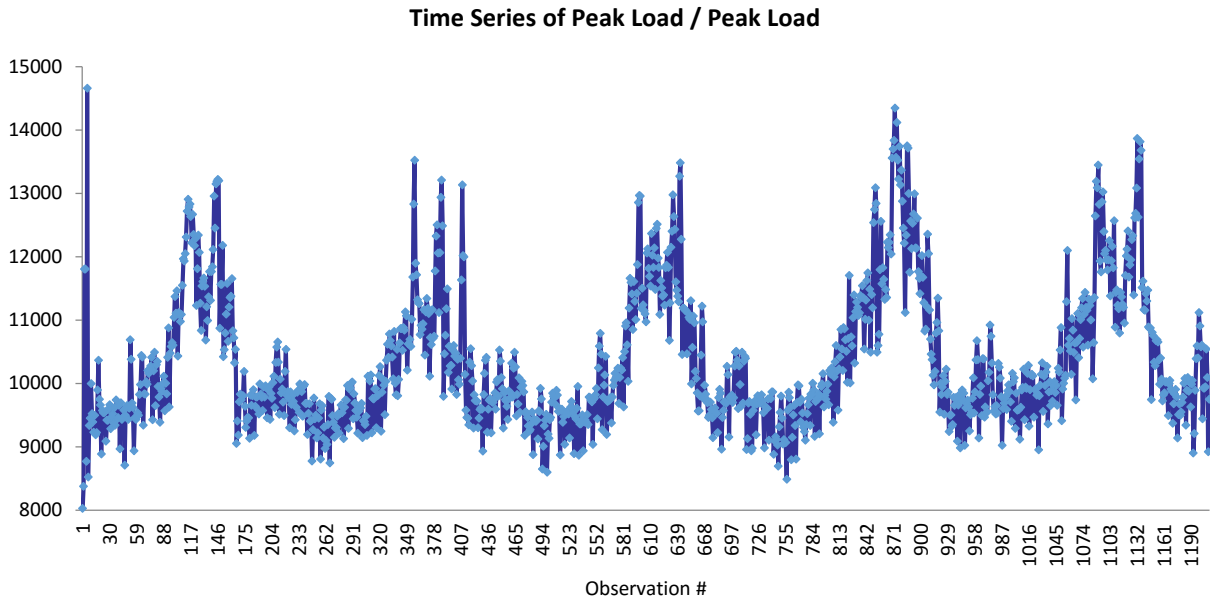


Figure-2.7: Peak Load versus Peak Temperature for Summer Season

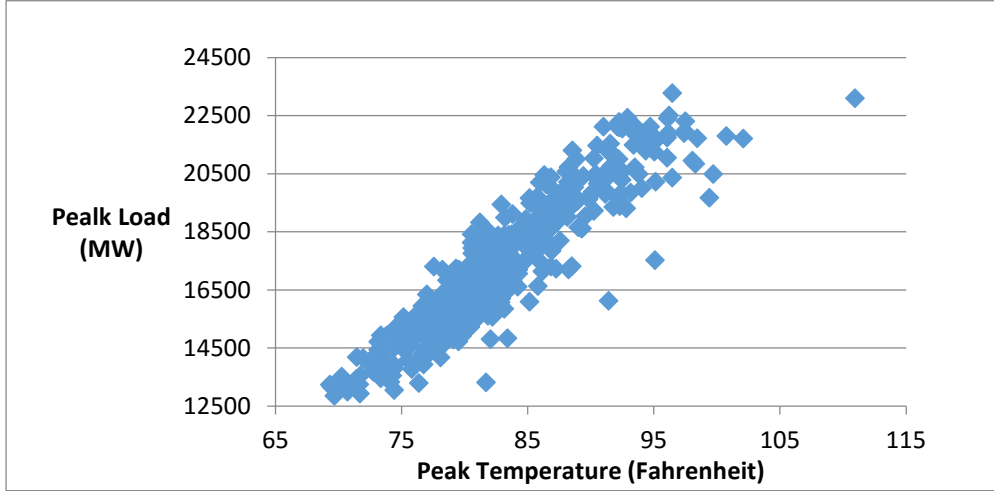


Figure-2.8: Peak Load versus Peak Temperature for Winter Season

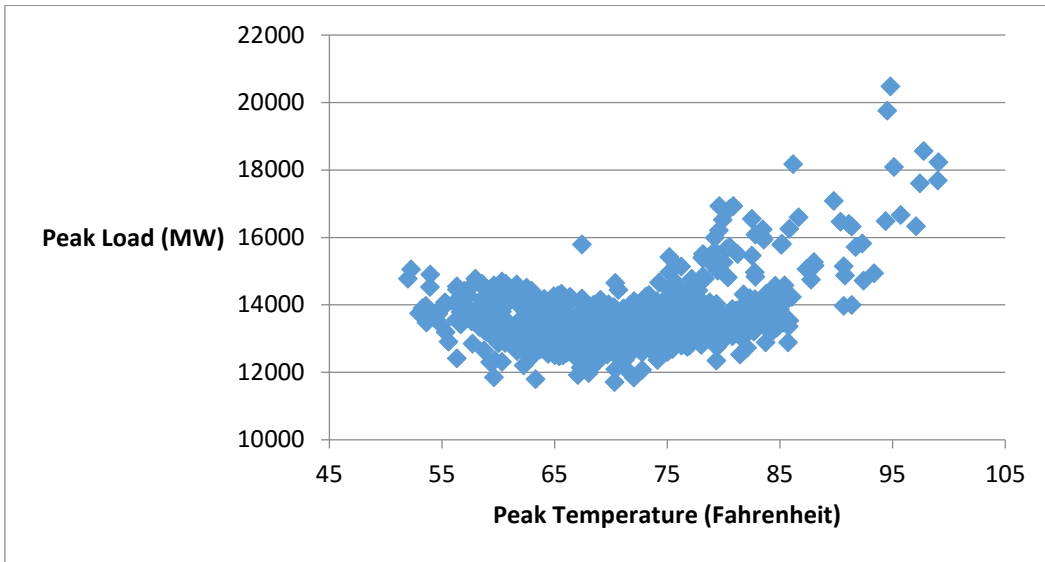


Figure-2.9: Peak Temperatures from 2011 to 2013 for the Counties of Los Angeles (top) and Riverside (bottom)

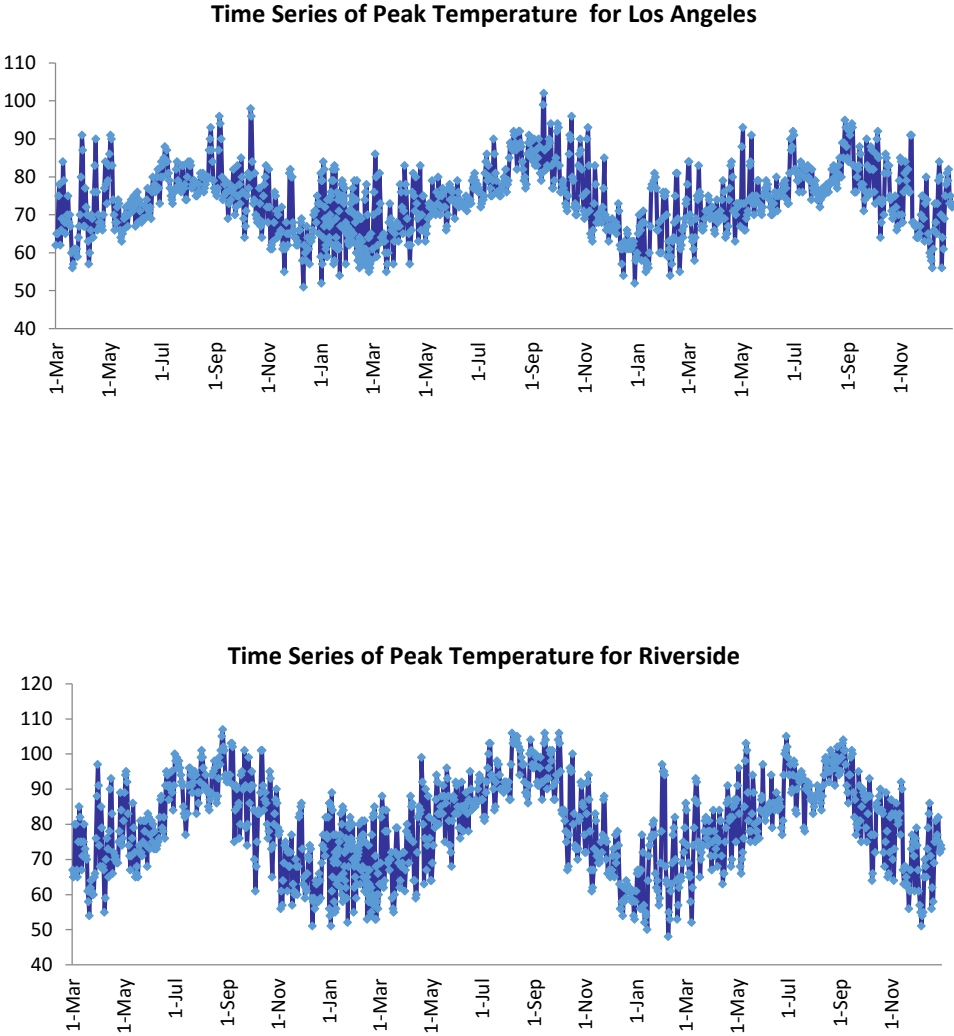


Figure-2.10: Actual and Deterministic Temperature Plots of Los Angeles (top) and Riverside (bottom)

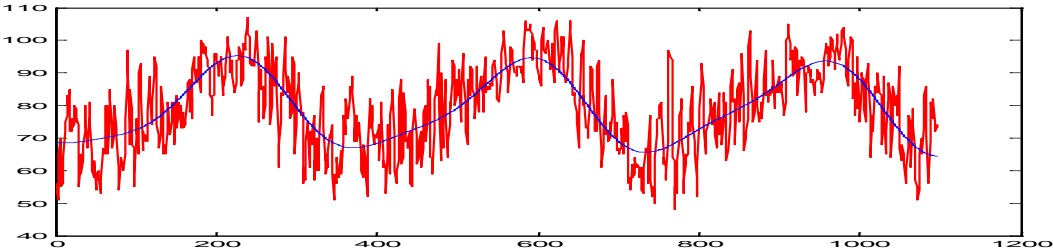
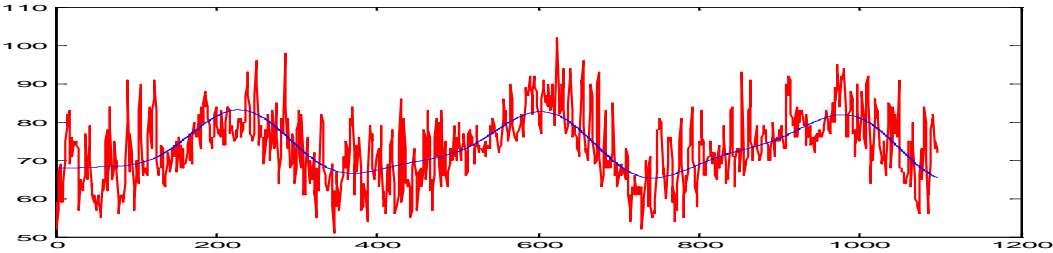


Figure-2.11: Serial Correlation and Sample Partial Autocorrelation Plots for Los Angeles (top) and Riverside (bottom)

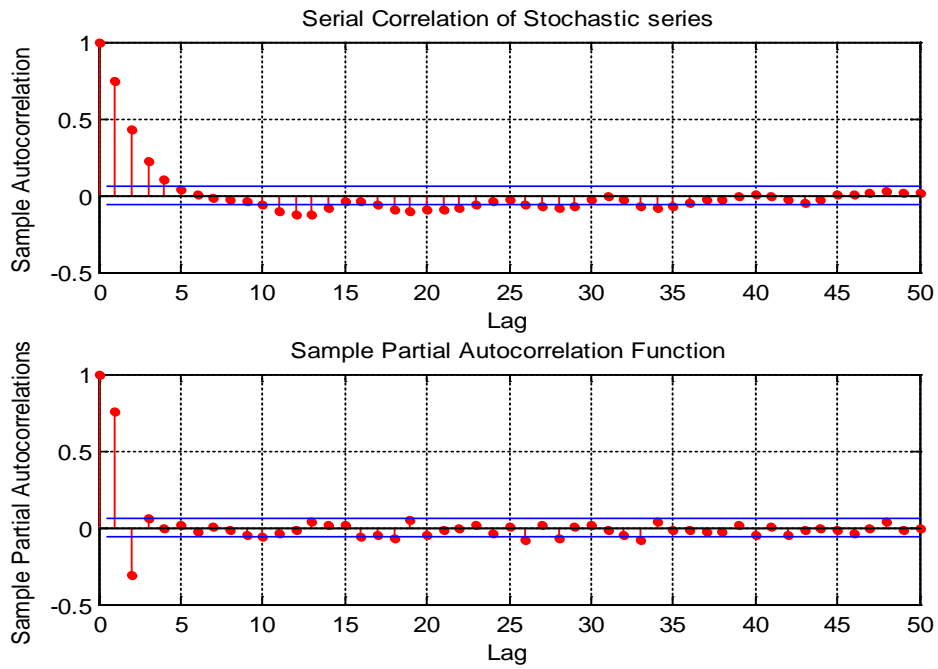
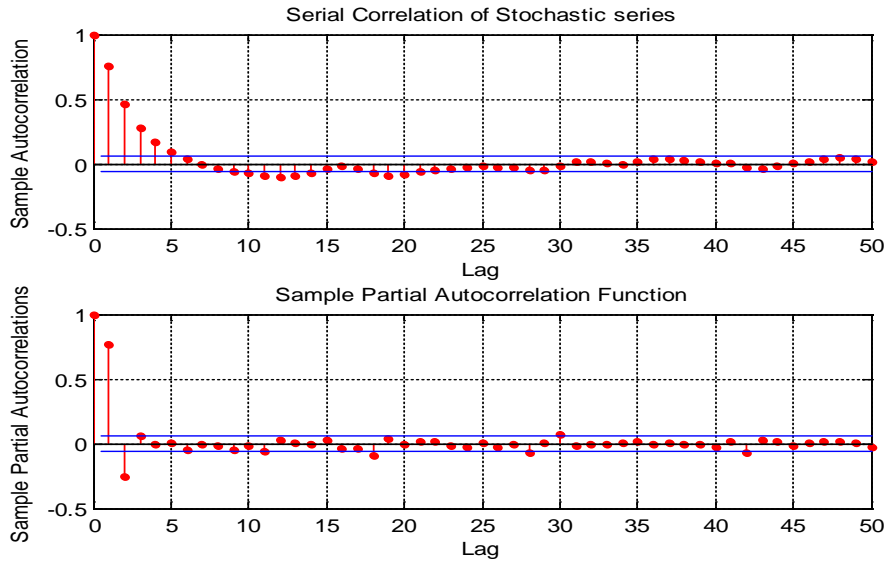


Figure-2.12: Regression Residuals and their Serial Correlation for Los Angeles (left) and Riverside (right)

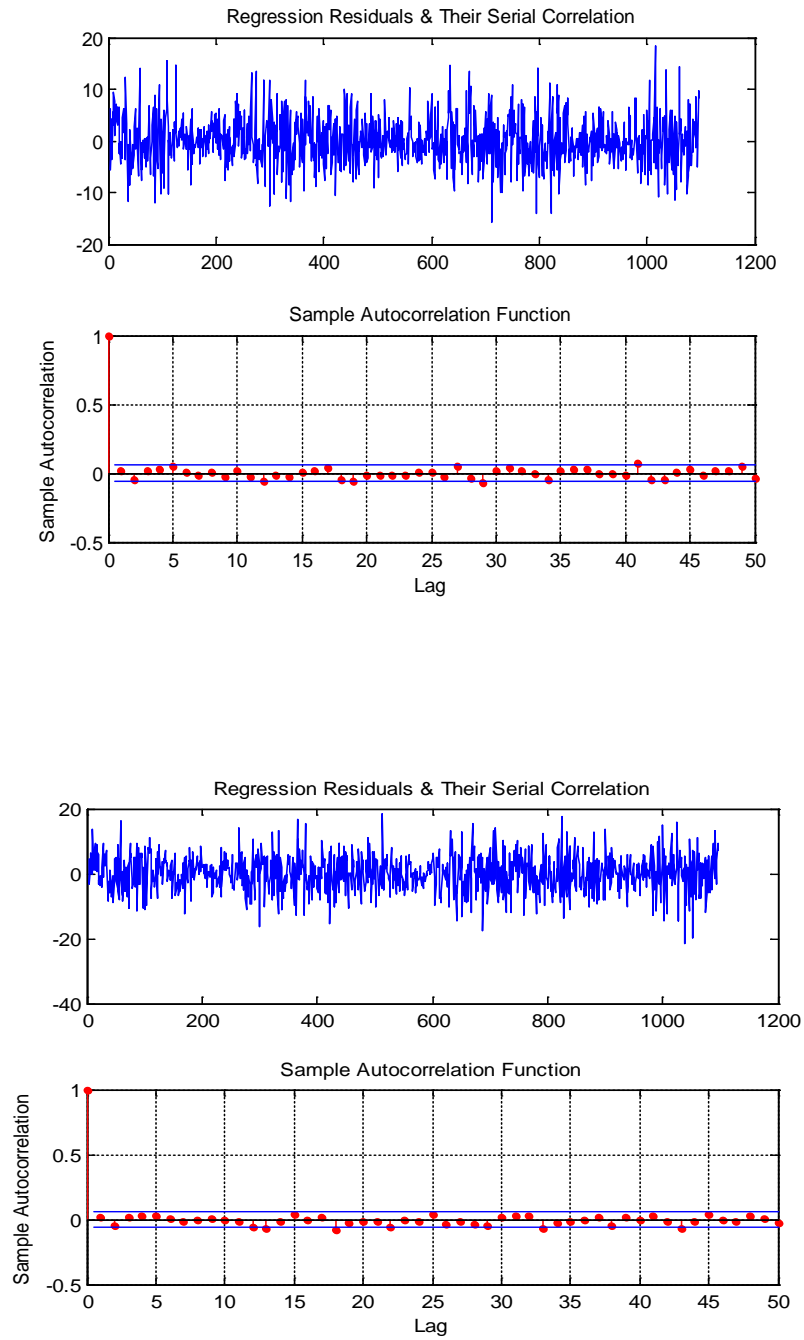


Figure-2.13: Original Temperature Time Series for 2010 and Four Simulations of Temperature Realizations

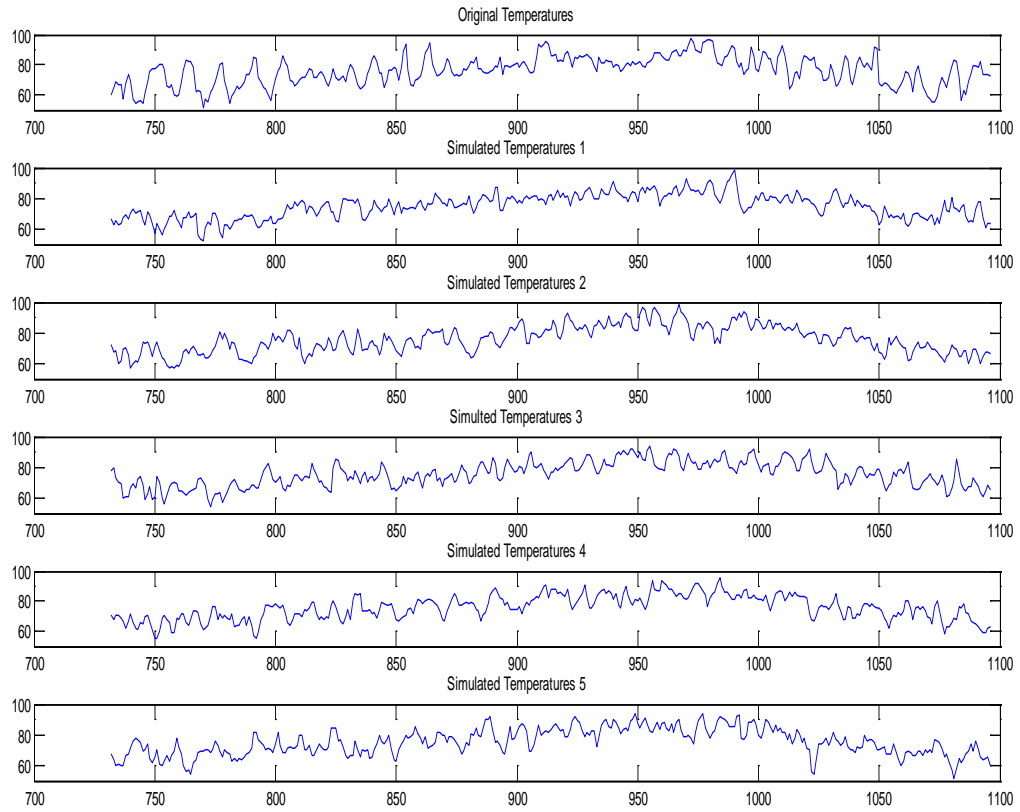


Figure-2.14: Peak Load Density

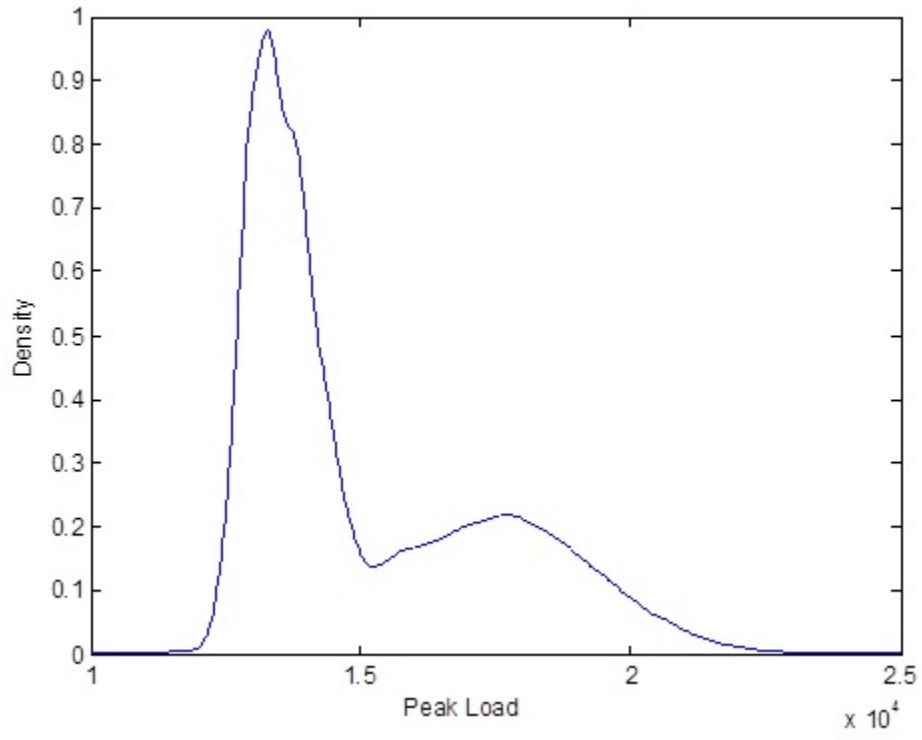


Table-2.1: Smoothing Parameters (DL: Day Light Saving)

Season	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13
T-Summer	1.16	2.10	2.19	2.01	1.65	1.48	1.40	1.10	1.30	1.44	1.21	1.33	1.36
L-Summer	0.29	0.33	0.41	0.31	0.56	0.63	0.66	0.65	0.70	0.74	0.49	0.50	0.41
T-WinterDL	3.27	2.37	2.09	2.55	1.53	1.72	2.07	1.87	2.41	0.57	0.79	2.23	2.85
L-WinterDL	0.15	0.15	0.23	0.31	0.36	0.47	0.58	0.75	0.85	0.59	0.53	0.59	0.48
T-Winter	2.42	3.54	4.5	1.7	3.07	2.54	2.46	2.15	2.27	2.85	2.35	2.00	2.01
L-Winter	0.13	0.21	0.1	0.17	0.08	0.16	0.23	0.28	0.24	0.19	0.13	0.14	0.19

Table-2.2: Short Term Forecasting Model Comparison

Model	Summer MAPE	Winter MAPE
Static Forecast	3.07%	2.56%
Neural Network	3.07%	2.64%
Dynamic Forecast	2.56%	2.18%
CAISO	2.57%	1.85%

Table-2.3: Multivariate Regression Results

Dependent variable: Peak Load

β	309.791*** (7.380)
μ_1	-241.985* (135.194)
μ_2	671.734*** (121.877)
μ_3	729.076*** (119.694)
α	-8,621.753*** (635.405)

Observations	421
R2	0.870
Adjusted R2	0.869
Residual Std. Error	871.578 (df = 416)
F Statistic	696.582*** (df = 4; 416)

=====
Note: *p<0.1; **p<0.05; ***p<0.01

Table-2.4: OLS Regression Results.

Dependent variable: $L_d - E(L_d|T_d)$

α_1	-468.566*** (87.023)
α_2	-696.907*** (92.130)
α_3	-959.118*** (84.073)
α_4	-902.082*** (83.294)
α_5	137.623 (84.468)
α_6	-239.011*** (86.601)
α_7	-444.735*** (87.542)
α_8	339.386*** (70.695)
α_9	425.731*** (68.429)
α_{10}	395.351*** (68.307)
α_{11}	362.436*** (68.937)
Observations	780
R2	0.323
Adjusted R2	0.313
F Statistic	33.371*** (df = 11; 769)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table-2.5: Covariance of hourly load and peak loads

Season	Hour 1	Hour 2	Hour 3	Hour 4	Hour 5	Hour 6	Hour 7	Hour 8	Hour 9	Hour 10	Hour 11	Hour 12	Hour 13
Summer	0.861	0.899	0.934	0.962	0.977	0.988	0.995	0.998	0.997	0.990	0.985	0.979	0.975
Winter	0.620	0.753	0.805	0.820	0.800	0.823	0.832	0.855	0.932	0.921	0.876	0.921	0.878

Table-2.6: Long Term Forecasting Model Comparison

Model	Summer (MAPE)	Winter (MAPE)
CAISO Peak Load (7DA)	6.1%	3.1%
Peak Load Prediction Model	3.5%	3.2%
CAISO Hourly Load (7DA)	6.6%	3.3%
Hourly Load Prediction Model	3.9%	4.6%

Table-2.7: Sine Model Parameters Estimates

Coefficient	α_1	β_1	δ_1	α_2	β_2	δ_2
Estimate(Los Angeles)	7.55	0.017	-2.22	2.27	0.033	0.20
Estimate(Riverside)	13.70	0.017	-2.41	3.16	0.034	-0.065

Table-2.8: Day Types

Day Type	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Type 7	Type 8	Type 9
Peak Load Range(GW)	[12-13]	[13-14]	[14-15]	[15-16]	[16-17]	[17,18]	[18,19]	[19,20]	[20,22]

Table-2.9: Monthly Probability Distribution of Day Types

Month	Type 1 [12,13]	Type 2 [13,14]	Type 3 [14,15]	Type 4 [15,16]	Type 5 [16,17]	Type 6 [17,18]	Type 7 [18,19]	Type 8 [19,20]	Type 9 [20,22]
January	0.14	0.84	0.02	0	0	0	0	0	0
February	0.2	0.8	0	0	0	0	0	0	0
March	0.74	0.26	0	0	0	0	0	0	0
April	0.55	0.45	0	0	0	0	0	0	0
May	0.36	0.61	0.03	0	0	0	0	0	0
June	0	0.11	0.21	0.26	0.22	0.14	0.06	0	0
July	0	0	0.04	0.1	0.18	0.24	0.22	0.13	0.08
August	0	0	0	0.04	0.11	0.2	0.25	0.24	0.16
September	0	0	0.05	0.11	0.21	0.24	0.21	0.12	0.06
October	0	0.43	0.48	0.09	0	0	0	0	0
November	0.1	0.79	0.11	0	0	0	0	0	0
December	0	0.83	0.17	0	0	0	0	0	0

Table-2.10: Day Type Load Profile Approximation

Day Type	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13
Type 1	11,686	11,961	12,162	12,191	12,140	12,123	12,009	11,841	11,614	11,568	11,732	12,437	12,405
Type 2	11,845	12,132	12,340	12,392	12,371	12,375	12,289	12,152	12,031	12,217	12,477	12,793	12,667
Type 3	12,511	12,889	13,187	13,328	13,383	13,506	13,537	13,570	13,865	14,098	13,936	13,857	13,702
Type 4	12,706	13,337	13,934	14,319	14,604	14,971	15,264	15,437	15,373	15,009	14,544	14,394	14,516
Type 5	12,984	13,734	14,433	14,916	15,302	15,809	16,182	16,439	16,395	15,925	15,425	15,184	15,052
Type 6	13,358	14,156	14,934	15,506	16,003	16,592	17,077	17,382	17,370	16,885	16,186	15,912	15,746
Type 7	13,645	14,523	15,424	16,118	16,727	17,425	17,947	18,345	18,374	17,906	17,197	16,678	16,464
Type 8	14,125	15,172	16,186	16,981	17,730	18,508	19,115	19,453	19,393	18,771	17,864	17,357	17,010
Type 9	14,911	16,129	17,327	18,355	19,265	20,125	20,720	20,969	20,834	20,218	19,361	18,847	18,561

CHAPTER 3

Certainty Equivalence Algorithm

3.1-Introduction

We consider the problem of a producer and provider of electricity that offers interruptible contracts to its industrial and commercial customers. The interruptible contracts that we consider are the ones offered by the Southern California Edison utility company. These programs consist of the Base Interruptible Program (BIP) and the Agricultural and Pumping Interruptible Program (API). Each program has multiple customers that are enrolled in. Customers are divided into groups. Each group consists of a set of customers enrolled in the same program that are close geographically to each other. Each enrolled customer can be interrupted at most once per day for duration of at most six hours. For the customer enrolled in the BIP program, there is a limit of 180 hours of total interruption per year while customers enrolled in the API program can be interrupted at most 25 times per year. The objective of the utility company is to minimize its electricity generation costs by deciding on a given day, in the case of an interruption which group to call as well as the timing and duration of the interruption for each of the interrupted groups. As previously discussed, the main difficulties in solving the dynamic program formulation of this problem stem from its dimensionality due to the large state space as well as the uncertainty in the electricity load demand. To address this problem, we solve the problem using a certainty equivalence approach. Using our load forecasting model results we solve on a daily basis a deterministic program which we refer to as the Deterministic Demand Response Program (DDRP). In the DDRP we model the problem of customer selection and timing of interruption as a nonlinear binary problem. We show that DDRP is NP-Hard. In order to solve this problem, we propose a heuristic that consists of first solving a nonlinear convex master problem and then a set of subproblems that determines the selection and

interruption of each group. To assess the quality of our heuristic we establish a lower bound to the DDRP by formulating a nonlinear convex problem. Our results show that our optimality gap is around 1%, 2% and 5% for small, midsize and large instances. We then use this heuristic in a certainty equivalence context in order to find a near optimal solution for our problem and test our certainty equivalence algorithm using data for the year 2014 and compare its performance with the lower bound established for the DDRP. There has been a large body of research in the literature on interruptible load contracts. The literature is divided into two broad categories: designing and pricing interruptible contracts or optimal execution of interruptible contracts. Our work falls in the latter category and is related to the work of Baldick et al (2006) and Goyal et al (2013). In their work, Baldick et al, consider the problem of a retailer with interruptible electricity contracts. In their work, they solve the problem of valuing these contracts as well as determining the optimal interruption strategy to maximize a retailer's profit by using stochastic dynamic programming and implementing a structural model to determine electricity prices, Goyal et al study the problem of a retailer with a number of interruptible contracts. The retailer's objective is to determine the near optimal interruption policy to minimize the expected l_β -norm of the observed load deviations from given thresholds and the contract execution cost over the planning horizon. Their algorithm is based on the Sample Average Approximation (SAA) dynamic program and the author also provide a sample complexity bound on the number of demand samples to be generated in order to obtain a $(1+\epsilon)$ approximate policy over the planning horizon. Compared to the papers cited above, our work is different in the following ways. First, while Baldick et al and Goyal et al solve the optimal execution problem using stochastic dynamic programming and SAA, our solution methodology is based on certainty equivalence which we implement by using a load forecasting model and a heuristic that solves the problem on a daily basis. Further, in their suggested algorithm,

Baldick et Al determine the total load to be used from each of the program they consider while Goyal et al determine the number of contracts to be used each day. In comparison, our approach allows us to determine which groups of customers to call in a given day as well as the timing and duration of interruption for each group. Given that we develop a heuristic for the DDRP, our paper also falls in the category of nonlinear binary programming. For a comprehensive review, the reader is referred to Cooper (1981). In this work we present a new methodology for solving such a problem by first clustering the groups of customers and then solving a master problem that determines the number of clusters to be called each day and time. In the second stage, we assign the groups to each day at each hour and time. The rest of this work is divided as follows: The article is organized as follows: in section 2, we present the deterministic formulation of our problem, in section 3 we discuss the corresponding nonlinear program relaxation to obtain a lower bound for the DDRP while in section 4 we present the heuristic. In section 5, we present results from our numerical study. In section 6 we present the certainty equivalence approach and show its performance through a numerical example. Finally, in section 8 we summarize our findings and provide future research directions.

3.2-The DDRP Model

The problem of group selection, timing as well as the duration of interruption for each group faced by the utility company in order to minimize its electricity generation costs can be expressed by the following nonlinear binary integer program.

$$\min_x \sum_{d \in D} \sum_{t \in T} a \left(u_{d,t} - \sum_{g \in G} \sum_{t' \in T'} \sum_{h \in H} I(t, t', h) l_g X_{d,g,t',h} \right)^2 + b \left(u_{d,t} - \sum_{g \in G} \sum_{t' \in T'} \sum_{h \in H} I(t, t', h) l_g X_{d,g,t',h} \right) \quad (5)$$

Subject to:

$$\sum_{t' \in T'} \sum_{h \in H} X_{d,g,t',h} \leq 1 \quad \forall d \in D, g \in G \quad (6)$$

$$\sum_{d \in D} \sum_{t' \in T'} \sum_{h \in H} h X_{d,g,t',h} \leq 180 \quad \forall g_1 \in G_1 \quad (7)$$

$$\sum_{d \in D} \sum_{t' \in T'} \sum_{h \in H} X_{d,g,t',h} \leq 25 \quad \forall g_2 \in G_2 \quad (8)$$

$$X_{d,g,t',h} \in \{0,1\} \quad \forall d \in D, g \in G, t' \in T', h \in H \quad (9)$$

The objective function consists of minimizing the hourly costs of electricity generations. Constraint (6) ensures that each group can be called at most once a day for a maximum of 6 consecutive hours. The left hand side of Constraint (7) imposes an upper bound on the yearly number of hours that a group in the BIP program can be interrupted. Constraint (8) limits the number of yearly interruptions of groups in the API program to 25. Finally, Constraint (9) represents the binary constraints.

Observe that the DDRP is a convex binary program since it consists of a convex objective function and linear constraints. By noting that the 0-1 Knapsack problem is NPC, it can be shown that DDRP is also NPC. Thus it is unlikely that real sized instances can be solved to optimality. We verify this in our computational results. Consequently, it is desirable to develop a heuristic to address this problem. The quality of the heuristic can be assessed by comparing it to a lower bound which we establish in the next section.

3.3-Aggregation and Lower Bound

Using the fact that the DDRP is convex, we formulate a lower bound by forming an aggregate nonlinear program. Denoting by $Y_{d,t',h}$ and $Z_{d,t',h}$ the amount of load reduction to be used on day d from time t' for h consecutive hours and let we define the Aggregate Load Reduction Problem (ALRP) problem as follows:

ALRP

$$\min_{W,Y,Z} \sum_{d \in D} \sum_{t \in T} \{aW_{d,t}^2 + bW_{dt}\} \quad (10)$$

Subject to:

$$\sum_{t' \in T'} \sum_{h \in H} I(t, t', h)(Y_{d,t',h} + Z_{d,t',h}) + W_{d,t} = u_{d,t} \quad \forall d \in D, t \in T \quad (11)$$

$$\sum_{t' \in T'} \sum_{h \in H} Y_{d,t',h} \leq \sum_{g_1 \in G_1} l_{g_1} \quad \forall d \in D, g_1 \in G_1 \quad (12)$$

$$\sum_{t' \in T'} \sum_{h \in H} Z_{d,t',h} \leq \sum_{g_2 \in G_2} l_{g_2} \quad \forall d \in D, g_2 \in G_2 \quad (13)$$

$$\sum_{d \in D} \sum_{t' \in T'} \sum_{h \in H} hY_{d,t',h} \leq 180 \sum_{g_1 \in G_1} l_{g_1} \quad \forall g_1 \in G_1 \quad (14)$$

$$\sum_{d \in D} \sum_{t' \in T'} \sum_{h \in H} Z_{d,t',h} \leq 25 \sum_{g_2 \in G_2} l_{g_2} \quad \forall g_2 \in G_2 \quad (15)$$

$$Y_{d,t',h}, Z_{d,t',h}, W_{d,t} \geq 0 \quad \forall d \in D, t' \in T', h \in H \quad (16)$$

Similar to the DDRP, the objective function consists of minimizing the total hourly electricity generation costs where the variable $W_{d,t}$ is defined in Constraint (11). Constraints (12) and (13) ensure that the total daily amount of load reduction from groups in the BIP and API programs respectively is no greater than the total load reduction of these groups. Constraint (14) imposes an upper bound on the yearly hourly weighted load reduction from the BIP program while Constraint (15) imposes an upper bound on the yearly load reduction from the API program. Finally, Constraint (16) represents the nonnegativity constraints,

Proposition 3.3.1: The ALRP is a lower bound of the DDRP.

Proof: Multiplying both sides of each constraint in the DDRP by the load impact of the corresponding group, and summing across the groups for each of the BIP and API programs leads to the aggregate daily and yearly constraints of the DDRP. Further, noting that $Y_{d,t',h}$ and $Z_{d,t',h}$ represent the terms $\sum_{g_1 \in G_1} l_{g_1} X_{d,g_1,t',h}$ and $\sum_{g_2 \in G_2} l_{g_2} X_{d,g_2,t',h}$ without the integrality constraints

respectively, we conclude that the ARLP in an aggregate problem of the DDRP and hence constitutes a lower bound for the latter.

Proposition 3.3.2: The lower bound problem is a convex problem.

Proof: The above claim follows directly from the fact the objective function of the lower bound problem is convex as it consists of the sum of two convex functions and the set of constraints is a convex set since it is formed of linear constraints.

As a result, the ALRP is a convex nonlinear program which can be readily solved by commercial solvers. In the next section we develop a heuristic that efficiently solves the DDRP and assess its performance.

3.4-DDRP Heuristic

3.4.1-Master Problem

In this section we develop a heuristic to obtain a feasible solution for the DDRP. The heuristic works first by solving a master problem whose output will determine the number of groups needed for each day at each hour. After determining the number of groups needed, we allocate the groups by determining the timing and duration of call of each one of them on a given day. Before solving the master problem, we form clusters by grouping the groups which load impacts are close in value. Hence, for a predefined load impact difference δ , if the absolute value of the difference between two load impacts is less than or equal to δ then the corresponding groups belong to the same cluster. We test the heuristic for a given value of δ in the numerical analysis section. After assigning the groups across clusters, the load impact of that cluster will be equal to the average of the load impact of its groups. The reason behind using the clusters will become clear in the heuristic used to solve the DDRP problem as it will provide a mean for decomposing the initial problem by cluster. To provide a precise definition of the heuristic, we define:

Indices:

C : Set of clusters.

c : The subscript for the set of clusters.

Parameters:

b_c : Number of BIP groups in cluster c .

a_c : Number of API groups in cluster c .

l_c : Load impact of cluster c .

Decision Variables:

$B_{d,c,t}$: The amount of load reduction from BIP groups of cluster c on day d at time t (Megawatts).

$A_{d,c,t}$: The amount of load reduction from API groups of cluster c on day d at time t (Megawatts).

$Q_{d,c,i,t'h}$: Number of groups called at time t' for h consecutive hours on day d for the BIP ($i=1$) or API ($i=2$) in cluster c .

The problem of determining the hourly number of groups from each cluster each day can be expressed by the following master problem which consists of a nonlinear convex program. We refer to the master problem as MP

MP

$$\min_{L,B,A} \sum_{d \in D} \sum_{t \in T} a(u_{dt} - L_{d,t})^2 + b(u_{dt} - L_{d,t}) \quad (17)$$

$$B_{d,c,t} \leq b_c l_c \quad \forall d \in D, c \in C, t \in T \quad (18)$$

$$A_{d,c,t} \leq a_c l_c \quad \forall d \in D, c \in C, t \in T \quad (19)$$

$$\sum_{t \in T} B_{d,c,t} \leq 6b_c l_c \quad \forall d \in D, c \in C \quad (20)$$

$$\sum_{t \in T} A_{d,c,t} \leq 6a_c l_c \quad \forall d \in D, c \in C \quad (21)$$

$$\sum_{d \in D} \sum_{t \in T} B_{d,c,t} \leq 180b_c l_c \quad \forall c \in C \quad (22)$$

$$\sum_{d \in D} \sum_{t \in T} A_{d,c,t} \leq 150a_c l_c \quad \forall c \in C \quad (23)$$

$$L_{d,t} = \sum_{c \in C} (B_{d,c,t} + A_{d,c,t}) \quad \forall d \in D, t \in T \quad (24)$$

$$L_{d,t}, B_{d,c,t}, A_{d,c,t} \geq 0 \quad \forall d \in D, c \in C, t \in T \quad (25)$$

As in the DDRP, the objective function consists of minimizing the total hourly electricity generation costs. The maximum load reduction that can be obtained from each cluster per hour is the aggregate of the load impacts of all its BIP and API groups. Constraints (18) and (19) impose this upper bound on the hourly load reduction from the groups of the BIP and API programs in cluster c respectively. Similarly, given that each group can be interrupted for at most 6 hours per day, Constraints (20) and (21) ensure that the daily load reduction from groups of the BIP and API programs in cluster c does not exceed $6b_cl_c$ and $6a_cl_c$ respectively. Constraints (22) and (23) represent the yearly available load reduction amounts for each cluster containing BIP and API groups respectively. Each group in the BIP program can be interrupted for a maximum of 180 hours and each group in the API program for 25 times for a maximum duration of 6 hours per interruption which is equivalent to 150 hours. Constraint (24) assigns the total load reduction on a given day and for a given hour as the sum of the load reductions across all clusters. Finally, Constraint (25) is the nonnegativity constraints. Notice that the master problem is a nonlinear convex program. We note that the master problem is neither an upper bound, nor a lower bound to DDRP. If we aggregate the above program by clusters and keep the load impact of each group to its original value, MP will be a relaxation of DDRP by following the logic outlined in Proposition 3.3.1. However, since the load impact of each group in cluster c is replaced by l_c , we cannot make such a conclusion.

3.4.2-Assignment Problem

Let $n_{d,c,t} = \left\lceil \frac{B_{d,c,t} + A_{d,c,t}}{l_c} \right\rceil$ be the number of groups from cluster c on day d at time t determined by the master problem. The next step consists of assigning groups from each cluster. Note that after solving the master problem the assignment problem can be solved for each cluster and thus

each assignment problem can be solved separately. Denoting by $G(c)$ the set of groups in cluster c and by $G_1(c)$ and $G_2(c)$ the set of BIP and API groups in cluster c respectively, we formulate the following assignment problem. Note that the subscript has been omitted for conciseness. Denoting by $V_{d,t}$ the deviation from the number of groups $n_{d,t}$, the assignment problem ASP is formulated and a quadratic binary problem.

ASP

$$\min_{X,V} \sum_{d \in D} \sum_{t \in T} V_{d,t}^2 \quad (26)$$

Subject to

$$\sum_{g \in G(c)} \sum_{t' \in T'} \sum_{h \in H} I(t, t', h) X_{d,g,t',h} + V_{d,t} = n_{d,t} \quad \forall d \in D, t \in T \quad (27)$$

$$\sum_{t' \in T'} \sum_{h \in H} X_{d,g,t',h} \leq 1 \quad \forall d \in D, g \in G(c) \quad (28)$$

$$\sum_{d \in D} \sum_{t' \in T'} \sum_{h \in H} h X_{d,g,t',h} \leq 180 \quad \forall g \in G_1(c) \quad (29)$$

$$\sum_{d \in D} \sum_{t' \in T'} \sum_{h \in H} X_{d,g,t',h} \leq 25 \quad \forall g \in G_2(c) \quad (30)$$

$$X_{d,g,t',h} \in \{0,1\} \quad \forall d \in D, t' \in T', h \in H \quad (31)$$

The above problem consists of allocating groups in each cluster across days and hours. The assignment is done by minimizing the squared deviations between the assigned groups and the proposed number of hourly groups each day as shown from the objective function and Constraint (27). The intuition behind minimizing the square of the deviations is to give higher priority for the days that have a higher number of clusters at a given time t . As previously mentioned in the DDRP, the assignment of groups in each cluster is done while ensuring that no group is called more than once a day as seen from Constraint (28) with a maximum duration of interruption of 180 hours per year for the groups in the BIP program and a maximum frequency of 25 interruptions for the ones in the API program as highlighted in Constraint (29) and Constraint (30) respectively.

Observe that the above problem is NP-Hard and thus it could be inefficient in solving large instances. We observe that the assignment problem can be solved more efficiently by first converting it to an integer program as this will preclude solving a quadratic binary program and then solving the resulting problem in two stages: an aggregate problem and then an allocation problem.

The conversion of the assignment problem to a mixed integer program is accomplished by noting from Constraint (24) that each deviation, $V_{d,t}$ can take integer values in the interval $R_{d,t} = [n_{d,t} - a_c - b_c, n_{d,t}]$ since the maximum number of assigned groups for a given cluster at each hour is at most $a_c + b_c$. We denote by $|R_{d,t}|$ the cardinality of the set $R_{d,t}$ and by $k_{d,j,t} \in R_{d,t}$ its possible integer elements for $j \in J = \{1, \dots, |R_{d,t}|\}$. Letting $\alpha_{d,j,t}$ be binary variables equal to 1 if $k_{d,j,t}$ is selected, 0 otherwise, we can reformulate the above problem as a mixed integer problem by replacing the objective function by:

$$\min_{\alpha, X, V} \sum_{d \in D} \sum_{t \in T} \sum_{j \in J} k_{d,j,t}^2 \alpha_{d,j,t} \quad (32)$$

and adding constraints (33) through (34)

$$\sum_{j \in J} k_{d,j,t} \alpha_{d,j,t} = V_{d,t} \quad \forall d \in D, t \in T \quad (33)$$

$$\sum_{j \in J} \alpha_{d,j,t} = 1 \quad \forall d \in D, t \in T \quad (34)$$

As seen from Constraint (33) and Constraint (34), only one value among the possible ones in the interval $R_{d,t}$ can be chosen while the objective function ensures that the objective function is equivalent to the one in ASP. We see from the above the rationale behind replacing the load impact of each group by the average load impact of its cluster since otherwise, the conversion of ASP to an MIP would not have been possible.

The goal behind the aggregate problem is to first obtain the number of groups required from the BIP and API programs and then assign groups from these programs in the allocation program. To this end, let $Q_{d,c,1,t',h}$ and $Q_{d,c,2,t',h}$ be the total number of groups called at time t' for h consecutive hours on day d from the BIP and API programs respectively. The aggregate problem AGP is formally states as follows:

AGP

$$\min_{\alpha, X} \sum_{d \in D} \sum_{t \in T} \sum_{j \in J} k_{d,j,t}^2 \alpha_{d,j,t} \quad (32)$$

Subject to:

$$\sum_{j \in J} k_{d,j,t} \alpha_{d,j,t} = V_{d,t} \quad \forall d \in D, t \in T \quad (33)$$

$$\sum_{j \in J} \alpha_{d,j,t} = 1 \quad \forall d \in D, t \in T \quad (34)$$

$$\sum_{i=1,2} \sum_{t' \in T'} \sum_{h \in H} I(t, t', h) Q_{d,c,i,t',h} + V_{d,t} = n_{d,t} \quad \forall d \in D, t \in T \quad (35)$$

$$\sum_{t' \in T'} \sum_{h \in H} Q_{d,c,1,t',h} \leq b_c \quad \forall d \in D \quad (36)$$

$$\sum_{t' \in T'} \sum_{h \in H} Q_{d,c,2,t',h} \leq a_c \quad \forall d \in D \quad (37)$$

$$\sum_{d \in D} \sum_{t' \in T'} \sum_{h \in H} h Q_{d,c,1,t',h} \leq 180 b_c \quad (38)$$

$$\sum_{d \in D} \sum_{t' \in T'} \sum_{h \in H} Q_{d,c,2,t',h} \leq 25 a_c \quad (39)$$

$$Q_{d,c,i,t',h} \in \mathbb{Z}^+, \alpha_{d,j,t} \in \{0,1\} \quad \forall d \in D, t \in T, t' \in T', h \in H, c \in C, j \in J, i = \{1,2\} \quad (40)$$

As shown in the formulation above, the aggregate problem determines the number of groups from each program to be called. Constraint (36) and Constraint (37) limit the number groups to be called each day from every program while constraints (38) and (39) impose upper bounds of the duration and number of calls for each group. The final step consists of allocating the groups from each cluster by solving the following allocation problem which we refer to as ALP.

ALP

$$\min_{M,X} \sum_d \sum_{t'} \sum_h h M_{d,t',h} \quad (41)$$

Subject to:

$$\sum_{g \in G(c)} X_{d,g,t',h} + M_{d,t',h} = Q_{d,G,1,t',h} + Q_{d,G,2,t',h} \quad \forall d \in D, t' \in T', h \in H \quad (42)$$

$$\sum_{t' \in T'} \sum_{h \in H} X_{d,g,t',h} \leq 1, d \in D \quad \forall g \in G \quad (43)$$

$$\sum_{d \in D} \sum_{t' \in T'} \sum_{h \in H} h X_{d,g_1,t',h} \leq 180 \quad \forall g_1 \in G_1(c) \quad (44)$$

$$\sum_{d \in D} \sum_{t' \in T'} \sum_{h \in H} X_{d,g_2,t',h} \leq 25 \quad \forall g_2 \in G_2(c) \quad (45)$$

$$X_{d,g,t',h} \in \{0,1\}, M_{d,t',h} \geq 0 \quad \forall g \in G, d \in D, t' \in T', h \in H \quad (46)$$

The objective function of the allocation problem consists of minimizing the hourly weighted deviation $M_{d,t',h}$ defined in Constraint (42). The remaining constraints (43) through (45) are the daily and yearly constraints of each of the groups in the BIP and API programs. Although solving the allocation problem in two stages will lead to a suboptimal solution, our numerical solution show that the resulting gap does not affect the quality of the solution.

The steps of the heuristics are summarized as follows:

1. Generate the clusters by using a certain cluster size.
2. Solve the MP and obtain the number of groups $n_{d,c,t}$ from each cluster for each time t on day d .
3. For each cluster c , solve the AGP for an obtain the number of BIP and API groups $Q_{d,c,1,t',h}$ and $Q_{d,c,2,t',h}$ respectively to be called on day d at time t' for h consecutive hours.
4. Using the values of $Q_{d,c,1,t',h}$ and $Q_{d,c,2,t',h}$, solve the ALP to assign groups from each cluster.
5. Compute the resulting electricity generation costs.

In the next section, we test the efficiency of our proposed heuristic by assessing the optimality gap between our heuristic's solution and our established lower bound across different instances.

3.5-Numerical Analysis

In this section, we test the efficiency of the heuristic by testing it on several instances. We consider a 50, 100 and 200-day horizon and test the problem on 100, 150 and 200 groups equally divided between the BIP and API programs. The number of hours and number of interruptions for the groups in the BIP and API programs are respectively (180,25) , (90,12) and (45,6) for the 200, 100 and 50 days horizon respectively. We let $a = 1.25$ and $b = 1$ for the electricity generation cost objective function. For each group, we consider the following ranges for the groups:

1. BIP Load Impact Range: [50MW, 300MW], API Load Impact Range: [10MW, 50MW].
2. BIP Load Impact Range: [50MW, 400MW], API Load Impact Range: [50MW, 100MW].

The first set of ranges was chosen based on what is observed in practice as groups enrolled in the API programs have a smaller load impact than the ones enrolled in the BIP program while the second range of load impacts was designed in order to test the performance of the heuristic in unconventional cases.

For each of the above ranges, we randomly generate 50, 75 and 100 load impacts. We consider a cluster size of 10 MW. The data was tested for the years of 2012 and 2013. Tables 2.1, 2.2 and 2.3 below summarize the average optimality gap of the heuristic from 20 simulations for each instance.

The code was written in Python and the optimization was done using the Gurobi solver. The tests were carried on a desktop with an Intel Core i5 processor with a CPU speed of 3.10 GHz. As seen from the tables, for instances with load impacts in the first set of ranges, the optimality gap varies between 1% and 3% while the optimality gap for the second set of ranges varies between 2% and 6%. As we notice, the heuristic performs well for real life instances even when the total number

of groups is as large as 200. Even for the second set of load ranges, the optimality gap is still acceptable given the large magnitude of the load impacts. The computational time of the heuristic depends on the number of days and approximately doubles for a given instance as does the number of days in the horizon with an average running time between 0.7 to 5min. We should note that a problem of 20 days with a total number of 10 groups was solved to optimality using Gurobi in around 3 minutes. Hence, based on the optimality gap and running time results of our numerical experiment, we can conclude that our heuristic is quite efficient.

3.6-Certainty Equivalent Approach:

To assess the performance of the certainty equivalence approach we test our heuristic and forecasting procedures using the hourly load electricity demand for the area served by SCE during the year 2014. We test our approach on a set of 200 groups equally divided between the BIP and API programs. The load impacts for the groups the API program were between 10 to 50 MW while the one for the groups in the BIP program were between 50 and 300 MW.

The dynamic program solution by certainty equivalence is summarized as follows:

- At the start of a given day , update the day ahead forecast for that day and compute the day ahead forecast using the short term forecasting model.
- Using the week ahead hourly temperature data, forecast the load profile for the corresponding days using the long term forecasting model.
- Update the number of day types for each of the months in the remaining horizon and replace the hourly load profile for each day type by its corresponding approximation.
- Solve the DDRP using the heuristic developed in Section 4 and determine the interruption schedule.
- Pick the interruption schedule of the current day.

- Update the remaining number of hours and interruptions of the selected groups in the BIP and API programs respectively.
- Repeat the above procedure for each day until the end of the horizon.

The optimality gap between the certainty equivalence approach and the lower bound was estimated to be equal to 2.6%. This result shows that our suggested approach could constitute a promising mechanism to be adopted in order to tackle such problem. This could be further validated if a robustness test can be developed in order to assess the worst and best performances of our certainty equivalence algorithm.

3.7-Conclusion

In this work, we presented a solution methodology for solving the problem of an electricity producer and supplier that offers interruptible contracts to its commercial and industrial customers. The challenges of this problem stems from the high number of interruption combinations to the enrolled customers as well as the uncertainty in the electricity demand in addition to the limited number of interruptions. We solve this problem using a certainty equivalence approach. Our approach consists of solving a deterministic program, the DDRP, from which we select the interruption schedule for that day, update the number and hours of interruptions of the selected customers and repeat the procedure on a daily basis. The input to the deterministic model is the electricity hourly load prediction for the short and long term horizon obtained from an electricity load forecasting model. Given that the DDRP is NP-Hard, we proposed an efficient heuristic for solving the DDRP and tested its efficiency. The certainty equivalence approach was tested for the year 2014 for the area served by the Southern California Edison electric utility company. There exists some areas of improvements and future research directions. As a possible improvement and as previously mentioned, a procedure could be developed in order to assess the

robustness of the certainty equivalence algorithm and assess its performance. Further, this problem could be solved by using a different approximate dynamic approach. This approach could consist of approximating the value functions in terms of their state space by using basis function as developed by Powell. Finally, this model could also be used in the design of interruptible contracts by determining the number of hours and interruptions for each enrolled customer.

Table 2.1: 50 Days Horizon Results

BIP Group Numbers	API Group Number	BIP Load Impact Range	API Load Impact Range	Optimality Gap (%)	Time (min)
50	50	[50,300]	[10, 50]	1.25	0.74
75	75	[50,300]	[10, 50]	1.89	0.93
100	100	[50,300]	[10, 50]	2.36	1.12
50	50	[50,400]	[50, 100]	2.41	0.81
75	75	[50,400]	[50, 100]	3.90	1.00
100	100	[50,400]	[50, 100]	5.41	1.19

Table 2.2: 100 Days Horizon Results

BIP Group Numbers	API Group Number	BIP Load Impact Range	API Load Impact Range	Optimality Gap (%)	Time (min)
50	50	[50,300]	[10, 50]	1.14	1.46
75	75	[50,300]	[10, 50]	1.76	1.84
100	100	[50,300]	[10, 50]	2.27	2.21
50	50	[50,400]	[50, 100]	2.06	1.64
75	75	[50,400]	[50, 100]	3.71	2.06
100	100	[50,400]	[50, 100]	5.29	2.45

Table 2.3: 200 Days Horizon Results

BIP Group Numbers	API Group Number	BIP Load Impact Range	API Load Impact Range	Optimality Gap (%)	Time (min)
50	50	[50,300]	[10, 50]	1.25	3.07
75	75	[50,300]	[10, 50]	1.89	4.00
100	100	[50,300]	[10, 50]	2.36	4.70
50	50	[50,400]	[50, 100]	2.41	3.39
75	75	[50,400]	[50, 100]	3.90	4.24
100	100	[50,400]	[50, 100]	5.41	5.12

References

- A. Cameron, P. Trivedi, *Microeconometrics Methods and Applications*, Cambridge, 2005.
- D. Bertsekas, *Dynamic Programming and Optimal Control, Volume I, Third Edition*. Athena Scientific.
- H. Alfares, M. Nazeeruddin. 2002. Electric load forecasting: Literature survey and classification of methods. *International Journal of Systems Science*. **33(1)**.
- H. Morita, T. Kase, Y. Tamura and S. Iwamoto. 1996. "Interval prediction of annual maximum demand using grey dynamic model", *International Journal Elect. Power Energy Syst.* **(18)7**.
- J. Nowicka-Zagrajek, R. Weron. 2002. Modeling electricity loads in California: ARMA models with hyperbolic noise. *Signal Processing*. **82**
- J. Vilar, R. Cao, G. Aneiros. 2012. Forecasting next-day electricity demand and price using nonparametric functional methods. *Electrical Power and Energy Systems*. **39**.
- M. Cooper. 1981. A Survey of Methods of Pure Nonlinear Integer Programming. *Management Science* **27(3)**
- M. Kuhn, K. Johnson, *Applied Predictive Modeling*, Springer 2013.
- M. Small, C. K. Tse. 2002. Applying the method of surrogates to cyclic time series. *Physica D*. **164**
- P. Brockwell, R. Davis, *Time Series: Theory and Methods*, Springer 1990.
- P. E. McSharry, S. Bouwman and G. Boloemhof. 2005. Probabilistic Forecasts of the Magnitude and Timing of Peak Electricity Demand. *IEE Trans. Power Syst.* **20(2)**.
- P. M. Robinson. 1988. Root-N Consistent Semiparametric Regression. *Econometrica*. **56 (4)**.
- P. Olinda, W. Charytoniuk, M. S. Chen. 1998. Nonparametric Regression Based Short-Term Load Forecasting. *IEE Trans. Power Syst.* **13(3)**.
- R Development Core Team, R: A Language and Environment for Statistical Computing R Foundation of Statistical Computing. Available: <http://www.R-project.org>.
- R. Hyndman, S. Fan. 2010. Density Forecasting for Long-Term Peak Electricity Demand. *IEE Trans. Power Syst.* **25(2)**.
- R. Baldick, S. Kolos, S. Tompaidis. 2006. Interruptible Contracts from an Electricity Retailer's Point of View: Valuation and Optimal Interruption. *Operations Research* **54(4)**
- V. Goyal, G. Iyengar, Z. Qiu. 2013. Near Optimal Execution Policies for Demand Response Contracts in Electricity Markets. *52nd IEE Conference on Decision and Control*.

Staff Report. Assessment of demand response and advanced metering: August 2006.
<http://www.FERC.gov>.