Title
Modeling, Analysis, and Control of Demand Response Resources

Author
Mathieu, Johanna L.

Publication Date
2012

Peer reviewed|Thesis/dissertation

Modeling, Analysis, and Control of Demand Response Resources

By

Johanna L. Mathieu

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

in

Engineering – Mechanical Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Assistant Professor Duncan S. Callaway, Co-Chair
Professor Ashok J. Gadgil, Co-Chair
Professor Alice M. Agogino
Associate Professor Alexandre M. Bayen

Spring 2012

Modeling, Analysis, and Control of Demand Response Resources

Abstract

Modeling, Analysis, and Control of Demand Response Resources

by

Johanna L. Mathieu

Doctor of Philosophy in Engineering – Mechanical Engineering

University of California, Berkeley

Assistant Professor Duncan S. Callaway, Co-Chair
Professor Ashok J. Gadgil, Co-Chair

While the traditional goal of an electric power system has been to control supply to fulfill demand, the demand-side can plan an active role in power systems via Demand Response (DR), defined by the Department of Energy (DOE) as "a tariff or program established to motivate changes in electric use by end-use customers in response to changes in the price of electricity over time, or to give incentive payments designed to induce lower electricity use at times of high market prices or when grid reliability is jeopardized" [29]. DR can provide a variety of benefits including reducing peak electric loads when the power system is stressed and fast timescale energy balancing. Therefore, DR can improve grid reliability and reduce wholesale energy prices and their volatility.

This dissertation focuses on analyzing both recent and emerging DR paradigms. Recent DR programs have focused on peak load reduction in commercial buildings and industrial facilities (C&I facilities). We present methods for using 15-minute-interval electric load data, commonly available from C&I facilities, to help building managers understand building energy consumption and 'ask the right questions' to discover opportunities for DR. Additionally, we present a regression-based model of whole building electric load, i.e., a baseline model, which allows us to quantify DR performance. We use this baseline model to understand the performance of 38 C&I facilities participating in an automated dynamic pricing DR program in California. In this program, facilities are expected to exhibit the same response each DR event. We find that baseline model error makes it difficult to precisely quantify changes in electricity consumption and understand if C&I facilities exhibit event-to-event variability in their response to DR signals. Therefore, we present a method to compute baseline model error and a metric to determine how much observed DR variability results from baseline model error rather than real variability in response. We find that, in general, baseline model error is large. Though some facilities exhibit real DR variability, most observed variability results from baseline model error. In some cases, however, aggregations of C&I facilities exhibit real DR variability, which could create challenges for power system operation. These results have implications for DR program design and deployment.

Emerging DR paradigms focus on faster timescale DR. Here, we investigate methods to coordinate aggregations of residential thermostatically controlled loads (TCLs), including air conditioners and refrigerators, to manage frequency and energy imbalances in power systems. We focus on opportunities to centrally control loads with high accuracy but low requirements for sensing and communications infrastructure. Specifically, we compare cases when measured load state information (e.g., power consumption and temperature) is 1) available in real time; 2) available, but not in real time; and 3) not available. We develop Markov Chain models to describe the temperature state evolution of heterogeneous populations of TCLs, and use Kalman filtering for both state and joint parameter/state estimation. We present a look-ahead proportional controller to broadcast control signals to all TCLs, which always remain in their temperature dead-band. Simulations indicate that it is possible to achieve power tracking RMS errors in the range of 0.26–9.3% of steady state aggregated power consumption. Results depend upon the information available for system identification, state estimation, and control. We find that, depending upon the performance required, TCLs may not need to provide state information to the central controller in real time or at all. We also estimate the size of the TCL potential resource; potential revenue from participation in markets; and break-even costs associated with deploying DR-enabling technologies. We find that current TCL energy storage capacity in California is 8–11 GWh, with refrigerators contributing the most. Annual revenues from participation in regulation vary from $10 to $220 per TCL per year depending upon the type of TCL and climate zone, while load following and arbitrage revenues are more modest at $2 to $35 per TCL per year. These results lead to a number of policy recommendations that will make it easier to engage residential loads in fast timescale DR.

To Ken.

Contents

List of Figures

List of Tables

Acknowledgments

A special thanks to my co-advisors Profs. Duncan Callaway and Ashok Gadgil, and the other members of my dissertation committee Profs. Alice Agogino and Alex Bayen. Also, thanks to my many collaborators: Sila Kiliccote at Lawrence Berkeley National Laboratory (LBNL), Phil Price at LBNL, Stephan Koch at Eidgenössische Technische Hochschule (ETH) Zürich, Mark Dyson at the University of California (UC) at Berkeley, Art Rosenfeld at LBNL, and Mary Ann Piette at LBNL. Thanks to everyone in the LBNL Demand Response Research Center, the Gadgil research group, and the EMAC research group for stimulating conversations and feedback.

This research was funded by a National Defense Science and Engineering Graduate Fellowship, a UC Berkeley Chancellor's Fellowship, and funds from the Power Systems Engineering Research Center (PSERC) Future Grid Initiative. Some of this work was conducted at the LBNL under United States Department of Energy Contract No. DE-AC02-05CH11231. The electric load data used in Chapters 2 and 3, was provided by Pacific Gas & Electric Company.

Abbreviations

ACF	autocorrelation function
APE	absolute percentage error
AR	autoregressive
CAISO	California Independent System Operator
CBP	capacity bidding program
CEC	California Energy Commission
C&I facilities	Commercial buildings and industrial facilities
COP	coefficient of performance
CPP	critical peak pricing
CT	compliance threshold
DBP	demand bidding program
DOE	Department of Energy
DLC	direct load control
DR	demand response
DRRC	Demand Response Research Center
EKF	extended Kalman filter
EMCS	energy management control system
ETH	Eidgenössische Technische Hochschule
EVM	energy variability metric
FDRD	facility demand response day
FERC	Federal Energy Regulatory Commission
GTA	global temperature adjustment
HVAC	heating, ventilation, and air conditioning
ILM	interruptible load management
LADWP	Los Angeles Department of Water and Power
LBNL	Lawrence Berkeley National Laboratory
LMP	locational marginal price
LOOCV	leave one out cross validation
MAPE	mean absolute percentage error
MPC	model predictive control
M&V	measurement and verification
nb	negative bound
OL	observed load
OLS	ordinary least squares
OS	observed shed
PACF	partial autocorrelation function
pb	positive bound
PBL	predicted baseline load

PLP	participating load pilot
PDP	peak day pricing
PG&E	Pacific Gas and Electric Company
PSERC	Power Systems Engineering Research Center
PV	photovoltaics
PVM	daily peak variability metric
QQ	quantile-quantile
RBL	real baseline load
REM	regulation energy management
RMS	root mean square
RS	real shed
RTP	real time pricing
RTU	rooftop unit
RVM	rebound variability metric
SAT	supply air temperature
SCE	Southern California Edison
SDG&E	San Diego Gas and Electric Company
SMUD	Sacramento Municipal Utility District
SVM	average demand shed variability metric
TCL	thermostatically controlled load
TMY	typical meteorological year
TOU	time of use
UC	University of California
UL	unmodeled load
VFD	variable frequency drive

Chapter 1

Introduction

The traditional goal of an electric power system has been to control the supply-side to fulfill demand; however, the demand-side can plan an active role in power systems via Demand Response (DR), defined by the Department of Energy (DOE) as "a tariff or program established to motivate changes in electric use by end-use customers in response to changes in the price of electricity over time, or to give incentive payments designed to induce lower electricity use at times of high market prices or when grid reliability is jeopardized" [29]. It is important to note that DR is not energy efficiency. Energy efficiency refers to actions taken to permanently reduce the energy consumption of goods and services, for example insulating a home, switching to more efficient appliances, and tuning a commercial heating, air conditioning, and ventilation (HVAC) system. DR entails *shifting* electricity use, for example, off peak, resulting in no net energy savings or *shedding* (i.e., curtailing) electricity use temporarily, for example, during peak hours, resulting in net energy savings but only for a small portion of the hours in a year.

DR programs can take a number of forms and provide a range of benefits to power systems [29, 9, 125, 2, 46]. Some examples are as follows:

- DR can reduce wholesale energy prices and their volatility. In systems without DR, demand is inelastic. Additionally, when a power system nears its generation capacity, supply becomes increasingly inelastic. The result is extreme wholesale electricity price volatility on days when system demand is high [8, 78].

- DR can reduce the need for power system infrastructure expansion. Power systems are sized so that they can provide electricity on the peak hour of the year. Through DR, the peak is reduced and new investments in power plants and transmission can be delayed.

- DR can limit the use of peaking power plants, i.e., peakers. Peakers are only used a small number of hours per year, and have high marginal costs [51] and are generally less efficient than other plants [39].

- DR can improve grid reliability [100]. For example, DR can provide emergency response to grid contingencies via ancillary services such as spinning reserve [42].
- DR can provide power system flexibility. Like generators and energy storage devices, it can be viewed as a resource that can 'provide energy' (via demand reductions) or provide services (via demand reductions *and* increases) to the grid.
- DR may be able to provide fast timescale energy balancing, especially important in power systems with high penetrations of intermittent renewable resources like wind and solar power [100, 19].

The goal of this work is to better understand the capabilities and constraints of DR resources. Using tools from statistics, controls, and optimization, we present methods to 1) model commercial buildings and industrial facilities (C&I facilities) and compute DR performance; 2) understand model error and quantify variability in responses to open-loop DR signals; 3) aggregate and control residential thermostatically controlled loads (TCLs)–such as air conditioners, heat pumps, electric water heaters, and refrigerators–to participate in fast timescale DR; and 4) estimate the resource size, revenues, and costs associated with TCL participation in fast timescale DR. The results of this work have implications for DR program design and development, and energy policy.

In the following sections, we present a brief background on DR and detail different types of DR. Then, we describe the organization of the rest of the dissertation. Detailed literature reviews on each research topic are presented within each chapter.

1.1 Background on DR

A brief history of DR is given in [22]. In summary, DR is not a new concept, but has been discussed since the deployment of the first electricity grids in the 1890s, especially with respect to time-differentiated electricity rates. Other DR concepts such as interruptible load management (ILM), mainly for industrial customers, and direct load control (DLC), mainly for residential customers, became popular in the 1970s. Around the same time, international energy crises lead to increased interest in demand side management and integrated resource planning, in which DR can play a part.

In the 1990s, many electricity systems in the U.S. started the process of deregulation/restructuring, moving from vertical integration to utility divestment in generation resources and competitive wholesale electricity markets. However, as the 2000-2001 California Energy Crisis showed, a competitive wholesale electricity market with an unresponsive demand-side can lead to problems of generator market power [8]. This spurred further interest in DR, for example, in California the Lawrence Berkeley National Laboratory (LBNL) Demand Response Research Center (DRRC) began research and pilot projects in 2004.

Many recent policies have aimed to eliminate barriers to DR including:

- 2005 Energy Policy Act, which eliminated some barriers for DR entry in energy, capacity, and ancillary services markets [27]
- Federal Energy Regulatory Commission (FERC) Order #719, which permits aggregators to bid DR on behalf of electricity customers [48]
- FERC Order #745, which says that DR must be compensated at the Locational Marginal Price (LMP) [49] (which is currently being challenged [80])
- FERC Order#755, which says that resources, including DR, must be paid for regulation on the basis of how well they can provide the service [50], e.g., faster ramping resources must be paid more

Increased investments in the 'smart grid' have lead to many new opportunities for DR. In fact, five out of the ten meanings of the term 'smart grid' listed by Morgan *et al.* [97] directly relate to DR: meters that can be read automatically, meters that can communicate to customers, time of day and time of use meters, control of customer loads, and selective load control. This new communications, sensing, and control infrastructure enables and enhances notification of DR signals and changing electricity prices, and customer responses to these signals/prices.

1.2 Types of DR

As the DOE DR definition explains, DR can take two main forms: time-differentiated electricity rates or incentive payments. Each of these forms has a large number of variants. In this section, we summarize the many types of DR.

1.2.1 Time-differentiated electricity rates

Time differentiated electricity rates, also known as price based DR programs [2] and time based DR programs, can be split into time of use (TOU) electricity rates and dynamic electricity rates. TOU rates are known to customers well in advance (i.e., months ahead) and encourage customers to shift electricity use to times when demand is usually low, for example, nighttime. The hours in a days are divided into some number of categories, for example, off-peak, part-peak, and peak, and electricity in each category is priced differently. Additionally, prices may change in different seasons. This is in contrast with the 'flat rates' normally seen by residential customers, in which the price of electricity is fixed throughout the day and year. In California, most large commercial and industrial customers have been on TOU rates for decades.

Dynamic electricity rates are not known to customers well in advance. For example, they may be published day ahead or day of, and high prices are used to signify high expected system load (often resulting from high or low expected outdoor air temperature). Dynamic

pricing can be further split into Critical Peak Pricing (CPP) and Real Time Pricing (RTP). In CPP programs, DR events are called on a small fraction of the days in a year and on those days electric rates are raised during peak hours to encourage both shifting and shedding. In California, all large commercial and industrial customers (greater than 200 kW peak) are defaulted to a CPP tariff, for example, Pacific Gas & Electric Company's (PG&E) Peak Day Pricing (PDP) tariff [106].

In RTP programs, electricity prices may change hour-to-hour and day-to-day Many economists advocate RTP tied directly to wholesale prices [9, 8, 12]. However, prices could also be a function of the value and cost of electricity in different time periods [29], which implies that they could encapsulate information other than wholesale prices. For instance, retail electricity prices could be higher during peak times (e.g., two times the wholesale price) to encourage power use reductions. Currently, RTP rates are uncommon. For example, in California only one utility offers an opt-in RTP tariff for large industrial customers, Southern California Edison (SCE) [114]. SCE uses a look-up table to determine next-day prices based on the next-day forecasted high temperature in downtown Los Angeles.

1.2.2 Incentive payements

As mentioned above, ILM and DLC programs have existed since the 1970s. In these programs, customers allow the program sponsor (e.g., utility, aggregator, etc.) to control their loads within some prior agreed-upon constraints in exchange for credits and/or incentive payments. Some work as been done to develop strategies to control large industrial plants through ILM [61, 25] and residential TCLs through DLC [11, 99] for peak load management.

Another way to achieve DR is to allow loads to participate in wholesale electricity markets. In Demand Bidding Programs (DBP) and Capacity Bidding Programs (CBP), loads or aggregations of loads offer demand reductions via price/quantity bids into energy and capacity markets. If their bids are accepted they must provide demand reductions at specified times. In some cases, loads can also participate in ancillary services markets. For example, in the Participating Load Pilot (PLP) program conducted by PG&E in 2009, C&I facilities submitted offers via price/quantity bids into the day-ahead non-spinning reserve market [71]. These bids were optimized by the California Independent System Operator (CAISO) together with supply-side bids/offers. Though treating loads symmetrically with generators in electricity markets is attractive, in practice, issues arise because demand bids are always relative to a baseline and baseline models are inaccurate. These issues will be detailed in Chapters 2 and 3.

1.2.3 Recent versus emerging DR

Most recent DR programs have focused on recruiting large C&I facilities rather than residential customers, which need to be aggregated together to achieve a measurable response. Moreover, large customers generally have energy management and control systems (EMCS)

Table 1.1: Recent vs. emerging DR.

	Recent DR	Emerging DR
Type of customer	Large commercial buildings and industrial facilities	Small commercial and residential buildings, and individual loads
Program goal	Peak load shedding, usually on the hottest summer afternoons	Shedding and shifting any day, any time
Actuation	Single DR signal: • manual or automated • centralized source • open-loop control	Continuous DR signal: • automated • centralized/decentralized source • feedback control
Timescale	Day-ahead to hour-ahead	Hour-ahead to real-time

which enable some level of DR automation. However, there are many advantages to engaging residential customers in DR (detailed in Chapter 5) and so we are beginning to see more work in this field. Table 1.1 details recent versus emerging DR programs. Emerging DR programs have expanded program goals, beyond peak load shedding. Moreover, through feedback control, emerging DR focuses on not only initiating a response, but also controlling the response.

Another difference between recent and emerging DR is the timescale. In recent DR programs, DR events are usually called day-ahead, e.g., via day-ahead prices (Figure 1.1). Emerging DR programs have much shorter timescales. For example, when loads participate in ancillary services markets they must respond to sub-hourly signals like 5-minute load following signals or 4-second automatic generation control (i.e., regulation) signals. At faster timescales, it is less practical to have 'a human in the loop,' so emerging DR proposals involve more automation and, possibly, DLC by the load manager. We present our DLC scheme in Chapter 4.

1.3 Organization of the Dissertation

This dissertation consists of four main chapters. Chapter 2 and 3 focus on analyzing the behavior of C&I facilities participating in recent DR programs, while Chapters 4 and 5 focus on emerging DR paradigms.

In Chapter 2, we present methods for using 15-minute-interval electric load data, commonly available from C&I facilities, to help building managers understand building energy consumption and 'ask the right questions' to discover opportunities for DR. Additionally, we present a regression-based model of whole building electric load, i.e., a baseline model,

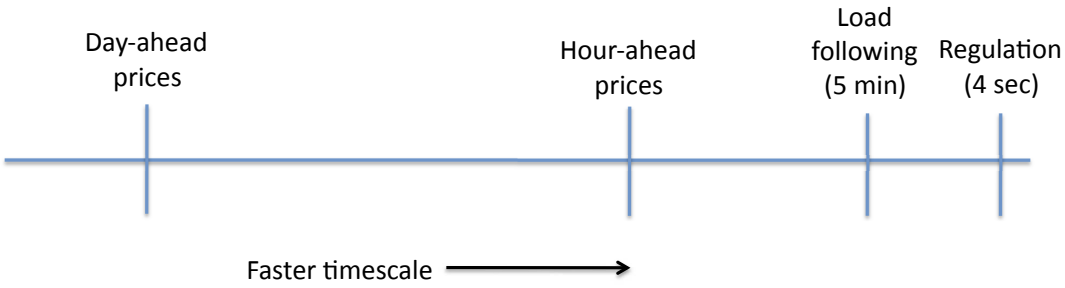


Figure 1.1: DR timescales.

which allows us to quantify DR performance. In Chapter 3, we use this baseline model to understand the performance of 38 C&I facilities participating in an automated dynamic pricing DR program in California. In this program, facilities are expected to exhibit the same response each DR event. However, baseline model error makes it difficult to precisely quantify changes in electricity consumption and understand if C&I facilities exhibit event-to-event variability in their response to DR signals. Therefore, we present a method to compute baseline model error and a metric to determine how much observed DR variability results from baseline model error rather than real variability in response.

In Chapter 4, we investigate methods to coordinate aggregations of residential TCLs to manage frequency and energy imbalances in power systems. We focus on opportunities to centrally control loads with high accuracy but low requirements for sensing and communications infrastructure. Specifically, we compare cases when measured load state information (e.g., power consumption and temperature) is available in real time; available, but not in real time; and not available. We present a Markov Chain model of load aggregations, describe how we applied Kalman filtering techniques for both state estimation and joint parameter and state estimation, and present a look-ahead proportional controller. We compare the results of our model-estimator-controller system to that of a simple proportional controller to demonstrate the value of the approach. In Chapter 5, we estimate the size of the TCL resource, potential revenue from participation in markets, and break-even costs associated with deploying DR-enabling technologies.

Chapter 2

Commercial Building Load Shapes & Baseline Models

This chapter presents methods for using 15-minute interval electric load data, commonly available from C&I facilities, to help building managers understand building energy consumption and 'ask the right questions' to discover opportunities for DR. Additionally, we present a regression-based model of whole building electric load, i.e., a baseline model, which allows us to quantify DR performance. We focus on recent DR programs. This chapter is largely based on [92].[1]

2.1 Chapter Introduction

Building managers often look for opportunities for energy cost savings through energy efficiency, electricity waste elimination (through anomoly detection, changing operational schedules, etc.); peak load management; and participation in DR tariffs and programs, which incentivize reduced electricity consumption during peak hours or when grid reliability is jeopardized. Analysis of 15-minute interval whole building electric load data is a good starting point for discovering opportunities to reduce energy costs through building energy management. These data are usually available to C&I customers. For example, PG&E collects 15-minute interval electric load data from all large C&I facilities (i.e., facilities with a maximum demand of more than 200 kW for three consecutive months) in its service territory. It uses these data to compute both energy costs and demand charges. Despite their availability, these data are not commonly used by building managers because the raw data are difficult to process and interpret. Therefore, more effective methods are needed to translate electric load data into actionable information.

[1] ©2011 IEEE. With permission from my co-authors: Phillip Price, Sila Kiliccote, and Mary Ann Piette. "Quantifying Changes in Building Electricity Use, With Application to Demand Response." *IEEE Transactions on Smart Grid*, Sept 2011.

In this chapter, we present methods for analyzing 15-minute interval electric load data from C&I facilities. Specifically, we describe graphical representations of electric load data, a regression-based electricity load model, and the definition of various parameters that characterize electric load as a function of time (i.e., the "load shape") and DR behavior. The technical contributions are as follows:

- We describe new ways of visualizing electric load data;
- we introduce a time-of-week indicator variable into regression models of electric load;
- we avoid the use of change-point models, but still capture a nonlinear relationship between outdoor air temperature and load, by dividing temperatures into many intervals and then fitting a piecewise linear and continuous temperature dependence;
- we define new parameters to characterize electric load shapes and DR behavior;
- and we apply the modeling methods to evaluate DR effectiveness.

In Figure 2.1, we present a framework for using electric load data to assess opportunities for DR, and evaluate the effectiveness of the DR strategies that were implemented. Traditionally, building managers develop DR strategies after only minimal analysis of their load data. They primarily work with their utility or aggregator to adapt DR strategies, such as those in [98], that have worked well in other buildings. Therefore, they may miss building-specific DR opportunities or may implement DR strategies that are ineffective in their buildings. Building managers do not typically approach DR strategy development systematically, in part because there are too few DR methods and tools [47].

Therefore, we propose that building managers use tools incorporating data analysis methods such as load shapes, standardized load parameters, parameter plots, and load prediction to analyze their facilities' current and historic load shape. These methods are described in Sections 2.3 and 2.4. With knowledge resulting from this analysis and knowledge of DR strategies that have worked well in other buildings, a building manager is able to develop a list of informed questions that help direct an evaluation of building operations, controls, systems, and end-uses. This process is explained in Section 2.5. The results of the evaluation allow the building manager to identify potential DR strategies that are specific to his or her building.

The effectiveness of DR strategies that have been executed is determined using load prediction models to estimate what load would have been on a DR event day if a DR event had not occurred (a literature review is presented in Section 2.4.1). There are few tools available to building managers that automate load prediction. Also, additional data analysis methods could enhance the interpretation of load prediction results. Therefore, we propose that building managers use tools incorporating load prediction, DR residuals, and DR parameters to evaluate DR effectiveness. These methods are detailed in Sections 2.4 and 2.6.

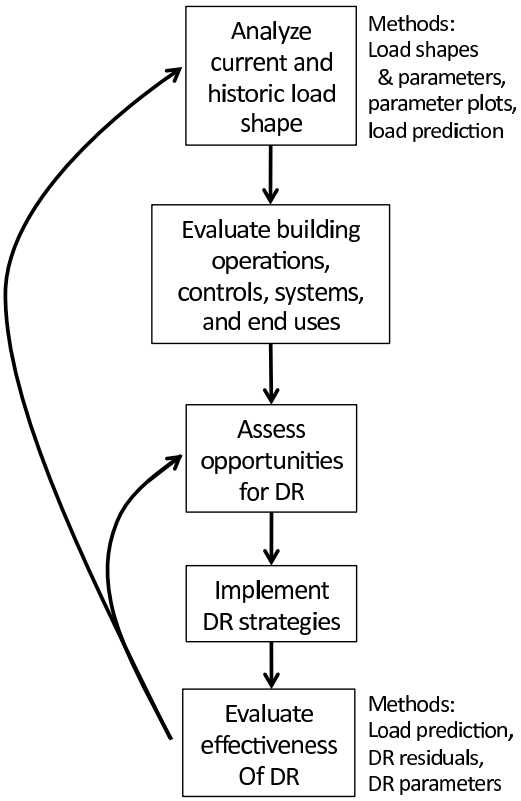


Figure 2.1: Framework to assess and evaluate DR opportunities. ©2011 IEEE

As Figure 2.1 shows, developing DR strategies should be an iterative process: identify a DR opportunity; implement and evaluate it; and, if desired, modify it to improve the response or move on to the next opportunity. Improving a building's DR strategies benefits the building by reducing its energy costs. It also benefits the grid [60]. Of course, these benefits must be weighed against DR costs [70].

A similar framework to that in Figure 2.1 could be applied to assess opportunities for any form of building energy management, but we focus on DR because more DR estimation tools and methods are needed [47], and the use of DR is expanding. In many utility service territories, C&I facilities are called to shed or shift load on hot summer afternoons when the electricity grid is stressed with high loads due to residential and commercial cooling.

In the future, facilities may be dispatched at any time, especially those participating in wholesale energy, capacity, and ancillary service markets (such as spinning reserve [42], non-spinning reserve [71], and regulation/load following [19]), and as DR is used to support the integration of intermittent renewable energy resources [19, 124]. Therefore, new analytical methods are needed to assess opportunities for and evaluate the effectiveness of "any day,

Table 2.1: Facilities used in examples. ©2011 IEEE

Facility	Type	Location	Weather Station(s)	Peak Load* (kW)	Load Factor*	Peak Load Intensity* (W/m^2)
County Bldg	Office	Martinez	Concord	543	0.33	44.6
Furniture Store	Retail	East Palo Alto	Palo Alto & Hayward	1344	0.50	48.2
Bakery	Industrial	Oakland	Oakland	620	0.61	66.1

*Computed for May - Sept 2008, with 15-minute-interval data.

any time" DR.

2.2 Data Sources

2.2.1 Electric load data

We use 15-minute interval whole building electric load data from 38 large C&I facilities (peak demand $\geqslant$200 kW for 3 consecutive months) in California that participated in PG&E's Automated CPP Program between 2006 and 2009. Table 2.1 shows the C&I facilities used in examples throughout this chapter. Appendix A.1 gives detailed information on each of the facilities. PG&E called CPP DR events on up to 12 summer business days (non-holiday, weekdays) per year when system-wide load was expected to be high, which, in California, usually occurs on hot summer days as a result of commercial and residential air conditioning. On DR days, electricity prices were raised to three times the normal price from 12 to 3 pm (moderate price period), and five times the normal price from 3 to 6 pm (high price period). In exchange for participating in the program, facilities paid lower energy prices on non-DR days. All 38 facilities used the OpenADR Communication Specification [107] to receive DR event notifications, which were provided by 3 pm the business day before the event. Each facility implemented a different set of pre-programmed DR strategies and executed the same strategies from event-to-event. Strategies included changes to HVAC system, light dimming/switching, and industrial process shedding/shifting [98].

2.2.2 Temperature data

From NOAA [101], we acquired hourly outdoor air temperature data for each facility from the nearest weather station. Unfortunately, some of the temperature data are spotty. We linearly interpolated the data to assign an approximate temperature to every 15-minute interval, though when six or more hours of data are missing we do not interpolate. In some cases, when the data for a station were particularly spotty, we have filled the holes with data from another nearby station. Temperature data for the aggregate populations were generated

by weighting and averaging data from the individual stations. Weights are determined by the number of facilities in the aggregate population associated with each station. For example, if N_1 facilities are associated with Station 1, which measured temperatures $T_1(t)$, and N_2 facilities are associated with Station 2, which measured temperature $T_2(t)$, and so on, then aggregate temperature, T_{agg}, at time step t is computed as follows:

$$T_{agg}(t) = \frac{\sum_{i=1}^{M} N_i T_i(t)}{\sum_{i=1}^{M} N_i} \quad , \tag{2.1}$$

where M is the total number of weather stations associated with facilities in the aggregate population. Table 2.1 shows the weather station(s) from which data were acquired for each of the facilities used in the examples. Detailed information about each of the weather stations and a data interpolation summary are given in Appendix A.1.

2.3 Visualizing Electric Load Data

It is possible to learn a significant amount about the performance of a building over time, and as compared to other buildings, by plotting electric load data in various ways [55, 54, 56]. Analyzing electric load data graphically can lead to more insights into building characteristics, operations, and use than can be gained by only analyzing summary statistics. In this section, we describe several ways of visualizing electric load data. We first plot time series data and describe what can be learned from these plots. We then define and plot a small set of parameters that are useful for describing load variation from one day to the next. These "parameter plots" may help identify aspects of load shapes more easily.

2.3.1 Load shapes

Figure 2.2 shows electric load versus time for the three example facilities. Striking differences between facilities are immediately apparent including differences in operating hours, daily load shape regularity, the magnitude of daytime versus nighttime loads, and the variation in load from one 15-minute-interval to the next (i.e., the smoothness of the load shape). In addition, day-to-day changes within facilities can be observed. For instance, all of the facilities have higher loads in the second week shown, likely due to higher outdoor air temperatures (and therefore more need for cooling) during the second week.

Many building load shapes share some features (Figure 2.3). Most buildings have a clear *base load*, attained during the night, below which the power consumption rarely falls. In the early morning, the HVAC system switches from nighttime to daytime operation, and, if the building interior warmed/cooled overnight, the HVAC system may turn on at high power to cool/warm the building. This results in a short-lived load spike called the *morning start-up*. As the morning continues, load increases with increased occupancy and, in the cooling season, with increased outdoor air temperature. At some point the building reaches its *peak*

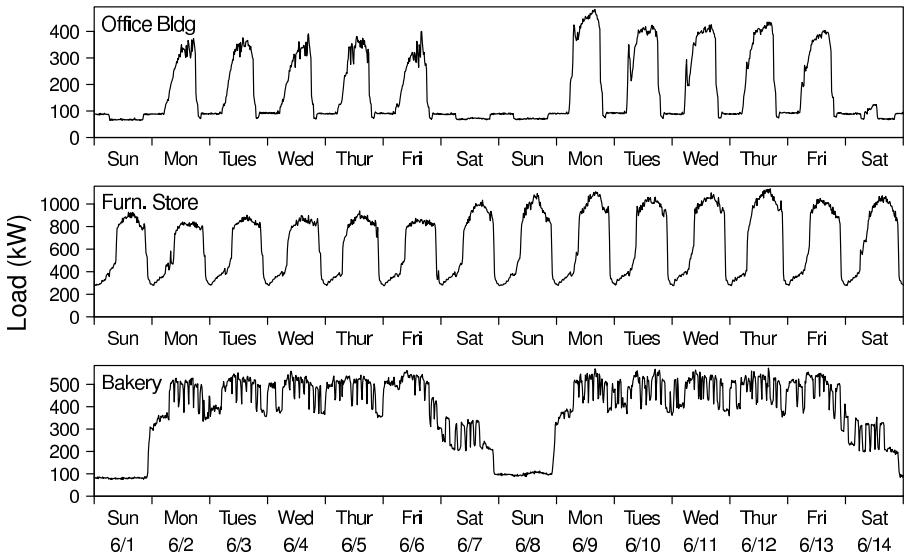


Figure 2.2: Electric load versus time. ©2011 IEEE

load for the day. Peak loads can be computed over any time interval (e.g., daily, seasonally, yearly), and are generally more variable than base loads. In the afternoon or evening, the HVAC system switches back to nighttime operation and the power consumption quickly decreases, a phenomenon called the *evening setback*. In some buildings, some excess over the base load persists into the evening leading to an *evening shoulder*.

Plotting time series load data and/or overlaying data from different time periods can be useful for noticing and characterizing changes in load shapes and their features. However, there are limitations to this approach:

1. For most facilities, energy consumption is a function of weather; however, we are generally interested in understanding changes in energy consumption that are not caused by weather variation. To deal with this issue, we present a method for weather-normalizing load data in Section 2.4.2.

2. Some phenomena may be difficult to recognize in plots of time series load data. For instance, noticing a tendency for load to increase gradually over a long period might be difficult, since this small trend will often be superimposed on seasonal variation and other features. Overlaying data separated by a year may reveal that load is higher now than a year ago, but will not reveal whether the change was gradual or abrupt.

3. Graphical approaches to understanding and comparing load shapes are useful only when people are able to devote time and effort to using them. Automated methods can potentially reduce this effort, though fully automating the *analysis* of time series data such as those in Figure 2.2 is a daunting task.

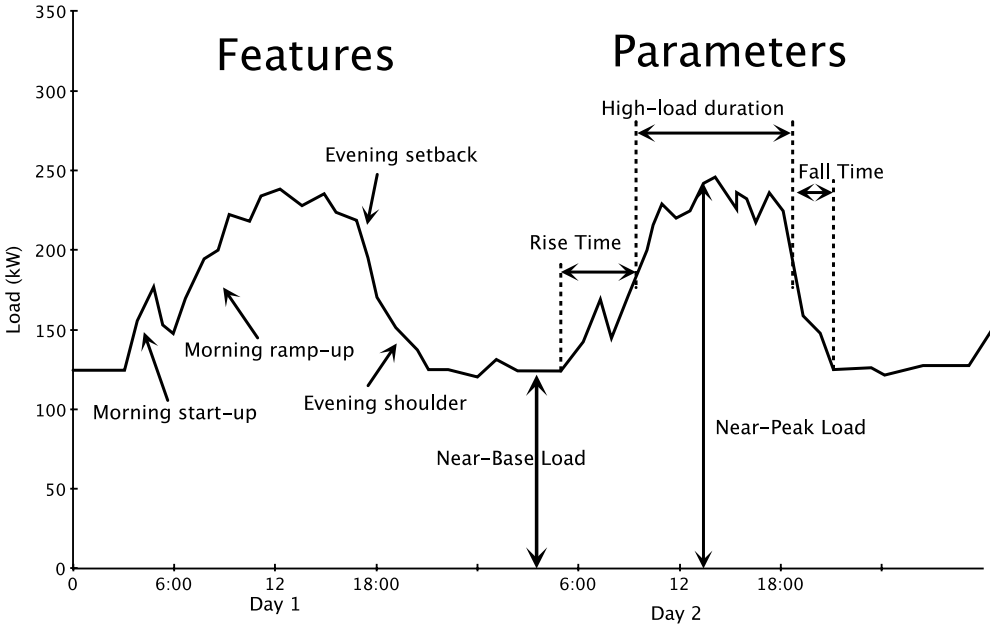


Figure 2.3: Load shape features and parameters. ©2011 IEEE

Table 2.2: Load shape parameter definitions. ©2011 IEEE

Load Shape Parameter	Definition
Near-Base Load (kW)	2.5^{th} percentile of daily load.
Near-Peak Load (kW)	97.5^{th} percentile of daily load.
High-Load Duration (hrs)	Duration for which load is closer to near-peak than near-base load.
Rise Time (hrs)	Duration for load to go from near-base load to start of high-load period.
Fall Time (hrs)	Duration for load to go from end of high-load period to near-base load.

2.3.2 Load shape parameters

As with any time series, load data invite the calculation of a wide variety of summary statistics. It is useful to distinguish between two kinds of summary statistics: those that summarize various aspects of the load and its variability, and those that summarize the amount of load or load variability that is not related to weather, i.e., weather-normalized summary statistics. In this section, we consider summary statistics that are *not* weather-normalized.

We recommend five parameters that are useful for describing load shapes (Figure 2.3 and Table 2.2). The value of each of the parameters can be calculated for each day and these values can be summarized (e.g., mean and standard deviation of each parameter).

We define the "near-base" and "near-peak" load since the base and peak load summarize

extreme events, which may differ qualitatively from events that are actually of interest. For instance, the power consumption during a power outage should not be considered a building's base load. As for the peak load, consider two buildings with the same power consumption profile that each have power usage spike for a full 15 minutes. The first building's spike occurs during a single 15-minute-interval, but the second building's spike is spread across two 15-minute-intervals. The measured peak load for the first building will be higher than that of the second building despite that the fact that they both consumed the same amount of power for the same amount of time. To help deal with these issues, we recommend using the 2.5^{th} and 97.5^{th} percentile of daily load instead of the minimum and maximum.

We define three time intervals, high-load duration, rise time, and fall time, to characterize how the load changes throughout the day. In practice, it is difficult to find definitions of these time intervals that yield consistent, easily interpretable results. The definitions in Table 2.2 work well if the load shape is something like Figure 2.3, but it do not produce useful numbers in some cases, such as for load shapes that do not vary substantially over time or load shapes that have multiple extreme maxima and minima during each day.

2.3.3 Parameter plots

Plotting the parameters defined in Table 2.2 can help us to recognize phenomena that we would be likely to miss if we only analyze plots of 15-minute-interval load data. For example, in Figure 2.4, we plot each facility's near-base and near-peak load for weekdays from May to Sept 2008. Examining these plots reveals several things:

- the office building's and furniture store's near-base loads were relatively constant, while the bakery's near-base load varied day-to-day;
- the bakery's near-peak load increased over the course of the summer; and
- the bakery's near-peak load did not vary significantly day-to-day, unlike that of the office building and furniture store.

While all of this information *could* be obtained from a plot of 15-minute-interval load data, the many layers of information present in such a plot make it very difficult to identify these trends.

Figure 2.5 shows an example of what one can learn from plotting a facility's high-load duration over time. Analyzing the plot we learn that, in early 2008, the furniture store's operating hours were extended by more than an hour each day, and became more uniform day-to-day. On an initial inspection of the store's 15-minute-interval load data we did not notice this change (though in retrospect it is visible). In comparison, when we inspected Figure 2.5, we immediately noticed the change and when it occurred. More examples of useful parameter plots can be found in [108].

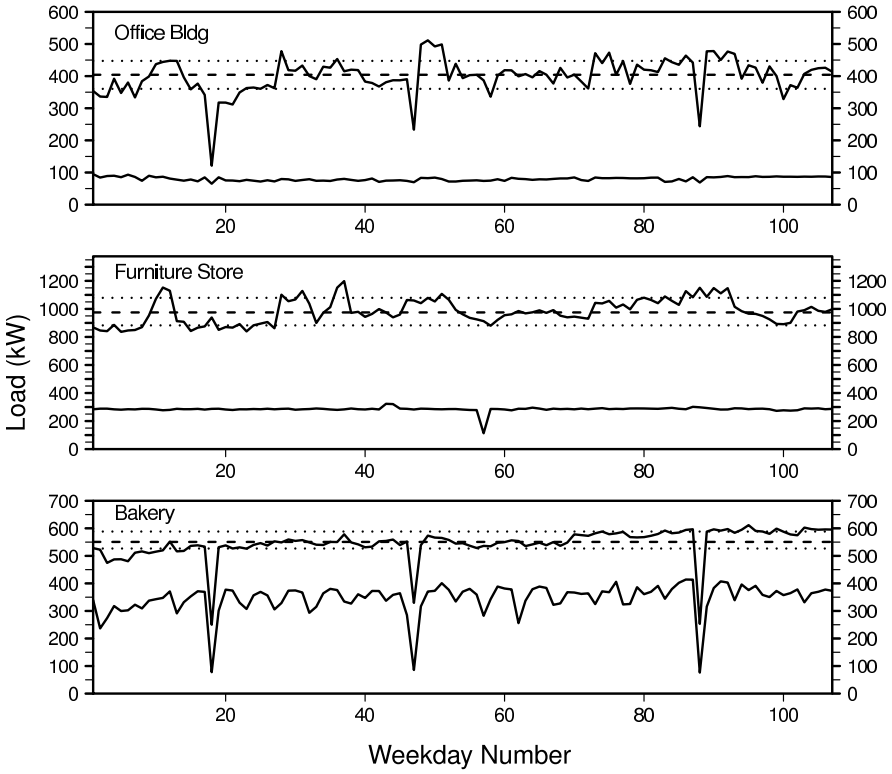


Figure 2.4: Near-base (lower curve) and near-peak load (upper curve) for each weekday from May to Sept 2008. Dotted and dashed lines show the 15^{th}, 50^{th}, and 85^{th} percentiles of the near-peak load. Significant dips in load for the office building and bakery are holidays (Memorial Day, Independence Day, and Labor Day).

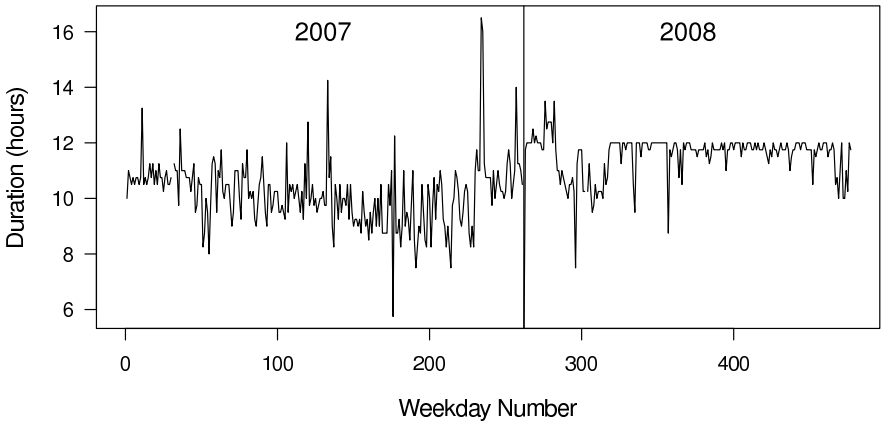


Figure 2.5: High-load duration for the furniture store for each weekday from Jan 2007 to Oct 2008. ©2011 IEEE

2.4 Predicting Electric Loads

In this chapter, we focus exclusively on understanding past electric load. We use statistical models that quantify the electric load as a function of time-of-week and outdoor air temperature. Following standard statistical terminology, we call the output of the model a "prediction" of the electric load, even if the events occurred in the past. Predictions of future load are referred to as "forecasts."

Electric load prediction is useful for comparing how a facility is currently performing to how it has performed in the past. Specifically, load prediction is used to:

- understand changes in a facility's electricity consumption patterns from one time period to the next;
- quantify the effectiveness of DR strategies;
- quantify the effectiveness of energy efficiency retrofits; and
- perform anomoly detection (by finding times when the building is not behaving as it has behaved in the past).

In each case, the predicted load is compared to the actual load. Importantly, the predicted load is computed under the same key conditions as those that lead to the actual load. For example, using actual weather data to compute the predicted load allows us to 'weather-normalize' the prediction; the remaining differences between the predicted and actual load are not weather-dependent.

Weather-normalization is especially important for facilities with significant temperature-dependent loads (e.g., cooling loads, electric heating loads). For example, in Figure 2.6

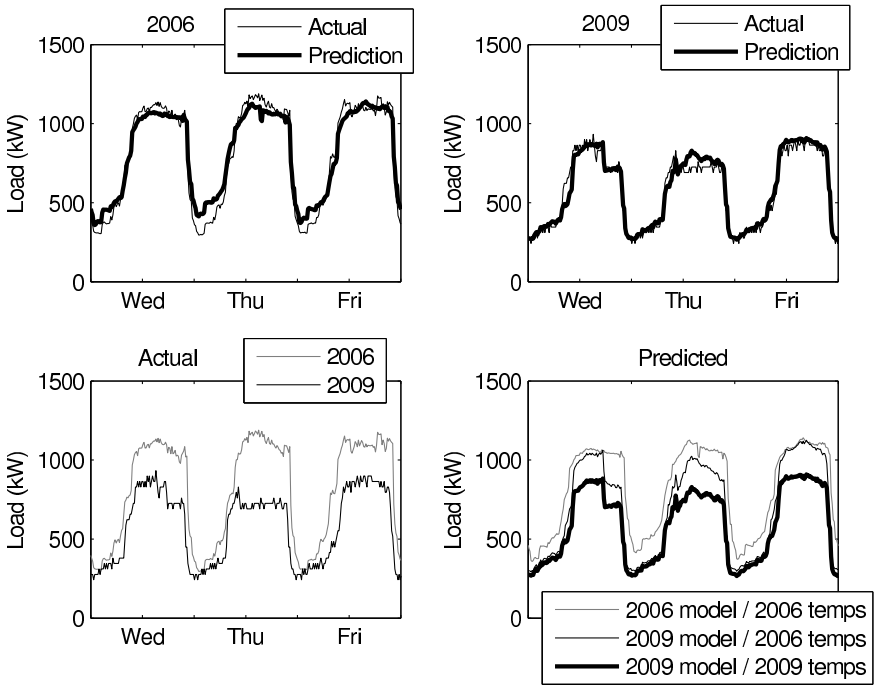


Figure 2.6: The importance of weather-normalization. The top plots compare actual and predicted load for each year, the bottom left plot compares actual load data across years, and the bottom right plot shows weather-normalized predictions. '2009 model / 2006 temps' refers to a load prediction made using a baseline model created with 2009 load data and 2006 outdoor air temperature data, etc. ©2011 IEEE

we show the furniture store's actual and predicted load for three days in July in 2006 and 2009. Predictions are computed using the load prediction method that will be introduced in Section 2.4.2. To give a sense for model accuracy, the top plots compare actual and predicted load for each year. The bottom left plot compares actual load data across years. From this plot, we learn that the facility used significantly less energy in 2009 than 2006. However, we would like to know how much of the difference is due to changes in equipment, operations, and use, and how much is simply due to weather. The bottom right plot shows weather-normalized predictions. Specifically, predictions from 2006 and 2009 are shown, as well as predictions that use the 2009 model but 2006 temperatures. Comparing the gray and thin black lines we can see the portion of the savings not due to weather, while comparing the thin and thick black lines we can see the portion of savings due to weather. Some, though not all, of the difference in daytime load is due to weather, while almost none of the difference in nighttime load is due to weather.

We next present a brief review of existing load prediction methods. Then, we propose a load prediction method that uses a time-of-week indicator variable and a piecewise linear and

continuous outdoor air temperature dependence. In addition, we briefly discuss the sources of error associated with load prediction.

2.4.1 Existing methods for load prediction

Electric utilities use simple methods to predict the electric loads of facilities that participate in DR programs. The predictions are called baselines because they provide a baseline against which demand reductions are computed. California electric utilities use methods such as averaging the electric use profiles of the three days with the highest energy usage out of the last ten business days, or averaging the electric use profiles of the last ten business days [28, 52]. To adjust for weather and other conditions on DR days, predictions are sometimes multiplied by a morning adjustment factor: if a building is using, say, 10% more electricity on the morning of a DR day than on other mornings, then its predicted afternoon load is increased by 10% as well. Unfortunately, this approach has a serious problem: if a facility shifts some of its load to the morning on a DR day, the result of the morning adjustment is an overestimate of the amount of energy the building would have consumed during the DR period. Essentially the building is credited twice for the same energy shift: once by using less energy than it would have if it did not participate in DR, and once by overestimating the amount of energy that it would have used.

More advanced load prediction methods have been developed for a variety of applications including estimating the effectiveness of energy efficiency retrofits and forecasting utility-scale electric loads. Claridge [26] discusses many approaches for using historical electric load data to model the electricity consumption of C&I facilities including linear regression models, calibrated simulations, Fourier series models, and neural network models (e.g., [65]). Granderson *et al.* [53] also describe several other methods for residential load prediction including non-linear models such as locally-weighted regressions, "bin" models in which load predictions are based on the average load for time periods that share the same bin as current conditions (e.g., weather and time-of-day), and nearest-neighbor models in which the current load is predicted to be the same as it was when previous conditions were closest to current conditions. Taylor *et al.* [120] compare several methods for forecasting utility-scale electric loads including exponential smoothing models, ARIMA models, neural network models (e.g., [104]), and regression with principal component analysis.

Here, we use linear regression models because – when constructed appropriately – they provide a good fit to load data in most buildings, their results are easy to interpret, they are easy to modify, and they present negligible computational burden. In addition, regression methods have performed well when compared against other load prediction methods [77, 57, 110]. In 1986, Fels [45] introduced the PRInceton Scorekeeping Method, or PRISM, a regression-based load prediction method to standardize the measurement of energy conservation savings. Heating degree-days and monthly electricity consumption are related through a simple piecewise linear regression model. Other innovative regression-based load prediction methods followed including methods using finer resolution (e.g., daily and hourly) electric

load data, change-point models [74], and multiple linear regressions [67, 66]. Kissock *et al.* [74] developed regression models specifically for commercial buildings, while Kissock and Eger [73] developed models for industrial buildings.

2.4.2 Load prediction method

We have developed a linear regression-based load prediction method that includes two novel features: a time-of-week indicator variable, and a piecewise linear and continuous outdoor air temperature dependence derived without the use of a change-point model or assumptions about when structural changes occur. Both of these of these features will be discussed in depth at the end of this section, after the load prediction method is introduced.

A facility's electric load is usually a function of both temperature and time-of-week, as shown in Figures 2.7 and 2.8.[2] Therefore, we consider both temperature and time-of-week in the regression model. Our method is as follows: we divide a week (Monday-Friday) into 15-minute-intervals (indexed by i), e.g., the first interval is from midnight to 12:15 on Monday morning, the second interval is from 12:15 to 12:30, and so on. A different regression coefficient for each time-of-week, α_i, allows each time-of-week to have a different predicted load. Additionally, we expect demand to be a piecewise linear and continuous function of outdoor air temperature, T, as described in [45, 74], for example, like that in Figure 2.9. When the outdoor temperature is high, cooling load will increase with temperature, and when the outdoor temperature is low, heating load will increase as temperature decreases (even when electricity is not used as the heat source electricity will be required to run pumps and fans when the building is heating). For some range of moderate temperatures, the load may be insensitive to temperature because neither cooling nor heating is needed (the temperature is said to be in the "dead-band"). Sometimes the outdoor air temperature may be so high that the cooling capacity cannot achieve the desired indoor temperature set point, at which point load is at the maximum possible AC load (maxed-out). Additional change-points are also possible (e.g., in facilities with two chillers, the second of which only turns on when the first is operating near capacity).

This nonlinear temperature effect can be modeled with a piecewise linear and continuous temperature-dependent load model. For each facility, we divide the outdoor air temperatures experienced by that facility into six equally-sized temperature intervals.[3] For example, if the minimum temperature experienced by the facility were 50°F and the maximum temperature experienced by the facility were 110°F, the temperature intervals would be 50–60°F, 60–70°F,

[2]Temperature and time-of-week are correlated: the highest temperatures generally occur in the afternoon and the lowest temperatures generally occur at night. Therefore, both time-of-week and temperature effects are superposed on each plot, and Figure 2.7 shows *total load* versus temperature, not *temperature-dependent load* versus temperature.

[3]Any number of intervals could be used, but we recommend using at least twice the expected number of change-points, but not so many as to cause over-fitting problems. Through trial and error, six bins were found to allow for enough change points and not cause over-fitting problems. This value is not optimized.

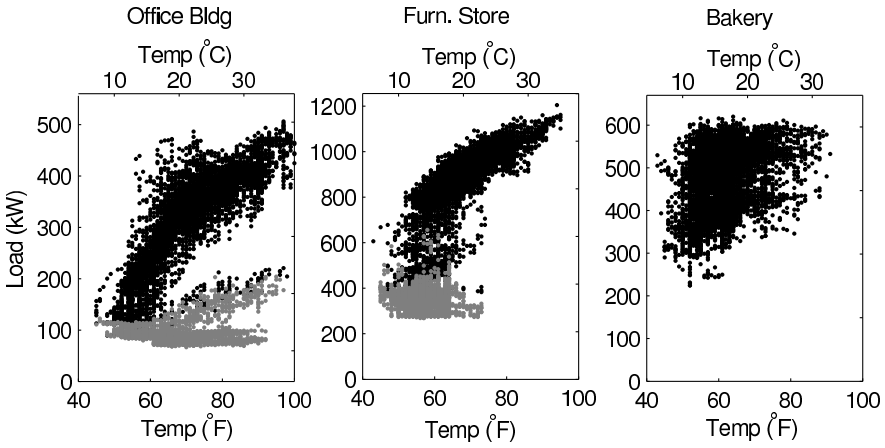


Figure 2.7: Scatter plot of load versus temperature, including both occupied mode load (black) and unoccupied mode load (gray). Data shown are from May-Sept 2008. ©2011 IEEE

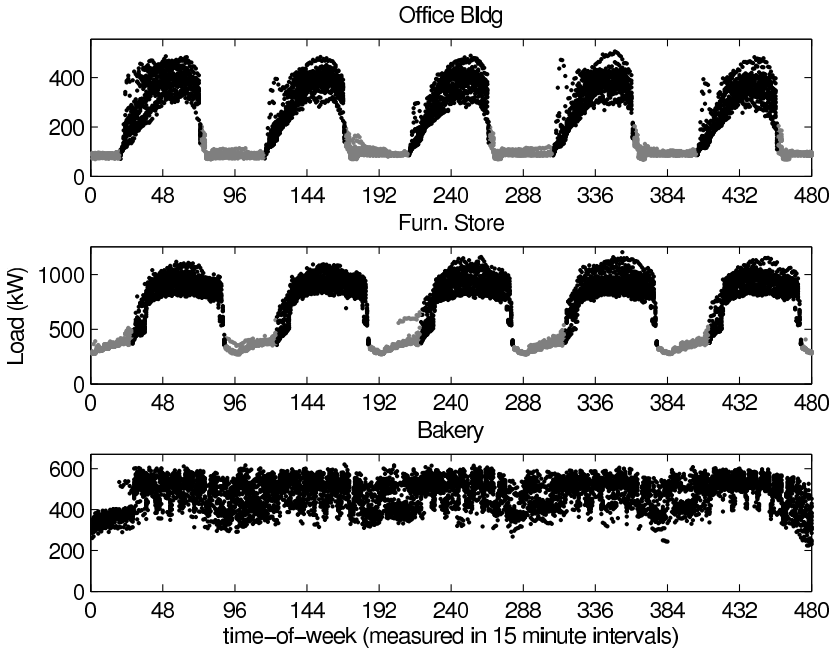


Figure 2.8: Scatter plot of load versus time-of-week, including both occupied mode load (black) and unoccupied mode load (gray). Data shown are from May-Sept 2008, so for each time-of-week there are 21-22 data points. ©2011 IEEE

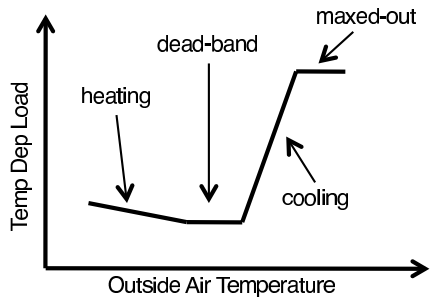


Figure 2.9: Temperature-dependent load. ©2011 IEEE

Table 2.3: Example of component temperature computation, for $B_1 = 10, B_2 = 20, ...B_5 = 50$ (in arbitrary temperature units). ©2011 IEEE

T	$T_{c,1}$	$T_{c,2}$	$T_{c,3}$	$T_{c,4}$	$T_{c,5}$	$T_{c,6}$
2	2	0	0	0	0	0
18	10	8	0	0	0	0
32	10	10	10	2	0	0
47	10	10	10	10	7	0
58	10	10	10	10	10	8

70–80°F, 80–90°F, 90–100°F, and 100–110°F. A temperature parameter, β_j with $j = 1...6$, is assigned to each outdoor air temperature interval.

To achieve piecewise linearity and continuity, the outside air temperature at time t (which occurs in time-of-week interval i), $T(t_i)$, is broken into six component temperatures, $T_{c,j}(t_i)$ with $j = 1...6$. Each $T_{c,j}(t_i)$ is multiplied by β_j and then summed to determine the temperature-dependent load. Let B_k ($k = 1...5$) be the bounds of the temperature intervals. Component temperatures are computed using the following algorithm:

1. Let B_k for $k = 1...5$ be the interior bounds of the temperature intervals.

2. If $T > B_1$, then $T_{c,1} = B_1$. Otherwise, $T_{c,1} = T$ and $T_{c,m} = 0$ for m=2...6 and algorithm is ended.

3. For $n = 2...4$, if $T > B_n$, then $T_{c,n} = B_n - B_{n-1}$. Otherwise, $T_{c,n} = T - B_{n-1}$ and $T_{c,m} = 0$ for $m = (n + 1)...6$ and algorithm is ended.

4. If $T > B_5$, then $T_{c,5} = B_5 - B_4$ and $T_{c,6} = T - B_5$.

An example computation is shown in Table 2.3.

The temperature parameters β_j are only used when a facility is operating in occupied mode since one would expect a facility's response to temperature would change at night. The start and end of the occupied mode are manually determined by looking at average load profiles on non-DR days. In Figures 2.7 and 2.8, data from occupied mode and unoccupied mode are differentiated (the bakery is never in unoccupied mode on weekdays). Analyzing Figure 2.7, it is clear that the office building and the furniture store exhibit different behavior in different modes. For all facilities, occupied load, L_o, is estimated as follows:

$$\hat{L}_o(t_i, T(t_i)) = \alpha_i + \sum_{j=1}^{6} \beta_j T_{c,j}(t_i). \tag{2.2}$$

To predict load when the building is in unoccupied mode, we use a single temperature parameter, β_u since we expect most facilities in the data set to be operating at or near the dead-band at night.[4] Unoccupied load, L_u, is estimated as follows:

$$\hat{L}_u(t_i, T(t_i)) = \alpha_i + \beta_u T(t_i). \tag{2.3}$$

The parameters α_i for $i = 1...480$, β_j for $j = 1...6$ and β_u are estimated using non-DR day load and temperature data with ordinary least squares. Each of the 487 parameters is physically meaningful: power use varies in each 15-minute interval in a week and varies as a function of outdoor air temperature. We use 15-minute interval data from May through September, so approximately 20 data points are available to estimate each α_i; hundreds or thousands to estimate each β_j; and thousands to estimate β_u. Applying the model to data from a much shorter interval, such as four or six weeks, would likely run into problems from "over-fitting," with parameter values being overly influenced by stochastic variability in the data.

We implemented the load prediction algorithm in MATLAB. Details on the implementation are given in Appendix A.2.

To test how well the load prediction method works we plot predictions on top of actual load data in Figure 2.10. We also include scatter plots in Figure 2.11. As can be seen in both figures, the prediction method works well for the office building and furniture store, but does not work as well for the bakery. The accuracy of the prediction is a function of how well the explanatory variables (time-of-week and outdoor air temperature) capture the power consumption of the facility. In the case of the bakery, we would expect that power consumption would be a function of the timing of industrial production processes, which is not captured in the model.

For the same facilities, we plot actual and predicted temperature-dependent load in both occupied and unoccupied model in Figure 2.12. The bakery's electric load is less correlated with temperature than the office building's or furniture store's electric loads.

[4] In climates where facilities transition between heating, dead-band, and cooling at night, it is advisable to use more temperature parameters to more accurately capture the load of the facility in unoccupied mode.

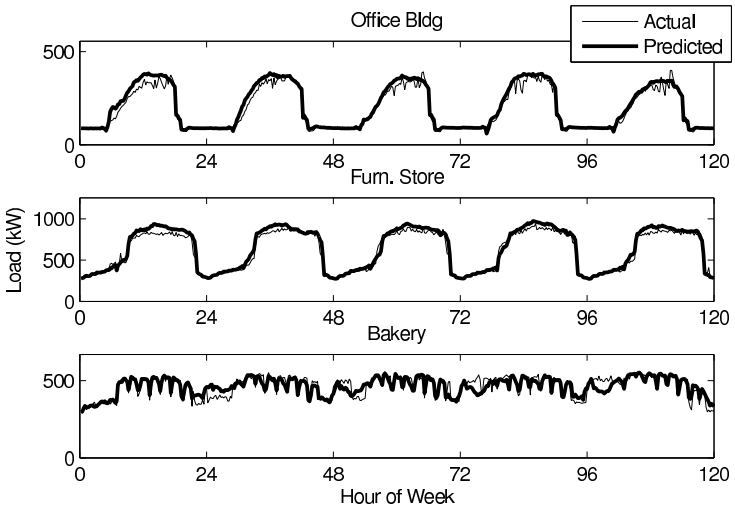


Figure 2.10: Actual versus predicted load time series for Monday-Friday, June 2-6, 2008. ©2011 IEEE

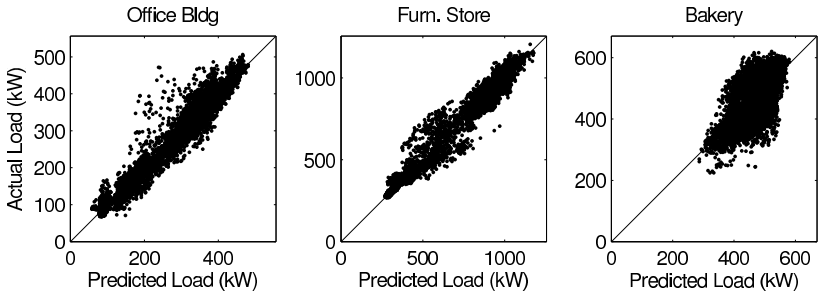


Figure 2.11: Actual versus predicted load scatter plots. Data shown are from May-Sept 2008. Significant outliers visible in the plot for the office building result from the model's inability to accurately predict the morning start-up. ©2011 IEEE

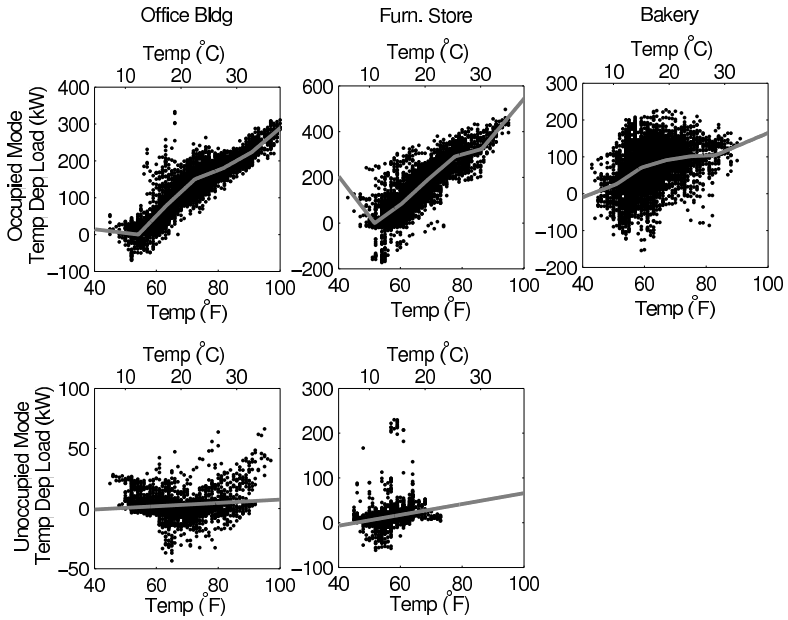


Figure 2.12: Actual (dots) and predicted (line) temperature-dependent load. Data shown are from May-Sept 2008. The gray lines are piecewise linear and continuous functions of temperature, as defined in Section 2.4.2. ©2011 IEEE

This load prediction method differs from existing methods in two ways. First, we use a time-of-week indicator variable, α_i. Most methods we have seen compute separate regressions for each time-of-day (e.g., [1] and Model 5 in [28]), but do not capture day-to-day load variation. Adding a weekday/weekend indicator variable [105] or day-of-week indicator variable [110] helps with this problem but effectively only shifts the daily load shape up or down as a function of the day-of-week; it does not allow for other day-to-day changes in the load shape, such as shorter operating hours on certain days of the week, or days of the week with consistently higher peaks. A time-of-week indicator variable solves this problem, improving model accuracy.

Second, this method is different because we avoid the use of change-point models. In much of the energy efficiency literature, change-point models are used to determine the outdoor air temperatures at which the building transitions from heating to the dead-band and from the dead-band to cooling (e.g., [74]). Our piecewise-linear modeling approach avoids the complexities of change-point models (such as the need for iterative regression) with no significant drawbacks. We divide the range of outdoor air temperatures experienced by the facility into six temperature intervals and allow the slope of the load versus temperature profile to be computed separately in each temperature interval, with the constraint that the predicted load must vary continuously as a function of temperature. We are able to do this because we are using high resolution load data (15-minute interval) and therefore have sufficient data to determine the linear temperature dependence in each temperature interval.

We explored several variations on the load prediction method, including adding parameters associated with other weather data, such as humidity, to the model. However, since all of the facilities analyzed in this study are located in low-humidity climates, we did not include humidity in the final model. We also experimented with different numbers of temperature intervals, and allowing the temperature coefficients to vary with time interval. Most of these changes did not substantially improve the model fit. Surprisingly to us, even adding predictive variables such as the overnight outdoor temperature (which we expected to help predict the load during the morning start-up) did not improve the fit very much in most cases. In the end, we settled on the model discussed above. Others are possible and may be better for some buildings or some situations.

2.4.3 Load prediction error

Reddy *et al.* [111] enumerate the many sources of error associated with using regression analysis to model building electric load. Our regression model residuals are autocorrelated and heteroscedastic and the regression parameters, α and β, are correlated. Therefore, it is very difficult to compute robust confidence intervals on the regression parameters (simply reporting the standard error associated with each regression parameter estimate, as in [45], underestimates the level of uncertainty). However, uncertainty on load predictions (as opposed to the regression parameters) can be approximated with the standard error, which can be computed at each interval, i. Computing robust confidence intervals on predictions

made on days that were not used to build the regression models (e.g., DR days) is more complicated and will be discussed in the next chapter.

2.5 Identifying DR Opportunities

In Table 2.4, we show how the methods presented in Sections 2.3 and 2.4 can help building managers generate a list of informed questions that will help them identify opportunities for DR. Specifically, we list five data analysis methods and explain each method's relevance to DR. We also describe what the building manager should look for when examining a plot or evaluating a parameter. Lastly, we list some example questions that may be generated by applying each method. While the example questions in Table 2.4 are fairly general, a building manager could generate much more specific questions after data analysis. For example, in examining Figure 2.4, one might ask: Why did the bakery's near-peak load increase during the summer, and could the additional load be curtailed during DR events or shifted outside of the DR period?

To answer these questions, the building manager will need to use the results of the data analysis to guide an evaluation of the facility's operations, controls, systems, and end uses (Figure 2.1). For example, to figure out why the bakery's near-peak load increased during the summer the building manager will first need to determine when the daily near-peak load occurs. Then, he or she can determine which loads run during those hours, how much power those loads consume, and how that has changed during the summer. With this knowledge, the building manager can determine if the additional load is curtailable or shiftable, or if it is essential to critical building operation.

The benefit of applying these methods is that the building manager is able to focus his or her evaluation on issues relating to DR (though the same methods could be used to generate questions related to energy efficiency, electricity waste elimination, or peak load management). The answers to the questions not only help building managers develop DR strategies but also help them pair their facilities with the right DR programs, since all DR programs have different requirements (e.g., minimum shed required, length of time to hold shed, predictability of shed, allowable times-of-day or days-of-week of DR events, frequency of DR events).

Table 2.4: Partial list of methods for helping to identify DR opportunities. ©2011 IEEE

Method	Relevance to DR	What to Look For?	Example Questions
Plot of daily load shape	Changes in electric load from hour-to-hour affect how much load a facility is able to shed.	How does load vary throughout the day?	Which equipment runs during each hour of the day? Which could be shed, shifted, or limited during a DR event?
Plot of near-peak and near-base load over time	Changes in electric load from day-to-day affect how much load a facility is able to shed.	How does load vary day-to-day?	Which equipment runs on the days with the highest loads? Which could be shed, shifted, or limited during a DR event?
Plot of high-load duration over time	Facilities can shed more load during hours when they are consuming more.	How long is the high-load duration, and how variable is it day-to-day and seasonally?	Which building systems/operations affect high-load duration and its variability?
Mean and standard deviation of the rise/fall time	Some facilities might have a hard time participating in DR while they are powering up/down.*	How long are the rise time and fall time, and how variable are they day-to-day?	Which building systems/operations affect rise/fall time and its variability?
Using load prediction to compare weather-normalized load shapes from two or more time periods	Variability in estimated DR sheds is a function of unmodeled load variability (discussed in Section 2.6).	How variable is the electric load over time (with weather held constant)?	Which loads are variable? Are the variable loads controllable (i.e., could they be shifted outside of DR periods)? Is load/shed predictability a requirement for the DR program?

*This is particularly relevant for facilities participating in DR to support the integration of intermittent renewable energy resources, such as wind and solar. These resources ramp in the morning and evening so facilities may be dispatched as they are powering up/down.

2.6 Evaluating DR Effectiveness

In this section, we explain how the load prediction method described in Section 2.4.2 can be used to quantify the effectiveness of DR strategies. In addition, we define the 'DR Residual' and a small set of parameters useful for characterizing the DR Residual.

To estimate the effectiveness of a facility's DR strategies, we considered adding parameters to the regression model to estimate the load shed during a DR event; however, this method would not allow us to understand shed-to-shed variability. Therefore, to analyze the effectiveness of a facility's DR strategies, we employ a 'predict and subtract' method:

1. Use the load prediction method described in Section 2.4.2 to develop a baseline model of electric load for the facility. Use only data from non-DR days to make the baseline model.

2. Acquire outdoor air temperature data from the DR day. Use the baseline model and the DR day temperatures to predict what load would have been on the DR day if the DR event had not been called.

3. For each time interval in the day, subtract the baseline prediction from the actual electric load on the DR day. We call the resulting vector the *DR Residual*. For facilities participating in ancillary services markets, the DR residual is called *pseudo-generation* [71].

We applied this method to data from each of the facilities in Table 2.1. Since all DR events were called on non-holiday weekdays in May - Sept 2008, each baseline model was constructed only with load data from non-holiday weekdays during the same period. Though all of these facilities participated in a traditional DR program (DR events were only called hot summer afternoons), the same methodology could be used to evaluate "any day, any time" DR.

In Figure 2.13, for each facility, we show the actual and baseline-predicted load for three DR days. In each case, the difference between the actual load and the baseline-predicted load is due to the facility's DR strategies and other DR event-related behavior change, and unmodeled load variability (i.e., explanatory variables in the model do not explain all of the components of the total load, resulting in model error as shown in Figures 2.10 and 2.11). Despite automation of DR strategies (described in Section 2.2.1), for each facility, there is variability in estimated DR sheds. Shed variability results from unmodeled load variability, and, possibly, factors such as occupant behavior change on DR days; changes in building operations, controls, equipment, and end-uses from event-to-event; and changes in shed capacity as a function of outdoor air temperature, occupancy, total facility load, and other variables. To fully understand shed-to-shed variability, we need to compute prediction error on DR days, which is a subject of current research.

Because of load/shed variability from one DR event to the next it is instructive to look at plots of averages. In Figure 2.14, for each facility, we plot the average baseline-predicted

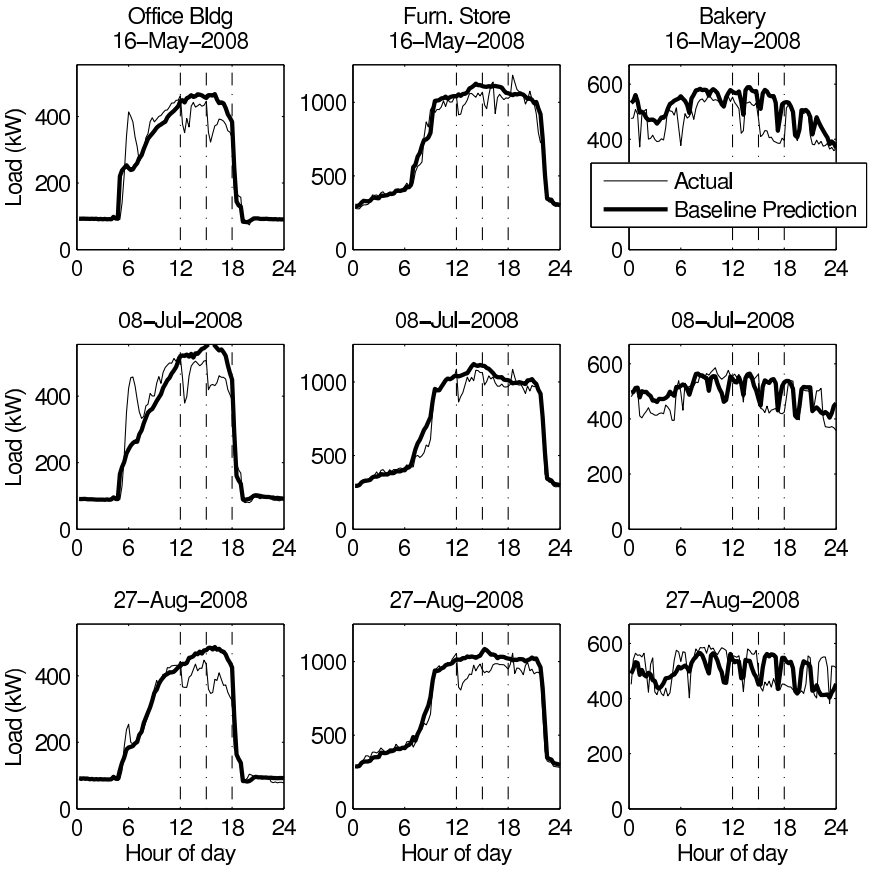


Figure 2.13: Actual and predicted load on three DR days (rows) for three facilities (columns). Vertical lines show the moderate and high price periods. ©2011 IEEE

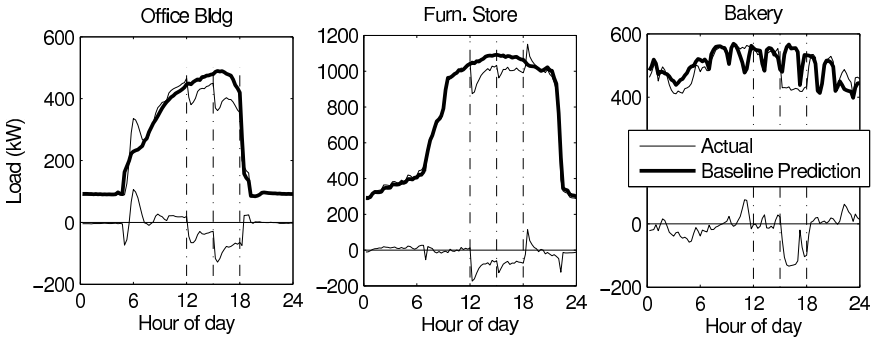


Figure 2.14: Average actual load, predicted load, and DR residual. Vertical lines show the moderate and high price periods. ©2011 IEEE

load, actual load, and DR residual for all eleven DR events days in 2008. The average DR residuals tell us many things including:

- the office building and furniture shed load during both the moderate and high price periods, while the bakery only sheds load during the high price period;
- the baseline model is unable to capture the office building's morning start-up; and
- the furniture store's load rebounds after the DR event.

A small set of parameters can be used to characterize DR residuals (Table 2.5 and Figure 2.15). The list is based on that developed by Mathieu *et al.* [91]. *Average demand shed* is important because it tells us, on average, how much load is shed during the DR event. *Intra-shed variability* is a way for us to capture some of the dynamics of the shed. If intra-shed variability is small relative to the average demand shed, the shed was held steady during the DR event, while, if it is large, the shed bounced around, increased, or decreased during the DR event. *Residual ramp time* tells us how quickly the facility sheds load, and *rebound* tells us how the facility behaves after the DR event. *Daily peak demand* and *daily energy* tell us how the DR event affects the facility's daily power and energy use.

These parameters can be computed for a certain DR event or an average DR event. In Table 2.6, for each facility, we give values for each parameter for each DR day shown in Figure 2.13 and for the average DR day shown in Figure 2.14. There is error associated with each parameter value, and the values for the means are more certain than the values for the individual events. For each facility, intra-shed variability is relatively high as compared to average demand shed. Certain DR strategies, such as changing HVAC set points, often lead to more intra-shed variability than strategies such as switching off lights or industrial processes. Both the county building and the furniture store can shed load within 15 minutes of the start of the DR period, while it takes the bakery 15-30 minutes. While all of the

Table 2.5: DR residual parameters and definitions. ©2011 IEEE

DR Residual Parameter	Definition
Average demand shed (kW)	Baseline minus actual average load during DR event.
Intra-shed variability (kW)	Standard deviation of demand shed during DR event.
Residual ramp time (min)	Duration for load to drop to average demand shed during DR event.
Rebound (kW)	Actual minus baseline average load in hour after DR event.
Daily peak demand (%)	Actual divided by baseline-predicted daily peak demand.
Daily energy (%)	Actual divided by baseline-predicted daily energy use.

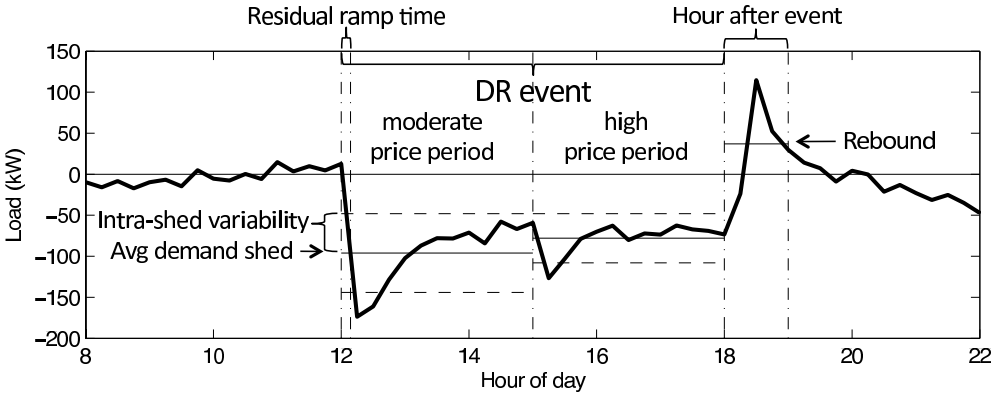


Figure 2.15: Several DR parameters shown on the furniture store's DR residual (from Figure 2.14). ©2011 IEEE

facilities use less *total* energy on most DR days, both the furniture store and the bakery may have higher *peak* demand on days when they participate in a DR event. Though a facility's DR day peak demand generally occurs outside of the DR event period (e.g., during the hour after the DR event, as is the case for the furniture store), these peaks affect demand charges.

2.7 Chapter Conclusion

We have presented methods for analyzing 15-minute interval electric load data from C&I facilities. Applying these methods could help building managers ask the right questions to discover opportunities to reduce electricity bills through DR, energy efficiency, electricity waste elimination, and peak load management. We have focused on DR because of a lack of existing methods and tools for identifying DR opportunities and estimating DR effectiveness, and because the use of DR is expanding.

A number of key findings from this study are as follows:

- Plotting time series electric load data is useful for understanding electricity consump-

Table 2.6: DR parameter values. ©2011 IEEE

Facility	Event Date	Average Demand Shed* (kW)	Intra-Shed Variability* (kW)	Residual Ramp Time* (min)	Rebound (kW)	Daily Peak Demand (%)	Daily Energy (%)
County Bldg	May 16	44 / 74	27 / 34	0-15 / 0-15	2	99	98
	July 8	56 / 93	39 / 28	0-15 / 0-15	5	93	99
	Aug 27	51 / 103	15 / 19	0-15 / 0-15	-15	92	93
	Mean**	42 / 88	21 / 25	0-15 / 0-15	-1	95	95
Furniture Store	May 16	73 / 67	33 / 62	0-15 / ***	50	105	97
	July 8	80 / 59	47 / 27	0-15 / ***	17	97	96
	Aug 27	124 / 83	50 / 42	0-15 / ***	-40	97	95
	Mean**	96 / 78	48 / 30	0-15 / ***	37	103	97
Bakery	May 16	— / 129	— / 34	— / 15-30	-7	95	90
	July 8	— / 82	— / 40	— / 15-30	-4	104	96
	Aug 27	— / 79	— / 37	— / 15-30	-68	105	101
	Mean**	— / 102	— / 39	— / 15-30	4	101	97

* Calculated for the moderate / high price period, except for the Bakery which did not shed load during the moderate price period.
**Calculated over all eleven events that occurred in the summer of 2008.
***The average demand shed during the moderate price period is higher than that during the high price period so it does not make sense to calculate this.

tion patterns and changes to those patterns, but results may be misleading if data from different time intervals are not weather-normalized.

- Parameter plots highlight key features of electric load data and may be easier to interpret than plots of time series data.
- A time-of-week indicator variable (as compared to time-of-day and day-of-week indicator variables) improves the accuracy of regression models of C&I facility electric load.
- A piecewise linear and continuous outdoor air temperature dependence can be derived without the use of a change-point model (which would add complexity to the modeling algorithm) or assumptions about when structural changes occur (which could introduce inaccuracy).
- DR shed variability is a function of unmodeled load variability, and, possibly, other factors such as occupant behavior change on DR days; changes in building operations, controls, equipment, and end-uses from event-to-event; and changes in shed capacity as a function of outdoor air temperature, occupancy, total facility load, and other variables.

- For measurement and evaluation accuracy, DR parameter values should be estimated at the end of the summer using all available data to build the baseline model, instead of directly after an individual event using only data before the event to build the baseline model.

- DR parameters values computed for an individual event are far less certain than values computed for an average event (e.g., the average of all events in a summer); therefore, it is good practice to report values associated with means along with values associated with individual events.

To assess and evaluate "fast DR," such as facilities providing ancillary services [71], which occur on time scales of seconds to minutes, high resolution (e.g., 5 minute and 4 second interval) electric load data are needed. Though we present methods for analyzing 15-minute interval electric load data, the same methods could be used to analyze higher resolution data.

In addition, some of the methods presented here could be used by building managers for real-time feedback control of DR resources. For example, at specific time intervals during a DR event, a baseline model built with data from non-DR days before the DR day could be used together with actual DR day temperature data to predict load. Subtracting the load prediction from the actual load results in the real-time DR residual, which can be used for feedback control [71]. A drawback to this approach is that the baseline model is built with fewer data, all from before the DR day. An existing barrier to this approach is that, while load and temperature data are often available in real-time, many existing building control systems are unable to handle these data in real-time.

The methods presented here could be translated into easy-to-use tools for building managers, helping them determine which DR strategies to implement and which DR programs to enroll in, and then evaluating the effectiveness of the DR strategies they have implemented. These tools could be integrated into real-time operating platforms to assist building managers in all forms of building energy management. Future work should include developing methods to quantify error in DR predictions, including an analysis of shed variability.

Chapter 3

Baseline Model Error & DR Parameter Variability

In this chapter, we use the baseline model introduced in the previous chapter to understand the performance of 38 C&I facilities participating in an automated dynamic pricing DR program in California. In this program, facilities are expected to exhibit the same response each DR event. However, baseline model error makes it difficult to precisely quantify changes in electricity consumption and understand if C&I facilities exhibit event-to-event variability in their response to DR signals. Therefore, we present a method to compute baseline model error and a metric to determine how much observed DR variability results from baseline model error rather than real variability in response. This chapter is largely based on [88][1] and [89].[2]

3.1 Chapter Introduction

Buildings are becoming increasingly important as active resources that support power system operations. Though buildings have played a small role in power systems operations in the past – either with relays that interrupt power to air conditioners and water heaters [117, 41], or by "voice dispatch" of large commercial and industrial loads [40] – recent Smart Grid investments are demonstrating the potential for buildings to become grid-interactive resources that are just as controllable as – or even more controllable than – electricity generators [20].

[1] ©2011 IEEE. With permission from my co-authors: Duncan Callaway and Sila Kiliccote. "Examining uncertainty in demand response baseline models and variability in automated responses to dynamic pricing." *IEEE CDC-ECC*, Dec 2011.

[2] ©2011 Elsevier. With permission from my co-authors: Duncan Callaway and Sila Kiliccote. "Variability in automated responses of commercial buildings and industrial facilities to dynamic electricity prices." *Energy and Buildings*, 2011.

In DR programs, power system operators can achieve system-wide demand reductions by providing financial incentives for buildings to change their electricity consumption patterns through both shifts in energy use and load reductions, or sheds. Buildings generally participate in DR by enrolling in dynamic electricity pricing programs or demand/capacity bidding programs. In dynamic pricing programs, buildings face high electricity prices during hours when the grid is stressed, encouraging them to shed load or shift energy use to less expensive hours. In capacity/demand bidding programs, buildings bid load reductions and, if called upon, shed load at certain times in exchange for payment. In this chapter, we focus on C&I facilities participating in a dynamic electricity pricing program. These facilities use the OpenADR Communication Specification [107] to receive DR event notifications from the utility, and during events they automatically execute pre-programmed DR strategies [127].

The central challenge we address in this chapter is that DR parameters, such as Average Demand Shed, Rebound, Daily Peak Demand, and Daily Energy (which we define in Table 3.2), must be measured relative to an estimate of how much electricity a facility would have consumed in the absence of the DR event. DR parameters are computed by subtracting a counterfactual baseline from the actual power consumption of the facility. Therefore, DR parameters will exhibit variability due to both baseline model error and real variability in the facility's response. We will use the following terms throughout the chapter:

1. **Unmodeled load variability**, or baseline model error, is load variability that is not captured by a baseline model and not due to a DR signal. Unmodeled load variability complicates DR programs that use baselines for financial settlement (e.g., demand/capacity bidding programs and programs in which loads participate in wholesale ancillary services markets). Moreover, even DR programs that do not use baselines for settlement (e.g., dynamic electricity pricing programs) use baselines for Measurement and Verification (M&V) and to calculate the cost-effectiveness of the DR programs [59].

2. **Real DR variability** is event-to-event variability in a facility's actual response, for example, due to building managers and/or occupants overriding pre-programmed DR strategies; broken equipment; and variability in responses as a function of occupancy, weather, and other variables.

3. **Observed DR variability** occurs as a result of the combination of unmodeled load variability and real DR variability.

Figure 3.1 illustrates the concepts of baseline model error and observed DR variability. In this figure, we plot the actual and baseline-predicted load for an office building on two DR days and one normal day. The left and middle plots show that responses to DR signals can seem variable – and may, in fact, be variable. The right plot demonstrates baseline model error.

The purpose of this chapter is to understand the variability of C&I facility responses to DR events. The question is important for two reasons. First, in order to efficiently

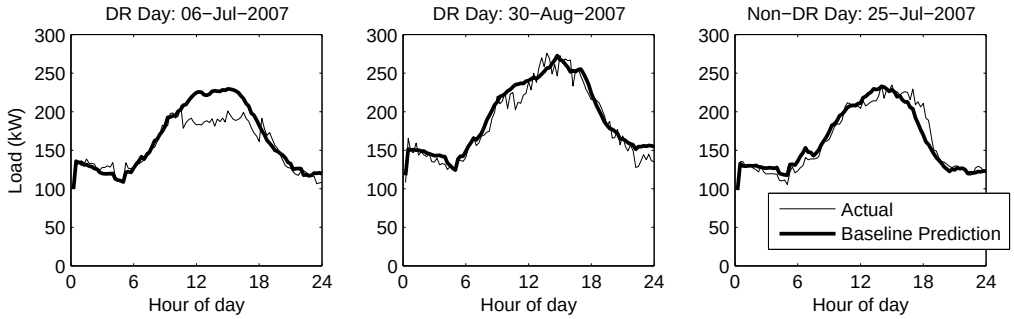


Figure 3.1: Actual and baseline-predicted demand for an office building on three different days during the summer of 2007. The left and middle plots show data from DR event days (the difference between the actual and the baseline prediction is a combination of the response to the DR signal and model error), while the right plot shows data from a normal day (the difference between the actual and the baseline prediction is model error). ©2011 IEEE

allocate generating resources, power system operators must predict how aggregations of facilities will respond on demand response days. If all observed DR variability resulted from unmodeled load variability, a power system operator could expect consistent DR behavior and would only need to deal with the usual amount of demand-side variability. However, if real DR variability is present, the DR program may create an extra burden of variability for the system operator to manage. This could require additional power system services (e.g., reserves). In extreme cases, real DR variability could result in significant deviations in grid frequency or expected power flow.

The second reason variability is important is because DR programs are evaluated on the basis of whether or not facilities (individually or in sum) appear to have reliable responses on DR days. M&V of utility DR programs, including those that do not use baselines for settlement, often include analyses of the DR performance (versus a baseline) of both individual facilities and aggregations of facilities [127]. Variability may affect the evaluation of the DR program and determinations about whether or not a facility is suitable for DR [13]. Moreover, observed DR variability in an individual facility affects how the facility perceives its own DR performance from event-to-event. A facility whose performance seems inconsistent from one event to another may be tempted to modify its DR strategy; however, the perceived inconsistency may have been caused by baseline model error.

Real DR variability is the most relevant measure for power system operators and DR program evaluators. However, real DR variability can only be estimated indirectly, by thoroughly characterizing unmodeled load variability and relating it to observed DR variability. Therefore, in this chapter, we first compute the error associated with DR parameter estimates (e.g., demand shed estimates) for 38 C&I facilities that participated in an automated dynamic electricity pricing program in California. We then construct a variability metric

that captures the relative importance of real DR variability versus unmodeled load variability, and compute this metric for all of the facilties. We find that most observed DR variability is the result of baseline model error.

A note on terminology: The DR community uses several different terms to denote the counterfactual power usage on DR days: baselines, predictions, and forecasts. In this chapter, we use the term 'baseline predictions' to refer to ex-post estimates of counterfactual power usage computed with regression parameters (identified with historical demand/temperature data) and *actual* temperature data for the purpose of M&V. We reserve the term 'forecast' for ex-ante estimates computed with *forecasted* temperature data, which we do not discuss in this chapter. We use the term 'DR parameter estimates' to refer to values, such as demand sheds, computed with actual demand data and baseline predictions. The DR community often refers to these values as 'DR calculations'; however, we prefer our terminology because it makes clear that the values are uncertain. The term 'DR parameter estimates' should not be confused with 'DR estimates,' engineering estimates of expected demand sheds.

The rest of this chapter is organized as follows: In Sections 3.2 and 3.3, we describe our data and baseline model. In Section 3.4, we explain our error analysis. Then, in Section 3.5, we present our results and discussion with respect to baseline model error and DR variability.

3.2 Methods

Fifteen-minute interval whole building electric load data and weather data are described in Sections 2.2.1 and 2.2.2. More details are provided in Appendix A.1.

In 2006, DR events were called separately in two geographic zones: Zone 1 included San Francisco and the San Francisco Peninsula, while Zone 2 included the rest of PG&E's service territory. Nine DR events were called in Zone 1 and eleven in Zone 2. In both 2007 and 2009, twelve events were called, while in 2008 eleven events were called. Several facilities participated in only a portion of the DR events in a year. If we knew that a facility did not participate in a certain DR event, we did not analyze data from that DR day.

Facilities' demand profiles change year-to-year due to equipment upgrades, changes in usage patterns, etc. To reduce the chance of creating baseline models with data from before and after significant structural changes only one year worth of data were used to create each model. In total, we have 87 facility-years worth of data (Table 3.1), where a facility-year is defined as one year of data for one facility. Twelve facility-years of available data were not analyzed because of significant structural changes visible in the data. Details are given in Appendix A.3.

Aggregate metrics were computed by summing power for all facilities for each year (and, in 2006, for each zone). We excluded facilities that did not participate in all of the DR events in a year and facility-years for which we were missing more than one week of data. In sum, nine facility-years were not included in the aggregate populations (hence the discrepancy in number of facilities between Tables 3.1 and 3.4). All aggregate results are computed from

Table 3.1: Number of facilities by year and facility type. ©2011 IEEE

Year (Zone)	Number of Events	Office Buildings	Industrial Facilities	Retail Stores	Retail with PV*	Prisons & Jails	Museums	**TOTAL**
2006 (1)	9	3	0	1	0	0	0	**4**
2006 (2)	11	6	0	1	0	1	1	**9**
2007	12	7	1	3	1	1	1	**14**
2008	11	12	8	2	3	1	1	**27**
2009	12	17	8	1	3	3	1	**33**
TOTAL	**45**	**17**	**8**	**7**	**6**	**4**	**87**	

*Retail stores with solar photovoltaics (PV).

baseline models built with the aggregate data, not the aggregate output of individual baseline modes. A detailed description of aggregate system synthesis is presented in Appendix A.5.

3.3 Baseline Model & DR Parameters

Electric utilities generally use simple models to determine baseline electric load on DR days for financial settlement and/or M&V. Many of these models involve averaging the daily electric demand over several days (e.g., those with the highest energy usage) before the DR day [52, 28]. Unfortunately, baseline models built by averaging can be biased. Regression-based baseline models, which are less likely to suffer from bias, have long been used for M&V by the energy efficiency community [45, 67, 74, 73] and are increasingly used for DR M&V [52, 28, 69, 10]. These models typically relate electric demand to weather and, sometimes, other relevant parameters. More sophisticated baseline modeling methods (e.g., neural networks) have been proposed, but are seldom used in practice.

We use the regression-based baseline model described in the previous chapter (Section 2.4.2) because it performs similarly to or better than baseline models commonly used for DR M&V. Therefore, the magnitude of the error associated with this model is comparable to or less than that associated with common baseline models. Using better baseline models not only allows one to compute more accurate DR parameter estimates, but also allows one to better determine if a facility exhibits real variability in its responses to DR events.

Since all 2006-2009 DR days were called May 1 to Sept 30, baseline models were constructed with non-DR day demand data during the same period. We did not use data from holidays, weekends, or days that appeared to have had power outages (i.e., days when the minimum power use is less than a percentage of the average minimum daily power use during the summer) to build the baseline models. Details on the MATLAB implementation are given in Appendix A.2.

The regression parameters α, β, and β_u are estimated with Ordinary Least Squares (OLS). We use the OLS estimator because, though it not 'best' (in a Gauss Markov sense)

due to autocorrelation and heteroscedasticity (see Section 3.4), it still produces unbiased regression coefficients [111, 112]. However, the standard errors associated with the regression coefficients are underestimated, so we do not use them.

The parameter estimates and temperatures on DR days are then used to predict demand on DR days. Four DR parameters (Table 3.2), computed from the baseline predicted demand and the actual demand, are used to characterize changes in electricity use on DR days. These parameters are computed with 15-minute interval data and 15-minute interval predictions. For example, average demand shed is computed by averaging the 15-minute interval demand data during the event, averaging the 15-minute interval demand predictions during the event, and computing the difference of the averages. These parameters are similar to those defined in the previous chapter; however, here we define Daily Peak Demand and Daily Energy slightly differently: as absolutes, not percentages.

Table 3.2: DR parameter definitions, meanings, and importance. ©2011 IEEE

Parameter	Definition	If this value is positive...	Importance
Average Demand Shed (kW)	Predicted minus actual average demand during the DR event.*	...the facility reduced power use during the event.	Key indicator for how well the facility performed.
Rebound (kW)	Actual minus predicted average demand in the hour after the DR event (6-7pm).	...the facility increased power use after the event.	Could affect a facility's demand charges; synchronized rebounds could create a new system-wide peak.
Daily Peak Demand (kW)	Actual minus predicted maximum demand on the DR day.**	...the facility had a higher demand peak than it would have if there was no DR event.	Could affect a facility's demand charges; will not affect the system-wide peak unless the individual peaks are synchronized.
Daily Energy (kWh)	Actual minus predicted total energy use on the DR day.	...the facility used more energy than it would have if there was no DR event.	Suggests whether energy shifting or shedding strategies predominate; implicated in understanding DR's effect on energy use and the environment, a research gap [23].

* The average demand shed is computed separately for each price period: 'Shed 1' refers to the moderate price period (12-3pm) and 'Shed 2' refers to the high price period (3-6pm).

** The actual and the baseline peak could happen at different times during the day.

3.4 Error Analysis

Most error analyses on regression-based baseline models use the standard errors associated with the regression coefficients [45, 73, 10]. However, these errors underestimate the true error due to a number of issues. First, the regression parameters are correlated. Specifically, time-of-week is correlated to temperature: the highest temperatures tend to occur in the afternoon and the lowest temperatures occur overnight. Second, the regression residuals are autocorrelated. In Figure 3.2, we show autocorrelation functions (ACF) and partial autocorrelation functions (PACF) computed with regression residuals from two facility-years. In both cases, the residuals are lag 1 autocorrelated, which is the case for all facility-years. In some cases, we find higher order autocorrelation. Other methods of detecting autocorrelation are described in Appendix A.4.1.

Third, the regression residuals are heteroscedastic. Specifically, we find that the variance of the regression residuals (referred to as the 'error variance') is a function of time-of-week. For a typical commercial building, error variance tends to be lower at night and higher during the day when fluctuating occupancy affects loads. For some facilities, the error variance is high during transition periods (e.g., when the facility is being populated in the morning). Figure 3.3 shows plots, created using (2.2) and (2.3), of error versus time-of-week. For the retail store, error is clearly a function of time-of-week, while for the office building, the effect is smaller. These results not only demonstrate heteroscedasticity, but also the importance of computing errors as a function of time-of-week. We have not computed error as a function of temperature or predicted demand because error does not seem to be a strong function of these variables as detailed in Appdenix A.4.2.

These issues suggest that one should use caution in interpreting the standard errors associated with the baseline model regression coefficients. Fortunately, as we will explain in the method description below, we do not need to calculate this in order to calculate the error associated with DR parameter estimates.

3.4.1 Method

The goal of our error analysis is to determine the error associated with each DR parameter estimate for each facility-year and each aggregation of facilities. Other studies have used regression residuals to generate baseline model error estimates [69]; however, regression residuals are self-influenced: the model is built and tested on the same data set. Therefore, error estimates generated with regression residuals underestimate the true error.

To avoid self-influence, we use a resampling technique called 'Leave One Out Cross Validation' (LOOCV). (More details on this choice are given in Appendix A.4.3). LOOCV is a type of K-fold cross validation, which involves randomly partitioning the data into K subsamples, reserving one subsample, building the model with data from the remaining subsamples, testing on the reserved subsample, and repeating this process for all K subsamples. The results for each subsample are combined resulting in an estimate of the prediction accuracy.

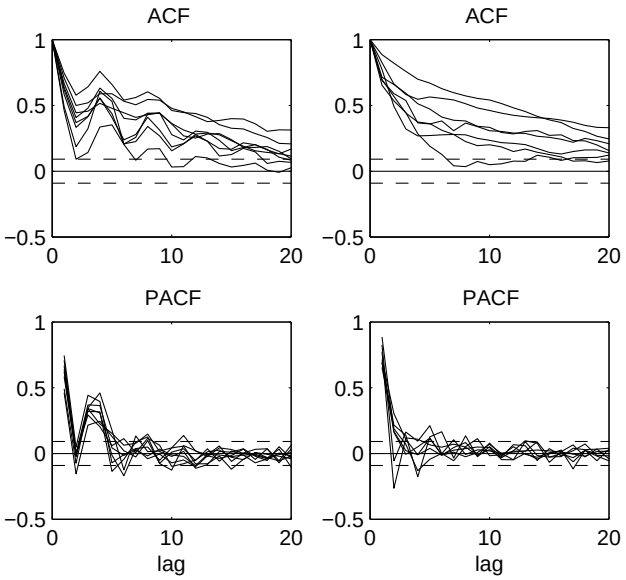


Figure 3.2: ACF and PACF computed with the regression residuals from an office building (left) and a retail store (right) in 2008. Each line was created with data from a week (Mon-Fri) in which there were no DR days, holidays, or power outage days. Dashed lines show the 95% confidence interval ($\pm 2/\sqrt{n}$, where n is the number of data points in the data set).

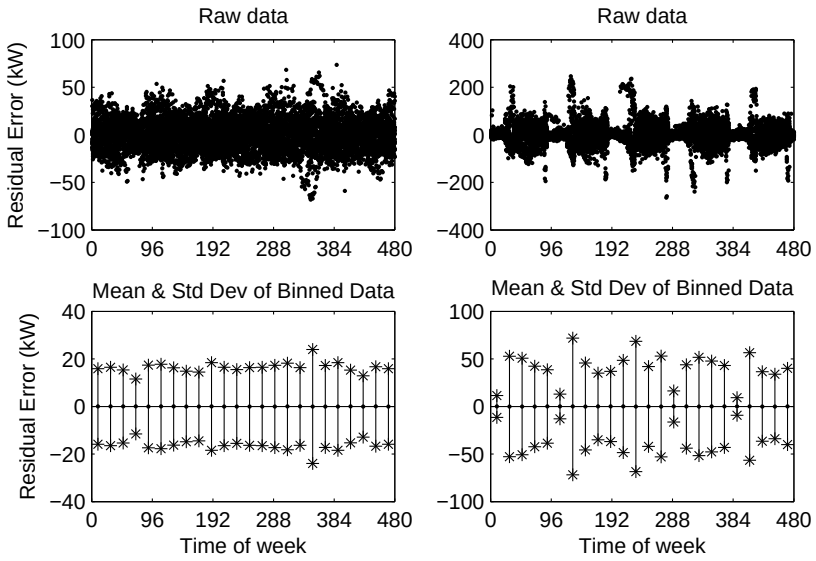


Figure 3.3: Error versus time-of-week for an office building (left) and a retail store (right) in 2008. ©2011 IEEE

In LOOCV, K is equal to the total number of observations, n. LOOCV is useful when n is small, though the technique is computationally intensive.

We treat the demand on each non-DR day as an observation. Therefore, n is equal to the number of non-DR days used to create the baseline prediction model ($\sim 90 - 95$ days per facility-year). We leave out one non-DR day, build the model with data from the rest of the non-DR days, predict the demand on the day that has been left out, compute the quantities associated with the DR parameters (e.g., average demand between 12 and 3 pm), compare the predictions to the actual quantities to generate an error observation, and repeat for each non-DR day. Since we consider error as a function of time-of-week, only residuals computed with data from Mondays are used to determine errors on Mondays, etc. Therefore, for each DR parameter for each day of week there are only $\sim 18-20$ error observations. It is difficult to determine the true error distribution with so few error observations (see Appendix A.4.4). Therefore, we assume that the error observations are normally-distributed and report error estimates as one standard deviation of the error observations.

We do not recommend using this error analysis method on baseline models parameterized with DR day data (e.g., morning adjustments [28]). For those models, this method will underestimate true model error if power use outside of the DR period is affected by the DR signal, which is common, especially for facilities that pre-cool, rebound, or otherwise shift energy use to the morning or evening on DR days.

3.4.2 Other sources of error

Ideally, an error analysis should quantify all possible sources of error associated with an estimate. The methodology described above accounts for demand/temperature measurement error; error resulting from the fact that the weather stations are not co-located with the facilities; error resulting from temperature data interpolation; and unmodeled load variation on days similar to those used to build the baseline model. There are two sources of error we have not quantified: over-fitting and extrapolation. DR days are generally called on the hottest days of the summer which means that, in some cases, baseline predictions are made with temperatures higher than those on non-DR days, resulting in extrapolation error; or experienced only a few times on non DR-days, resulting in over-fitting error. Over-fitting/extrapolation error could increase error variance and/or introduce bias in predictions made with high outdoor air temperatures.

For 26% of our DR day baseline predictions, the highest temperature on the DR day is greater than the highest temperature used to build the baseline model. We found that model error associated with extrapolated baseline predictions is comparable to that associated with non-extrapolated baseline predictions (see Appendix A.4.5). Also, as described previously, error variance does not appear to be a strong function of temperature, even the highest temperatures used to build baseline models (i.e., temperatures for which we might expect possible over-fitting). A more detailed analysis of over-fitting error is given in Appendix A.4.6. Other baseline models, such as those that use fewer data to build the model, may be more susceptible to over-fitting/extrapolation error. Baseline models that model a load as a purely linear function of temperature may suffer from extrapolation error. If over-fitting/extrapolation error is present, the method presented in Section 3.4.1 will underestimate DR parameter error. However, if error variance is not a function of temperature and consequently over-fitting/extrapolation introduces only bias, over-fitting/extrapolation error will not affect the accuracy of the DR parameter variability metrics (introduced in Section 3.5.2).

3.5 Results & Discussion

3.5.1 DR parameter errors

The error analysis method presented in Section 3.4.1 allows us to assign error estimates to DR parameter estimates. In Figure 3.4, we show DR parameter and error estimates for all 2009 facility-years and the 2009 aggregate population. In most cases, the error estimates are large relative to the DR parameter estimates. For example, on average, across all years, the error associated with Shed 1 is approximately $\pm 120\%$ of the parameter values and the error associated with Shed 2 is approximately $\pm 180\%$ of the parameter values. In addition, observed DR parameter variability is often large. However, given the magnitude of the error estimates, we would expect some observed DR parameter variability.

This interpretation of Figure 3.4 illustrates how including error estimates along with DR parameter estimates allows us to draw the right conclusions from the data. Without error estimates, it would be easy to classify a facility with observed shed variability as a variable shedder, and, therefore, conclude that such a facility is difficult to control. However, if the error associated with that facility's shed estimates is large, then it is possible that the response is actually consistent and we are simply unable to measure the exact response because of baseline model error.

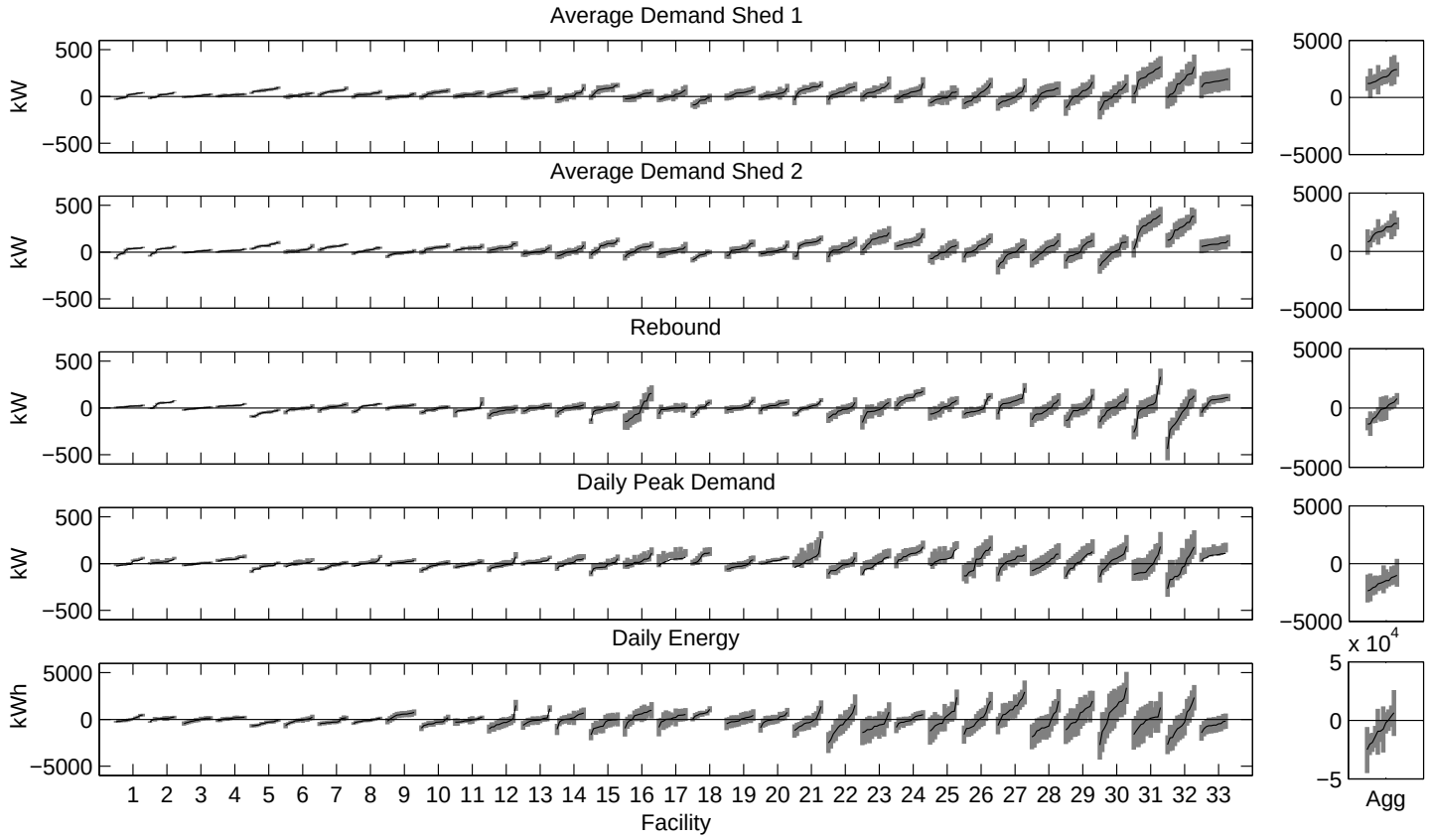


Figure 3.4: DR parameter estimates (black) and error estimates (grey) for all 2009 facility-years. Facilities are arranged in order of smallest to largest mean error for Average Demand Shed 1. For each facility and each DR parameter, parameters are arranged in order of smallest to largest. Results for the 2009 aggregate population are shown on the right.

There are several other things to learn from Figure 3.4. Some facilities that shed power during DR events consume less energy on DR days, while some do not, meaning that they shift load outside of the DR period. We also find that the Daily Peak Demand is often biased low, because regression-based baseline models tend to under-predict maximum values (i.e., outliers). The aggregate population results demonstrate that DR works: the aggregated facilities shed power during DR events and reduce the peak demand on DR days, despite the fact that individual facilities may become peakier. Also, on average, the aggregated facilities exhibit almost no rebound and save some energy on DR days, indicating that there is some net curtailment–the facilities do not simply shift all load outside of the DR period.

We do not discuss the statistical significance of the DR parameter estimates because the error estimates are not confidence intervals. Since a facility's DR behavior from one DR event to the next is not independent, Bayesian techniques should be used to not only determine appropriate confidence intervals, but also pinpoint DR parameter estimates. This would involve pooling information across DR events (i.e., using knowledge about a facility's behavior during one DR event to help us predict its behavior during another DR event). We do not tackle this here because we are interested in using the error estimates to assess DR parameter variability, not statistical significance.

3.5.2 DR parameter variability

Observed DR parameter variability has two possible sources: unmodeled load variability and real parameter variation. For example, consider the Average Demand Shed. We generally observe shed variability from one DR event to the next. We would like to know if observed shed variability is a result of real shed variability (i.e., a facility curtails a different amount from event-to-event) or if it results from unmodeled load variability (i.e., baseline model error). If observed shed variability results exclusively from unmodeled load variability, then we can expect consistent responses and the system operator need only deal with the usual level of demand-side variability. If real shed variability exists, the system operator may require additional reserves to deal with more demand-side variability than usual.

We first derive a metric, the Average Demand Shed Variability Metric (SVM), to discern between unmodeled load variability and real DR parameter variation. Similar derivations yield metrics for each DR parameter: the Rebound Variability Metric (RVM), Daily Peak Demand Variability Metric (PVM), and Daily Energy Variability Metric (EVM). Then, we present DR parameter variability metric results for the individual facility-years and the aggregate populations, respectively.

SVM derivation

On a DR day, the *Observed Load* (OL) is equal to the *Real Baseline Load* (RBL) minus the *Real Shed* (RS):

$$OL = RBL - RS \quad . \tag{3.1}$$

Neither the real baseline load nor the real shed can be measured. The real baseline load is estimated with the *Predicted Baseline Load* (PBL). The difference between the real baseline load and the predicted baseline load is the *Unmodeled Load* (UL):

$$UL = RBL - PBL \quad . \tag{3.2}$$

To compute the *Observed Shed* (OS), the predicted baseline load is subtracted from the observed load:

$$OS = OL - PBL = UL - RS \quad . \tag{3.3}$$

Our goal is to determine the variance of the real shed. Therefore, we take the variance of (3.3), which results in:

$$\mathrm{Var}(OS) = \mathrm{Var}(UL) + \mathrm{Var}(RS) - 2\mathrm{Cov}(UL, RS) \quad . \tag{3.4}$$

We can estimate $\mathrm{Var}(OS)$ by taking the variance of the $9-12$ observed sheds and $\mathrm{Var}(UL)$ by taking the variance of the ~ 95 error observations (since DR events can occur on any weekday, error observations are used without regard to day-of-week). Therefore, we define the shed variability metric (SVM) as:

$$\begin{aligned} \mathrm{SVM}: &= \mathrm{Var}(OS) - \mathrm{Var}(UL) \\ &= \mathrm{Var}(RS) - 2\mathrm{Cov}(UL, RS) \quad . \end{aligned} \tag{3.5}$$

While the SVM does not tell us the exact value of $\mathrm{Var}(RS)$ due to the complicating covariance term, it does tell us if real shed variability likely exists or not. Also, since $\mathrm{Var}(RS) \geqslant 0$, the SVM may tell us something about the sign of the covariance term. If the covariance term is positive, then as unmodeled load increases, real shed increases. This could occur when the equipment that drives the unmodeled load is also the equipment that is curtailed. Alternatively, if the covariance term is negative, then as unmodeled load increases, real shed decreases. This could occur when load is higher than predicted, electricity consuming services are in high demand, and occupants/building operators override automated DR strategies; or when load is higher than predicted, the HVAC system is operating at or beyond its maximum capability, and consequently a reduction in HVAC setpoint has a limited effect.

Individual facility-years

To compare facilities by SVM, we normalize the measurements of the unmodeled load and the observed shed such that $\mathrm{Var}(UL) = 1$. Therefore, the minimum value of SVM is -1 (i.e., when $\mathrm{Var}(OS) = 0$). Each DR parameter variability metric is normalized similarly.

Histograms showing DR parameter variability metrics for the 87 facility-years are shown in Figure 3.5. To understand what these histograms tell us about real parameter variability, we can compare them to distributions generated for the case when real parameter variability is zero. If real parameter variability were zero, the covariance term would also be zero,

Table 3.3: Number and percentage of facility-years with variability metrics inside and outside the 95% confidence bounds. ©2011 IEEE

Metric	Inside Bounds	Outside Bounds		
		Below	Above	TOTAL
SVM1	65 (75%)	8 (9%)	14 (16%)	22 (25%)
SVM2	62 (71%)	4 (5%)	21 (24%)	25 (29%)
RVM	62 (71%)	2 (2%)	23 (26%)	25 (29%)
PVM	71 (82%)	6 (7%)	10 (11%)	16 (18%)
EVM	69 (79%)	6 (7%)	12 (14%)	18 (21%)

* Percentages do not always add properly due to rounding.

resulting in a DR parameter variability metric of zero. However, we are unable to compute the 'true' values of the DR parameter variability metrics because we can only estimate observed parameter variance from ~ 11 observations. Assuming that the observations are normally-distributed, we would expect the distribution of observed parameter variances to follow a scaled χ^2 distribution with $N-1$ degrees of freedom [122]:

$$\frac{(N-1)x}{\sigma^2} \sim \chi^2_{N-1} \quad , \tag{3.6}$$

where x is the sample variance, N is the number of observations, and σ^2 is the true variance. Therefore, the expected variability metric distributions for the case when real variability is zero is that given in (3.6), shifted left by 1 (resulting from the subtraction of $\text{Var}(UL) = 1$ in (3.5)). These distributions (for $N = 11$) are plotted in Figure 3.5. One caveat associated with these results is that we have assumed that we know the 'true' value of $\text{Var}(UL)$, though, in reality, it is an estimate (computed from ~ 95 observations). When we normalize the measurements of the unmodeled load and the observed shed such that $\text{Var}(UL)$=1, any error in our estimate of $\text{Var}(UL)$ will affect our estimate of $\text{Var}(OS)$, which, in turn, affects our estimate of the SVM.

If none of the facility-years exhibited real parameter variability then we would expect only 5% of facilities to fall outside of the 95% confidence bounds. However, for each parameter, we find that substantially more than 5% of the facility-years fall outside of the bounds (Table 3.3). This implies that some facility-years exhibit real parameter variability. Facilities with disproportionally positive variability metrics likely exhibit real parameter variability. Facilities with disproportionally negative variability metrics likely exhibit positive covariance and, subsequently, real parameter variability. For the remainder of the facility-years, any observed parameter variability may simply result from model error and sampling.

Through simulation we find that, in order to achieve the distributions shown in Figure 3.5, it is likely that a number of facility-years have large real parameter variability, while the majority of facility-years have little to no real parameter variability (see Appendix B.1). Also, it is likely that for the vast majority of facility-years the covariance term is positive

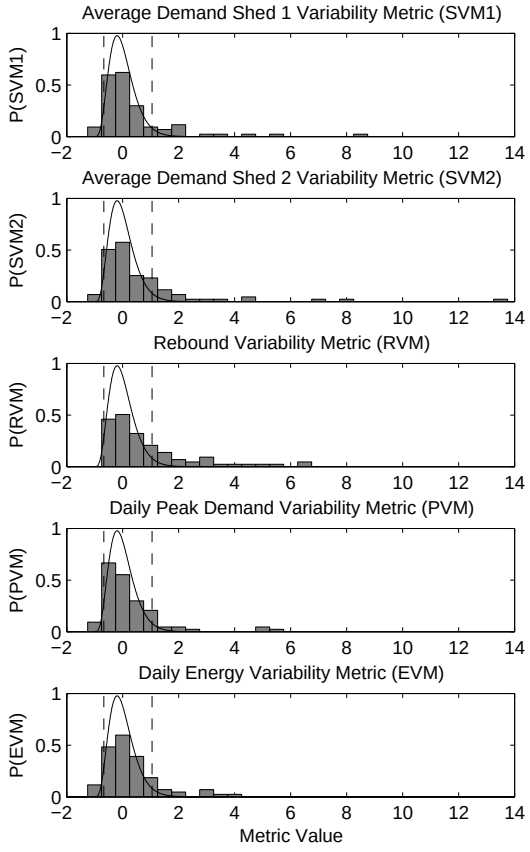


Figure 3.5: Histograms showing DR parameter variability metrics for the 87 facility-years. Solid lines show the expected distributions if real parameter variability were zero and $N = 11$ (dashed lines show the 95% confidence interval). Disproportionally positive variability metrics result from real parameter variability. Disproportionally negative variability metrics result from negative covariance and, subsequently, real parameter variability.

which implies that as unmodeled load increases, real shed increases, which is consistent with intuition. Additionally, we find that all combinations of the variability metrics are positively correlated. Several combinations have a population correlation coefficient, $\rho_{x,y}$, greater than 0.5: SVM1 and SVM2 ($\rho_{x,y} = 0.76$); SVM1 and PVM ($\rho_{x,y} = 0.59$); and SVM2 and PVM ($\rho_{x,y} = 0.56$).

The FERC has called for better understanding of responses to dynamic prices as a function of customer type [47], so we attempted to disaggregate parameter variability results by facility attributes including facility type, HVAC system type, DR strategy, and shed size (Appendix B.2). Results were inconclusive because of the small number of facility-years in the data set. It was particularly difficult to disaggregate the facilities by DR strategy because many facilities use more than one strategy. Therefore, we were unable to determine what kinds of facilities have more or less variable DR parameters. In an effort to do this, we are in the process of acquiring a larger data set.

Aggregate populations

DR parameter variability metrics for each aggregation of facilities are shown in Table 3.4. For each variability metric, we have computed the two-sided p-value under the null hypothesis that there is no real parameter variability. Therefore, real parameter variability likely exists when p-values are small. Surprisingly, the aggregate populations exhibit a wide range of variability metrics, similar to that seen for the individual facility-years. We would expect more real DR parameter variability in smaller aggregate populations. For example, in 2006 Zone 2 (8 facilities), we find likely real variability in each DR parameter. However, we also find likely real variability in both the Average Demand Shed 1 and the Daily Peak Demand in 2009 (32 facilities). Real variability in the aggregate could result from unmodeled correlation across facilities and/or large variable facilities dominating the aggregate results.

Table 3.4: DR parameter variability metrics computed for the aggregate populations. Bold values indicate p-values ⩽0.05.

Year (Zone)	Facilities (Peak*)	Shed 1		Shed 2		Rebound		Daily Peak Demand		Daily Energy	
		SVM1	p-value	SVM2	p-value	RVM	p-value	PVM	p-value	EVM	p-value
2006 (1)	4 (2.7 MW)	**-0.819**	**(0.01)**	-0.269	(0.67)	0.077	(0.75)	-0.386	(0.47)	**-0.737**	**(0.04)**
2006 (2)	8 (8.4 MW)	**3.039**	**(<0.01)**	**3.399**	**(<0.01)**	**1.044**	**(0.05)**	**1.131**	**(0.04)**	**4.578**	**(<0.01)**
2007	13 (11.7 MW)	0.579	(0.21)	-0.117	(0.90)	-0.454	(0.32)	-0.531	(0.24)	-0.210	(0.78)
2008	21 (14.6 MW)	-0.210	(0.72)	-0.142	(0.86)	**1.295**	**(0.02)**	-0.217	(0.71)	0.163	(0.62)
2009	32 (26.9 MW)	**-0.696**	**(0.03)**	-0.331	(0.46)	0.304	(0.43)	**-0.702**	**(0.04)**	-0.227	(0.69)

*Peak demand computed for May 1 - Sept 30.

3.6 Chapter Conclusion

We have developed a method to determine the error associated with DR parameter estimates. We find that this error is often large and so DR parameter estimates reported without error estimates may be misleading. For example, we may classify a steady shedder as a variable shedder and, therefore, judge the facility to be poorly controlled when, in fact, baseline model error simply prevents us from measuring consistent sheds. Since DR parameter estimates have error, all calculations derived with these estimates, including cost effectiveness estimates, also have error. Future research should explore the degree to which DR parameter error affects cost/benefit analyses on DR programs and technologies.

Observed DR variability in both individual facilities and aggregations of facilities affects the perception of persistent and reliable demand sheds. However, we find that observed DR variability is driven, in large part, by baseline model error, not real DR variability. For most facilities, observed DR variability can likely be explained by baseline model error alone; however, a number of facilities likely exhibit high real DR variability. In addition, most facilities exhibit a positive correlation between unmodeled load and real shed, which implies that the equipment that drives baseline model error is the equipment that is curtailed during DR events. We were unable to discern any relationship between response variability and facility attributes, possibly due to the small number of facilities.

Variability metrics computed for the aggregate populations show that in some cases the aggregate population likely exhibits real DR variability, which has implications for utilities with respect to M&V of DR programs. It also has implications for the system operator. If the aggregate response is not consistent, the system operator may have to deal with more demand-side variability than exists on non-DR days and, therefore, will need to procure more power systems services. In extreme cases, DR variability could result in significant deviations in grid frequency or expected power flow. More research is needed to understand DR variability in aggregate populations composed of facilities executing manual DR strategies, as they may exhibit even more variability than populations composed of facilities executing automated strategies.

The DR signal considered here, a critical peak pricing signal, is open-loop since the prices do not change in response to changes in load (though the signal is often implemented in individual facilities as closed-loop indoor air temperature control). Our results would be different if a closed-loop DR signal were used. Specifically, we would expect less DR variability, which could mitigate some of the issues we have described. This is an important subject of future research.

Chapter 4

TCL Modeling, State Estimation & Control

In this chapter, we shift our focus from recent DR programs to emerging DR paradigms. Specifically, we investigate methods to coordinate aggregations of residential TCLs to manage frequency and energy imbalances in power systems. We focus on opportunities to centrally control loads with high accuracy but low requirements for sensing and communications infrastructure. This chapter is largely based on [75],[1] [86],[2] and [90].[3]

4.1 Chapter Introduction

Increasingly, electric loads are participating in DLC and DR programs to improve electric grid reliability and reduce wholesale electricity prices [41, 126]. Traditional research in this area has focused on developing strategies that enable loads to *decrease* power use in the event of loss of generation or high prices, e.g., [99, 61, 129]. Recent research has explored incorporating local load states into control decisions to enable nondisruptive load reductions [7, 109]. This chapter also focuses on nondisruptive control, but for the purpose of delivering services such as load following and regulation by both decreasing and *increasing* power use over short time scales. The need for these services is likely to grow with increasing production from wind and solar generators, which are expected to increase frequency deviations and energy imbalance [84]. TCLs, such as refrigerators, air conditioners, and electric water heaters, are excellent candidates for providing these services because they are capable of storing thermal energy, much like a battery stores chemical energy [20]. We note that services could also be provided

[1] With permission from my co-authors: Stephan Koch and Duncan Callaway. "Modeling and control of aggregated heterogeneous thermostatically controlled loads for ancillary services." *PSCC*, Aug 2011.

[2] ©2012 IEEE. With permission from my co-author: Duncan Callaway. "State estimation and control of heterogeneous thermostatically controlled loads for load following." *HICSS*, Jan 2012.

[3] With permission from my co-authors: Stephan Koch and Duncan Callaway. "State estimation and control of electric loads to manage real-time energy imbalance" *(in review)*.

by energy storage devices [43], variable speed wind turbines [4, 94], and other forms of generation. However, in relation to these other options, TCLs may have advantages because the amount of balancing capacity they can potentially deliver far exceeds that available from wind generators and expected amounts of storage in the near term.

This chapter has two central objectives. The first is to extend our previous work [75, 86] to develop a modeling and control framework that enables non-disruptive control of TCLs for fast DR applications like load following and regulation. We present an analytical approach to include TCL parameter heterogeneity in reduced form TCL population models. This is in contrast to previous work that has either addressed parameter heterogeneity by simulation [83, 96], or assumed homogeneity [19, 82, 5, 79]. The modeling framework in this chapter is also novel in its general formulation of the application of ON/OFF control signals.

The second objective of this chapter is to gain insight into the level of sensing and communications infrastructure that is required to enable fast DR. The studies referenced above assume that power measurements are available from all loads for real-time feedback control. Here, we relax that assumption since it is currently expensive to integrate real time power measurement telemetry for ancillary services into existing utility SCADA systems: Pacific Gas and Electric Company spent more than $140,000 per load in a 2009 study [15] and Southern California Edison estimates current costs to be about $70,000 per measurement point [115]. Though costs may decline in the future, at present these costs are too high to measure power consumption at individual light commercial and residential facilities. Therefore, for smaller loads, aggregated power measurements are necessary. Though decentralized control may reduce communications requirements [58, 118, 62], here we focus on centralized control to preserve visibility and controllability in the control room. We will examine the effect of limited sensing and communications on the performance of aggregations of TCLs participating in 5-minute energy markets (i.e., load following). The analysis and results are similar for loads participating in regulation, following an automatic generation control signal.

Figure 4.1 shows the information hierarchy we consider. The local level consists of controllable TCLs that may be metered to transmit information to a central controller (the global level). The semi-global level consists of distribution substations, where one can measure the power consumed on feeders and estimate the state of controlled TCLs. We will assume all decisions are made at the global level (whether by a system operator or load aggregator responding to system operator commands), and explore several different scenarios of information available at the global level from various levels in the hierarchy.

One of the key aims of this chapter is increased insight into how much centralized information is needed for real-time feedback control of loads. In the most aggressive case, we find that it is possible to estimate aggregated load states and parameters on a distribution system with a single measurement of aggregate power consumption at the distribution substation. Our results suggest that the amount of infrastructure needed to build a "smart grid" that engages loads in power system services with high fidelity may be relatively small.

Section 4.2 describes our modeling framework, Section 4.3 presents four scenarios for infrastructure and communications, Section 4.4 details our methodology, Section 4.5 presents

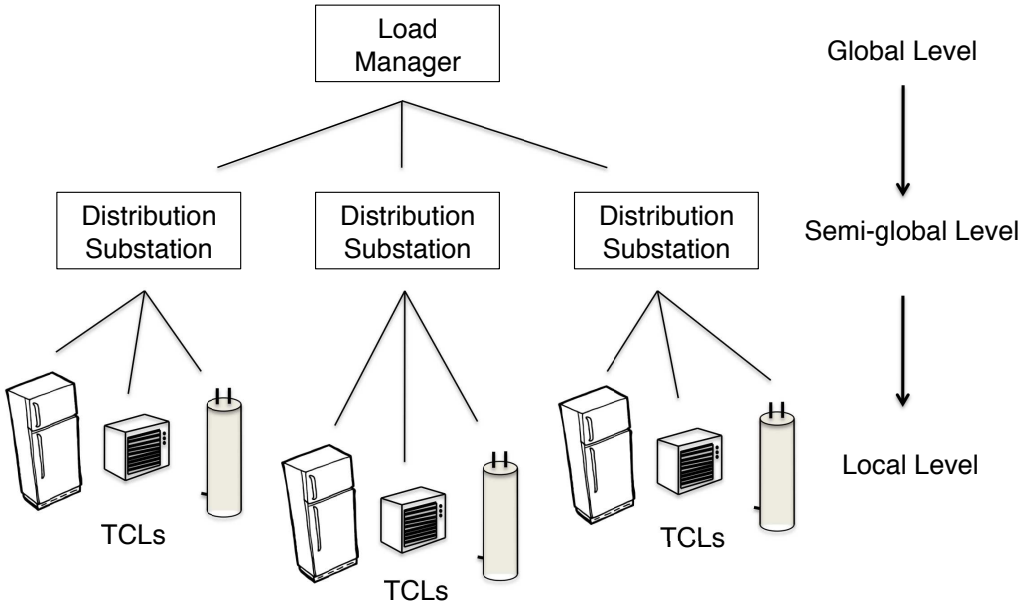


Figure 4.1: Electric power system information hierarchy. ©2012 IEEE

our simulation results, and Section 4.6 concludes.

4.2 Modeling Framework

This section describes the models used for TCL simulation. Because we are primarily interested in applications that manage frequency and energy imbalance in power systems, our modeling framework neglects reactive power flows and voltage drop on distribution feeders. Load control strategies could affect reactive power flows and voltage drop along distribution feeders, and this could in turn affect power consumption and duty cycles of voltage dependent loads. We investigated this effect with a simple model of a distribution feeder with impedance and voltage dependent loads and found that neglecting voltage effects introduces an error that is unlikely to be more than 2% of real power delivered to the distribution substation. Given the small effect and the fact that control strategies can be designed to compensate for it, for simplicity, we will not explicitly model voltage effects here. However, all distribution impacts must eventually be managed for the successful implementation of any load control paradigm.

4.2.1 Individual TCL model

We model the temperature state evolution of an individual TCL with a discrete time difference equation commonly used in the literature, e.g., [19, 30], and verified against real populations of TCLs [30, 81, 95, 37]:

$$\theta_i(k+1) = a_i\theta_i(k) + (1-a_i)(\theta_{\mathrm{a},i}(k) - m_i(k)\theta_{\mathrm{g},i}) + \epsilon_i(k) \tag{4.1}$$

where $\theta_i(k)$ is the internal temperature of TCL i at time step k, θ_a is the ambient temperature, and ϵ is a noise process. The dimensionless TCL parameter a_i is defined as follows:

$$a_i = e^{-h/(C_iR_i)} \tag{4.2}$$

where C is a TCL's thermal capacitance, R is its thermal resistance, and h is the simulation time step. θ_g is the temperature gain when a TCL is ON:

$$\theta_\mathrm{g} = R_iP_{\mathrm{trans},i} \tag{4.3}$$

where P_trans is a TCL's energy transfer rate, which according to our conventions is positive for cooling TCLs and negative for heating TCLs. P_i is the power consumed by TCL i when it is on:

$$P_i = \frac{|P_{\mathrm{trans},i}|}{\mathrm{COP}_i} \tag{4.4}$$

where COP is its Coefficient of Performance (COP). We focus on TCLs with hysteretic ON/OFF local control within a dead-band. Therefore, the local control variable m is a dimensionless discrete variable equal to 1 when the TCL is ON and 0 when the TCL is OFF. For cooling TCLs, it is defined as follows:

$$m_i(k+1) = \begin{cases} 0, & \theta_i(k+1) < \theta_{\mathrm{set},i} - \delta_i/2 \\ 1, & \theta_i(k+1) > \theta_{\mathrm{set},i} + \delta_i/2 \\ m_i(k), & \text{otherwise} \end{cases} \tag{4.5}$$

where θ_set is the temperature setpoint and δ is the dead-band width. For heating TCLs, the position of the 0 and 1 are switched. Though more detailed TCL models are available (e.g., [76, 116]), this model captures the key local dynamics, and is reducible to an aggregated framework.

4.2.2 Plant: The TCL population

This study uses a simulated plant. Specifically, we simulated thousands of TCLs using (4.1). Within each time step we advance TCLs through the temperature space without regard to their temperature dead-band. At each time step, we switch ON/OFF TCLs that have moved outside of the dead-band. Switches for individual TCLs operating autonomously occur on

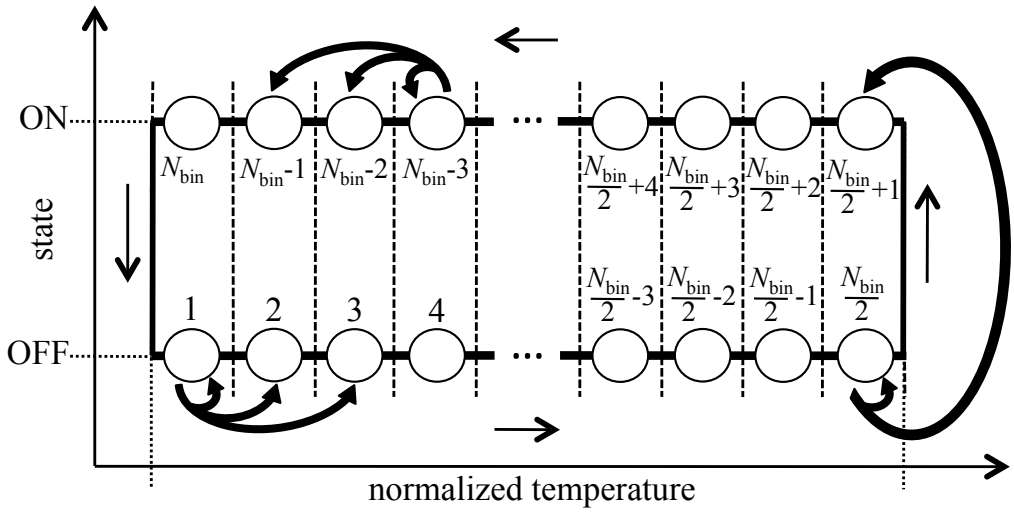


Figure 4.2: State bin transition model for a cooling TCL.

the order of once every 10 minutes. Since we set $h = 2$ seconds in all simulations, the switching dynamics do not differ substantially from those of a continuous time system.

It is important to note that all control actions have been applied directly to the plant. The reduced form TCL population model, described in the next subsection, was used for designing state estimators and computing control signals only.

4.2.3 Reduced form TCL population model

We developed a discrete time, linear, time-invariant system model of heterogeneous TCL populations. The advantage of this model is that it admits a range of system analysis, state estimation, and control techniques, which allows us to achieve better tracking performance than model-free control algorithms, as detailed in Section 4.5. Each TCL's temperature dead-band is mapped to a 'normalized temperature dead-band' that is divided into $\frac{N_{\text{bin}}}{2}$ temperature intervals (Figure 4.2). For any given temperature a TCL may be ON or OFF, so we divided each temperature interval into two state bins, resulting in a total of N_{bin} state bins. Each state bin contains the fraction of TCLs within that state and each TCL moves through the normalized temperature state space at a particular rate. Consider thousands of TCLs, each at a different point in the state space, each moving at a different rate. We can model the movement of the TCL population by computing the *probability* that a randomly chosen TCL transitions from one state bin to another over one time step. An $N_{\text{bin}} \times N_{\text{bin}}$ Markov Transition Matrix (MTM) can be derived by computing transition probabilities from/to each bin. The transpose of the MTM is the $\boldsymbol{A}$-matrix commonly used in control

applications, defined by:

$$\boldsymbol{x}(k+1) = \boldsymbol{A}\boldsymbol{x}(k), \tag{4.6}$$

where $\boldsymbol{x}$ is the $N_{\text{bin}} \times 1$ state bin vector, which contains the fraction of TCLs in each bin.

We next provide the analytical foundation for the $\boldsymbol{A}$-matrix. Though it is possible to model heterogeneity in more than just one parameter, for simplicity, we model heterogeneity only in C (and therefore a) and ignore ϵ (later we consider full parameter heterogeneity). Therefore, (4.1) becomes:

$$\theta_i(k+1) = a_i\theta_i(k) + (1-a_i)(\theta_{\text{a}} - m_i(k)\theta_{\text{g}}). \tag{4.7}$$

Consider a group of TCLs that are at the same temperature θ_{start} and either all ON or all OFF. The probability that a randomly chosen TCL from this group will move to θ_{end} in one time step is:

$$\text{P}(\theta_{\text{end}}|\theta_{\text{start}}) = \text{P}(a_i) \tag{4.8}$$

where one can use (4.1) to solve for a_i in terms of θ_{start} and θ_{end}. Assuming $\text{P}(a_i)$ is independent of temperature, the probability that a randomly chosen TCL will move from θ_{start} to anywhere in the n^{th} bin, $\theta_n \leqslant \theta < \theta_{n+1}$, in one time step is:

$$\text{P}(\theta_n \leqslant \theta_{\text{end}} < \theta_{n+1}|\theta_{\text{start}}) = \int_{a_1}^{a_2} p(a)\,\text{d}a \tag{4.9}$$

where $p(\cdot)$ is the probability density function, and one can use (4.1) to solve for a_1 and a_2 in terms of θ_{start}, θ_n, and θ_{n+1}:

$$a_1 = \frac{\theta_{\text{a}} - \theta' - m\theta_{\text{g}}}{\theta_{\text{a}} - \theta_{\text{start}} - m\theta_{\text{g}}} \qquad a_2 = \frac{\theta_{\text{a}} - \theta'' - m\theta_{\text{g}}}{\theta_{\text{a}} - \theta_{\text{start}} - m\theta_{\text{g}}}.$$

Here, $\theta' = \theta_n$ and $\theta'' = \theta_{n+1}$ if the TCL is traversing from low temperatures to high temperatures, or $\theta' = \theta_{n+1}$ and $\theta'' = \theta_n$ if the TCL is traversing from high temperatures to low temperatures.

Assuming TCLs are uniformly distributed in any temperature bin and $p(a)$ is independent of temperature, the probability that a randomly chosen TCL will move from the m^{th} to the n^{th} state bin in one time step is:

$$\begin{aligned}&\text{P}(\theta_n \leqslant \theta_{\text{end}} < \theta_{n+1}|\theta_m \leqslant \theta_{\text{start}} < \theta_{m+1}) \\ &\qquad = A_{n,m} = \int_{\theta_m}^{\theta_{m+1}} \int_{a_1}^{a_2} p(a)\,\text{d}a\,\text{d}\theta_{\text{start}}.\end{aligned} \tag{4.10}$$

This integral can not be solved analytically (Appendix C), so for each combination of starting and ending bins, we evaluate (4.10) numerically to generate the analytically-derived $\boldsymbol{A}$-matrix. Again, we assume TCLs switch only at discrete time steps, and so within a time step TCLs may move outside of the dead-band. We assume that each TCL that moves

Table 4.1: Partially heterogenous air conditioner parameters.

Parameter	Meaning	Value*
θ_{set}	temperature setpoint	20°C
δ	temperature dead-band width	0.5°C
θ_{a}	ambient temperature	32°C
R	thermal resistance	2°C/kW
C	thermal capacitance	1.5-2.5 kWh/°C
P_{trans}	energy transfer rate	14 kW
COP	coefficient of performance	2.5

*Values for a 250 m^2 house adapted from [19], except C which was derived using data from [63] assuming that thermal capacitance scales with floor space.

outside of the dead-band at some time between k and $k+1$ is switched at $k+1$. However, at $k+1$ these TCLs are still outside of the dead-band. To keep track of them, we expand the temperature range of bins 1 and $\frac{N_{\text{bin}}}{2}+1$ in Figure 4.2 to $-\infty$ and ∞, respectively.

If we allow R and P_{trans} to vary across TCLs, the $\boldsymbol{A}$-matrix becomes harder to derive. However, for arbitrary heterogeneity, we can identify the $\boldsymbol{A}$-matrix using *real* or *simulated* full state information (temperature and ON/OFF state) from a sample of the TCL population. This is possible with Hidden Markov Model parameter estimation.

We analytically derived the $\boldsymbol{A}$-matrix for a system of TCLs with the partially heterogeneous TCL parameters, listed in Table 4.1. We considered the case where the TCL parameter, a, is uniformly distributed between $a_{\min}$ and $a_{\max}$ (computed from the parameters in the table):

$$p(a) = \begin{cases} \frac{1}{a_{\max}-a_{\min}} & \text{if } a_{\min} \leqslant a \leqslant a_{\max} \\ 0 & \text{otherwise} \end{cases} \tag{4.11}$$

We also identified the $\boldsymbol{A}$-matrix using state information from a plant composed of 1,000 TCLs (with the same partially heterogenous TCL parameters) to count TCL transitions from one bin to the next over 1,800 time steps. We compare the eigenvalues of the analytically-derived and identified matrices in Figure 4.3. For $N_{\text{bin}} > 2$, the eigenvalues form an ellipse on the complex plane. When $N_{\text{bin}} = 2$, both eigenvalues are real. For each number of state bins, the eigenvalues of the analytically derived and identified matrices are nearly identical. We use identified matrices in subsequent simulations.

4.2.4 Centralized control of the TCL population

The previous subsection described the dynamics of TCL populations as governed by local hysteresis control only. In this section, we present a centralized control framework. Consider

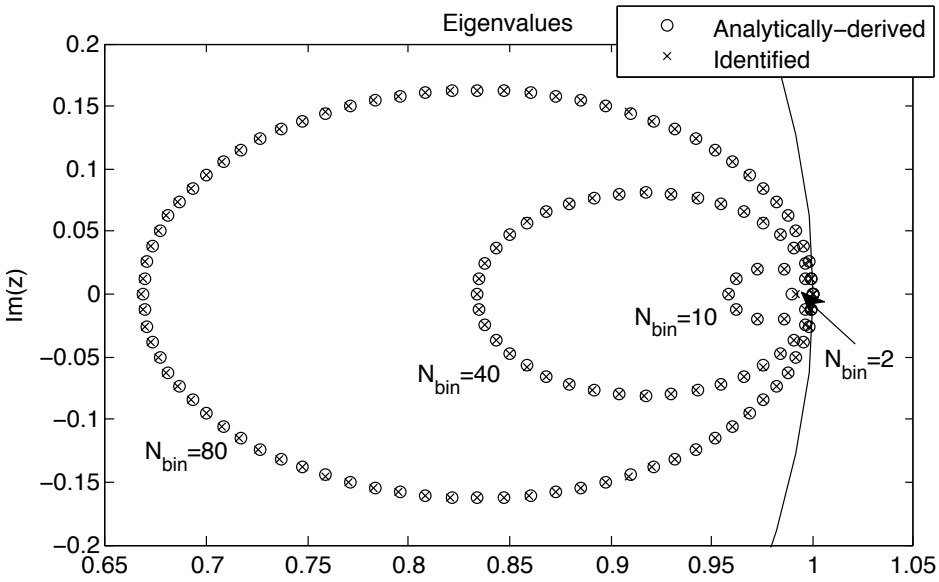


Figure 4.3: Comparison of the eigenvalues of the analytically-derived and identified $\boldsymbol{A}$-matrices, computed for the partially heterogeneous system. The solid line shows the unit circle.

the state space model:

$$\boldsymbol{x}(k+1) = \boldsymbol{A}\boldsymbol{x}(k) + \boldsymbol{B}\boldsymbol{u}(k) \tag{4.12}$$

$$\boldsymbol{y}(k) = \boldsymbol{C}\boldsymbol{x}(k) \tag{4.13}$$

where the vector $\boldsymbol{x}$ contains the fraction of TCLs in each bin and $\boldsymbol{u}$ and $\boldsymbol{y}$ are the input and output vectors respectively.

We define $\boldsymbol{u}$ to be an $\frac{N_{\text{bin}}}{2} \times 1$ vector that only influences a TCL's ON/OFF state (i.e., a TCL's temperature state is not directly controllable). The absolute value of each element of $\boldsymbol{u}$ is the fraction of the total population of TCLs to switch. Positive elements of $\boldsymbol{u}$ switch TCLs ON, while negative elements switch TCLs OFF. We can constrain $\boldsymbol{u}$ so as to switch only TCLs in certain bins to, for example, minimize compressor short cycling, as described in Section 4.4.4.

Since each TCL that switches out of an ON bin switches into the corresponding OFF bin (and vice versa), $\boldsymbol{B}$ is structured as follows:

$$\boldsymbol{B} = \left[\begin{array}{ccc} -1 & & 0 \\ & \ddots & \\ 0 & & -1 \\ \hline 0 & & 1 \\ & \iddots & \\ 1 & & 0 \end{array}\right]. \tag{4.14}$$

This control design ensures that each TCL will stay within its dead-band, though TCLs may switch more frequently than they would in the absence of control.

If the only real time measurement is TCL aggregate power, $P_{\text{total,meas}}$, y is a scalar and $\boldsymbol{C}$ is a row vector as follows:

$$\boldsymbol{C} = \underbrace{\bar{P}_{\text{ON}}\, N_{\text{TCL}}}_{c_p} \Big[\underbrace{0, \cdots, 0}_{\frac{N_{\text{bin}}}{2}}, \underbrace{1, \cdots, 1}_{\frac{N_{\text{bin}}}{2}}\Big] \tag{4.15}$$

where N_{TCL} is the number of TCLs in the population and $\bar{P}_{\text{ON}}$ is the mean power consumption of TCLs in the ON state. Note that $\bar{P}_{\text{ON}}$ may not equal the mean power of all TCLs in the population since TCLs with lower rated power may spend proportionally more time in the ON state. $\bar{P}_{\text{ON}}$ can be computed if all TCL parameters, ambient temperatures, and dead-bands are known, or if ON/OFF state information is available and the aggregate steady state power consumption of the population, $P_{\text{total,ss}}$, is known.

For the reference case, described in Section 4.3, we assume full state information from all TCLs is available in real time. In that case, $\boldsymbol{C}$ becomes an $(N_{\text{bin}} + 1) \times N_{\text{bin}}$ matrix:

$$\boldsymbol{C}_{\text{ref}} = \left[\begin{array}{c} \boldsymbol{I}_{N_{\text{bin}} \times N_{\text{bin}}} \\ \boldsymbol{C} \end{array}\right]. \tag{4.16}$$

Therefore, the first N_{bin} rows of the $(N_{\text{bin}}+1)\times 1$ $\boldsymbol{y}$-vector are the measured states and the last row is the aggregate power measurement. Though not considered here, one could use $\boldsymbol{C}_{\text{ref}}$ if full state information is available in real time from a subset of TCLs.

The pair $[\boldsymbol{A},\ \boldsymbol{B}]$ is not controllable: the controllability matrix is of rank $n-1$ because the controller cannot drive all states to zero (the fraction of TCLs in each bin must sum to one). However, for both $\boldsymbol{C}$ and $\boldsymbol{C}_{\text{ref}}$, aggregate power can be tracked and the system is observable.

4.2.5 Model performance

We first compared the autonomous behavior of the reduced form model to that of the plant. Starting all TCLs in the same bin (i.e., all TCLs are synchronized), the aggregate power consumed by the TCL population exhibits an oscillatory decay towards the steady state power consumption of the TCL population. The oscillatory decay is related to TCL parameter heterogeneity: homogeneous populations exhibit undamped oscillations, and oscillations decay faster with increasing parameter heterogeneity as TCLs spread out over the temperature state space. In the reduced form model, there is a negative correlation between the number of bins and the damping of the oscillation. This finding is consistent with [5], where a high number of bins (200) was chosen in order to yield a non-decaying oscillation corresponding to a homogeneous system. Interestingly, a reduced form model built with a large number of bins produces non-decaying behavior, even when the model has been identified from a heterogeneous parameter set. Thus, for heterogeneous populations, the number of bins can be seen as a tuning variable used to achieve good matching to the decay behavior. However, for tightly controlled systems, getting the number of bins right may not be very important.

To evaluate model performance under more realistic conditions, we compared the forced behavior of the reduced form model to that of the plant. Using the TCL parameters listed in Table 4.2, we simulated a 10,000 TCL plant. We randomly drew each parameter from uncorrelated uniform distributions between the minimum and maximum values shown in the table, and used simulated state information to identify the $\boldsymbol{A}$-matrix and $\bar{P}_{\text{ON}}$. Two randomly generated open-loop control sequences (drawn from uniform distributions) were used to force the population: high forcing, in which up to 12.5% of the TCLs were switched in one time step, and low forcing, in which up to 2.5% of the TCLs were switched in one time step. The control was applied such that TCLs in bins nearer to the dead-band were switched preferentially. We then evaluated the ability of the model to predict the aggregate power consumption of the plant. We assume that the model knows the state perfectly when the prediction horizon is zero and gains no additional state information over time.

In Figure 4.4, we plot root mean square (RMS) prediction error (normalized by the steady state power consumption of the TCL population) versus prediction horizon, and compare the results to those generated with a persistence model. Several cases, using models with different numbers of state bins, are shown. The mismatch between the output of the plant and that of the reduced form model is due, at least in part, to the assumption that the

Table 4.2: Fully heterogenous air conditioner parameters.

Parameter	Meaning	Value*
θ_{set}	temperature setpoint	15–25°C
δ	temperature dead-band width	0.25–1°C
θ_{a}	ambient temperature	32°C
R	thermal resistance	1.5–2.5°C/kW
C	thermal capacitance	1.5–2.5 kWh/°C
P_{trans}	energy transfer rate	10–18 kW
COP	coefficient of performance	2.5

*Values for a 250 m^2 house adapted from [19], except C which was derived using data from [63] assuming that thermal capacitance scales with floor space.

parameter distribution of TCLs is independent of temperature. In actuality, the parameter distribution of TCLs is a function of initial condition, simulation time, and temperature. We find that, for low forcing, models with two state bins perform the worst, though beyond that the number of state bins does not seem to significantly affect the accuracy of aggregate power predictions. However, for high forcing, models with more state bins produce more accurate predictions. The plot also shows that control actions that require long prediction horizons may not perform well. We have restricted ourselves to using the model to predict aggregate power in the next two seconds, though, in the future, we plan to explore control schemes that require longer prediction horizons.

4.3 Scenarios for Infrastructure and Communications

Different levels of infrastructure and communications lead to different system set-ups and control performance. In this section, we first discuss several options for system identification, estimation, and aggregate power measurement. We then describe the four scenarios that we investigate in this chapter.

4.3.1 Options

System identification options

We can identify the system with real or simulated full state information from all or a subset of the TCLs. To do this, we need either 1) state measurements, $\boldsymbol{x}_{\text{meas}}$, and $P_{\text{total,ss}}$ or 2) knowledge of the TCL parameters, ambient temperatures, and dead-bands (i.e., DR participants could provide their TCL parameters when they enroll in a program). If this information is unavailable, we can identify the $\boldsymbol{A}$-matrix using a joint parameter and state estimation technique, such as an Extended Kalman Filter (EKF).

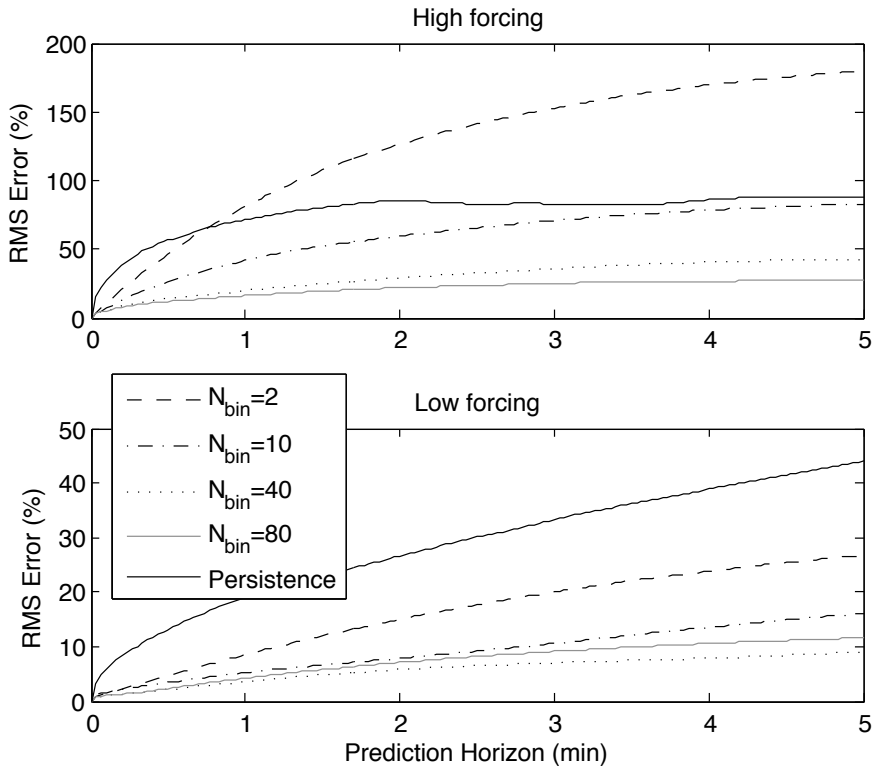


Figure 4.4: Aggregate power prediction RMS error versus prediction horizon.

State estimation options

If we have already identified the system, a state estimator, such as a Kalman Filter, can be used to estimate the states based on knowledge of $\boldsymbol{u}$, $P_{\text{total,meas}}$, and, possibly, $\boldsymbol{x}_{\text{meas}}$. If we have not identified the system, we can use a joint parameter and state estimation technique.

Aggregate power estimation options

We can estimate aggregate power by metering all or a subset of the TCLs to report ON/OFF state to the central controller in real time. We can also estimate it at the distribution substation by subtracting a counterfactual substation power forecast (i.e., a forecast created with models/data that do not include the effects of the DR program) from the actual power recorded at the substation. The remaining signal is a combination of the control response and the forecast noise. Adding $P_{\text{total,ss}}$ to this signal results in a noisy measurement of the aggregate power consumption of the TCL population. This method of computing the control response is similar to that used to assess DR sheds in traditional DR programs, except that in these programs one usually subtracts a counterfactual ex-post building power prediction from the actual power recorded [28].

4.3.2 Four scenarios

The four scenarios are composed of combinations of the above options, as described in Table 4.3. Block diagrams describing the cases are shown in Figure 4.5. In all scenarios, a controller takes in a desired trajectory, $P_{\text{total,des}}$, along with the current aggregate power estimate, $P_{\text{total,est}}$, and state estimate, $\boldsymbol{x}_{\text{est}}$, to produce the input vector, $\boldsymbol{u}$. We translate $\boldsymbol{u}$ into $\boldsymbol{u}_{\text{rel}}$, an $N_{\text{bin}} \times 1$ vector of *switch probabilities* so that we can broadcast the same control vector to all TCLs.

Using broadcast control vectors, as opposed to addressing each TCL individually, ensures that the computational burden is low. In each time step, TCLs are switched only ON or OFF. Therefore, the switch probability associated with one state bin in each temperature interval is 0% and the switch probability associated with the other is computed by dividing the relevant entry of $\boldsymbol{u}$ by the relevant entry of $\boldsymbol{x}_{\text{est}}$, which nonlinearizes the control:

$$u_{\text{rel},j} = \begin{cases} \left[\min\left(\frac{u_j}{x_{\text{est},j}}, 1\right)\right]^+ \text{ for } j = 1, ..., \frac{N_{\text{bin}}}{2} \\ \left[\min\left(\frac{-u_{N_{\text{bin}}-j+1}}{x_{\text{est},j}}, 1\right)\right]^+ \text{ for } j = \left(\frac{N_{\text{bin}}}{2}+1\right), ..., N_{\text{bin}} \end{cases} \tag{4.17}$$

where $[\cdot]^+$ sets negative values to zero. At each time step, each TCL receives $\boldsymbol{u}_{\text{rel}}$; determines its current state and, subsequently, which element of $\boldsymbol{u}_{\text{rel}}$ to act upon; generates a random number to determine if it should switch or not; and, possibly, switches ON/OFF.

In Scenarios 1–3, system identification is performed offline using either the TCL parameters or $\boldsymbol{x}_{\text{meas}}$ and $P_{\text{total,ss}}$, and a Kalman Filter is used for state estimation (Figure 4.5a).

In Scenario 1, the full state is measured perfectly in real time (so $\boldsymbol{C}$ is defined as in (4.16)), while in Scenarios 2–3 it is not measured in real time (so $\boldsymbol{C}$ is defined as in (4.15)). Despite perfect state measurements, we have found it beneficial to use a Kalman Filter in Scenario 1 because the TCL population model is imperfect. Using the imperfect model together with the imperfect state estimates, rather than the perfect state measurements, to predict aggregate power in the next time step results in better control performance.

The only difference between Scenarios 2 and 3 is the way in which TCL aggregate power is measured in real time. In Scenario 4, the only information available offline is c_p, so system identification and state estimation are performed in real time with an EKF (Figure 4.5b).

Each scenario requires different infrastructure. The minimum infrastructure required is shown in Table 4.3. In Scenarios 2 and 3, additional infrastructure is required if system identification is performed with state measurements, instead of TCL parameters. Specifically, TCLs must be equipped with temperature sensors and measurement/knowledge of their ON/OFF state, though they do not need to transmit this information in real time.

Table 4.3: Scenarios.

	System Identification	Information Available in Real Time	State Estimation	Minimum Infrastructure Required
Scenario 1	Offline, using TCL parameters or $\boldsymbol{x}_{\text{meas}}$ and $P_{\text{total,ss}}$	Temperature and power consumption from all TCLs, measured perfectly	Full state measured; Kalman Filter	TCL-level low-latency two-way data connection, TCL temperature sensor, TCL power measurement/knowledge, local decision making*
Scenario 2	Same as Scenario 1	ON/OFF state information from 100% or 30% of TCLs	Full state not measured; Kalman Filter	TCL-level low-latency two-way data connection, TCL ON/OFF state measurement/knowledge, local decision making*
Scenario 3	Same as Scenario 1	Distribution area power consumption and forecasts, assuming forecast error standard deviations of 5% or 10% of the the substation load	Same as Scenario 2	TCL-level low-latency one-way data connection, substation-level low-latency one-way data connection, substation power measurement, local decision making*
Scenario 4	In real time, using an EKF to identify the $\boldsymbol{A}$-matrix; assumes knowledge of c_p	Same as Scenario 3	EKF	Same as Scenario 3

*Local decision making capabilities are required to translate the control input vector into actions.

4.4 Methodology

4.4.1 Simulation parameters and system identification

We parameterized the TCL population using the TCL parameters in Table 4.2. We randomly drew each parameter from uncorrelated uniform distributions between the minimum and maximum values shown in the table, and ϵ from a normal distribution with mean 0 and standard deviation 5×10^{-4}.

We assumed a distribution substation load of 17 MVA (in California, the typical size range for substations is 15–45 MVA) and a coincident TCL load of 15% of the distribution substation load, which was chosen as a balance between having so few loads that the state estimation methods would not work and having so many loads that it would be difficult for an aggregator to recruit them all within one distribution area. To satisfy these assumptions, we used 1,000 TCLs per distribution substation. For 10,000 TCL populations, we aggregated the results of ten individual substations. The TCLs within each substations are not identical, though we draw the TCL parameters from the same distributions.

For Scenarios 1–3, we identified the $\boldsymbol{A}$-matrix by simulating a population of TCLs and using full state information from all of the TCLs to count the number of TCL transitions from each starting bin to each ending bin over 1800 time steps, and normalizing the resulting matrix such that each column summed to one. For Scenario 4, we identified the $\boldsymbol{A}$-matrix using an EKF, which will be described in Section 4.4.3. In all scenarios, we compute $\bar{P}_{\mathrm{ON}}$ by dividing the aggregate power use of the population by the number of TCLs in the ON state at each time step, and then finding the mean.

4.4.2 State estimation

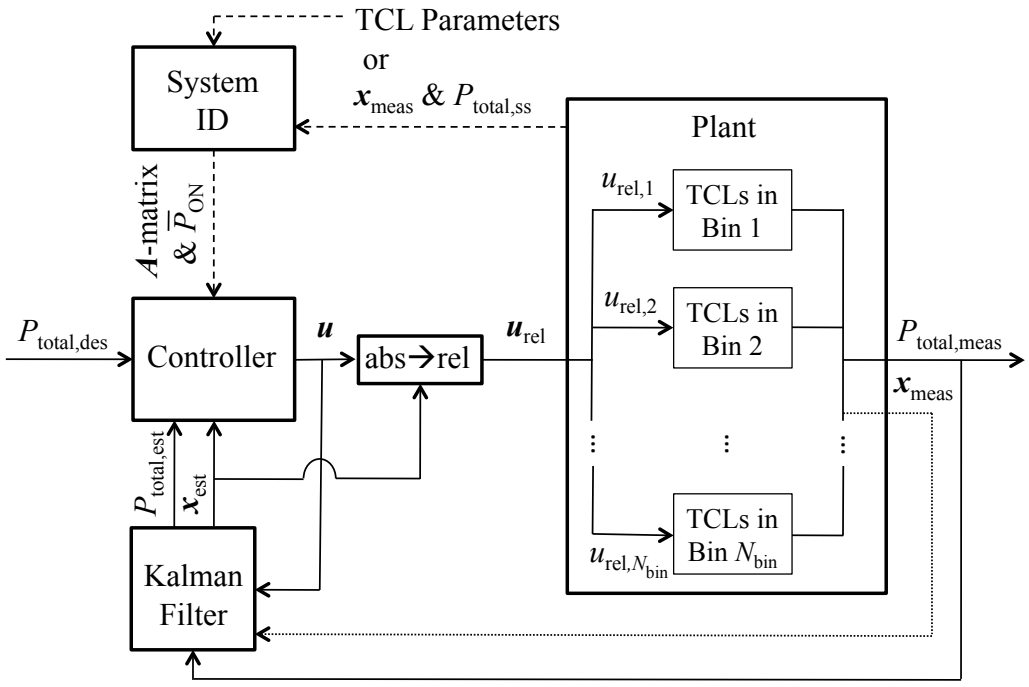
For Scenarios 1–3, we designed a standard Kalman Filter [119] using the MATLAB function `kalman` on the identified system:

$$\boldsymbol{x}(k+1) = \boldsymbol{A}\boldsymbol{x}(k) + \boldsymbol{B}\boldsymbol{u}(k) + \boldsymbol{B}_{\omega}\boldsymbol{\omega}(k) \tag{4.18}$$

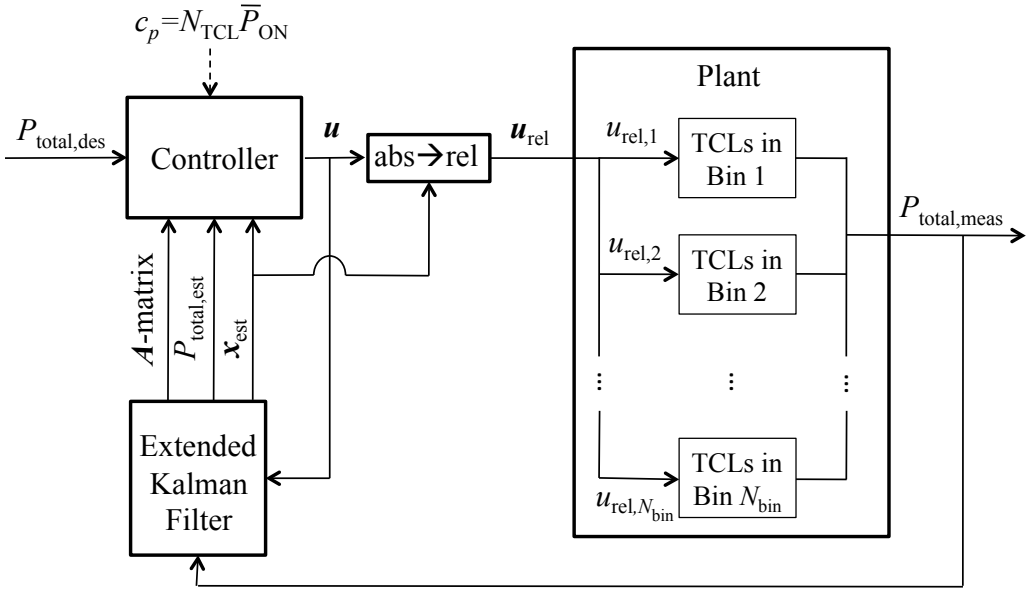
$$\boldsymbol{y}(k) = \boldsymbol{C}\boldsymbol{x}(k) + \boldsymbol{\nu}(k) \tag{4.19}$$

where $\boldsymbol{B}_{\omega} = \boldsymbol{I}_{N_{\mathrm{bin}} \times N_{\mathrm{bin}}}$, $\boldsymbol{\omega}$ is an $N_{\mathrm{bin}} \times 1$ process noise vector, and $\boldsymbol{\nu}$ is a measurement noise vector the same size as $\boldsymbol{y}$: $(N_{\mathrm{bin}} + 1) \times 1$ in Scenario 1 and a scalar in Scenarios 2 and 3.

The $N_{\mathrm{bin}} \times N_{\mathrm{bin}}$ process noise covariance matrix, $\boldsymbol{Q}_{\mathrm{KF}}$, was computed with the residuals between the state values associated with the plant and those predicted by the model. Therefore, 'process noise' includes both noise associated with individual TCLs and plant-model mismatch. The noise was modeled as white since the Kalman Filter assumes white noise. In reality, plant-model mismatch results in colored process noise, making the Kalman Filter sub-optimal. Since we would never expect to know $\boldsymbol{Q}_{\mathrm{KF}}$ perfectly, for each case, we constructed two system parameterizations (different draws of TCL parameters, noise, etc.



(a) Scenarios 1–3



(b) Scenario 4

Figure 4.5: Block diagrams for each scenario. Dashed lines indicate offline actions. In Scenario 1, the state is measured perfectly in real time (dotted line), while in Scenarios 2–3, it is not.

from the same distributions), and use one to compute $\boldsymbol{Q}_{\text{KF}}$ and the other, along with $\boldsymbol{Q}_{\text{KF}}$, to assess tracking performance.

For Scenario 1 and Scenario 2 (100% metering), the aggregate power measurement noise variance, R_{KF}, was set to zero. However, in some cases, the function `kalman` was unable to build a convergent Kalman Filter with $R_{\text{KF}} = 0$, so the value was increased to up to 10 kW^2. For Scenario 2 (30% metering), R_{KF} was computed with the residuals between the system output associated with the plant and the measured system output. The measurement noise was also modeled as white, though in reality it too is colored, making the Kalman Filter sub-optimal. Again, we use one system parameterization to compute R_{KF} and another to assess tracking performance. For Scenario 3, R_{KF} was computed from the assumed distribution substation power forecast error standard deviation: 5% and 10% of the distribution substation load. R_{KF} is held constant in each simulation run. In reality, the distribution power forecast error variance could change over time. Therefore, we model the hypothetical changes in variance by having it follow a sinusoid with a period equal to the length of the simulation run, between 0.5 and 1.5 of R_{KF}.

In practice, we may not have adequate information to compute $\boldsymbol{Q}_{\text{KF}}$ and R_{KF} as described. In that case, we could compute $\boldsymbol{Q}_{\text{KF}}$ and R_{KF} through simulation, iterative tuning, or by measuring full state information and aggregate power consumption perfectly from a small population of TCLs (e.g., in a pilot program) and then extrapolating the results to the larger population.

The accuracy of the Kalman Filter estimates for the various scenarios is shown in Figure 4.6. Each scenario is carried out with a population of 1,000 TCLs using Controller 1, which will be described in Section 4.4.4. The figure shows that estimates of states closer to the edge of the dead-band are more prone to error. This is because the accuracy of the Kalman Filter is a function of the accuracy of the reduced form TCL population model. In turn, the reduced form model is a function of the data used to build it. The model is best at predicting states for systems forced similarly to the system used to identify the model. Here, the model was identified with data from an unforced plant and then applied to forced plants. In this case, the model predictions of states closer to the edge of the dead-band (especially bins 1 and $\frac{N_{\text{bin}}}{2} + 1$) are more prone to error. Randomly forcing the system and using the state trajectory to identify the model does not improve the ability of the model to predict the states of a system controlled to participate in load following because the two state trajectories are completely different. One could implement an adaptive scheme, for example, use a preliminary model to control the system to participate in load following, use the state trajectory to identify a better model, and iterate. Such an approach merits further investigation, but it may be difficult to implement in practice.

4.4.3 Joint parameter and state estimation method

Scenario 4 requires joint parameter and state estimation. Treating the entries of the $\boldsymbol{A}$-matrix as unknown states that do not vary over time, we can derive nonlinear state/output

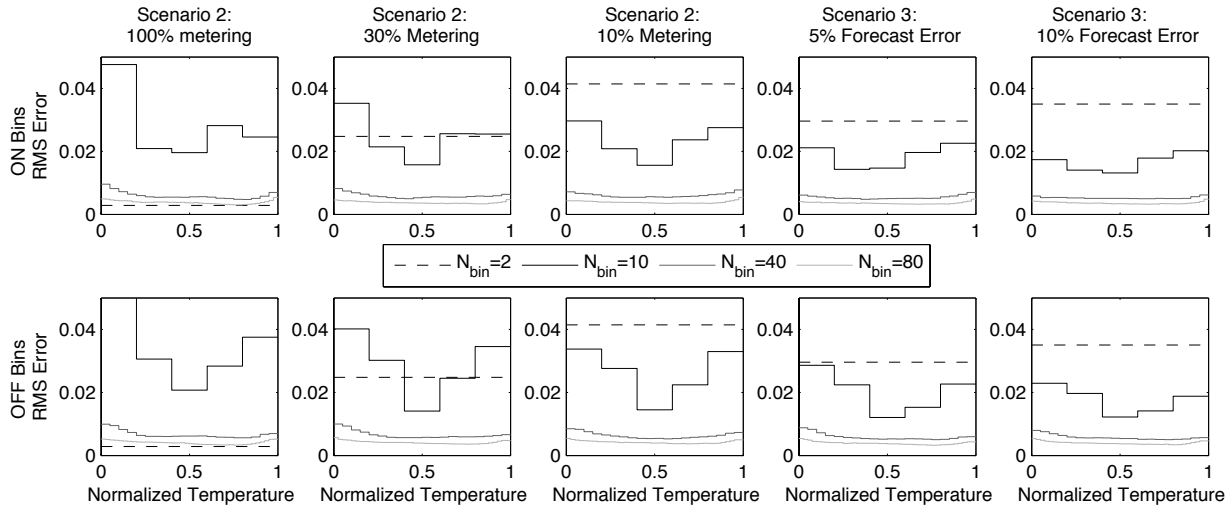


Figure 4.6: Accuracy of the Kalman Filter estimates.

equations and use an EKF for state (and therefore parameter) estimation. We attempted to estimate both the $\boldsymbol{A}$-matrix and c_p online; however, the estimator was unable to converge to the true value of c_p. Therefore, this approach requires measuring, deriving, or estimating c_p.

Consider the $N_{\text{bin}} = 2$ system:

$$\boldsymbol{x}(k+1) = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix} \boldsymbol{x}(k) + \begin{bmatrix} -1 \\ 1 \end{bmatrix} u(k) \tag{4.20}$$

$$y(k) = c_p \begin{bmatrix} 0 & 1 \end{bmatrix} \boldsymbol{x}(k) \tag{4.21}$$

where $A_{1,1}$, $A_{1,2}$, $A_{2,1}$, and $A_{2,2}$ are unknown parameters. Since the number of TCLs is fixed, we know that $A_{1,1} + A_{2,1} = 1$, $A_{1,2} + A_{2,2} = 1$, and $x_1 + x_2 = 1$. Treating $A_{2,1}$ and $A_{2,2}$ as states, we can derive the following nonlinear state/output equations:

$$x_2(k+1) = A_{2,1}(1 - x_2(k)) + A_{2,2}\, x_2(k) + u(k) \tag{4.22}$$

$$A_{2,1}(k+1) = A_{2,1}(k) \tag{4.23}$$

$$A_{2,2}(k+1) = A_{2,2}(k) \tag{4.24}$$

$$y(k) = c_p\, x_2(k). \tag{4.25}$$

Provided the system is locally observable, we can use an EKF to estimate the three parameters/states and compute the remaining parameter/states from the results. To check for local observability, we employ the method detailed in [3], which requires forming the discrete time, nonlinear, local observability matrix, Θ. Θ must be full rank for the system

to be locally observable. For our system, Θ is a 3×3 matrix defined as follows:

$$\Theta = \begin{bmatrix} \nabla y(k) \\ \nabla y(k+1) \\ \nabla y(k+2) \end{bmatrix} \tag{4.26}$$

where ∇y is the gradient of the output equation with respect to each of the states/parameters at time step k. For forced systems, to compute $y(k+n)$, where $n > 0$, in terms of $x_2(k)$, $A_{2,1}(k)$, and $A_{2,2}(k)$, we assume a series of inputs $u_1, ..., u_n$. We find that the $N_{\text{bin}} = 2$ system is locally observable everywhere, except when:

$$\det(\Theta) = A_{2,1} + u_1 - x_2 - A_{2,1}\, x_2 + A_{2,2}\, x_2 = 0. \tag{4.27}$$

We used an EKF to identify parameters and states in the $N_{\text{bin}} = 2$ system. We were unable to get the system to converge for more complicated systems (see Appendix D). More research is needed to determine if other nonlinear state estimation methods might be appropriate. Also, it may be possible to derive high order systems using parameters identified for the $N_{\text{bin}} = 2$ system along with knowledge and/or assumptions about the TCL parameters.

To implement the EKF, we have adapted the algorithm implemented in [21], which computes the Jacobian of $\boldsymbol{A}$ and $\boldsymbol{C}$ through complex step differentiation and then implements the discrete time Kalman Filter equations on the linearized system. We chose $\boldsymbol{Q}_{\text{KF}}$ iteratively and think of it as tuning parameter. R_{KF} is computed as in Scenario 3.

4.4.4 Controller design

We designed two simple controllers. Each entails first computing the total fraction of TCLs to switch either ON or OFF in the next time step, u_{goal}, defined as follows:

$$u_{\text{goal}}(k) = K \frac{P_{\text{total,des}}(k+1) - P_{\text{total,pred}}(k+1)}{N_{\text{TCL}} \bar{P}_{\text{ON}}} \tag{4.28}$$

where K is the control gain and $P_{\text{total,pred}}$ is the predicted aggregate power. If the state estimates are near perfect, then the aggregate power estimate is near perfect and K should be one. However, if there is significant error in the state estimates, $K = 1$ can result in high frequency oscillations. Therefore, for Scenario 1 and Scenario 2 (100% metering), we set $K = 1$, and for the other cases we selected K through iterative tuning to minimize RMS error.

The two controllers differ in how u_{goal} is divided amongst the bins. We assume that control actions can not force TCLs that are outside of the dead-band to switch. Since bins 1 and $\frac{N_{\text{bin}}}{2} + 1$ may contain TCLs that are outside of the dead-band, we do not apply control to these bins in systems with $N_{\text{bin}} > 2$. In systems with $N_{\text{bin}} = 2$, control actions are applied to TCLs regardless of their temperature state, generally resulting in too few TCLs switched.

Controller 1 divides u_{goal} equally amongst all of the allowable OFF or ON bins depending upon the sign of u_{goal}. Since the division of u_{goal} does not take into account the number of TCLs in each bin, a bin may be called on to switch more TCLs than are actually in that bin, resulting in a 100% switch probability and too few TCLs switched.

Controller 2 preferentially switches TCLs in bins closer to the edge of the dead-band, with the goal of switching TCLs that are about to 'naturally' switch, therefore minimizing the total number of times a single TCL is switched over time. Additionally, after switching off, an air conditioner, heat pump, or refrigerator must remain off for a certain length of time to allow liquid refrigerant to exit the compressor, and Controller 2 decreases the chance of compressor short cycling. The controller uses the current state estimates to assign fractions of TCLs to switch sequentially to the bin closest to the dead-band, the following bin, and so on, until all of u_{goal} has been assigned. Error in state estimates results in either too few or too many TCLs switched.

4.4.5 Benchmark controller

We have benchmarked our results against the results of a simple proportional controller. The switch probability at each time step was computed as follows:

$$u_{\text{rel}}(k) = K_{\text{P}} \frac{P_{\text{total,des}}(k) - P_{\text{total,meas}}(k)}{N_{\text{TCL}} \bar{P}_{\text{ON}}} \tag{4.29}$$

where K_{P} is the proportional gain, which was selected through iterative tuning to minimize RMS error. If $u_{\text{rel}} > 0$, TCLs are switched ON and if $u_{\text{rel}} < 0$, TCLs are switched OFF. We assume that, at each time step, each TCL receives u_{rel}; determines if it is ON or OFF and, subsequently, whether it may need to act; generates a random number to determine if it should switch or not; and, possibly, switches. Note that the benchmark controller may cause TCL compressors to short cycle, and so the tracking results reported Table 4.5 are optimistic.

4.4.6 Market signal and performance metrics

$P_{\text{total,des}}$ was designed to mimic CAISO 5-minute market signals, described in [14]. Uncorrelated random desired power values uniformly distributed between $\frac{3}{4}$ and $\frac{5}{4}$ times the steady state power consumption of the TCL population were chosen every 5 minutes. The TCLs are expected to ramp linearly from one desired value to the next, reaching the desired power value exactly halfway into the 5-minute period. To participate in the 5-minute market, generators provide bids including total capacity and ramp rates. A TCL population's total capacity could be 0 to $\sum_{i=1}^{N_{\text{TCL}}} P_i$; however, here we have used a smaller range because it reduces the probability that the population will deviate from steady state long enough to saturate (i.e., all TCLs at either the high or low end of the dead-band). TCL populations have essentially no ramping constraints.

We computed power tracking performance over 1800 time steps, always starting the system at approximately steady state. To evaluate tracking performance, we used two metrics: (1) the RMS power tracking error as a percentage of the steady state power consumption of the population and (2) the Compliance Threshold (CT). The CT is relevant to the CAISO, which defines non-compliance as deviation from desired power values by a specific threshold for more than n_{nc} consecutive intervals [14]. After a period of non-compliance, a resource can become compliant by coming within the threshold for n_{c} of consecutive intervals [14]. We define the CT as the minimum power value at which the TCL population is compliant, and set $n_{\text{nc}} = 3$ and $n_{\text{c}} = 1$, which are the values the CAISO initially uses for a new resource.

4.5 Simulation Results

We present the results of the model-estimator-controller system in Table 4.4 and the benchmark controller in Table 4.5. Note that the values of K and K_{P} should not be directly compared. Figure 4.7 shows the tracking performance of 1,000 TCL populations in each scenario. We have subtracted $P_{\text{total,ss}}$ from both the desired trajectory and the aggregate power consumption of the TCL population, so the results shown are zero mean deviations from steady state. Figure 4.8 shows state/parameter convergence, for Scenario 4 (5% forecast noise).

Table 4.4: Model-estimator-controller performance results.

Case	N_{TCL}	K	$N_{\text{bin}} = 2$* Ctrlr 1/2		$N_{\text{bin}} = 10$ Ctrlr 1		$N_{\text{bin}} = 10$ Ctrlr 2		$N_{\text{bin}} = 40$ Ctrlr 1		$N_{\text{bin}} = 40$ Ctrlr 2		$N_{\text{bin}} = 80$ Ctrlr 1		$N_{\text{bin}} = 80$ Ctrlr 2	
			RMS (%)	CT (kW)	RMS (%)	CT (kW)	RMS (%)	CT (kW)	RMS (%)	CT (kW)	RMS (%)	CT (kW)	RMS (%)	CT (kW)	RMS (%)	CT (kW)
Scenario 1:																
Reference Case	1,000	1.0	0.91	15	0.61	9	0.59	13	0.59	20	0.57	7	0.59	10	0.54	22
	10,000	1.0	0.59	180	0.25	49	0.29	35	0.26	74	0.30	31	0.26	29	0.31	69
Scenario 2:																
100% Metering	1,000	1.0	0.95	20	0.75	8	1.0	5	0.69	8	0.76	21	0.66	18	0.75	7
	10,000	1.0	0.59	162	0.31	32	0.36	80	0.30	10	0.41	90	0.29	67	0.45	129
30% Metering	1,000	0.70–0.99	5.0	102	5.0	62	4.8	93	4.8	42	4.5	38	4.4	71	4.8	137
	10,000	0.49–0.90	2.7	1,037	2.2	541	2.4	549	1.8	369	2.2	409	1.8	167	2.1	224
Scenario 3:																
5% Error**	1,000	0.31–0.90	7.6	156	5.9	15	6.9	117	5.2	72	6.3	180	4.9	74	6.8	97
	10,000	0.31–0.90	7.0	1,768	5.0	531	6.2	725	4.1	528	5.4	1,310	3.9	448	6.1	1,059
10% Error**	1,000	0.30–0.85	9.2	174	6.8	251	8.7	281	6.1	147	7.4	213	5.4	46	8.5	180
	10,000	0.30–0.85	9.3	1,675	6.6	1,707	8.7	2,256	4.9	543	6.6	1,651	4.4	819	7.5	2,309
Scenario 4:																
5% Error**	1,000	0.05	7.1	59												
	10,000	0.05	6.6	763												
10% Error**	1,000	0.02	8.8	167												
	10,000	0.02	8.1	1,445												

* For $N_{\text{bin}} = 2$, both controllers produce the same control input, so we only report one set of results.

** Distribution substation power forecast error standard deviation.

Table 4.5: Proportional controller performance results.

Case	N_{TCL}	K_{P}	RMS (%)	CT (kW)
100% Metering	1,000	6.5	1.1	29
	10,000	7.5	0.68	109
30% Metering	1,000	6.3	4.3	56
	10,000	6.5	1.5	220
5% Error*	1,000	0.3	9.1	189
	10,000	0.3	8.6	2,352
10% Error*	1,000	0.08	10.5	174
	10,000	0.08	10.2	1,879

* Distribution substation power forecast error standard deviation.

The results show that the model-estimator-controller system performs better than the proportional controller in all cases except Scenario 2 (30% metering). In Scenario 2 (30% metering), the measurement noise is highly autocorrelated so the Kalman Filter does not work well. Some slight performance improvements are possible by tuning R_{KF}; however, even then, the proportional controller performs better. A different state estimation technique that does not assume Gaussian white noise would likely perform better.

Generally, as less information is available for system identification, state estimation, and control, the tracking performance degrades. An exception is Scenario 4 in which the TCLs perform better than the TCLs in Scenario 3, 2 bins. This indicates that there is some value in the $\boldsymbol{A}$-matrix not being fixed. In all other cases (10, 40, and 80 bins), Scenario 3 performs better than Scenario 4, indicating the value in additional bins. In fact, comparing the performance of Controller 1 across systems with different numbers of bins, we find that systems with more bins almost always produce better tracking results. Systems with more bins more accurately predict aggregate power in the next time step, provided the TCL population model and state estimates are accurate.

Error in Scenario 1, the reference case, results from both our imperfect model and our imperfect method of applying the control. Our model is imperfect because we assume that the probability of a_i is independent of temperature and that TCLs in each bin are uniformly distributed over the temperature range of the bin. Also, as described previously, the model is best at predicting states for systems forced similarly to the system used to identify the model. The control is applied imperfectly because it is converted into switch probabilities.

In each case, increasing the number of TCLs in the population reduces tracking error (except one anomaly: Scenario 3, 10% forecast error). However, aggregating distribution substations does not help as much as we would expect if errors were independent across substations. If that was the case, we would expect the standard deviation of the tracking error to increase as $\sqrt{P_{\text{total}}}$ since the variance of independent random variables adds; however, we find that the standard deviation of the tracking error increases almost linearly, as shown in Figure 4.9. This is likely due to the fact that we are sending the same signal to TCLs in

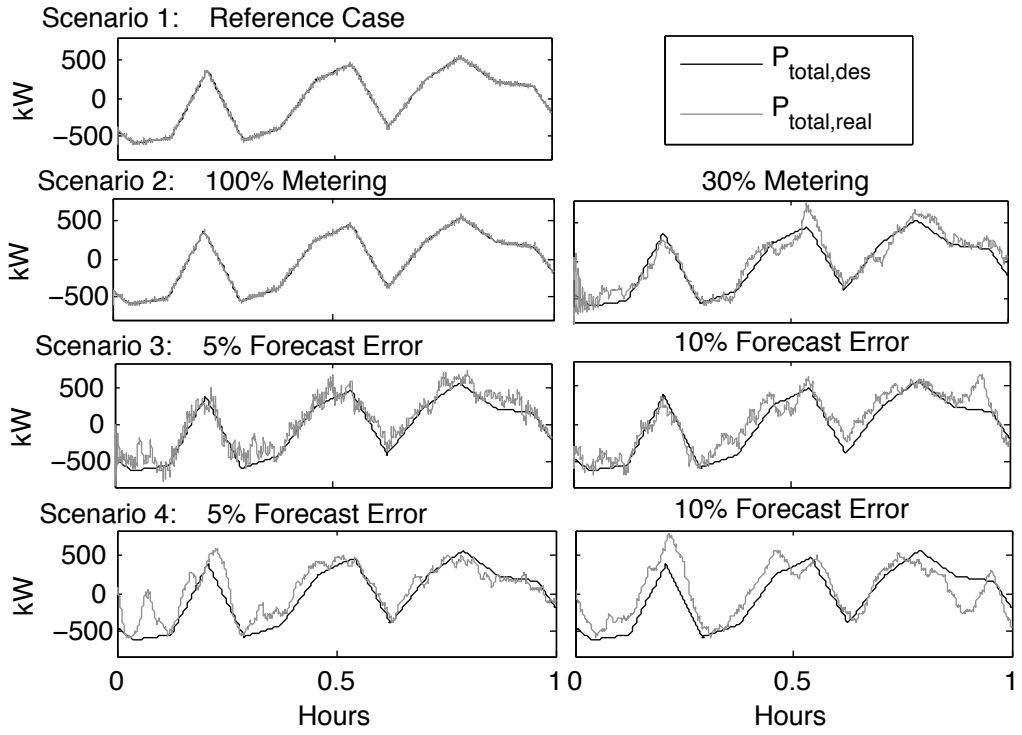


Figure 4.7: Tracking performance for Scenarios 1–3 (N_{bin}=40, Controller 1) and Scenario 4 (N_{bin}=2) with N_{TCL}=1,000.

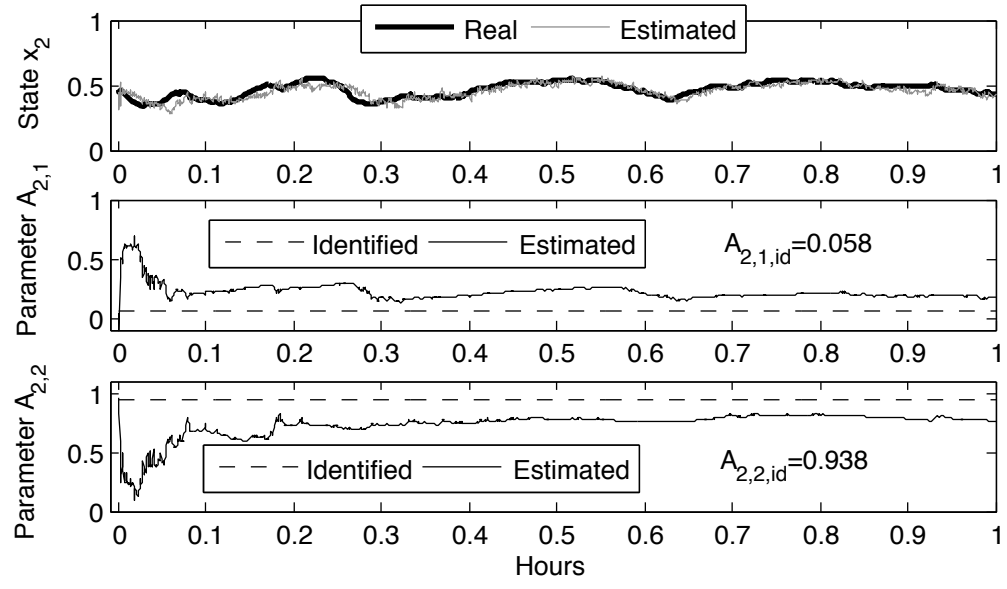


Figure 4.8: State/parameter convergence for Scenario 4 (5% forecast error) with 1,000 TCLs.

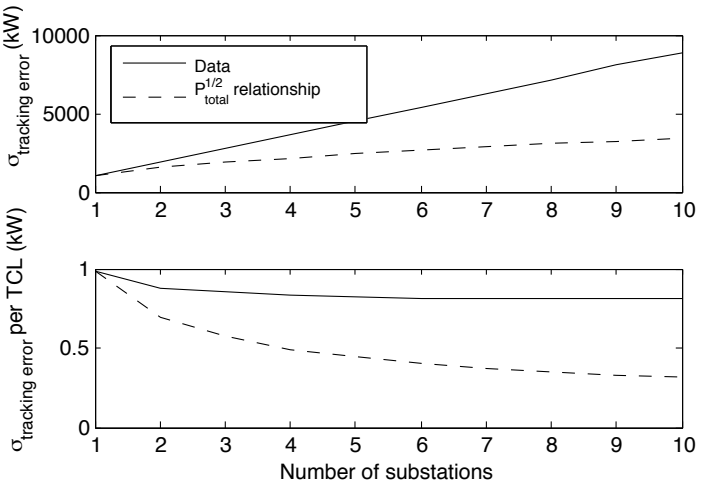


Figure 4.9: Standard deviation of tracking error versus number of distribution stations for Scenario 4 assuming 10% forecast error. ©2012 IEEE

each substation and when the TCLs in one substation area are performing poorly because they are near saturation, all are.

The results also show that Controller 2 generally performs worse than Controller 1. While both controllers use state estimates to predict aggregate power in the next time step, Controller 2 also uses state estimates to *allocate the control*, meaning its more sensitive to bad state estimates. Moreover, Controller 2 forces the system further from steady state by preferentially switching devices in bins closer to the dead-band, and this may act to increase plant-model mismatch. Despite increased RMS tracking error, for TCLs with compressors, it is advisable to use Controller 2 to minimize TCL switching and compressor short cycling. We computed the distribution of lengths of off times for both controllers and found that while Controller 1 results in some high-frequency switching (off times as low as 2-seconds), Controller 2 decreases off times much less, avoiding short cycling.

The CTs do not vary as systematically as the RMS error results because the values are computed with only 12 data points – the aggregate power at the halfway point of each 5-minute interval. There is no existing performance metric used to determine if a resource is ramping linearly. Since TCLs may not ramp linearly, as seen in Figure 4.7, it would be worthwhile to develop such a metric.

A number of assumptions affect our results. We assume that the ambient temperature is constant and the TCL population does not change over time. Therefore, we hold the model constant (except in Scenario 4). In a real system, the TCL population could change over time, possibly necessitating ongoing re-identification of the model. In the future, we plan to study how sensitive the performance results are to system variability. Additionally, in Scenario 3 and 4 we assume that we have access to unbiased distribution substation load forecasts. Biased noise would cause the Kalman Filter estimates to degrade, affecting TCL

population performance. This is another area for future research.

4.6 Chapter Conclusion

We have used a state bin transition model, state estimation techniques, and simple controllers to investigate how various levels of sensing and communications infrastructure affect the power tracking performance of populations of heterogeneous TCLs. Minimizing sensing and communications requirements reduces costs associated with DR, specifically real-time telemetry currently estimated to cost around $70,000 per meter [115]. We find that it is possible to control loads with very little information, though tracking performance generally degrades as less information is available. We have compared our results to that of a tuned proportional controller and found that our model-estimator-controller system performs better than the proportional controller, except when aggregate power measurement noise is significantly autocorrelated.

These findings provide encouraging evidence that, depending on engineering and policy objectives, the cost to instrument the "smart grid" may quite low. Indeed the required communications infrastructure may be no more substantial than a simple broadcast receiver at each load plus a low power transmitter and conventional SCADA system in each distribution substation. The appeal of this approach is that it can be implemented directly on top of existing distribution companies' efforts to modernize their networks, with little additional investment in infrastructure.

In the future, we plan to explore other methods for estimation and the use of more intelligent controllers. In this chapter, we have analyzed systems without communications delays or bandwidth constraints. Concepts from networked control systems and model predictive control (MPC) may be useful for further investigating these scenarios and assigning control actions for periods when communication is not possible. An MPC framework could also help minimize excessive TCL switching and keep the TCL population closer to steady state.

Chapter 5

TCL Resource, Revenues & Costs

In this chapter, we continue our we our investigation into emerging DR resources. Specifically, we estimate the size of the TCL resource, potential revenue from participation in markets, and break-even costs associated with deploying DR-enabling technologies. This chapter is largely based on [93][1] and [87].[2]

5.1 Chapter Introduction

To reduce greenhouse gas emissions many states have implemented renewable portfolio standards that require a certain percentage of electricity generation to come from renewable sources. Both wind and solar photovoltaics are expected to comprise a significant portion of new renewables; however, both technologies produce variable and uncertain power. As a result, system operators will need to procure more regulation and load following [84].

Rather than using existing and new generation units to provide the additional services needed, it may be more cost-effective and/or environmentally beneficial to provide these services with alternative technologies, namely energy storage devices (e.g., batteries, flywheels, compressed gas, and pumped hydro) and DR. We focus on DR for two reasons: First, it has significant potential for near-term deployment and impact. Second, it has been identified as a crucial component of the portfolio of energy technologies needed to achieve massive de-carbonization of our energy supply system [23].

One generally thinks of large C&I facilities as the best candidates for traditional DR programs because these customers consume large amounts of electricity and, subsequently, are able to curtail a substantial amount of power when a DR event is called. Moreover, it is

[1] With permission from my co-authors: Mark Dyson, Duncan Callaway, and Art Rosenfeld. "Using residential electric loads for fast demand response: The potential resource and revenues, the costs, and policy recommendations." to appear in *ACEEE*, Aug 2012.

[2] With permission from my co-author: Duncan Callaway. "The value of real-time data in controlling electric loads for demand response." *CMU Conference on the Electricity Industry*, Mar 2012.

simpler for the utility or aggregator to interact with a small number of large customers than a large number of small customers. However, there are several benefits to using aggregations of small residential loads to provide power system services. Specifically, these loads:

- can provide more reliable responses in aggregate than small numbers of large loads [18, 72];
- are spatially distributed;
- employ simple local controls, which both facilitates faster responsiveness [18] and makes them easy to model for state estimation and control purposes [75, 86]; and
- display continuous, not discrete, control responses in aggregate as opposed to large loads which generally employ DR strategies such as stepping down industrial processes, HVAC loads, or lighting [98].

TCL have perhaps the most potential of all types of residential loads because they operate within a hysteretic ON/OFF temperature dead-band. Adjusting TCL temperature within the dead-band adjusts the thermal energy stored in the TCL–much like a battery stores chemical energy–without a noticeable effect on end-use function [20]. Therefore, carefully designed load control schemes should allow us to control TCLs to track a desired trajectory (e.g., market signal or automatic generation control signal) while still performing their stated function.

In this chapter, we use models of TCL populations to estimate the size of the TCL resource, possible financial rewards, and costs associated with deploying enabling infrastructure. Additionally, we present a number of policy recommendations, which, if implemented, would allow loads to participate in energy and ancillary services markets more fully and effectively.

5.2 Model

Using the individual TCL model described in Section 4.2.1, we simulate thousands of individual TCLs to create a TCL population model (previously, referred to as the 'plant') and use the population model to perform the analysis described in the subsequent sections. As before, we assume that, in each time step, TCLs move through the temperature space without regard to their temperature dead-bands. Therefore, some TCLs may leave their dead-band within a time step. We assume that TCLs that leave their dead-band between k and $k+1$ are switched by their local controllers at $k+1$. Additionally, we assume that, through direct load control, we are able to switch TCLs ON/OFF, except TCLs that are outside of their dead-band. Therefore, TCLs within their dead-band are in the 'controllable subspace' (c) and TCLs outside of their dead-band are in the 'uncontrollable subspace' (uc). See Figure 5.1.

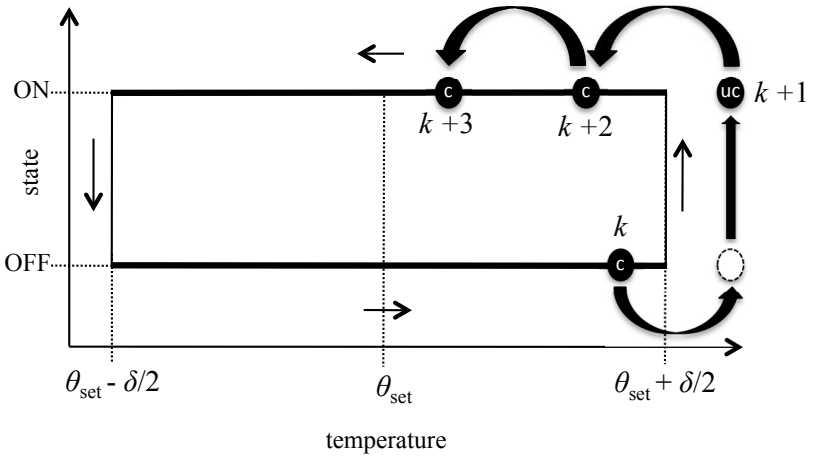


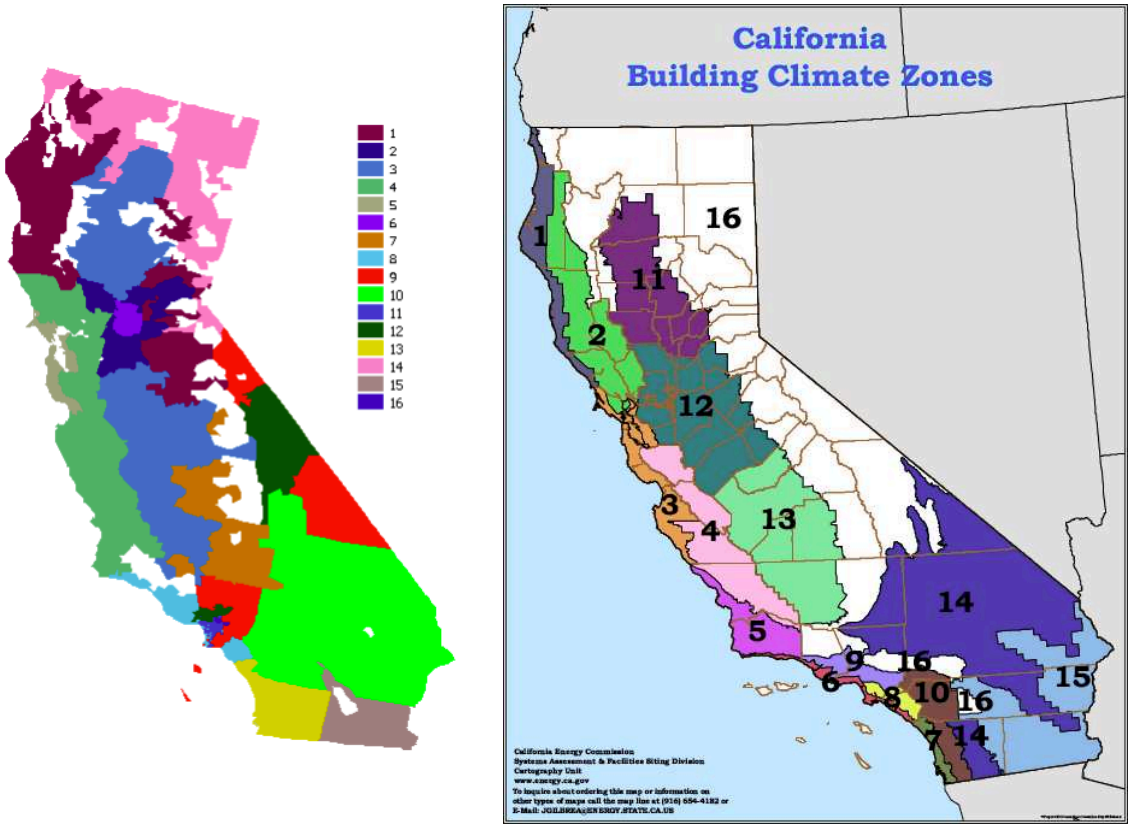
Figure 5.1: A cooling TCLs temperature dead-band and motion in discrete time.

We use the proportional controller described in Section 4.4.5 to track trajectories. Note that the proportional controller does not honor minimum compressor on/off times so TCLs with compressors (ACs, heat pumps, and refrigerators) may short cycle. Managing this issue, which may be possible with the use of the model-estimator-controller proposed in the previous chapter, specifically Controller 2 (Section 4.4.4), is a subject of future research.

As before, at each time step, the central controller broadcasts a new switch probability, u_{rel}, computed with (4.29), to all TCLs. If the value is negative, all TCLs that are ON must switch OFF with probability $-u_{\text{rel}}$, and if the value is positive, all TCLs that are OFF must switch ON with a probability u_{rel}.

5.3 Data Sources

We estimated the number of central air conditioners (ACs), refrigerators, central heat pump space heaters (referred to as simply 'heat pumps'), and electric resistance water heaters (referred to as simply 'water heaters') in five utility districts in California–PG&E, SCE, San Diego Gas and Electric (SDG&E), Sacramento Municipal Utility District (SMUD), and Los Angeles Department of Water and Power (LADWP)–using appliance saturation rates from the 2010 California Residential Appliance Saturation Survey [103] and forecasts for the number of households in 2012, and extrapolated to 2020, from the California Energy Commission (CEC) [85]. For ACs and heat pump heaters, we used data by CEC forecasting climate zone (1–13), since power consumption varies as a function of outdoor air temperature. No saturation rate data is available for climate zone 6 (SMUD) so we used saturation rate data for climate zone 2 (PG&E Sacramento Area). We used typical meteorological year (TMY) outdoor air temperature data from EnergyPlus [31]. Since the California building climate zones (Figure 5.2b) used by EnergyPlus are different than the CEC forecasting



(a) CEC Forecasting Climate Zones [68]. (b) California Building Climate Zones [24].

Figure 5.2: California climate zones. Source: CEC (reprinted with permission).

climate zones (Fig 5.2a), we mapped each CEC zone to a California building zone (Table 5.1). We also used interval locational marginal prices (LMPs) from CAISO [16] and weather data from NOAA [102] to compute energy cost savings through energy arbitrage.

5.4 Resource Potential Analysis

The resource potential of a population of TCLs depends upon the number TCLs and the TCLs' parameters. In this section, we use the TCL population model to compute the energy and power capacity of populations of heterogeneous TCLs. TCL parameters, given in Table 5.2 were derived or estimated from various sources (see Appendix E). To create a TCL population, we parameterize individual TCL models with TCL parameters (denoted by *) randomly drawn from uniform distributions between the maximum and minimum shown in the table. Other parameters are computed from the randomly drawn parameters.

Table 5.1: Mapping of CEC forecast climate zones to California building climate zones.

CEC Forecasting Climate Zone	California Building Climate Zone
1: PG&E North Coast Mountain	1: Arcata
2: PG&E Sacramento Area	12: Sacramento
3: PG&E Central Valley	12: Sacramento
4: PG&E East Bay*	2: Santa Rosa
5: PG&E San Francisco	3: Oakland
6: SMUD	12: Sacramento
7: SCE San Joaquin	13: Fresno
8: SCE LA Basin Coast	6: Los Angeles
9: SCE LA Basin Inland	10: Riverside
10: SCE Inland Empire	14: China Lake
11: LADWP	10: Riverside
12: LADWP	10: Riverside
13: SDG&E	7: San Diego

*This is not what Bay Area residents typically think of as the East Bay. See Figure 5.2a.

We verified the TCL parameters by comparing the model-predicted mean annual energy consumption per appliance to actual California data [35]. To derive the model prediction for ACs and heat pumps, we used the appliance and temperature data described above to simulate TCL populations in each climate zone for one year. Our current water heater model does not include water draw profiles–the effect of draws is currently captured by R–so we have less confidence in our duration curves for water heaters (Figure 5.4) than for other appliances. Also, though power consumption of a compressor (i.e., AC, refrigerator, heat pump) changes as a function of ambient temperature [123] and the COP of a TCL changes as a function of ambient temperature, we have not modeled these effects here. These are important topics for future research.

5.4.1 Resource potential of 1,000 TCLs

For each type of TCL, we calculate the energy and power capacity of 1,000 heterogeneous TCLs. We start each system at steady state and then force it so as to turn on all devices in each time step, until all TCLs are at the edge of the dead-band and we lose control. (Note, that in practice you would never want to force all TCLs to the edge of the dead-band; they would track a set point without dead-band and switch far too often.) Integrating the power curve, referenced to the steady state power consumption when all TCLs are at the edge of the dead-band , we can find the aggregate energy increase, e_{inc}. The aggregate power increase, p_{inc}, is the difference between the steady state power consumption and the power consumption after the first control action is applied. Similarly, we can compute the aggregate energy decrease, e_{dec}, and the aggregate power decrease, p_{dec}, by turning off all devices in each time step until all TCLs are at the edge of the dead-band. The aggregate

Table 5.2: TCL parameter assumptions and mean annual energy consumption. (Detailed description of sources in Appendix E.)

Parameter	ACs	Refrigerators	Heat Pumps	Water Heaters
ambient temperature, θ_a (°C)	variable	20	variable	20
dead-band width, δ (°C)	0.25–1.0	1–2	0.25–1.0	2–4
temperature set point, θ_{set} (°C)	18–27	1.7–3.3	15–24[a]	43–54
thermal resistance, R (°C/kW)	1.5–2.5	80–100	1.5–2.5	100–140
thermal capacitance, C (kWh/°C)	1.5–2.5	0.4–0.8	1.5–2.5	0.2–0.6
thermal time constant, RC (h)	2.25–6.25	32–80	2.25–6.25	20–84
energy transfer rate, P_{trans}	10–18	0.2–1.0	(-25.2)–(-14)	(-5)–(-4)
coefficient of performance, COP (-)	2.5	2	3.5	1
power consumption, P, (kW)	4–7.2	0.1–0.5	4–7.2	4–5
model-predicted mean annual energy consumption per appliance, 2012 (kWh)	1,610	858	4,700	2,100
actual California mean annual energy consumption per appliance, 2005 (kWh) [35]	1,637	867	N/A	2,121

[a] Though, we assume that people do not turn on their heat pumps if the mean daily temperature is > 15°C.

energy capacity, e_{max}, is defined as the sum of edec and e_{inc}. Similarly, the aggregate power capacity, p_{max}, is defined as the sum of p_{dec} and p_{inc}.

While the e_{max} is similar to a battery's energy capacity, p_{max} is not directly comparable to a battery's power capacity. While a battery can output the same power in each time step until the battery is discharged, the power capacity of a TCL population changes as a function of the TCL population state. Specifically, as TCLs move to the uncontrollable subspace (i.e., outside of the dead-band), we cannot switch TCLs, so a TCL population's power capacity is a function of the number of TCLs outside of the dead-band and the power consumption of those TCLs. TCLs are more likely to be outside of the dead-band as the system is forced further from steady state.

Results for each type of TCL are reported in Table 5.3. Since results for ACs and heat pumps are a function of temperature, as shown in Figure 5.3, we present the maximum and minimum capacities and the mean annual capacity in three representative climate zones. As Figure 5.3 shows, energy and power capacities are highest at moderately high and moderately low temperatures, when all TCLs are heating/cooling. Capacities decrease at the highest and lowest temperatures when TCLs struggle to cool/heat the space and are therefore unavailable for load control.

5.4.2 TCL resource potential in California

To understand the total TCL resource potential in California, we examined three scenarios:

(A) 2012 base scenario

Table 5.3: Estimates of the hourly energy and power capacity for 1,000 heterogeneous TCLs.

	Aggregate Energy Capacity (kWh)	Aggregate Power Capacity (kW)
ACs		
Max	490	5,500
Mean of CEC Zone 5 (San Francisco)	30	380
Mean of CEC Zone 6 (Sacramento)	90	1,100
Mean of CEC Zone 10 (LA Basin Inland)	120	1,500
Min	0	0
Heat Pumps		
Max	360	5,500
Mean of CEC Zone 5 (San Francisco)	190	3,000
Mean of CEC Zone 6 (Sacramento)	170	2,600
Mean of CEC Zone 10 (LA Basin Inland)	120	1,800
Min	0	0
Combined AC/Heat Pumps		
Max	490	5,500
Mean of CEC Zone 5 (San Francisco)	220	3,400
Mean of CEC Zone 6 (Sacramento)	260	3,700
Mean of CEC Zone 10 (LA Basin Inland)	240	3,300
Min	0	0
Water Heaters	1,200	4,300
Refrigerators	430	300

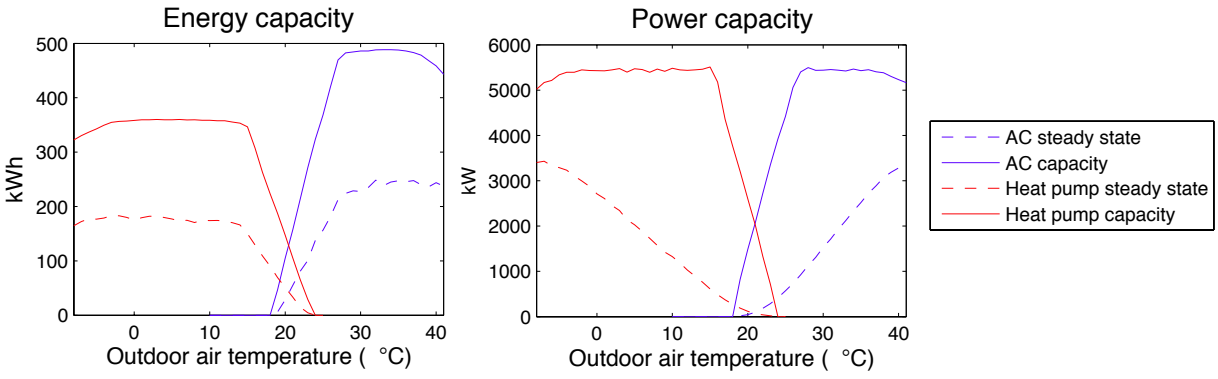


Figure 5.3: Energy and power capacity of 1,000 heterogeneous ACs or heat pumps.

(B) 2020 base scenario, with increased numbers of households, the same appliance saturation rates as 2012, and increased COPs (by 0.5)

(C) 2020 electrification scenario, which is the same as the 2020 base scenario except with increased saturation rates for heating (we assume 30% of gas water heaters and 30% of gas space heaters are electrified).

For each scenario, we calculated the hourly TCL energy and power capacities, based on total appliance ownership and weather data by climate zone. Resource duration curves, similar to load duration curves, are shown in Figure 5.4. These graphs show the DR resource available in each hour of the year, sorted by hourly sum; thus, the y-axis shows the DR capacity that exists for at least the number of hours shown on the x-axis. For example, power capacity of all DR resources in Scenario A is at least 9 GW for every hour of the year, and exceeds 20 GW for at least 2,000 hours. Comparing Scenarios A and B, we can see that the effect of COP improvements outweighs the effect of additional TCLs. In Scenario C, water heaters and heat pumps play a far larger role.

5.5 Revenue Potential Analysis

There are three ways that controlled TCL populations could profit: direct participation in existing markets, energy arbitrage, and participation in new markets for non-generator resources. We discuss the first two below.

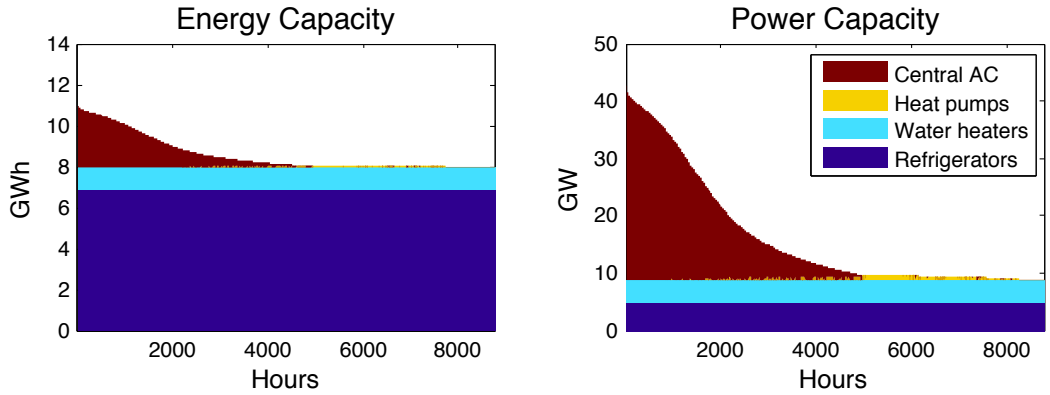
5.5.1 Direct participation in existing markets

Assuming TCL populations can be compared directly to energy storage devices, they could expect to earn $785-2,010/kW by participating in regulation or $600-1,000/kW by participating in load following over 10 years [44].[3]. Though regulation is a more lucrative market, the potential is smaller; specifically, the 10-year maximum market potential in California is estimated to be 80 MW for regulation and 2,900 MW for load following [44].

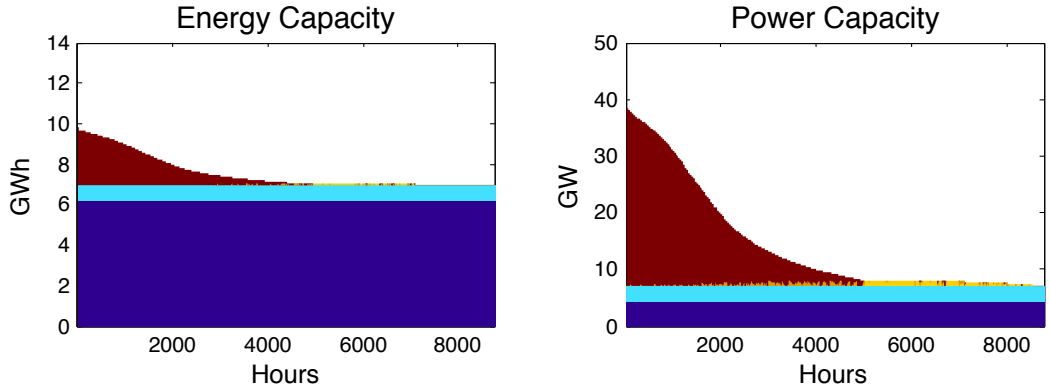
Considering both power and energy constraints, we can compute a TCL populations 'regulation power capacity' and 'load following power capacity.' The revenues quoted above assume certain discharge durations (i.e., the amount of time to discharge at rated capacity before recharging): 15-30 minutes for regulation and 2–4 hours for load following. Note that the regulation discharge duration is consistent with CAISO's Regulation Energy Management (REM) functionality, which allows non-generator resources to bid their 15-minute capacity [17]. Therefore, we compute the regulation power capacity to be a TCL populations 15-minute capacity and the load following power capacity to be a TCL populations 2-hour

[3]These are present values, assuming 2.5% inflation and 10% discount rate, which imply a nominal rate of 12.75%.

A. 2012 Base Scenario



B. 2020 Base Scenario



C. 2020 Electrification Scenario

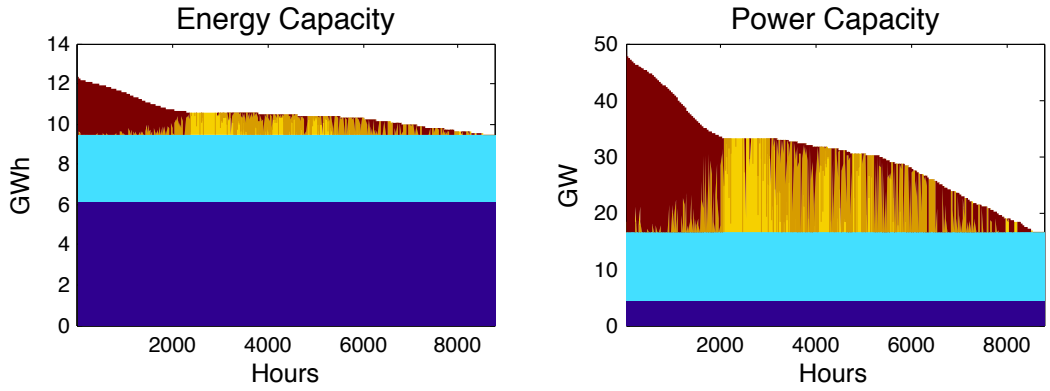


Figure 5.4: Resource duration curves for three scenarios.

capacity. Assuming a TCL population must provide the same amount of up and down regulation or load following, the regulation power capacity is the minimum of $e_{\max}/4$, p_{inc}, and p_{dec} and the load following power capacity is the minimum of $2 \times e_{\max}$, p_{inc}, and p_{dec}. Capacity and revenue results are reported in Table 5.4. For ACs and heat pumps, we use mean power and energy values computed over one year, and analyze the results by climate zone.

Table 5.4: Potential revenues from participation in regulation and load following

	Regulation			Load Following		
	Power Capacity, 1,000 TCLs (kW)	Present Value of Revenue over 10 yrs, 1,000 TCLs	Mean Nominal Revenue, ($/yr/TCL)	Power Capacity, 1,000 TCLs (kW)	Present Value of Revenue over 10 yrs, 1,000 TCLs	Mean Nominal Revenue ($/yr/TCL)
ACs						
CEC Zone 5	37	$52,000	$9.40	15	$12,000	$2.20
CEC Zone 6	220	$310,000	$57.00	45	$36,000	$6.60
CEC Zone 10	310	$430,000	$79.00	60	$48,000	$8.80
Heat Pumps						
CEC Zone 5	600	$840,000	$150.00	95	$76,000	$14.00
CEC Zone 6	650	$910,000	$170.00	85	$68,000	$12.00
CEC Zone 10	410	$570,000	$100.00	60	$48,000	$8.80
Combined ACs/Heat Pumps						
CEC Zone 5	640	$890,000	$160.00	110	$88,000	$16.00
CEC Zone 6	880	$1,200,000	$220.00	130	$100,000	$19.00
CEC Zone 10	720	$1,000,000	$180.00	120	$96,000	$18.00
Water Heaters	240	$330,000	$61.00	240	$190,000	$35.00
Refrigerators	97	$140,000	$25.00	97	$78,000	$14.00

All values given to two significant figures. Revenues computed with values from [44].

5.5.2 Energy arbitrage

In this section, we analyze the extent to which TCLs could save money by buying energy in CAISO's 5-minute energy market preferentially during low-price intervals and using the loads storage capacity to avoid purchasing energy during high-priced intervals, i.e., energy arbitrage. Unlike in the previous chapter, here, the TCL population is not sent a control signal to follow. Instead, TCL populations must compute their own trajectory, with the goal of minimizing their total energy costs. We employ a hierarchical approach consisting of three layers of control:

1. a linear program (LP) which solves for the optimal mean power usage in each 5-minute time interval based on a dynamical model of the system, physical system constraints, noisy forecasts of the 5-minute prices, and noisy forecasts of future temperatures, which dictate the available resource capacity for heating/cooling equipment

2. a TCL population controller which controls TCLs to follow the optimal trajectory

3. a population of TCLs (each with a hysteric local controller), each modeled with a difference equation

This approach can be thought of as similar to approaches commonly used in aviation path planning in which simplified models are used to compute trajectories and more precise and complicated models are used for control, e.g., [64].

We assume that DR energy arbitrage does not affect market prices, so our results are valid only if TCLs constitute a small fraction of the total market. The savings we calculate can best be thought to accrue to the load-serving entity (i.e., electric utility) that is purchasing the marginal energy on the spot market. New contracts between customers or aggregators and utilities would be required to distribute some or all of these savings to the owners of the appliances participating in arbitrage.

Method to compute optimal power consumption over time

The energy state (i.e., state of charge), e, of a TCL population in each time step, k, is related to past mean power usages, p:

$$(e_{k+1} - e_{\text{baseline},k+1}) = (e_k - e_{\text{baseline},k}) + (p_k - p_{\text{baseline},k})\Delta T \tag{5.1}$$

where ΔT is the length of each time step. p_{baseline} and e_{baseline} are the power usage and the energy state in the absence of control. These are functions of both the current ambient temperature *and* the past ambient temperatures. For example, an air conditioner does not immediately turn on in the morning when the outdoor air temperature rises above the TCL dead-band because the house has stored 'coolth' from the night.

e and p are positive and bounded by the aggregate capacities:

$$0 \leqslant e_k \leqslant e_{\max,k} \tag{5.2}$$
$$0 \leqslant p_k \leqslant p_{\max,k} \tag{5.3}$$

where $e_{\max}$ and $p_{\max}$ are also functions of both the current ambient temperature and the past ambient temperatures.

Our goal is to find $p^* = p_1, ...p_N$, which minimizes total energy costs over N time steps:

$$J = l_1 p_1 \Delta T + l_2 p_2 \Delta T + ... + l_N p_N \Delta T \tag{5.4}$$

subject to (5.1)-(5.3), where l_k is the energy price in each interval. This problem can be solved as a linear program (LP):

$$\min \quad f'x \tag{5.5}$$
$$\text{s.t} \quad Ax \leqslant b \tag{5.6}$$
$$\text{and} \quad x \geqslant 0 \tag{5.7}$$

where $x = [p_1, ..., p_N, e_1, ..., e_N]'$, $f = [l(1)\Delta T, ..., l(N)\Delta T, 0, ..., 0]$, and the constraints are given in (5.1)-(5.3).

In this preliminary implementation, we computed p_{baseline}, $e_{\max}$, and $p_{\max}$ as a function of only the current ambient temperature. Specifically, we assume TCLs are available for control if the current ambient temperature is above the TCL dead-band. We use the *linprog* function in MATLAB to implement the LP as follows:

1. Determine $e_{\max}$ and $p_{\max}$ for each feasible temperature and store the results.
2. Formulate f with price forecasts.
3. Formulate (5.1) in matrix form.
4. Use temperature forecasts to formulate (5.2) and (5.3).
5. Apply *linprog* to find x^*, the first half of which is p^*.
6. If iterating, save only the first control, $p(k)^*$, and return to Step 3 on the next time step.

Once we determine p^* (i.e., the control trajectory) for the whole year, we use the proportional controller described in Section 4.4.5 to track the trajectory. In the future, we plan to investigate more sophisticated approaches, e.g., [6].

Cases investigated

We used 2010 interval LMPs from MERCED_1_N001 [16], and temperature data from Merced 23 WSW weather station [102] to compute savings through energy arbitrage. Merced is in CEC climate zone 3, which experiences temperatures similar to those in CEC climate zone 6. We examined three cases:

(A) base case, in which we assume we have perfect temperature and price forecasts over 24 hours and so we use a 24-hour prediction horizon with no iteration

(B) iterative base case, in which we assume we have perfect temperature and price forecasts over the next hour and so we use a 1-hour prediction horizon, and at each interval we run the LP to generate p^*, though we implement only $p(1)$

(C) iterative noisy case, which is the same as the iterative base case except with noisy temperature and price forecasts.

In the last case, we model the temperature forecast error in the next hour as Gaussian white noise with standard deviation β, and we model the price forecast error, ω, with an autoregressive model of order one (i.e., AR(1)-process):

$$\omega_{k+1} = \gamma\omega_k + \alpha\epsilon_k \tag{5.8}$$

where ϵ is Gaussian white noise with standard deviation 1, and $\omega(1)= 0$. Varying α, β, and γ, we can understand how performance changes with forecast error.

Results

Figure 5.5 shows an example optimal control trajectory for approximately one day (Case C, assuming $\alpha = 40\$/\text{MWh}$, $\beta = 3°C$, and $\gamma = 0.3$). While the results of the LP predict significant energy cost savings, the TCL population is unable to track the signal closely resulting in lower energy savings (Figure 5.6). For example, considering the same error statistics used to generate Figure 5.5, the LP predicts 15% annual energy costs savings while the TCL population is only able to achieve about 9% annual energy cost savings. The discrepancy between the LP-predicted savings and the actual savings is in part because we do not take into account temperature history when solving for p_{baseline}, $e_{\max}$, and $p_{\max}$ causing us to significantly over estimate the resource in the mornings and somewhat underestimate the resource in the evenings. It is also because the proportional controller is unable to switch the correct number of TCLs in each time step, and this issue is further compounded by the fact that at each time interval (especially when the system is forced to the min or max) some TCLs are outside of the dead-band and therefore uncontrollable. It is likely that a more sophisticated control approach, such as that described in Section 4.4.4, could lessen these issues. Though using that approach would require a time-varying model.

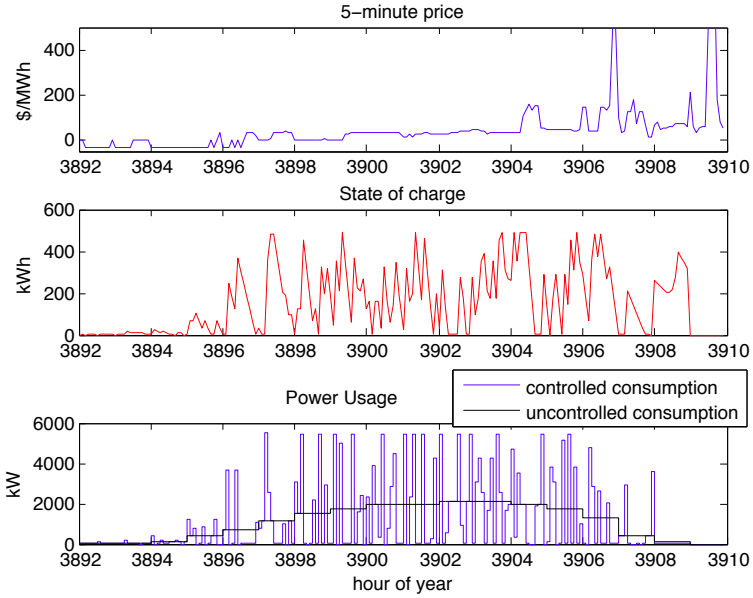


Figure 5.5: Optimized state of charge and control trajectory, given 5-minute energy prices and 1,000 air conditioners.

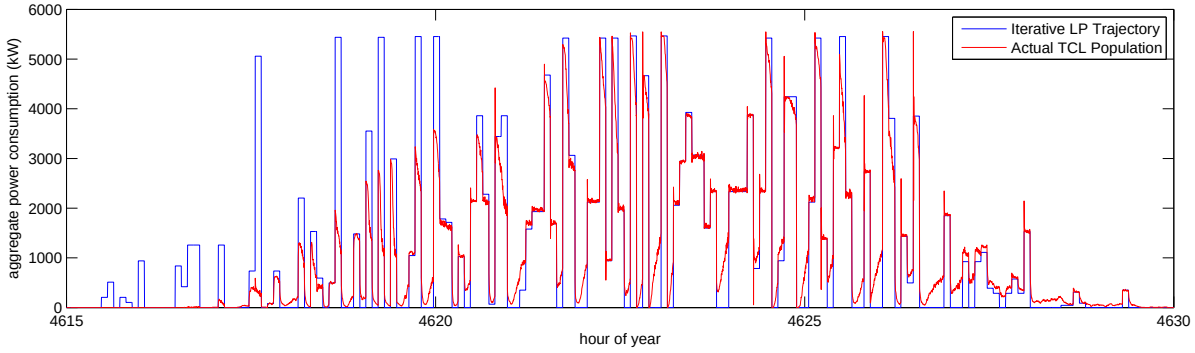


Figure 5.6: TCL population tracking performance: 1,000 air conditioners.

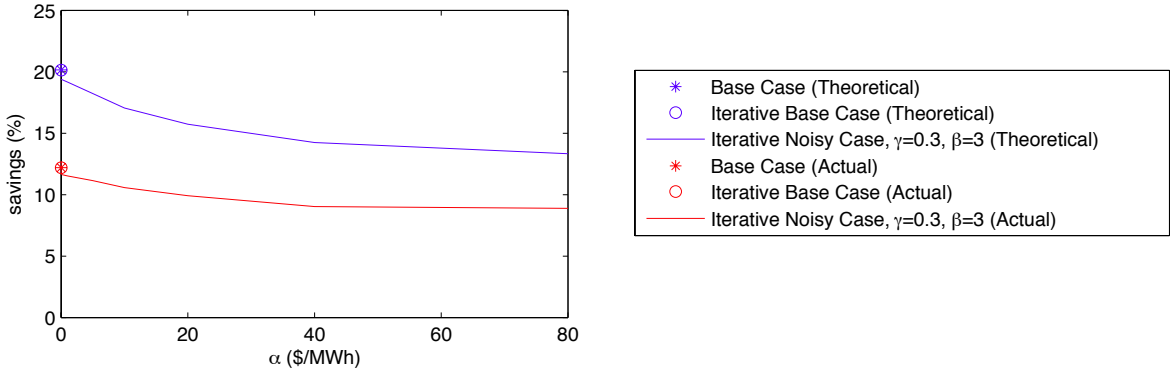


Figure 5.7: Effect of price forecast error on AC energy cost savings.

Table 5.5: Wholesale energy cost savings through arbitrage (Merced).

	Mean Energy Costs, No control ($/yr/TCL)	Mean Arbitrage Energy Costs, Base Case ($/yr/TCL)	Savings, Base Case (%)	Mean Arbitrage Energy Costs , Iterative Noisy Case ($/yr/TCL)	Savings, Iterative Noisy Case (%)
ACs	$92.85	$81.52	12.2%	$84.46	9.0%
Heat Pumps	$266.94	$239.74	10.2%	$249.11	6.7%
Combined	$359.79	$321.26	10.7%	$333.57	7.3%
Water Heaters	$81.80	$48.23	41.0%	$68.41	16.4%
Refrigerators	$33.44	$24.39	27.1%	$28.92	13.6%

Figure 5.7 shows the effect of price forecast error on energy cost savings. We plot both the savings predicted by the LP (theoretical) and the savings computed after controlling the TCL population to follow the LP-generated control trajectory (actual). As α increases the savings decrease, but even for large values of α savings are possible. Similar analyses are possible for γ and β, though we do not expect β to be very large and we find that varying γ does not significantly affect the results.

In Table 5.5, we show the actual mean nominal energy costs along with mean arbitrage energy costs and savings for each type of TCL. Case 2 results are very similar to Case 1 so we do not show them. For Case 3, we set $\alpha = \$40/\text{MWh}$, $\beta = 3\text{C}$ for ACs and heat pumps and 0°C for water heaters and refrigerators, and $\gamma = 0.3$. For control, we set $K_{\text{P}} = 5$ for ACs and heat pumps and $K_{\text{P}} = 2$ for water heaters and refrigerators.

5.6 Cost of Enabling Technologies

The cost of enabling technologies depends upon the system set-up. In this section, we list the components required to control populations of TCLs, and then use the potential revenues derived in the previous section to compute per-TCL break-even capital and annual costs.

5.6.1 Required components for TCL control

In our work, we assume that, in each time step, the same control signal is broadcast to all loads, and each load acts upon the control signal based upon its current ON/OFF state. Therefore, the following components are required TCL control:

- low latency broadcast communication system
- hardware/software that enables the loads to:
  - receive control signals
  - determine their current state i.e., whether they are currently ON or OFF
  - make decisions based on both the current control signal and their current state about whether or not they should switch ON or OFF
  - override the local control and switch (provided they are within their dead-band)

Additionally, equipment is required to provide the central controller with the following information at each time step:

1. the aggregate power consumption of the TCLs,
2. the number of devices available for control, and
3. the mean power consumption of the TCLs that are ON.

While #2 and #3 can be estimated based on system models, #1 requires real-time data. In the previous chapter, we described several ways to estimate the aggregate power consumption of the TCLs: with information from each of the loads, some of the loads, or the distribution substation. The method chosen is a tradeoff between the desired accuracy of the aggregate power measurement and the costs. For example, getting information from each of the loads results in a very accurate measurement of aggregate power, but requires low-latency data connections from each of the loads to the central controller. On the other hand, getting information from the distribution substation simply requires one low-latency data connection and a distribution substation power meter; however, we would need to estimate the aggregate power consumption of the TCLs, which make up a small fraction of the total substation load, using load forecasts, forecast noise models, TCL population models, and state estimation methods.

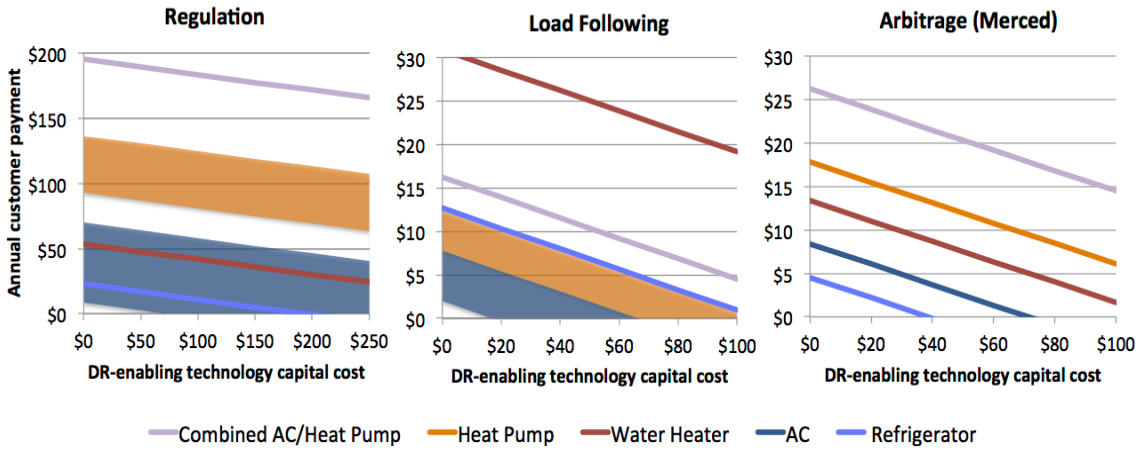


Figure 5.8: Per-TCL capital and annual costs required to break even.

Other costs associated with TCL control include hardware/software installation costs, re-occurring operations and maintenance costs, and financial compensation to electricity customers participating in these programs.

5.6.2 Break-even cost points

The costs of enabling infrastructure would vary dramatically depending upon location and equipment choices. Therefore, we identified break-even cost points for DR by comparing the revenue sources discussed above to potential ranges of capital costs (e.g., DR-enabling technology) and annual costs (e.g., annual customer payments). For capital costs up to $250 per TCL, we calculated the annualized capital cost using a lifetime of 20 years and a real discount rate of 10%. Then for each TCL and for each type of revenue stream (regulation, load following, and arbitrage), we calculated the maximum annual cost per TCL that would make total annual costs equal to the potential revenue. Results are presented in Figure 5.8. If the total installation and annual participant costs required to enable a TCL to participate in regulation, load following, or arbitrage fall below and to the left of the lines in Figure 5.8, then TCLs can potentially make money in these markets. For ACs and heat pumps providing regulation and load following, we present ranges bounded by values computed with the mean per-TCL earning potential from CEC Zone 10 (LA Basin Inland) and CEC Zone 5 (San Francisco). For combined ACs/heat pumps providing regulation and load following, we present values for CEC Zone 6 (Sacramento) since all three zones had very similar results. For arbitrage, we use values from the Merced analysis. Note that even the most cost-effective battery technologies being considered for grid regulation [38] have capital/annual costs much above those shown in Table 5.5.

5.7 Chapter Conclusion

We find that existing TCLs could provide a substantial portion of the fast timescale reserves required in power systems with high penetrations of intermittent renewables. Specific results depend upon the type of TCL and, for ACs and heat pumps, the climate that the TCL operates in. Comparing the arbitrage energy cost savings in Table 5.5 to the potential revenues from direct participation in energy markets in Table 5.4 (results for CEC Zone 6 are most relevant), we find that savings/revenues for arbitrage and load following are approximately the same for ACs and heat pumps while refrigerators and water heaters could earn more through load following. However, in each case, savings/revenues are only valid if TCLs constitute a small portion of the market; otherwise their actions could reduce market prices, decreasing potential revenues. All TCLs could earn much more by participating in regulation, though as mentioned before, the regulation market is relatively small.

The key to inclusion of DR in energy markets is the resolution of a number of policy barriers. Based on our findings, we have several policy recommendations. First, we recommend the design of new energy and ancillary services market products suited to loads, which do not have the same characteristics and constraints as generators. For example, our estimates of potential revenues from participation in existing markets are based on assumptions about TCL discharge durations because TCLs, just like energy storage devices, must follow zero mean control trajectories. While CAISO has proposed the REM functionality for regulation, a similar mechanism does not exist for load following. New market products could leverage the advantages of TCLs as well. For example, TCLs do not have substantial ramp rates and therefore can provide much faster responses to control signals.

Our second policy recommendation is the design and adoption of communications and appliance standards that will make it easier to engage residential loads in DR. It is much less expensive to manufacture TCLs with the ability to communicate necessary information with the load aggregator/utility and compute control actions based on control signals than to retrofit existing TCLs (though that would also be a key component of large-scale DR deployment). Good communication standards should minimize privacy and security issues. Additionally, appliance standards increasing TCL dead-bands would increase energy storage capacities. We find that doubling the dead-band approximately doubles the energy capacity.

Chapter 6

Conclusions

In future power systems, DR will be one of many available resources. Understanding how DR's capabilities and constraints compare to existing and new generation, and energy storage devices will help us optimize its use in ways that are minimally- or non-disruptive to the consumer. This dissertation has tackled a variety of important research topics aimed at this goal. In this chapter, we conclude with a summary of our key findings, a discussion of research implications, and a list of future research topics.

6.1 Key Findings

In Chapters 2 and 3, we focused on recent DR programs. We found that methods to visualize 15-minute interval electric load data, load parameters, and baseline models can help us 'ask the right questions' to determine opportunities for DR and evaluate DR performance. We also examined baseline model error and variability in open-loop control responses from one DR event to the next. We presented a method to compute baseline model error and a metric to determine how much observed DR variability results from baseline model error rather than real variability in response. We found that, in general, baseline model error is large. Though some facilities exhibit real DR variability, most observed variability results from baseline model error. However, in some cases, aggregations of C&I facilities exhibit real DR variability, which could create challenges for power system operation.

In Chapters 4 and 5, we investigated emerging DR paradigms. Our goal was to understand how access to offline and real-time data affects our ability to centrally control TCL aggregations to participate in load following. We developed a Markov Chain model of heterogeneous TCL aggregations and used Kalman filtering techniques for state estimation and joint parameter and state estimation. We also developed a look-ahead proportional controller and compared its performance to that of a simple proportional controller, demonstrating the value of the proposed model-estimator-controller approach. Simulations showed that it is possible to achieve power tracking RMS errors in the range of 0.26–9.3% of steady state

aggregated power consumption, and that results depend upon the information available for system identification, state estimation, and control. We found that, depending upon the performance required, TCLs may not need to provide state information to the central controller in real time or at all. We also attempted to answer a number of key questions: How big is the aggregated TCL resource? Will the revenues from participation in regulation, load following, and arbitrage cover the costs of DR-enabling technologies and customer participation payments? We found that current TCL energy storage capacity in California is 8–11 GWh, with refrigerators contributing the most. Annual revenues from participation in regulation vary from $10 to $220 per TCL per year depending upon the type of TCL and climate zone, while load following and arbitrage revenues are more modest at $2 to $35 per TCL per year. These results led to a number of policy recommendations that aim to make it easier to engage residential loads in fast timescale DR.

6.2 Research Implications

The results presented in Chapters 2 and 3 have important implications for DR program design and deployment. While we have shown that baseline models are very useful for M&V, especially when estimating DR performance over several DR events and averaging the results, we have also shown that baseline model error can be very large. Unfortunately, many DR programs continue to use baselines, generally simpler and less accurate than the one we propose here, for financial settlement purposes. On top of that, baselines built with historical load data can be gamed [52]. We recommend using DR programs that do not require baselines for financial settlement, such as dynamic pricing programs. However, if the utility and/or ISO sees value in demand and capacity bidding programs that require baselines for financial settlement, there are solutions such as 'build-your-own (BYO) baselines' [9] in which loads would purchase a baseline and DR sheds would be computed with respect to that baseline.

Another important implication resulting from the research in Chapter 3 is that DR event actuation strategies should be changed if utilities and ISOs value predictable and consistent demand sheds. Specifically, utilities should shift from open loop control to feedback control. However, for certain types of DR programs, such as CPP programs, it may not make sense to put too much value on event-to-event consistency since the goal of these programs is massive system-wide load shedding via price signals, not energy and frequency balancing, which require more precise DR actions. Emerging DR programs, such as the one proposed in Chapter 4, can provide this value to the system.

Chapters 4 and 5 have helped us to understand the role that residential loads can play in the emerging 'smart grid.' In Chapter 4, we focused on understanding requirements for sensing and communications when controlling TCLs for fast timescale energy balancing. Minimizing infrastructure and communications not only helps minimize system costs, but also helps avoid privacy and security issues associated with advanced metering and data

transmission. Some customers may be adverse to allowing their TCLs to transmit real-time data to a central controller, so strategies that leverage distribution substation level data are appealing. The key, then, is enabling TCLs to receive broadcast control signals, process those signals, and respond appropriately. Perhaps, the simplest way of doing this is through the design of new communications and appliance standards.

In Chapter 5, we estimated the size of the TCL resource and it is natural to ask: how could you increase TCL energy and power capacity? One answer is to increase electrification rates of certain types of appliances like water and space heaters, which is desirable for decarbonization of our energy supply [23]. However, there are ways to increase energy capacity without increasing electric energy use. For example, we have found that by doubling a TCL's dead-band you can double its energy capacity. While most consumers do not change dead-band settings, which are pre-set by manufacturers, this could be a powerful way for consumers to choose how much energy capacity they want to provide to the grid. It is also possible to improve energy capacity by adding insulation and thermal mass, e.g., chilled/hot water systems, ice storage systems, and even geothermal wells.

6.3 Future Research Topics

While this work has answered some important research questions, there are a number of important topics for future research including:

- Improved load modeling of C&I facilities. Better models would mean less baseline model error, giving us a better understanding of DR parameter variability. Promising options might combine statistical (e.g., Bayesian) and physical approaches.
- Development of feedback control strategies to ensure consistent control responses in C&I facilities, if consistency is valued by the DR program.
- Development of better state estimation techniques to handle non-Gaussian noise and nonlinearities associated with TCL aggregation models.
- Analysis of the control performance of TCL aggregations given communications constraints and delays.
- Time varying load models of TCL aggregations, for example models that explicitly handle changing ambient temperatures.
- Quantification of the emissions impacts of DR used for load following and regulation. DR would displace some of the power plants currently providing these services, but lifecycle environmental impacts are currently unclear.

- Pilot studies demonstrating real system behavior. Here, we have focused on simulation studies for TCL aggregations, but pilot studies will be essential for industry and consumer acceptance.
- Analysis of DR opportunities in other contexts, for example, in emerging regions where residential appliance adoption and investment in commercial HVAC systems are on the rise.

Bibliography

[1] H. Akbari. "Validation of an algorithm to disaggregate whole-building hourly electrical load into end uses". In: *Energy* 20.12 (1995), pp. 1291–1301.

[2] M.H. Albadi and E.F. El-Saadany. "A summary of demand response in electricity markets". In: *Electric Power Systems Research* 78 (2008), pp. 1989–1996.

[3] F. Albertini and D. D'Alessandro. "Remarks on the observability of nonlinear discrete time systems". In: *Proceedings of Conference on System Modelling and Optimization*. Prague, Czech Republic, 1995, pp. 155–162.

[4] O. Anaya-Lara et al. "Contribution of DFIG-based wind farms to power system short-term frequency regulation". In: *IEE Proc Gener, Transm. Distrib.* 153.2 (2006), pp. 164–170.

[5] S. Bashash and H.K. Fathy. "Modeling and control insights into demand-side energy management through setpoint control of thermostatic loads". In: *American Control Conference*. 2011, pp. 4546–4553.

[6] A. Bemporad, F. Borrelli, and M. Morari. "Model predictive control based on linear programming–The explicit solution". In: *IEEE Transactions on Automatic Control* 47.12 (2002), pp. 1974–1985.

[7] K. Bhattacharyya and M.L. Crow. "A fuzzy logic based approach to direct load control". In: *IEEE Transactions on Smart Grid* 11.2 (1996), pp. 708–714.

[8] S. Borenstein. "The trouble with electricity markets: Understanding California's restructuring disaster". In: *Journal of Economic Perspectives* 16.1 (2002), pp. 191–211.

[9] S. Borenstein, M. Jaske, and A. Rosenfeld. *Dynamic pricing, advanced metering and demand response in electricity markets*. Tech. rep. CSEMWP-105. University of California Energy Institute: Center for the Study of Energy Markets, 2002.

[10] S.D. Braithwait, D.G. Hansen, and J.D. Reaser. *2009 load impact evaluation of California statewide critical-peak pricing rates for non-residential customers: Ex post and ex ante report*. Christensen Associates Energy Consulting CALMAC Study ID: SDG0244.01. Apr. 2010.

[11] W.J. Burke and D.M. Auslander. "Low-frequency pulse width modulation design for HVAC compressors". In: *Proceedings of ASME 2009 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference*. San Diego, CA, Sept. 2009.

[12] J. Bushnell, B.F. Hobbs, and F.A. Wolak. *When it comes to demand response is FERC is own worst enemy?* Tech. rep. CSEMWP-191. University of California Energy Institute: Center for the Study of Energy Markets, Aug. 2009.

[13] CAEC. *Highly volatile-load customer study for Southern California Edison, Pacific Gas and Electric company, and San Diego Gas & Electric*. Tech. rep. Christensen Associates Energy Consulting, Oct. 2010. URL: https://www.pge.com/regulation/DemandResponse2009-2011-Projects/Other-Docs/Joint/2010/DemandResponse2009-2011-Projects_Other-Doc_Joint_20101027-01.pdf.

[14] CAISO. *Business practice manual for market operations, Version 18*. Tech. rep. California Independent System Operator Business Practice Manuals Library, 2011. URL: https://bpm.caiso.com/bpm/bpm/version/000000000000129.

[15] CAISO. *California ISO participating load pilot project report*. Tech. rep. RM07-19-001 CAISO Order 719 compliance filing regarding assessment of smaller demand resources providing ancillary services. California Independent System Operator, 2010.

[16] CAISO. *Open access same-time information system site*. California Independent System Operator. 2012. URL: http://oasis.caiso.com.

[17] CAISO. *Regulation energy management draft final proposal*. Tech. rep. California Independent System Operator, Jan. 2011. URL: http://www.caiso.com/Documents/RevisedDraftFinalProposal-RegulationEnergyManagement-Jan13_2011.pdf.

[18] D.S. Callaway. "Can smaller loads be profitably engaged in power system services." In: *Proceedings of IEEE Power and Energy Society General Meeting*. Detroit, MI, 2011.

[19] D.S. Callaway. "Tapping the energy storage potential in electric loads to deliver load following and regulation, with application to wind energy". In: *Energy Conversion and Management* 50 (2009), pp. 1389–1400.

[20] D.S. Callaway and I.A. Hiskens. "Achieving controllability of electric loads". In: *Proceedings of the IEEE* 99.1 (2011), pp. 184–199.

[21] Y. Cao. *Learning the extended Kalman filter*. Tech. rep. MATLAB Central: File Exchange, 2008. URL: http://www.mathworks.com/matlabcentral/fileexchange/18189.

[22] P. Cappers, C. Goldman, and D. Kathan. "Demand response in U.S. electricity markets: Empirical evidence". In: *Energy* 35 (2010), pp. 1526–1535.

[23] CCST. *California's energy future: The view to 2050*. Tech. rep. California Council on Science and Technology, 2011. URL: http://ccst.us/publications/2011/2011energy.pdf.

[24] CEC. *California building climate zone map*. California Energy Commission. 2011. URL: http://www.energy.ca.gov/maps/renewable/building_climate_zones.html.

[25] C.S. Chen and J.T. Leu. "Interruptible load control for Taiwan power company". In: *IEEE Transactions on Power Systems* 5.2 (1990), pp. 460–465.

[26] D.E. Claridge. "A perspective on methods for analysis of measured energy data from commercial buildings". In: *Journal of Solar Energy Engineering* 120 (1998), p. 150.

[27] 109th Congress. *Energy Policy Act of 2005*. Aug. 2005. URL: http://www1.eere.energy.gov/femp/pdfs/epact_2005.pdf.

[28] K. Coughlin et al. "Statistical analysis of baseline load models for non-residential buildings". In: *Energy and Buildings* 41.4 (Apr. 2009), pp. 374–381.

[29] DOE. *Benefits of demand response in electricity markets and recommendations for achieving them*. Tech. rep. Department of Energy Report to the US Congress, 2006. URL: http://www.oe.energy.gov/DocumentsandMedia/congress_1252d.pdf.

[30] P. Du and N. Lu. "Appliance commitment for household load scheduling". In: *IEEE Transactions on Power Systems* 2.2 (2011), pp. 411–419.

[31] EERE. *EnergyPlus energy simulation software: Weather data*. US Department of Energy, Energy Efficiency and Renewable Energy. 2012. URL: http://apps1.eere.energy.gov/buildings/energyplus/cfm/weather_data.cfm.

[32] EERE. *Estimating appliance and home energy usage*. US Department of Energy, Energy Efficiency and Renewable Energy. 2012. URL: http://www.energysavers.gov/your_home/appliances/index.cfm/mytopic=10040.

[33] EERE. *Lower water heater temperature for energy savings*. US Department of Energy, Energy Efficiency and Renewable Energy. 2012. URL: http://www.energysavers.gov/your_home/water_heating/index.cfm/mytopic=13090.

[34] EERE. *Results and methodology of the life-cycle cost analysis for residential water heater efficiency standards*. Tech. rep. US Department of Energy, Energy Efficiency and Renewable Energy, 1998.

[35] EIA. *Consumption and expenditures, residential energy consumption survey (RECS)*. US Energy Information Agency. 2005. URL: http://www.eia.gov/consumption/residential/data/2005/.

[36] EIA. *Housing characteristics, residential energy consumption survey (RECS)*. US Energy Information Agency. 2009. URL: http://www.eia.gov/consumption/residential/data/2009/.

[37] S. El-Férik, S.A. Hussain, and F.M. Al-Sunni. "Identification of physically based models of residential air-conditioners for direct load control management". In: *Proceedings of 5th Asian Control Conference*. Melbourne, Australia, 2004.

[38] EPRI. *Electricity energy storage technology options: A white paper primer on applications, costs, and benefits*. Tech. rep. 1020676. Electric Power Research Institute, 1020676, 2010.

[39] EPRI. *The green grid: Energy savings and carbon reductions enabled by a smart grid*. Tech. rep. EPRI TR-1016905. Electric Power Research Institute, June 2008.

[40] ERCOT. *Load Participation in the ERCOT Nodal Market*. Tech. rep. Electricity Reliability Council of Texas Wholesale Market Subcommittee Demand-Side Working Group, June 2010.

[41] J.H. Eto et al. *Demand response spinning reserve demonstration*. Tech. rep. LBNL-62761. Lawrence Berkeley National Laboratory, 2007.

[42] J.H. Eto et al. "Innovative developments in load as a reliability resource". In: *Proceedings of the IEEE PES Winter Meeting*. Vol. 2. 2002, pp. 1002–1004.

[43] J. Eyer and G. Corey. *Energy storage for the electricity grid: Benefits and market potential assessment guide*. Tech. rep. SAND2010-0815. Sandia National Laboratory, 2010.

[44] J. Eyer and G. Corey. *Energy storage for the electricity grid: Benefits and market potential assessment guide*. Tech. rep. SAND2010-0815. Sandia National Laboratory, SAND2010-0815, 2010.

[45] M. Fels. "PRISM: an introduction". In: *Energy and Buildings* 9.1-2 (1986), pp. 5–18.

[46] FERC. *A national assessment of demand response potential*. Tech. rep. Federal Energy Regulatory Commission, Prepared by the Brattle Group; Freeman, Sullivan & Co.; and Global Energy Partners, June 2009. URL: `www.ferc.gov/legal/staff-reports/06-09-demand-response.pdf`.

[47] FERC. *National action plan on demand response*. Tech. rep. Federal Regulatory Energy Commission, Prepared with the support of The Brattle Group, GMMB, Customer Performance Group, Definitive Insights, and Eastern Research Group, June 2010. URL: `http://www.ferc.gov/legal/staff-reports/06-17-10-demand-response.pdf`.

[48] FERC. *Order No. 719, Wholesale competition in regions with organized electric markets*. Federal Energy Regulatory Commission. 2008. URL: `http://www.ferc.gov/whats-new/comm-meet/2008/101608/E-1.pdf`.

[49] FERC. *Order No. 745, Demand response compensation in organized wholesale energy markets*. Federal Energy Regulatory Commission. 2011. URL: `http://www.ferc.gov/EventCalendar/Files/20110315105757-RM10-17-000.pdf`.

[50] FERC. *Order No. 755, Frequency regulation competition in organized wholesale power markets*. Federal Energy Regulatory Commission. 2011. URL: `http://www.ferc.gov/whats-new/comm-meet/2011/102011/E-28.pdf`.

[51] H. Fraser. "The importance of and active demand side in the electricity industry". In: *The Electricity Journal* 14.9 (2001), pp. 52–73.

[52] M.L. Goldberg and G.K. Agnew. *Protocol development for demand response calculation–findings and recommendations*. Tech. rep. CEC 400-02-017F. California Energy Commission (KEMA-XENERGY), 2003.

[53] J. Granderson et al. *Building energy information systems: State of the technology and user case studies*. Tech. rep. LBNL-2899E. Lawrence Berkeley National Laboratory, 2009.

[54] J. Haberl, R. Sparks, and C. Culp. "Exploring new techniques for displaying complex building energy consumption data". In: *Energy and Buildings* 24.1 (1996), pp. 27–38.

[55] J. Haberl et al. "Improving the paradigm for displaying complex building energy consumption data." In: *Proceedings of ASME International Solar Energy Conference*. Washington, DC, Apr. 1993, pp. 475–485.

[56] J.S. Haberl and M. Abbas. "Development of graphical indices for viewing building energy data: Part I-II". In: *Journal of Solar Energy Engineering* 120 (1998), pp. 156–167.

[57] J.S. Haberl and S. Thamilseran. "The great energy predictor shootout. II: Measuring retrofit savings". In: *ASHRAE Journal* 40.1 (1996), pp. 49–56.

[58] D. Hammerstrom et al. "Standardization of a hierarchical transactive control system". In: *Proceedings of the Grid Interop* 9 (2007).

[59] D.G. Hansen. *2009 load impact evaluation and cost effectiveness tests of California statewide automated demand response programs*. Christensen Associates Energy Consulting CALMAC Study ID: SDG0246.01. Sept. 2010.

[60] G. Heffner. *Demand response valuation frameworks paper*. Tech. rep. LBNL-2489E. Lawrence Berkeley National Laboratory, 2009.

[61] K.Y. Huang, H.C. Chin, and Y.C. Huang. "A model reference adaptive control strategy for interruptible load management". In: *IEEE Transactions on Power Systems* 19.1 (2004), pp. 683–689.

[62] M.D. Ilić et al. "A possible engineering and economic framework for implementing demand side participation in frequency regulation at value". In: *Proceedings of IEEE PES General Meeting*. Detroit, MI, 2011.

[63] R. Judkoff et al. "Buildings in a test tube: Validation of the short-term energy monitoring (STEM) method". In: *Proceedings of ASES National Solar Conferences Form.* Washington, DC, 2001.

[64] M. Kamgarpour, V. Dadok, and C. Tomlin. "Trajectory generation for aircraft subject to dynamic weather uncertainty". In: *IEEE Conference on Decision and Control*. Atlanta, GA, 2010.

[65] S. Karatasou, M. Santamouris, and V. Geros. "Modeling and predicting building's energy use with artificial neural networks: Methods and results". In: *Energy and Buildings* 38.8 (2006), pp. 949–958.

[66] S. Katipamula. "Great energy predictor shootout II: Modeling energy use in large commercial buildings". In: *ASHRAE Transactions* 102 (1996), pp. 397–404.

[67] S. Katipamula, T.A. Reddy, and D.E. Claridge. "Multivariate regression modeling". In: *Journal of Solar Energy Engineering* 120 (1998), p. 177.

[68] KEMA-XENERGY. *California statewide residential appliance saturation survey*. California Energy Commission CEC400-04-009. June 2004.

[69] S. Kiliccote, M.A. Piette, and D. Hansen. "Advanced controls and communications for demand response and energy efficiency in commercial buildings". In: *Proceedings of Second Carnegie Mellon Conference in Electric Power Systems: Monitoring, Sensing, Software and Its Valuation for the Changing Electric Power Industry*. Pittsburgh, PA, Jan. 2006.

[70] S. Kiliccote et al. "Installation and commissioning automated demand response systems". In: *Proceedings of 16th National Conference on Building Commissioning*. Newport Beach, CA, 2008.

[71] S. Kiliccote et al. "Using open automated demand response communications in demand response for wholesale ancillary services". In: *Proceedings of Grid Interop Forum*. Denver, CO, 2009.

[72] B. Kirby. *Spinning reserve from responsive loads*. Tech. rep. ORNL/TM-2003/19. Oak Ridge National Laboratory, 2003.

[73] J.K. Kissock and C. Eger. "Measuring industrial energy savings". In: *Applied Energy* 85.5 (2008), pp. 347–361.

[74] J.K. Kissock, T.A. Reddy, and D.E. Claridge. "Ambient-temperature regression analysis for estimating retrofit savings in commercial buildings". In: *Journal of Solar Energy Engineering* 120 (1998), p. 168.

[75] S. Koch, J.L. Mathieu, and D.S. Callaway. "Modeling and control of aggregated heterogeneous thermostatically controlled loads for ancillary services". In: *Proceedings of the Power Systems Computation Conference*. Stockholm, Sweden, Aug. 2011.

[76] J. Kondoh, N. Lu, and D.J. Hammerstrom. "An evaluation of the water heater load potential for providing regulation service". In: *IEEE Transactions on Power Systems* 26.3 (2011), pp. 1309–1316.

[77] J.F. Kreider and J.S. Haberl. "Predicting hourly building energy use: The great energy predictor shootout–Overview and discussion of results". In: *ASHRAE Transactions* 100.2 (1994), pp. 1104–1118.

[78] D. Krischen. "Demand-side view of electricity markets". In: *IEEE Transactions on Power Systems* 18.2 (2003), pp. 520–527.

[79] S. Kundu et al. "Modeling and control of thermostatically controlled loads". In: *Proceedings of* 17^{th} *Power Systems Computation Conference*. Stockholm, Sweden, 2011.

[80] Latham and Watkins LLP. *RTOs and ISOs Prepare to Comply with FERC's Demand Response Compensation Rule; Questions Raised Concerning FERC's Jurisdiction*. Latham's Clean Energy Law Report. 2011. URL: `http://www.cleanenergylawreport.com/energy-regulatory/rtos-and-isos-prepare-to-comply-with-fercs-demand-response-compensation-rule-questions-raised-concer/`.

[81] S. Lefebvre and C. Desbiens. "Residential load modeling for predicting distribution transformer load behavior, feeder load and cold load pickup". In: *Int. J. Elec. Power* 24 (2002), pp. 285–293.

[82] N. Lu and D.P. Chassin. "A state queuing model of thermostatically controlled appliances". In: *IEEE Transactions on Power Systems* 19.3 (2004), pp. 1666–1673.

[83] N. Lu, D.P. Chassin, and S.E. Widergren. "Modeling uncertainties in aggregated thermostatically controlled loads using a state queueing model". In: *IEEE Transactions on Power Systems* 20.2 (2005), pp. 725–733.

[84] Y.V. Makarov et al. "Operational impacts of wind generation on California power systems". In: *IEEE Transactions on Power Systems* 24.2 (2009), pp. 1039–1050.

[85] L. Marshall and T. Gorin. *California energy demand 2008-2018: Staff revised forecast*. Tech. rep. California Energy Commission, CEC-200-2007-015-SF2, 2007.

[86] J.L. Mathieu and D.S. Callaway. "State estimation and control of heterogeneous thermostatically controlled loads for load following". In: *Proceedings of the Hawaii International Conference on Systems Science*. Wailea, HI, Jan. 2012, pp. 2002–2011.

[87] J.L. Mathieu and D.S. Callaway. "The value of real-time data in controlling electric loads for demand response". In: *Carnegie Mellon University Conference on the Electricity Industry: Data Driven Sustainable Energy Systems*. Pittsburgh, PA, 2012.

[88] J.L. Mathieu, D.S. Callaway, and S. Kiliccote. "Examining uncertainty in demand response baseline models and variability in automated responses to dynamic pricing". In: *Proceedings of the IEEE Conference on Decision and Control and European Control Conference CDC-ECC, 2011*. 2011, pp. 4332–4339.

[89] J.L. Mathieu, D.S. Callaway, and S. Kiliccote. "Variability in automated responses of commercial buildings and industrial facilities to dynamic electricity prices". In: *Energy and Buildings* 43 (2011), pp. 3322–3330.

[90] J.L. Mathieu, S. Koch, and D.S. Callaway. "State estimation and control of electric loads to manage real-time energy imbalance". In: *TBD (in review)* (2012).

[91] J.L. Mathieu et al. "Characterizing the response of commercial and industrial facilities to dynamic pricing signals from the utility". In: *Proceedings of ASME 2010 4th International Conference on Energy Sustainability*. Phoenix, AZ, May 2010.

[92] J.L. Mathieu et al. "Quantifying changes in building electricity use, with application to demand response". In: *IEEE Transactions on Smart Grid* 2.3 (2011), pp. 507–518.

[93] J.L. Mathieu et al. "Using residential electric loads for fast demand response: The potential resource and revenues, the costs, and policy recommendations". In: *Proceedings of the ACEEE Summer Study on Buildings (to appear)*. Pacific Grove, CA, Aug. 2012.

[94] J.M. Mauricio et al. "Frequency regulation contribution through variable-speed wind energy conversion systems". In: *IEEE Transactions on Power Systems* 24.1 (2009), pp. 173–180.

[95] A. Molina et al. "Implementation and assessment of physically based electrical load models: application to direct load control residential programmes". In: *Int. J. Elec. Power* 150.1 (2003), pp. 61–66.

[96] A. Molina-Garcia, F. Bouffard, and D.S. Kirschen. "Decentralized demand-side contribution to primary frequency control". In: *IEEE Transactions on Power Systems* 26.1 (2011), pp. 411–419.

[97] M.G. Morgan et al. *The many meanings of 'smart grid'*. Tech. rep. Paper 22. Carnegie Mellon University Department of Engineering and Public Policy, 2009. URL: `http://repository.cmu.edu/epp/22`.

[98] N. Motegi et al. *Introduction to commercial building control strategies and techniques for demand response*. Tech. rep. LBNL-59975. Lawrence Berkeley National Laboratory, 2007.

[99] N. Navid-Azarbaijani and M.H. Banakar. "Realizing load reduction functions by aperiodic switching of load groups". In: *IEEE Transactions on Power Systems* 11.2 (1996), pp. 721–727.

[100] NERC. *Long-term reliability assessment: 2008-2017*. Tech. rep. North American Electric Reliability Corporation Report, Oct. 2008. URL: `http://www.nerc.com/files/LTRA2008.pdf`.

[101] NOAA. *National Climatic Data Center*. National Oceanic and Atmospheric Administration. 2009. URL: `http://www7.ncdc.noaa.gov/CDO/dataproduct`.

[102] NOAA. *National Climatic Data Center: US Climate Reference Network Data Set: Merced 23 WSW*. National Oceanic and Atmospheric Administration. 2010. URL: http://www1.ncdc.noaa.gov/pub/data/uscrn/products/hourly02/2010/CRNH0202-2010-CA_Merced_23_WSW.txt.

[103] C. Palmgren et al. *2009 California residential appliance saturation survey*. Tech. rep. Prepared for the California Energy Commission by KEMA, CEC-200-2010-004, 2010.

[104] D.C. Park et al. "Electric load forecasting using an artificial neural network". In: *IEEE Transactions on Power Systems* 6.2 (1991), pp. 442–449.

[105] L. Pedersen, J. Stang, and R. Ulseth. "Load prediction method for heat and electricity demand in buildings for the purpose of planning for mixed energy distribution systems". In: *Energy and Buildings* 40.7 (2008), pp. 1124–1134.

[106] PG&E. *Peak day pricing*. Pacific Gas and Electric Company. 2012. URL: http://www.pge.com/pdp/.

[107] M.A. Piette et al. *Open automated demand response communications specification (Version 1.0)*. Tech. rep. CEC-500-2009-063. California Energy Commission, PIER Program, Apr. 2009.

[108] P. Price. *Methods for quantifying electric load shape and its variability*. Tech. rep. LBNL-3713E. Lawrence Berkeley National Laboratory, 2010.

[109] B. Ramanathan and V. Vittal. "A framework for evaluation of advanced direct load control with minimum disruption". In: *IEEE Transactions on Power Systems* 23.4 (2008), pp. 1681–1688.

[110] R. Ramanathan et al. "Short-run forecasts of electricity loads and peaks". In: *International Journal of Forecasting* 13.2 (1997), pp. 161–174.

[111] T.A. Reddy, J.K. Kissock, and D.K. Ruch. "Uncertainty in baseline regression modeling and in determination of retrofit savings". In: *Journal of solar energy engineering* 120 (1998), p. 185.

[112] D.K. Ruch, J.K. Kissock, and T.A. Reddy. "Model identification and prediction uncertainty of linear building energy use models with autocorrelated residuals". In: *Proceedings of ASME International Solar Energy Conference*. Apr. 1993, pp. 465–473.

[113] Sam. *Sam's fridge data*. University of California, Berkeley, ARCH.244: The Secret Life of Buildings. 2010. URL: http://www.ced.berkeley.edu/courses/fa10/arch244/?p=249.

[114] SCE. *Real-time pricing RTP-2 general service*. Southern California Edison. 2012. URL: http://asset.sce.com/Documents/Shared/RTP2.pdf.

[115] SCE. *Southern California Edison company 2012-2014 demand response program portfolio*. Tech. rep. Application No. A.11-03-003 filed before the Public Utilities Commission of California. Southern California Edison, 2011.

[116] K.P. Schneider, J.C. Fuller, and D.P. Chassin. "Multi-state load models for distribution system analysis". In: *IEEE Transactions on Power Systems* 26.4 (2011), pp. 2425–2433.

[117] Y.S. Sherif and S.S. Zahir. "Communication systems for load management". In: *IEEE Transactions on Power Apparatus and Systems* PAS-104.12 (1985), pp. 3329–3337.

[118] J.A. Short, D.G. Infield, and L.L. Freris. "Stabilization of grid frequency through dynamic demand control". In: *IEEE Transactions on Power Systems* 22.3 (2007), pp. 1284–1293.

[119] D. Simon. *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*. Wiley-Interscience, 2006.

[120] J.W. Taylor, L.M. De Menezes, and P.E. McSharry. "A comparison of univariate methods for forecasting electricity demand up to a day ahead". In: *International Journal of Forecasting* 22.1 (2006), pp. 1–6.

[121] The Learning Channel. *What is the ideal temperature for a refrigerator?* 2012. URL: `http://tlc.howstuffworks.com/home/question121.htm`.

[122] G.G. Vining. *Statistical Methods for Engineers*. Duxbury Press, 1998.

[123] WECC. *AC unit model specifications*. Tech. rep. Western Electricity Coordinating Council Load Modeling Task Force, 2007. URL: `http://www.wecc.biz/committees/StandingCommittees/PCC/TSS/MVWG/032708/Lists/Minutes/1/AC%20Unit%20Model%20Specifications.pdf`.

[124] H.J. Wellinghoff et al. "Creating regulatory structures for robust demand response participation in organized wholesale electric markets". In: *Proceedings of ACEEE Summer Study on Energy Efficiency in Buildings*. Pacific Grove, CA, 2008.

[125] J. Wellinghoff and D.L. Morenoff. "Recognizing the importance of demand response: The second half of the wholesale electric market equation". In: *Energy Law Journal* 28.2 (2007), pp. 389–419.

[126] G. Wikler et al. *2007 Auto-DR program: Task 13 deliverable: Auto-DR assessment study*. Tech. rep. Pacific Gas & Electric Company, 2007.

[127] G. Wikler et al. *Pacific Gas & Electric company 2007 Auto-DR program: Task 13 deliverable: Auto-DR assessment study*. Tech. rep. Pacific Gas and Electric Company, Dec. 2007. URL: `http://drrc.lbl.gov/pubs/pge-auto-dr-assessment-study.pdf`.

[128] H.L. Willis. In: *Electrical Transmission and Distribution Reference Book*. Raleigh, NC: ABB Power T&D Company, 1997. Chap. Characteristics of distribution loads.

[129] L. Yao, W.C. Chang, and R.L. Ten. "An iterative deepening genetic algorithm for scheduling of direct load control". In: *IEEE Transactions on Power Systems* 20.3 (2005), pp. 1515–1421.

Appendix A

Data Processing & Analysis Methods

A.1 Data

A list of the 38 California facilities (87 facility-years) analyzed is given in Table A.1. The weather station (and, in some cases, the backup weather station) used for each facility is also given. Details on each weather station are listed in Table A.2. Table A.3 gives a sense for how often temperature data from each weather station (in each year) were interpolated. If the temperature data had been collected every 15 minutes for the whole summer (May 1–Sept 30, which is 153 days) there would be 14,688 data points. If the data had been collected every hour there would be 3,672 data points. The variable d refers to the difference in time-stamp (measured in hours) between temporally adjacent temperature measurements. The last five columns of the table list the number times we observe a certain differences in time-stamps. For example, in the column labeled '$2 \leqslant d < 3$,' we list the number of times the difference in temporally adjacent time stamps is between two and three hours.

A.2 Load Prediction Method Implementation

For each facility-year, 15-minute interval temperature and demand data from May 1–Sept 30 were imported into MATLAB. Data from weekends and three holidays (Memorial Day, Independence Day,[1] and Labor Day) were removed.[2] The data was then split into DR days and non-DR days. If it was known that a facility did not participate in a certain DR event, that day was not tagged as a DR day for that facility-year. Moreover, data from that day was not used to build the baseline model for that facility-year, since there was some concern that non-participating facilities may be testing their DR strategies on DR days. Six facility-years contain such exceptions.

[1] In 2009, Independence Day was celebrated on July 3rd because July 4th was a Saturday.

[2] Facility 3 was always closed on Mondays. Therefore, for all 4 of its facility-years, data from Mondays were removed and DR events that occurred on Mondays were not analyzed.

Table A.1: All facilities analyzed including weather station.

	Facility Type	Location	Weather Station	Back-up Weather Station	Years of Data
1	Office	Fremont	Hayward		4
2	Office (w/ data center)	Concord	Concord		4
3	Museum	Oakland	Oakland		4
4	Office	Martinez	Concord		3
5	Office	Martinez	Concord		4
6	Prison/Jail	Martinez	Concord		4
7	Office	San Jose	San Jose	Hayward	1
8	Office	Foster City	San Carlos	Hayward	4
9	Office	Foster City	San Carlos	Hayward	4
10	Office	Foster City	San Carlos	Hayward	4
11	Retail	East Palo Alto	Palo Alto	Hayward	3
12	Office	Rocklin	Mather		1
13	Retail (w/ solar)	Hayward	Hayward		3
14	Retail	Emeryville	Oakland		3
15	Manufacturing (Bakery)	Oakland	Oakland		2
16	Retail (w/ solar)	Antioch	Concord		3
17	Retail (w/ solar)	Bakersfield	Bakersfield		3
18	Manufacturing	Milipitas	San Jose	Hayward	2
19	Manufacturing	Milipitas	San Jose	Hayward	2
20	Office	Milipitas	San Jose	Hayward	2
21	Manufacturing	Milipitas	San Jose	Hayward	2
22	Manufacturing	Milipitas	San Jose	Hayward	2
23	Manufacturing	Milipitas	San Jose	Hayward	2
24	Manufacturing	Milipitas	San Jose	Hayward	1
25	Manufacturing	Milipitas	San Jose	Hayward	2
26	Manufacturing	Milipitas	San Jose	Hayward	2
27	Office	Dublin	Livermore		2
28	Office (w/ data center)	Dublin	Livermore		2
29	Office	Foster City	San Carlos	Hayward	2
30	Office	Foster City	San Carlos	Hayward	2
31	Office	San Jose	San Jose	Hayward	1
32	Office	San Jose	San Jose	Hayward	1
33	Prison/Jail	San Jose	San Jose	Hayward	1
34	Office	San Jose	San Jose	Hayward	1
35	Office	San Jose	San Jose	Hayward	1
36	Office	San Jose	San Jose	Hayward	1
37	Office	San Jose	San Jose	Hayward	1
38	Prison/Jail	Milipitas	San Jose	Hayward	1

Table A.2: Weather station details.

USAF	WBAN ID	Station Name	Latitude	Longitude	Elevation
723840	23155	Bakersfield/Meadows	35.434 N	119.056 W	150.0 m
724833	23206	Mather Field	38.554 N	121.285 W	29.3 m
724927	23285	Livermore Municipal	37.694 N	121.817 W	121.0 m
724930	23230	Oakland/Metrop. Oak	37.755 N	122.221 W	26.8 m
724935	93228	Hayward Air Term	37.659 N	122.121 W	14.3 m
724936	23254	Concord/Buchanan	37.992 N	122.052 W	18.3 m
724937	23289	Palo Alto Airport	37.467 N	122.117 W	2.0 m
724938	93231	San Carlos Airport	37.517 N	122.250 W	1.0 m
724946	93232	San Jose/Reid/Hillv	37.333 N	121.817 W	41.0 m

Table A.3: Temperature data interpolation summary.

Weather Station	Year	Facilities*	Data Points	$d < 1$	$1 \leqslant d < 2$	$2 \leqslant d < 3$	$3 \leqslant d < 6$	$d \geqslant 6$
Bakersfield	2007	1	4205	2962	1217	17	4	4
	2008	1	4222	3017	1187	8	5	4
	2009	1	4309	3082	1214	7	3	2
Mather	2006	1	2941	1483	1080	260	110	7
Livermore	2008	2	4246	3174	1059	9	2	1
	2009	2	3993	2845	1133	10	2	2
Oakland	2006	1	4034	2924	1098	7	3	1
	2007	3	4082	3016	1027	28	6	4
	2008	3	4101	3028	1054	12	3	3
	2009	2	4246	3174	1059	9	2	1
Hayward	2006	2	4217	3130	1074	7	4	1
	2007	1	4202	3131	1045	13	6	6
	2008	2	4073	3009	1047	6	4	7
	2009	2	4286	3222	1050	6	2	5
Concord	2006	4	3804	2599	1193	6	4	1
	2007	5	3753	2587	1120	34	5	6
	2008	5	3688	2497	1165	13	6	6
	2009	4	3881	2701	1168	9	2	1
Palo Alto	2006	1	2068	1025	751	119	19	153
	2007	1	1180	895	650	120	60	154
	2008	1	2105	1140	711	90	8	155
San Carlos	2006	3	2117	1086	758	103	15	154
	2007	3	1772	841	560	141	73	156
	2008	5	1890	953	613	129	39	155
	2009	5	2041	1043	694	119	30	155
San Jose	2006	1	2240	1120	821	128	16	154
	2008	8	2109	1058	722	141	30	157
	2009	17	2200	1059	846	119	21	154

* Number of facilities for which we use this station/year as the primary weather station.

Next, days with power outages were removed so that they would not be used to create the baseline prediction models. Similarly, DR days with power outages were not analyzed. Power outages are difficult to define because of the way that 15-minute interval meters work. The meters measure energy consumed (kWh) during a 15-minute interval and convert the measured value into average power (kW) over that interval. Therefore, if a power outage occurs within one 15 minute interval and lasts less than 15 minutes the average power during that interval will not be zero. Therefore, 'power outage days' could not be defined as days in which the minimum power consumption is zero. Instead, 'power outage days' were defined as days in which the minimum power consumption during that day is less than 50% of the average minimum daily power consumption during the summer. All data was screened to ensure the usefulness of this rule. In two cases, the rule was modified:

1. Facilities with solar PV can have zero electric demand so we were unable to determine when power outages occurred. Therefore, for these facilities, we did not remove any 'power outage days.'

2. For one facility with a particularly volatile minimum daily power consumption we changed the rule so that 'power outage days' are defined as days in which the minimum power consumption during that day is less than 25% of the average minimum daily power consumption during the summer. Otherwise, many 'normal-looking' days would have been declared 'power outage days.'

All toll, few 'power outage days' were removed. On average 0.39 (out of 94 to 97) non-DR days and 0.16 (out of 9 to 12) DR days were removed from a facility-year worth of data. The maximum number of non-DR days removed from a single facility-year of data was 4, and the maximum number of DR days removed from a single facility-year of data was 3.

The remaining non-DR days were then split up by day of week. Each facility-year was left with ~ 20 Mondays, ~ 20 Tuesdays, etc. that could be used to create the baseline prediction model. Non-DR day electric demands were stacked, all Mondays first, then Tuesdays, etc. Non-DR day temperatures were stacked similarly.

The regression equations, (2.2) and (2.3), can be written in matrix form:

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \epsilon \tag{A.1}$$

where x is the parameter vector; y is the output vector (electric demand); ϵ is the error; and A is composed of time-of-week indicators, occupied mode temperature components, and unoccupied mode temperatures. Specifically, the first 96 columns of matrix A are time-of-week indicators for Mondays (there are 96 15-minute intervals in a day), the second 96 columns are time-of-week indicators for Tuesdays, etc. In total, the first 480 columns are time-of-week indicators. The next 6 columns, columns 481–486, contain occupied mode component temperatures (and zeros when the facility is in unoccupied mode). The last

column of matrix A, column 487, contains unoccupied mode temperatures (and zeros when the facility is in occupied mode).

Let n be the number of non-DR days used to create the baseline prediction model. The matrices/vectors $\boldsymbol{A}$ ($96n \times 487$), $\boldsymbol{x}$ (487×1), and $\boldsymbol{y}$ ($96n \times 1$) look as follows:

$$\boldsymbol{A} = \begin{bmatrix}
1 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & \dots & [T^1_{c,1}(t_1), 0] & \dots & [T^1_{c,6}(t_1), 0] & [0, T^1_u(t_1)] \\
0 & 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & \dots & [T^1_{c,1}(t_2), 0] & \dots & [T^1_{c,6}(t_2), 0] & [0, T^1_u(t_2)] \\
0 & 0 & 1 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & \dots & [T^1_{c,1}(t_3), 0] & \dots & [T^1_{c,6}(t_3), 0] & [0, T^1_u(t_3)] \\
\vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & & \vdots & & \vdots & \vdots \\
1 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & \dots & [T^2_{c,1}(t_1), 0] & \dots & [T^2_{c,6}(t_1), 0] & [0, T^2_u(t_1)] \\
0 & 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & \dots & [T^2_{c,1}(t_2), 0] & \dots & [T^2_{c,6}(t_2), 0] & [0, T^2_u(t_2)] \\
0 & 0 & 1 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & \dots & [T^2_{c,1}(t_3), 0] & \dots & [T^2_{c,6}(t_3), 0] & [0, T^2_u(t_3)] \\
\vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & & \vdots & & \vdots & \vdots \\
0 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & \dots & [T^1_{c,1}(t_{97}), 0] & \dots & [T^1_{c,6}(t_{97}), 0] & [0, T^1_u(t_{97})] \\
0 & 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 & 0 & 0 & \dots & \dots & [T^1_{c,1}(t_{98}), 0] & \dots & [T^1_{c,6}(t_{98}), 0] & [0, T^1_u(t_{98})] \\
0 & 0 & 0 & \dots & 0 & 0 & 1 & \dots & 0 & 0 & 0 & \dots & \dots & [T^1_{c,1}(t_{99}), 0] & \dots & [T^1_{c,6}(t_{99}), 0] & [0, T^1_u(t_{99})] \\
\vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & & \vdots & & \vdots & \vdots \\
0 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & \dots & [T^2_{c,1}(t_{97}), 0] & \dots & [T^2_{c,6}(t_{97}), 0] & [0, T^2_u(t_{97})] \\
0 & 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 & 0 & 0 & \dots & \dots & [T^2_{c,1}(t_{98}), 0] & \dots & [T^2_{c,6}(t_{98}), 0] & [0, T^2_u(t_{98})] \\
0 & 0 & 0 & \dots & 0 & 0 & 1 & \dots & 0 & 0 & 0 & \dots & \dots & [T^2_{c,1}(t_{99}), 0] & \dots & [T^2_{c,6}(t_{99}), 0] & [0, T^2_u(t_{99})] \\
\vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & & \vdots & & \vdots & \vdots \\
0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & \dots & [T^1_{c,1}(t_{193}), 0] & \dots & [T^1_{c,6}(t_{193}), 0] & [0, T^1_u(t_{193})] \\
0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 1 & 0 & \dots & \dots & [T^1_{c,1}(t_{194}), 0] & \dots & [T^1_{c,6}(t_{194}), 0] & [0, T^1_u(t_{194})] \\
0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 1 & \dots & \dots & [T^1_{c,1}(t_{195}), 0] & \dots & [T^1_{c,6}(t_{195}), 0] & [0, T^1_u(t_{195})] \\
\vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & & \vdots & & \vdots & \vdots \\
0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & \dots & [T^2_{c,1}(t_{193}), 0] & \dots & [T^2_{c,6}(t_{193}), 0] & [0, T^2_u(t_{193})] \\
0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 1 & 0 & \dots & \dots & [T^2_{c,1}(t_{194}), 0] & \dots & [T^2_{c,6}(t_{194}), 0] & [0, T^2_u(t_{194})] \\
0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 1 & \dots & \dots & [T^2_{c,1}(t_{195}), 0] & \dots & [T^2_{c,6}(t_{195}), 0] & [0, T^2_u(t_{195})] \\
\vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & & \vdots & & \vdots & \vdots
\end{bmatrix}$$

$$\boldsymbol{x} = \begin{bmatrix}
\alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_{97} \\ \alpha_{98} \\ \alpha_{99} \\ \vdots \\ \alpha_{193} \\ \alpha_{194} \\ \alpha_{195} \\ \vdots \\ \vdots \\ \beta_1 \\ \vdots \\ \beta_6 \\ \beta_u
\end{bmatrix}$$

$$\boldsymbol{y} = \begin{bmatrix} D^1(t_1) \\ D^1(t_2) \\ D^1(t_3) \\ \vdots \\ D^2(t_1) \\ D^2(t_2) \\ D^2(t_3) \\ \vdots \\ D^1(t_{97}) \\ D^1(t_{98}) \\ D^1(t_{99}) \\ \vdots \\ D^2(t_{97}) \\ D^2(t_{98}) \\ D^2(t_{99}) \\ \vdots \\ D^1(t_{193}) \\ D^1(t_{194}) \\ D^1(t_{195}) \\ \vdots \\ D^2(t_{193}) \\ D^2(t_{194}) \\ D^2(t_{195}) \\ \vdots \end{bmatrix}$$

Here, $T^k_{c,j}(t_i)$ indicates the j^{th} temperature component at time t_i for the k^{th} week.[3] In $\boldsymbol{A}$, we write $[T^k_{c,j}(t_i),\, 0]$ to show that the matrix entry is $T^k_{c,j}(t_i)$ if the facility is in occupied mode, or zero if the facility is in unoccupied mode. Similarly, we write $[0,\, T^k_u(t_i)]$ to show that the matrix entry is zero if the facility is in occupied mode, or $T^k_u(t_i)$ if the facility is in unoccupied mode. $D^k(t_i)$ indicates the demand at time t_i for the k^{th} week.

To estimate the parameter vector $\boldsymbol{x}$, we use the OLS estimator:

$$\hat{\boldsymbol{x}} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{y}. \tag{A.2}$$

Specifically, we use the MATLAB function *regress*. The benefits of using *regress* instead of computing $\hat{\boldsymbol{x}}$ directly are that the function eliminates any dependent rows and it ignores missing data. This last feature is important; significant portions of the demand and temperature data are missing (missing data is represented by NaN in MATLAB).

A.3 Baseline Model Fit

The baseline prediction model in (2.2) and (2.3) fits the data from different facility-years to differing degrees. In general, even when the model fit is poor, we apply the model to

[3]Here, a 'week' is unconventionally defined: The first week contains the first non-DR day Monday that has not been removed from the data set (due to a holiday or power outage), the first non-DR day Tuesday, etc. Therefore, in this definition, a 'week' doesn't necessarily contain days that occurred chronologically.

estimate the DR parameters; though poor model fit means high model error, resulting in high uncertainty on the DR parameter estimates. However, we do not apply the model if we observe significant structural changes in electricity usage in a certain facility-year. Though DR parameter estimates for the 'average DR event' would not be biased, DR parameter estimates for individual DR events could be significantly biased.

To check for structural changes in each facility-year (original we had 99 facility-years worth of data from 44 facilities), we plotted the regression residuals chronologically and then looked for both step changes and slow increases/decreases over time. For most facility-years, some clear changes and/or trends are visible in the residuals. We only 'threw out' facility-years if the changes/trends seemed particularly large as compared to those of other facility-years. In total, we did not analyze 12 facility-years. Causes of significant changes/trends include the following:

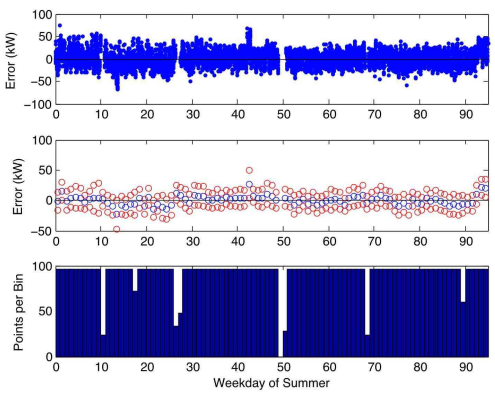
1. facility closure due to construction/maintenance for several weeks of the summer,
2. installation of solar PV mid-summer,
3. slow increase in power consumption throughout the summer indicative of the gradual population of a new building, and
4. schools running at part load during part of the summer (mid-June to August).

Plots of regression residuals versus time for two facility-years that will be used in later examples are given in Figures A.1a and A.1b. The top plots are scatter plots of residuals versus time. For the middle plots, we bin the residuals and plot the mean (blue circle) and standard deviation (red circle). The bottom plots show the number of points per bin. For both of these facility-years, the model fits reasonably well. Figure A.1c shows an example of a facility-year that was 'thrown out' because of an obvious structural change. Figures A.1d-A.1f show examples of marginal cases: facility-years that were not thrown out despite the fact that changes/trends are obvious. In each of these cases, the changes/trends were deemed small enough to ignore.

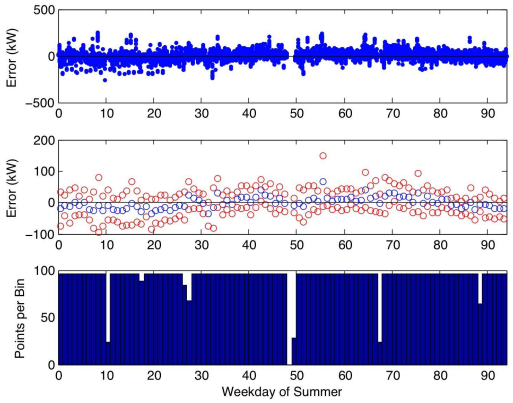
A.4 Baseline Model Error Analysis

A.4.1 Autocorrelation

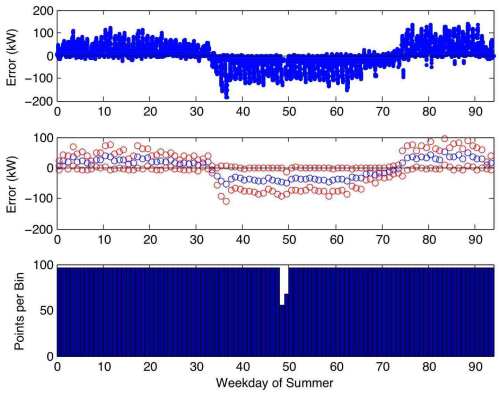
Two simple ways of testing for autocorrelation are to apply the Durbin-Watson Test, which tests for lag one autocorrelation, or plot the ACF and PACF. First, we attempted the Durbin-Watson Test using the MATLAB function *dwtest*. Unfortunately, the function was unable to produce an answer, likely because our $\boldsymbol{A}$ matrix ('design matrix') is too sparse. Therefore, we plotted the ACFs and PACFs. Results are given in Section 3.4.



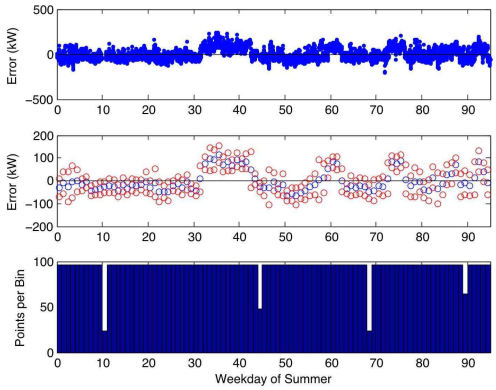
(a) Facility 1, 2008.



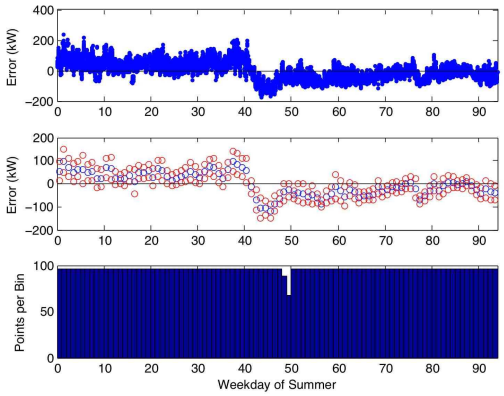
(b) Facility 11, 2008.



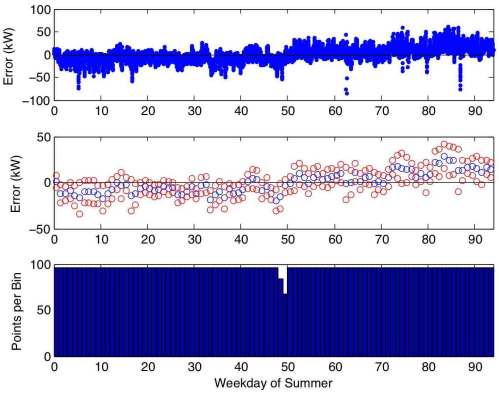
(c) School, 2009 (not analyzed): Structural change.



(d) Facility 6, 2008: Jumpy.



(e) Facility 23, 2009: Abrupt change.



(f) Facility 29, 2009: Slow change in consumption.

Figure A.1: Example plots of regression residuals versus time. The middle plot shows mean error (blue circles) and standard deviation (red circles).

A.4.2 Heteroscedasticity

Two common methods of checking for heteroscedasticity are the White Test and the Breush-Pagan Test with the Koenker modification (BPK Test). We attempted both tests using MATLAB user-written function *testhet*,[4] available on the MATLAB website. Using the function to perform the BPK Test produces an output that demonstrates that the regression residuals are almost certainly heteroscedastic (with the null hypothesis of homoscedasticity, the BPK Test produces a p-value of 3.1×10^{-13}). Unfortunately, the function is unable to perform the White Test on our data, again likely because our A matrix is too sparse.

Therefore, to understand how the variance of the regression residuals (referred to as 'error variance') changes as a function of the input and output variables, we analyzed error as a function of time-of-week, temperature, actual load, and predicted load (all of which are correlated). (To investigate model over-fitting, we repeated this analysis with the LOOCV residuals, detailed in Section A.4.6.)

1. **Error versus time-of-week:** For most facilities, error variance is clearly a function of time-of-week. For a typical commercial building, error variance tends to be fairly low at night and higher during the day. This is because during the day fluctuating occupancy affects HVAC and lighting loads in addition to plug loads. For some facilities the error variance is particularly high during the transition periods (when the facility is being populated in the morning or de-populated in the evening) since the timing and the rate of the transition can vary day-to-day (e.g., when a retail store changes its hours), making it difficult to accurately predict these transitions.

   For 20 facility-years, we created scatter plots of error versus time-of-week. We then binned the errors into 24 equally-sized time-of-week bins and computed the mean error, standard deviation of the errors, and the number of points per bin. Two examples are given in Section 3.4. For Facility 11, error is clearly a function of time-of-week, while for Facility 1, error does not seem to be a function of time-of-week.

2. **Error versus temperature:** For the same facility-years, we also created scatter plots of error versus temperature. Again, we then binned the errors into 24 equally-sized temperature bins and computed the mean error, standard deviation of the errors, and the number of points per bin. Two examples are given in Figure A.2. Error variance does not seem to be a clear function of temperature.

3. **Error versus actual demand:** We also created scatter plots of error versus actual demand (examples in Figure A.3). In some cases, the number of points per bin varied greatly and we worried that the number of points per bin could affect the computation of the mean error and standard deviation of the error. Therefore, we also binned the error into 24 demand bins, each with an equal number of demand observations. We

[4] http://www.mathworks.com/matlabcentral/fileexchange/24722

computed the mean error and standard deviation of the errors in each bin (examples in Figure A.4).

Both methods of binning the data show that error variance does not seem to be a clear function of actual demand.[5] However, error bias does: at high demand the model under-predicts and at low demand the model over-predicts, as one would expect with any regression model. The correlation between error and actual demand is a function of the model fit, the magnitude of the prediction error, etc.

Since separate equations are used to compute occupied demand and unoccupied demand, errors from when the facility is operating in occupied mode follow a different trend than errors from when the facility is operating in unoccupied mode. This is especially visible in the scatter plots for Facility 1 (Figures A.3a and A.4a). For this facility, unoccupied mode errors occur at actual demands of $\sim 110 - 180$ kW and the correlation between unoccupied mode error and actual demand is strong (there is a clear slope between error and actual demand). Occupied mode errors occur at actual demands of $\sim 180 - 230$ kW and the correlation between occupied mode error and actual demand is weaker (shallower slope), but still present.

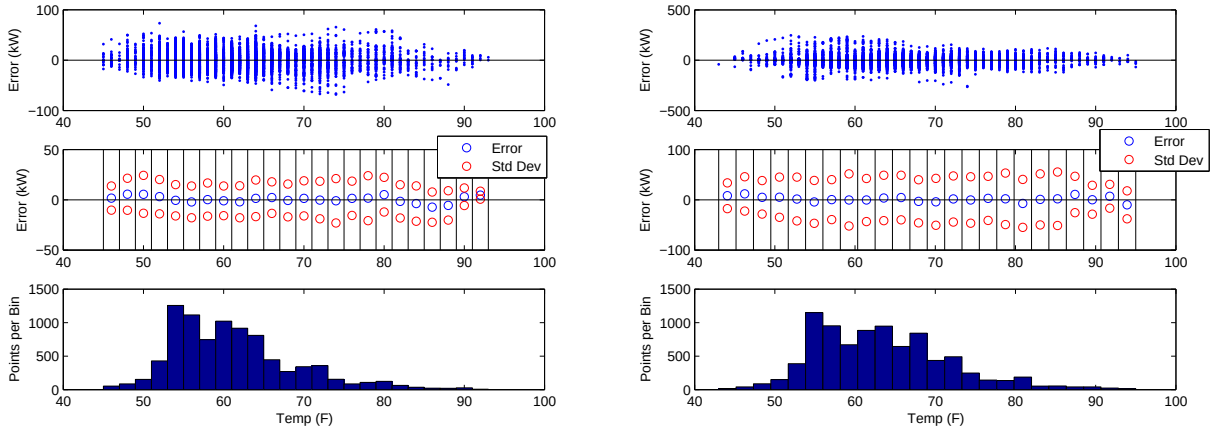
4. **Error versus predicted demand:** Lastly, we created scatter plots of error versus predicted demand (examples in Figure A.5). Unlike actual demand, predicted demand is not correlated to error. Also, there is no clear relation between error variance and predicted load (see previous footnote).

As a result of this analysis, we chose to only consider error variance as a function of time-of-week and ignore error variance as function of temperature, actual load, and predicted load. Note that if we had chosen to tackle two or more of these issues our analysis would have become very difficult since all of these input and output variables are correlated. A future study could consider these issues in more depth.

A.4.3 Error analysis method

The goal of the error analysis was to determine the prediction error associated with each DR parameter for each facility-year. We assume that there is no measurement error associated with the temperature or actual demand measurements. Therefore, prediction error is purely a result of how well our model predicts the variability of the facility.

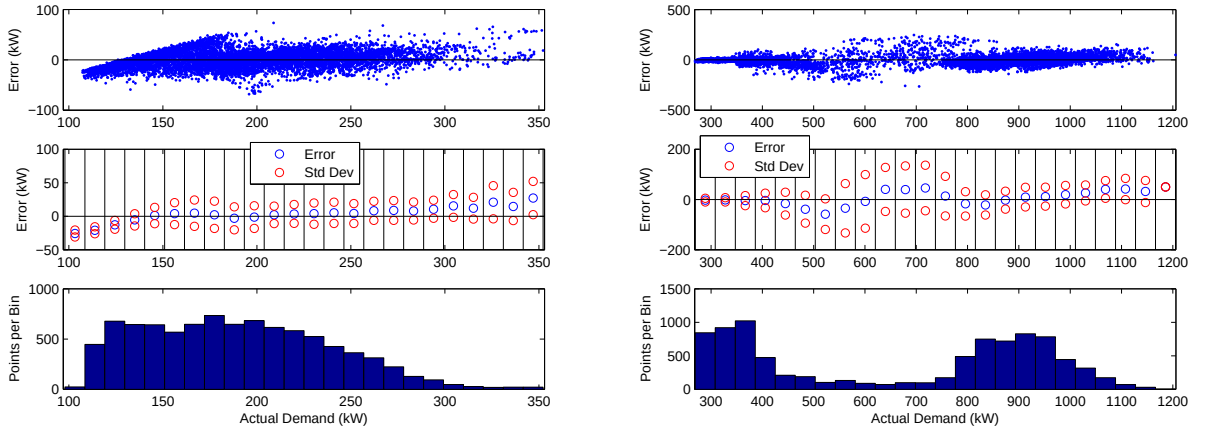
[5] For Facility 11, it appears as though error variance is higher at mid-range demands ($\sim 400 - 800$ kW). The histogram tells us that the vast majority of the time this facility is either 'all on' or 'all off.' Therefore, transitions through mid-range demands occur rapidly. We know that this facility changed its operating hours during the summer (shifting the timing of the transitions) and that accounts for higher error variance at mid-range demands. Though we do not consider error variance as a function of actual/predicted demand, we consider error variance as a function of time-of-week, and so this issue is taken into account in our analysis.



(a) Facility 1 in 2008.

(b) Facility 11 in 2008.

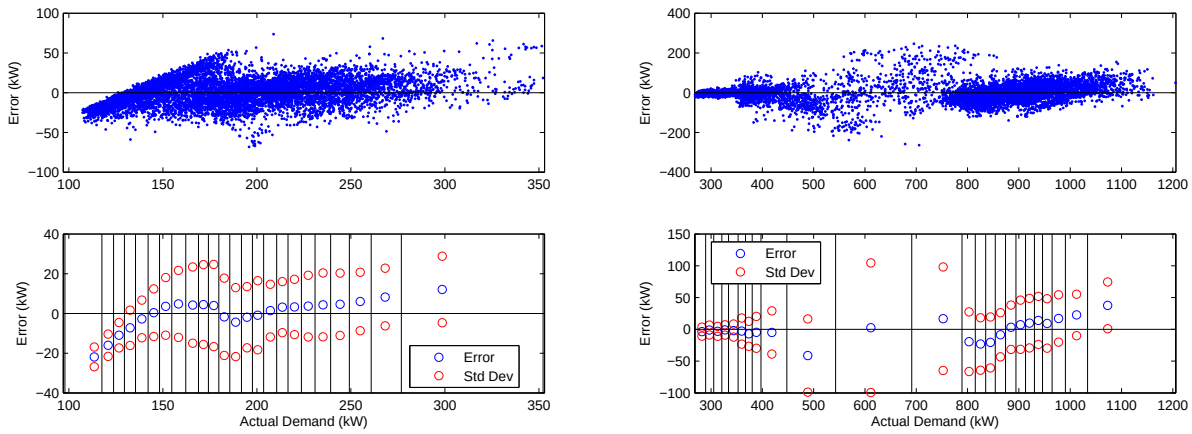
Figure A.2: Two example plots of error vs. temperature.



(a) Facility 1 in 2008.

(b) Facility 11 in 2008.

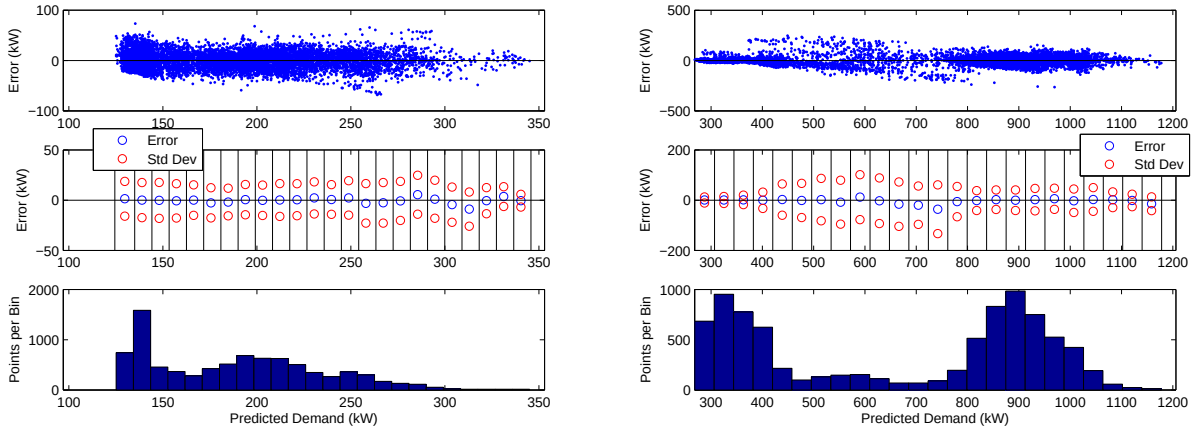
Figure A.3: Two example plots of error vs. actual demand (with equal size demand bins).



(a) Facility 1 in 2008.

(b) Facility 11 in 2008.

Figure A.4: Two example plots of error vs. actual demand (with an equal number of demand observations per bin).



(a) Facility 1 in 2008.

(b) Facility 11 in 2008.

Figure A.5: Two example plots of error vs. predicted demand.

Our error analysis was performed using non-DR day data only. We considered the following options:

1. Reserve a random 25% of the data as a 'testing set' and build the model with the remaining 75% of the data ('training set'). Analyze the performance of the model on the testing set.

2. Building the model with all of the data and analyze the regression residuals.

3. Use LOOCV, in which a single observation (testing set) is reserved for validation of the model, which is built with the rest of the observations (training set). This is repeated for each observation. Then, the LOOCV residuals (actual minus predicted demand) are analyzed. The final model, used for prediction on DR days, is built with all of the data.

As described above, we consider error variance as a function of time-of-week. Therefore, our error analysis is limited by the fact that there are a small number of weeks (~ 22) in each facility-year and many weekdays (DR days, holidays, 'power outage days') have been removed from the data set (leaving us with $\sim 18-20$ weeks of data). For example, Option #1 would involve reserving ~ 5 weeks of data as a testing set. Analyzing the performance of the model on the testing set would yield ~ 5 error observations per time-of-week: not enough observations to develop a robust error distribution.

Option #2 is attractive because there are an equal number of regression residuals and observations. At each time-of-week there are $\sim 18-20$ regression residuals, which can be treated as error observations. This is a much more reasonable number of observations to use to develop an error distribution. However, the problem with this option is that the training set and the testing set are the same, and the model performs best on the data used to build it due to self-influence. Therefore, the error estimates would be optimistic.

Option #3 avoids self-influence and allows us to retain the same number of error observation per time-of-week ($\sim 18-20$) as option #2. LOOCV is a type of K-fold cross validation (also called rotation estimation), which is similar to other resampling techniques such as jackknifing, bootstrapping, and permutation tests, all of which are commonly used in statistics. K-fold cross validation involves randomly partitioning the data into K subsamples, reserving one subsample, building the model with data from the remaining subsamples, testing on the reserved subsample, and repeating this process for all K subsamples. The results for each subsample are combined (e.g., the error statistics/distributions are averaged) resulting in an estimate of the prediction accuracy. Ten-fold cross validation is commonly recommended/used [ref: Refaeilzadeh et al. 2009]. In LOOCV, K is equal to the total number of observations, n. LOOCV is useful when n is small because LOOCV produces more error estimates than K-fold cross validation (when K$< n$). However, LOOCV is the most computationally intensive form of K-fold cross validation because the the model is built n times.

Given the pros and cons of each option, we ruled out Option #1 and then implemented both Option # 2 and #3 for comparison. To perform LOOCV, we treat the demand on each non-DR day as an observation. Therefore, n is equal to the number of non-DR days used to create the baseline prediction model ($\sim 90 - 95$ days). We leave out one non-DR day, build the model with data from the rest of the DR days, predict the demand on the day that has been left out, compute the LOOCV residual, and repeat for each non-DR day. Since we consider error variance versus time-of-week, only residuals computed with data from Mondays are used to determine errors on Mondays, residuals computed with data from Tuesdays are used to determine errors on Tuesdays, and so on. Therefore, for each time-of-week there are $\sim 18 - 20$ residuals.[6]

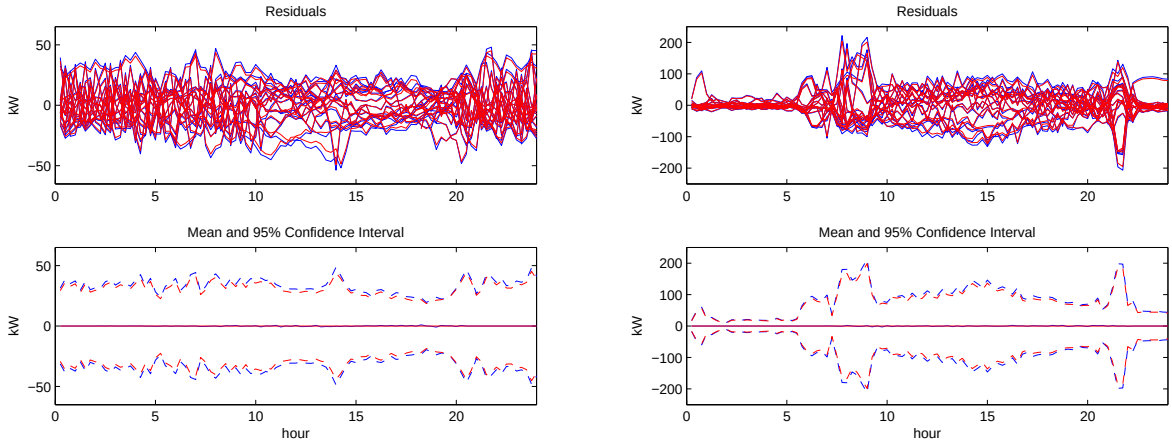
To compare the regression residuals to the LOOCV residuals, we plotted time series residuals (examples in Figure A.6) and the Absolute Percentage Error (APE) and the Mean Absolute Percentage Error (MAPE) (examples in Figure A.7). For clarity, the figures show residuals and APE/MAPE for Monday only. Since the regression residuals are self-influenced, their magnitude is slightly smaller than that of the LOOCV residuals. Therefore, LOOCV produces slightly larger MAPE and 95% confidence intervals (though the means are approximately the same, as expected).

We also used both the regression residuals and the LOOCV residuals to compute the error on the DR parameters. Again, only error observations computed with data from Mondays are used to determine the error distributions for Monday, etc. Therefore, for each DR parameter, error distributions differ by day-of-week. That means that the day-of-week on which each DR day falls determines which error distribution is applied.

Examples of error observations computed with LOOCV residuals and regression residuals are shown in Figure A.8 (again, we only show data for Monday).[7] Again, errors computed with LOOCV are slightly larger in magnitude than those computed with the regression residuals. Based on these results, we chose Option #3 instead of Option #2 because, though the results are very similar, LOOCV provides a more realistic (and conservative) estimate of the error variance.

[6] An alternative to treating demand on each non-DR day as an observation is to treat each week (composed of potentially non-chronological non-DR days, as defined in a previous footnote) as an observation. That would require building the model $\sim 18 - 20$ times instead of $\sim 90 - 95$. However, it would also result in building each model with less information on the facility's temperature dependence (since data from all times-of-week are used to compute the temperature parameters).

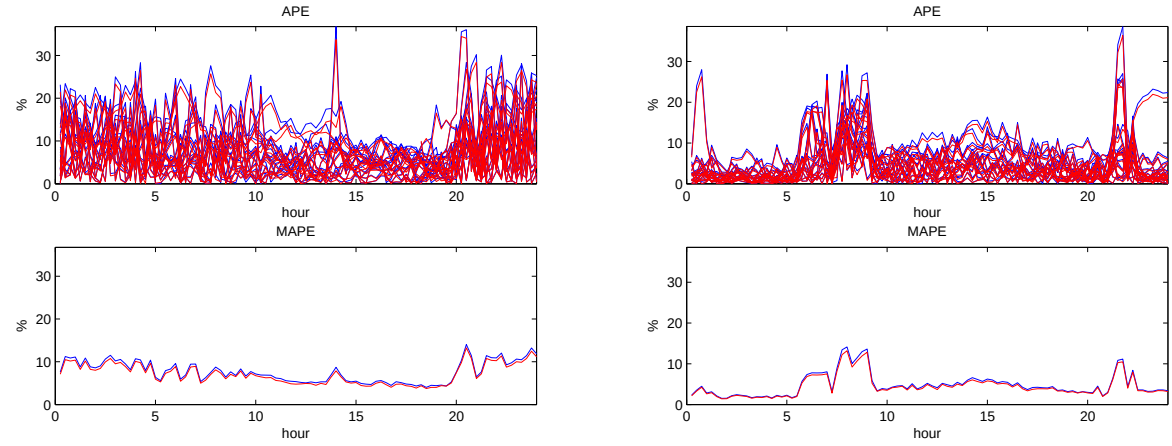
[7] In Figure A.8, the error observations have been sorted in ascending order, and plotted versus observation number. Ideally, these would be plotted in histogram form but with so few observations histograms are hard to interpret. Notice that in several cases, we have less than 18 error observations. This is because, on some DR days, temperature data is not available and so we are unable to compute predicted demand, residuals, or DR parameter errors.



(a) Facility 1 in 2008.

(b) Facility 11 in 2008.

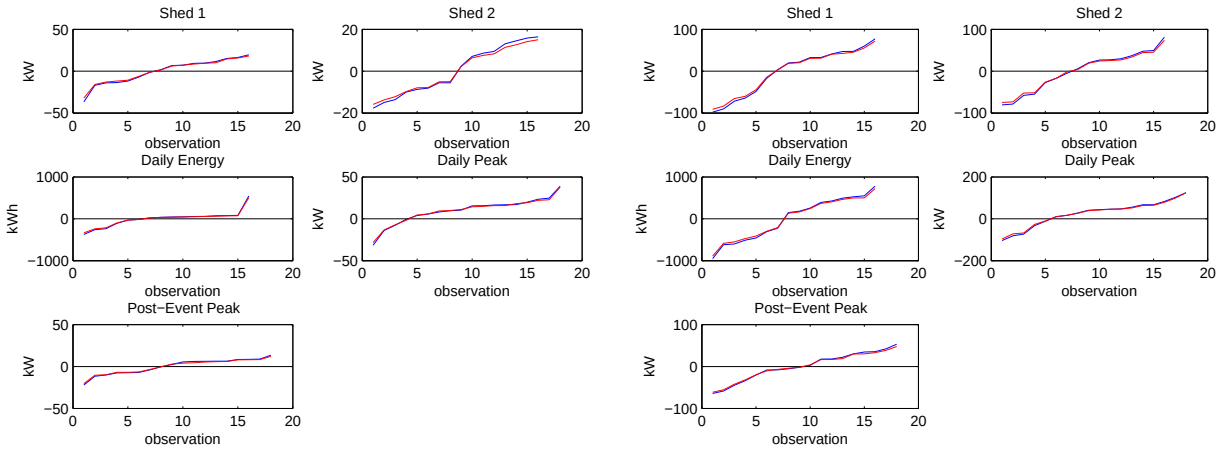
Figure A.6: Two example plots of regression residuals (red) and LOOCV residuals (blue) for Monday. The 95% confidence intervals are computed assuming that the residuals are normally distributed at each time (therefore, two standard deviations are used).



(a) Facility 1 in 2008.

(b) Facility 11 in 2008.

Figure A.7: Two example plots of APE/MAPE for Monday, computed with regression residuals (red) and LOOCV residuals (blue).



(a) Facility 1 in 2008.

(b) Facility 11 in 2008.

Figure A.8: Two example plots of DR parameter errors for Monday, computed with regression residuals (red) and LOOCV residuals (blue).

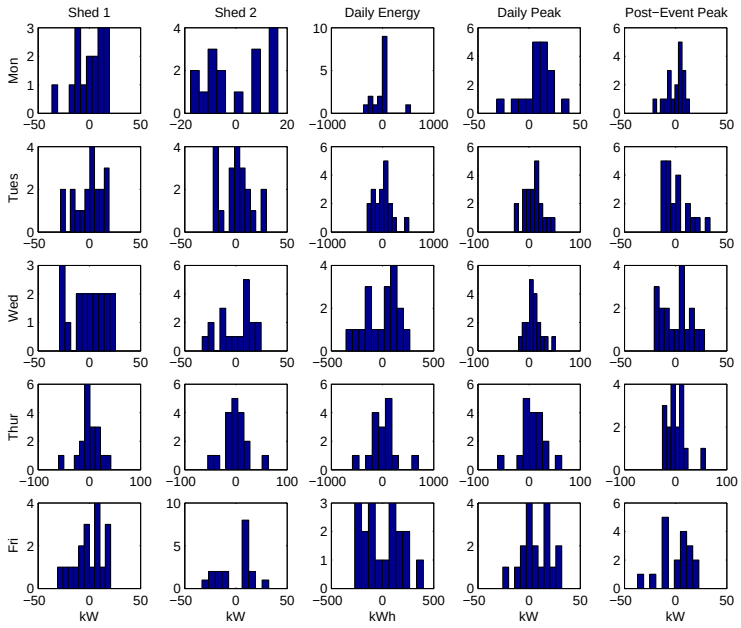
A.4.4 Error distributions

Unfortunately, with less than 20 error observations for each DR parameter for each day-of-week, it is non-trivial to choose an error distribution. Two options are to assume that the error distribution is normal,[8] or to use the observed error distribution. With so few observations, the real error distribution is likely different than the observed error distribution, so we do not use the observed error distribution. Unfortunately, twenty observations is too few to use a histogram or quantile-quantile (QQ) plot to verify that the error distribution is normal. However, for a number of facility-years we plotted histograms (examples in Figure A.9) and QQ plots (examples in Figure A.10) of error observations to determine if the normality assumption is blatantly unreasonable. By examining these plots, we found that the normality assumption does not seem unreasonable, so we assume a normal error distribution.

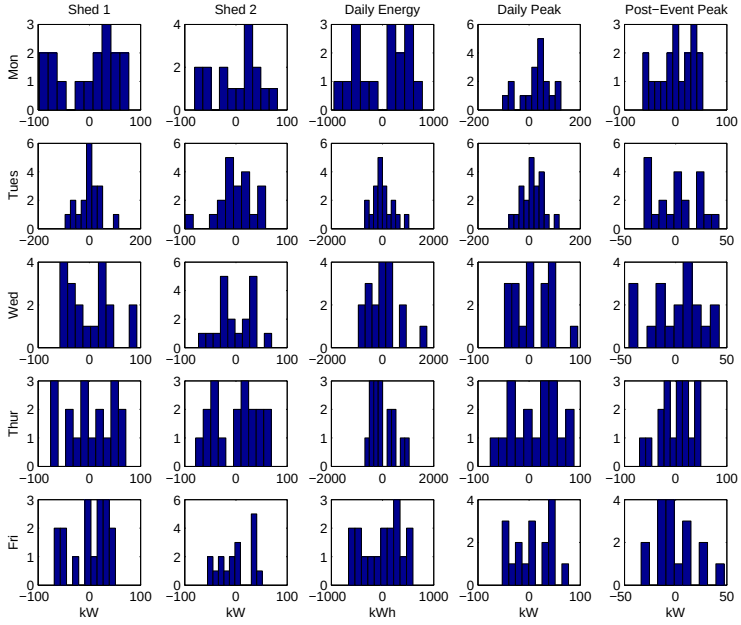
A.4.5 Extrapolation error

Predicting demand on DR days requires using DR day temperatures, which are often higher than the highest temperature used to build the baseline prediction model. While in Section A.4.2 we found that error variance does not seem to be a function of temperature, we only

[8] We also briefly looked at other common distribution (Exponential, Extreme Value, Lognormal, Rayleigh, and Weibull). Only the Extreme Value distribution is relevant to our data set since it contains both positive and negative numbers. The data did not seem to fit the extreme value distribution better than the normal distribution.

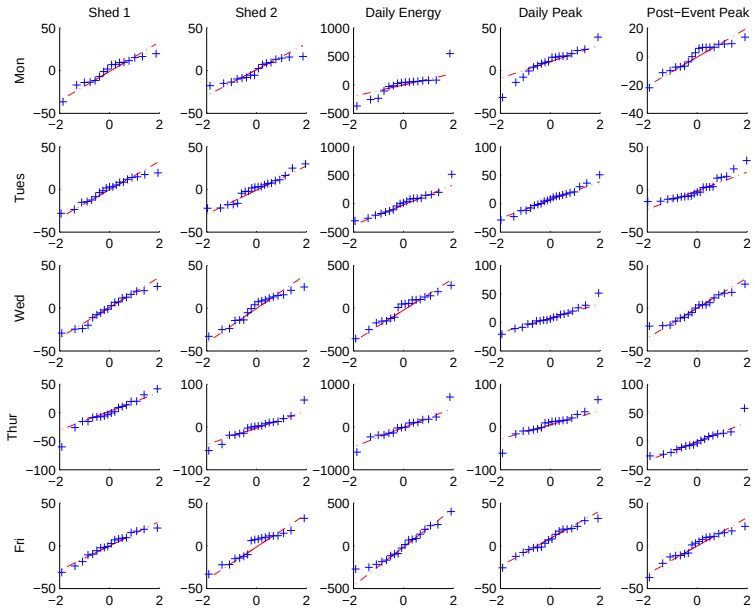


(a) Facility 1 in 2008.

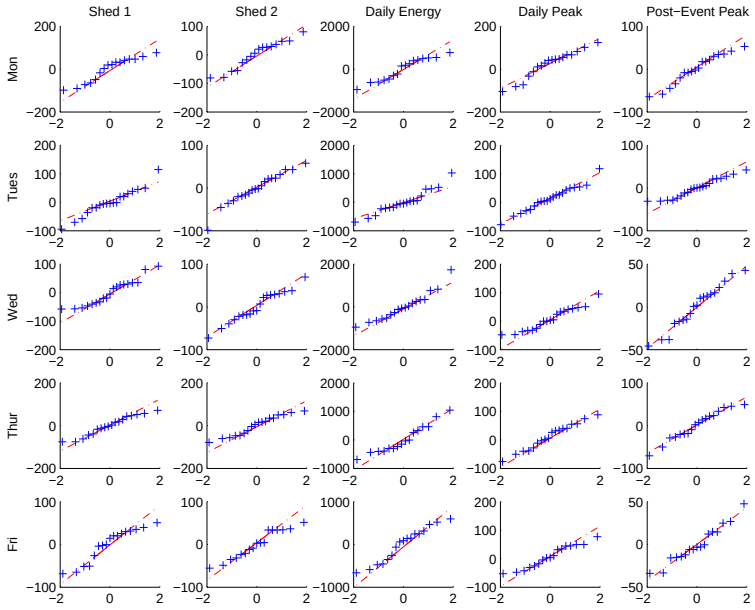


(b) Facility 11 in 2008.

Figure A.9: Two example plots of histograms of error observations for each DR parameter for each day of week.

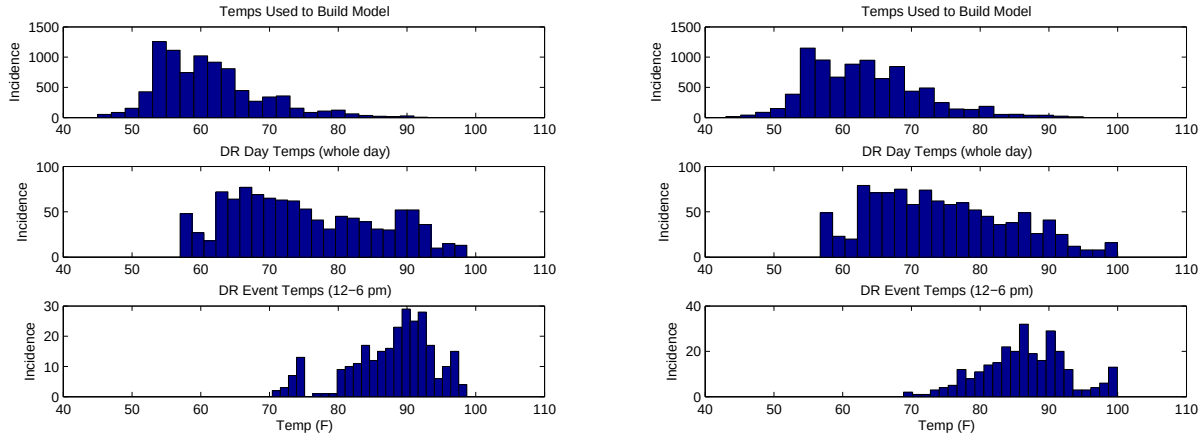


(a) Facility 1 in 2008.



(b) Facility 11 in 2008.

Figure A.10: Two example plots of QQ plots of error observations for each DR parameter for each day of week. In each case, sample data quantiles are plotted versus standard normal quantiles.



(a) Facility 1 in 2008.

(b) Facility 11 in 2008.

Figure A.11: Two examples of temperature histograms.

explored the range of temperatures used to build the model. Therefore, we also investigated extrapolation error, specifically the error associated with predicting demand using higher temperatures than those used to build the model.

To understand the degree of extrapolation for each facility-year, we plotted histograms of temperatures used to build the model, DR day temperatures, and DR event temperatures (temperatures between 12 and 6 pm on DR days). Two examples are shown in Figure A.11. Note that the vertical scales are not the same. Unsurprisingly, average DR days are hotter than average non-DR days. Also, DR events occur during the hottest hours of the day (12 –6 pm). In some cases, DR day temperatures (generally those occurring during the DR event) are higher than the highest non-DR day temperatures. This means that some of our demand predictions are extrapolated.

To understand the magnitude of the extrapolation issue, we analyzed all data to determine what percent of facility-DR-days (defined as one DR day for one facility-year) were extrapolated. We call a facility-DR-day extrapolated if the maximum temperature on the DR day is greater than the maximum temperature used to build the baseline prediction model. Table A.4 summarizes the number of extrapolated facility-DR-days. A total of 26% of facility-DR-days were extrapolated.

To estimate the error caused by extrapolation, we considered the following options:

1. Use demand/temperature data from the year before a facility started participating in the CPP program (or, if they are no longer in the program, use data from the year after they stopped participating in the CPP program). Call the 12 hottest days 'proxy DR days.' Then, use the methods presented above to develop a baseline prediction model (without data from the 'proxy DR days') and perform an error analysis. Use

Table A.4: Extrapolated facility-DR-days (FDRD).

Year (Zone)	Total FDRD	Extrapolated FDRD	% Extrapolated FDRD
2006 (1)	36	0	0%
2006 (2)	94	47	50%
2007	167	3	1.8%
2008	284	122	43%
2009	392	71	18%
TOTAL	937	243	26%

the baseline model to predict demand on the 'proxy DR days.' Compare the error of predicting on 'proxy DR days' to the error of predicting on non-DR days. If the error distributions are similar then assume that extrapolation does not introduce error. If they are different, quantify the error of extrapolation.

2. Use demand/temperature data from comparable facilities that do not participate in DR, and perform the same analysis as explained in #1.

3. Use our existing demand/temperature data, but remove the data from CPP days. Then, call the 12 hottest days 'proxy DR days' and perform the same analysis as explained in #1.

To use Option #1, we would need to ensure that the facility did not participate in a different DR program the year before they participated in the CPP program, or, if they did, we would need to take into account the specifics of that DR program when performing our analysis. Using this option we assume that lessons we learn about extrapolation error in one year apply to other years. Using Option #2, we assume that lessons we learn about extrapolation error from facilities that do not participate in DR apply to facilities that do participate in DR. Option #3 avoids the assumptions required of the other options. However, it involves building the model with fewer data (specifically, removing ∼9-12 days worth of data) and assuming that the error of extrapolation to 'proxy DR day' temperatures, which would generally be lower than CPP day temperatures, is equivalent to the error of extrapolation to CPP day temperatures.

In the end, we performed a variation on the first option. We analyzed demand data from three facilities that previously participated in the CPP Program and, in 2009, moved to a different DR program, the Participating Load Pilot (PLP) Program. Details about each facility are shown in Table A.5. During the summer, several PLP events were called, and different facilities participated in each PLP event (each facility participated in 3-6 events). For the purpose of our analysis, we tagged each day that a facility participated in a PLP event as a holiday so that these days were not used to build the baseline models used for our extrapolation study (referred to as 'ES baseline models'). That means that ES baseline models were built with less data than regular baseline models (though more data than the

Table A.5: Extrapolation study.

Facility Number	Days Used to Build Model	Highest Temp (Non-DR Days)	CPP Days Analyzed	Highest Temp (CPP Days)	Extrapolated CPP Days
4	88	100°F	12	100°F	0
11	91	91°F	12	97°F	2
17	89	86°F	11*	92 °F	3

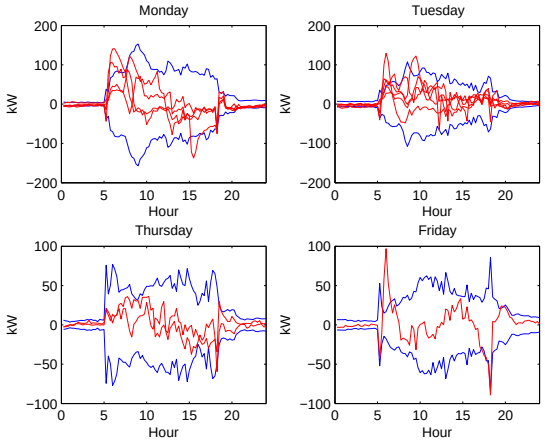
*On one day, the utility called both a PLP event for this facility and a CPP event so the data from that day were not analyzed.

ES baseline models in Option #3). See Table A.5 for the number of days used to build each model. Instead of choosing the 12 hottest days as 'proxy DR days,' we used the actual 2009 CPP days (except in one case when, for one facility, a PLP event was called on a CPP day so the data from that day were not analyzed). Because we use the actual 2009 CPP days, which are not always called on the hottest days, not all DR day predicted demands are extrapolated. Table A.5 lists the highest temperature used to build the model, highest temperature on the CPP days, and the number of CPP days when the predicted demand is extrapolated.

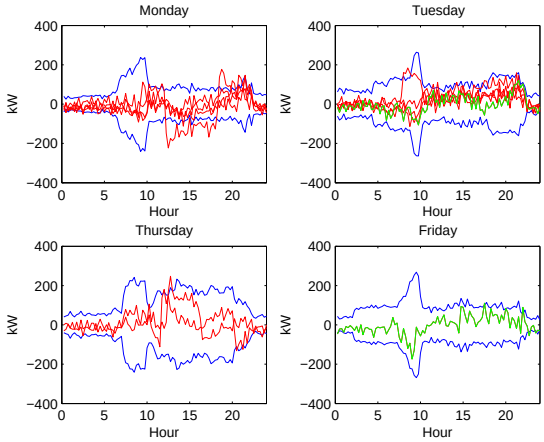
Figure A.12 shows extrapolation study residuals (actual CPP day demand minus baseline-predicted CPP day demand) by day-of-week. In 2009, there were no CPP events on Wednesdays so Wednesdays are not shown. Residuals from non-extrapolated CPP days are shown in red, while residuals from extrapolated CPP days are shown in green. For each day of week, in blue, we plot 95% confidence intervals (assuming a normal distribution) computed from the regression residuals. If the extrapolation study residuals and the regression residuals are similarly distributed, we would expect the extrapolation study residuals to be outside of the 95% confidence interval about 5% of the time.

Unfortunately, we find that, on several CPP days, the extrapolation study residuals are outside of the 95% confidence interval a significant portion of the time. Looking closely at the shape of the extrapolation study residuals themselves, it appears as though all three facilities executed their CPP event DR strategies on the first CPP day of the summer (which occurred on a Monday) despite the fact that they were no longer participating in the CPP program. It also appears as though Facility 17 may have executed its CPP event DR strategy on the second CPP day of the summer (which occurred on a Tuesday). Note that none of these 'suspect' extrapolation study residuals are extrapolated. In any case, they should not be used to inform the extrapolation error.

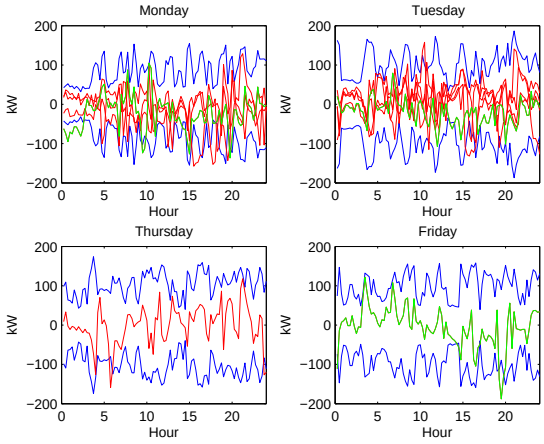
Figure A.13 shows DR parameter errors by day-of-week computed from the regression residuals (blue) and extrapolation study residuals (red for extrapolated CPP days, green for non-extrapolated CPP days). The blue lines are 95% confidence intervals (assuming a normal distribution) computed from the DR parameter errors, which were computed from the regression residuals. Again, Wednesdays are not shown. In this figure, there are several instances when the DR parameter errors computed with the extrapolation study residuals are



(a) Facility 4 in 2009.



(b) Facility 11 in 2009.



(c) Facility 17 in 2009.

Figure A.12: Extrapolation study residuals by day-of-week: 95% confidence intervals (blue), non-extrapolated CPP day residuals (red), and extrapolated CPP day residuals (green).

outside of the 95% confidence interval. All but one of these outliers are consistent with the facilities executing their CPP day DR strategies on the first CPP day of the summer (and, for Facility 17, the second CPP day of the summer). The only unexplained outlier is one of Facility 4's post-event peak errors on Thursday (note that it occurred on a non-extrapolated CPP day). If the DR parameter error distribution computed from the regression residuals was similar to that computed from the extrapolation study residuals, we would expect the DR parameter errors computed with the extrapolation study residuals to be outside of the 95% confidence interval 5% of the time. Therefore, one unexplained outlier is not inconsistent with the distributions being similar.

Our extrapolation study shows that extrapolation error many be small-to-neligible, so we ignore extrapolation error. However, we have only scraped the surface of what could be done to quantify extrapolation error. We have performed our analysis on only three facilities and very few (5) extrapolated days. With such a small data set we are not able to demonstrate that the DR and non-DR day error distributions are *similar*, only that they are *not inconsistent*. A future study, using a larger data set, could do a more careful analysis to quantify the error of extrapolation.

Note that in this section we have only considered the error due to extrapolation; a related issue is model over-fitting when using a small number of data points to build part of the model (e.g., modeling a facility's temperature dependence at high temperatures when the amount of data taken at high temperatures is low). This issue is discussed in the following section.

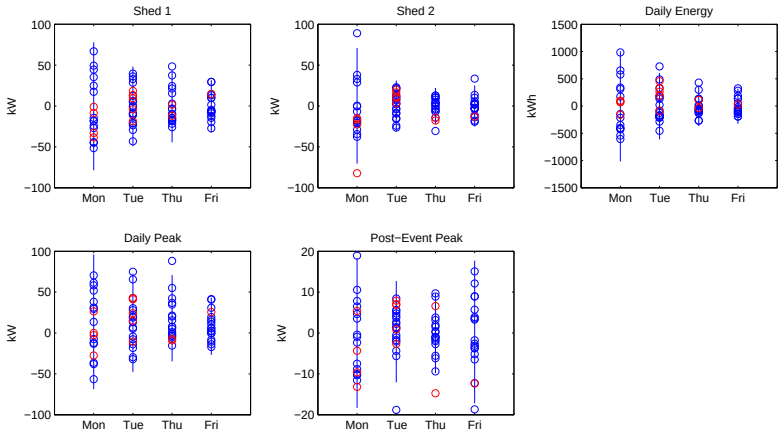
A.4.6 Over-fitting error

We repeated the analysis in Section A.4.2 using the LOOCV residuals. The goal of this analysis was to understand model over-fitting when we use a small number of data points (p) to build part of the model. For example, for both Facility 1 and 11, there are few data from the lowest/highest temperatures and the highest actual/predicted demands, and for Facility 11, there are few data from the mid-range actual/predicted demands. We hoped that using the LOOCV residuals would give us a more accurate estimate of the error distributions associated with these 'low-p bands.'

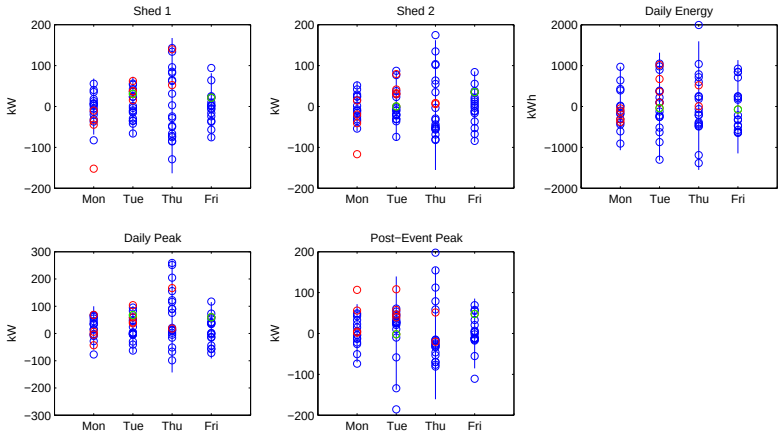
Example plots, directly comparable to the plots in Section A.4.2 , are given in Figures A.14, A.15, A.16, A.17, and A.18. The differences between these plots and those in given in Section A.4.2 are subtle. A few differences are worth noting are as follows:

1. In Figure A.15a, the LOOCV errors in the highest temperature bin have less variance but more bias than the comparable regression residual errors.

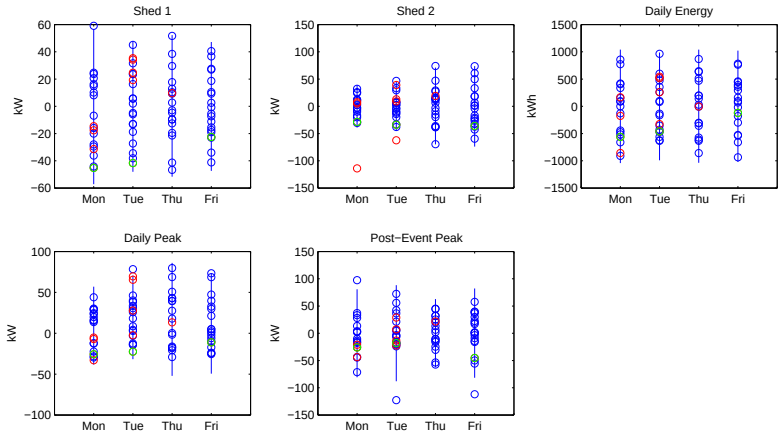
2. In Figure A.15b, the LOOCV errors in the lowest temperature bin have less variance but more bias than the comparable regression residual errors.



(a) Facility 4 in 2009.

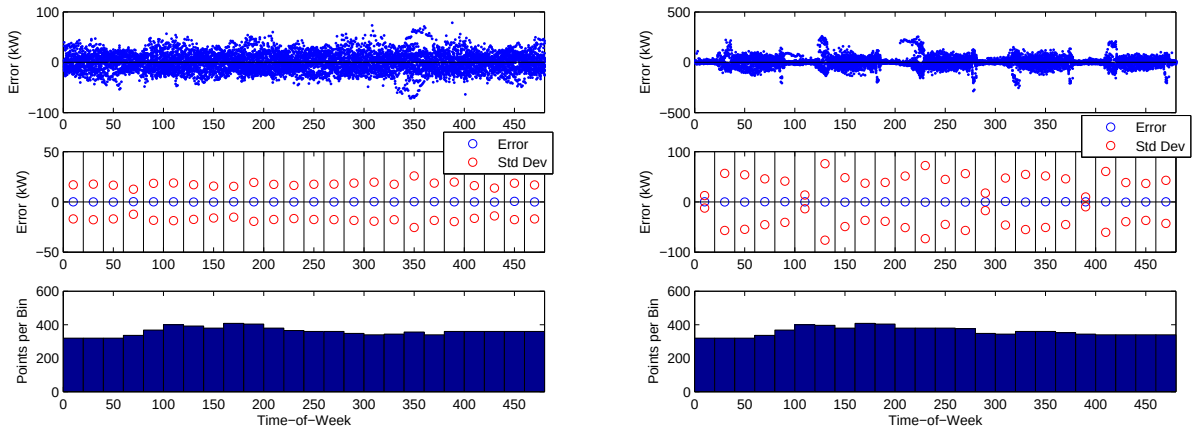


(b) Facility 11 in 2009.



(c) Facility 17 in 2009.

Figure A.13: Extrapolation study DR parameter errors by day-of-week: DR parameter errors (blue circles), 95% confidence intervals (blue line), non-extrapolated CPP day DR parameter errors (red circles), and extrapolated CPP day DR parameter errors (green circles).



(a) Facility 1 in 2008.

(b) Facility 11 in 2008.

Figure A.14: Two example plots of LOOCV error vs. time-of-Week.

3. In Figure A.17b, the LOOCV errors in the mid-range actual demand bins ($\sim 400-800$ kW) have more variance than the comparable regression residual errors (note that the y-axis scale is different in the Figures A.4b and A.17b).

4. In Figure A.18a, the LOOCV errors in the lowest predicted demand bin have more bias than the comparable regression errors.

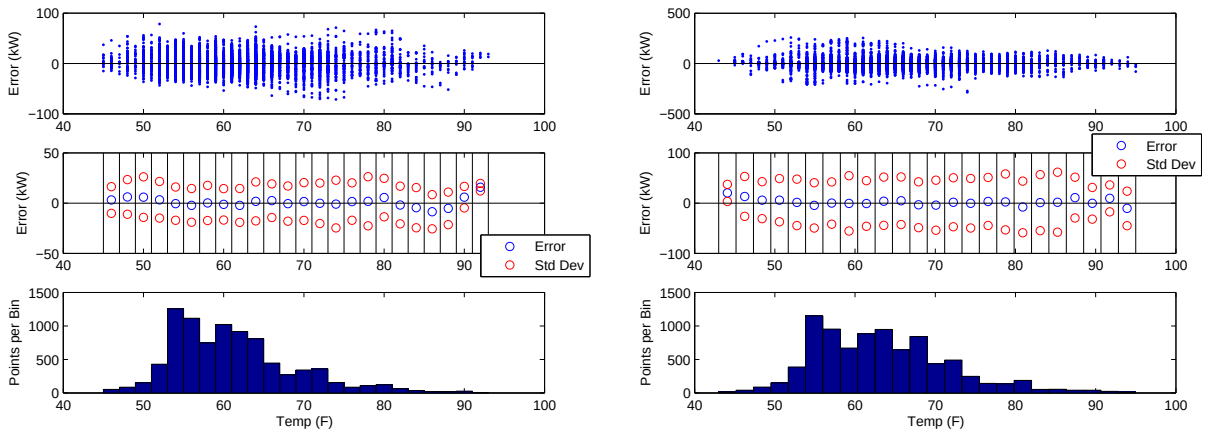
5. In Figure A.18b, the LOOCV errors in the highest predicted demand bin have less variance but more bias than the comparable regression residual errors.

In total, this analysis tells us that, while some model over-fitting exists, it does not seem to be a significant issue.

A.5 Aggregate System Synthesis

Several issues arose when trying to analyze the aggregate system. One issue was how to sum the load data. If all the data sets were complete this would be trivial; however, this is not the case. For some facility-years, we have spotty load data. This is not a problem when building a baseline model for a particular facility-year since the regression algorithm simply ignores missing data. However, to create the aggregate load we have the choice two options: (#1) ignore missing data ($x + NaN = x$) or (#2) propagate missing data ($x + NaN = NaN$).[9] There are pros/cons to both options. Using the first option, we preserve more data, but,

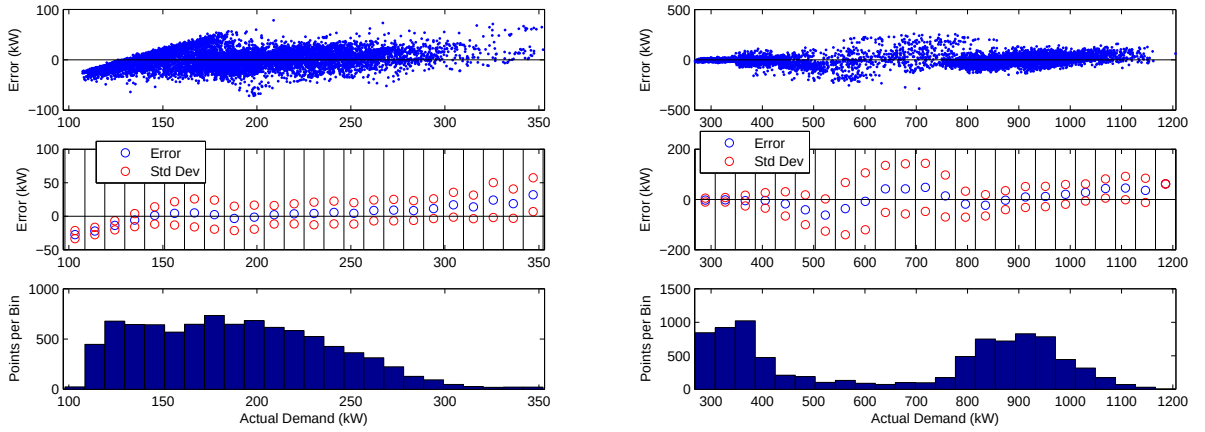
[9]NaN = not a number, MATLAB uses NaN to indicate missing data



(a) Facility 1 in 2008.

(b) Facility 11 in 2008.

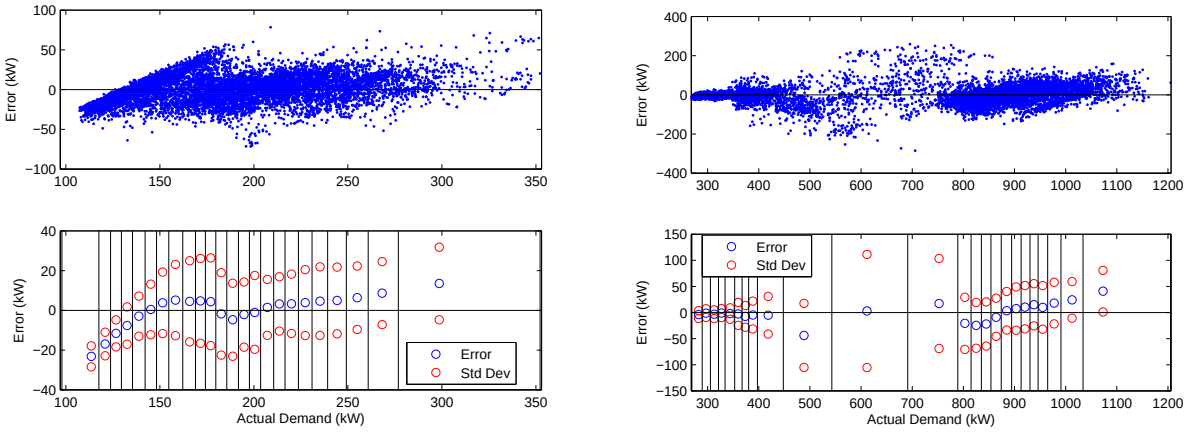
Figure A.15: Two example plots of LOOCV error vs. temperature.



(a) Facility 1 in 2008.

(b) Facility 11 in 2008.

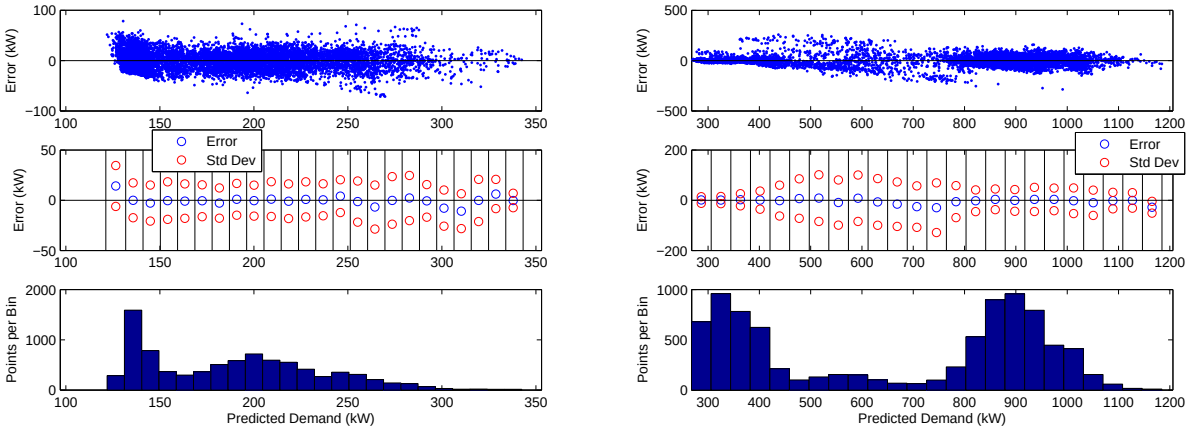
Figure A.16: Two example plots of LOOCV error vs. actual demand (with equal size demand bins).



(a) Facility 1 in 2008.

(b) Facility 11 in 2008.

Figure A.17: Two example plots of LOOCV error vs. actual demand (with an equal number of demand observations per bin).



(a) Facility 1 in 2008.

(b) Facility 11 in 2008.

Figure A.18: Two example plots of LOOCV error vs. predicted demand.

in some time intervals, we sum data from only a portion of the facilities. Using the second option, in each time interval, we sum data from all of the facilities, but we may end up with particularly spotty aggregate load data, especially if missing data are not aligned in time. We tried both options and found that Option #1 produces undesirable results when load data from high-demand facilities are missing. Therefore, we use Option#2. This method works well as long as large amounts of data are not missing from any individual facility-year. Unfortunately, in 2008, we were missing more than one week of data from each of three facilities. Therefore, we did not include them in the aggregate system. In other years, no facilities were excluded from the aggregate system due to missing data.

Another issue was how to compute the temperatures that would be used to build the baseline model and predict DR day demand. For each year, we took a weighted average of the temperature data from the nine weather stations listed in Table A.2.[10] Specifically, for each year, we weighted the data from each station according to the number of facilities in the aggregate system associated with that station [could be better to weight by the mean/peak demand associated with each station]. Again, were were presented with two options: (#1) ignore missing data ($mean(x, y, NaN) = mean(x, y)$) or (#2) propagate missing data ($mean(x, y, NaN) = NaN$). Initially, it seemed like the obvious choice was Option #1 since temperatures collected by the different weather stations do not vary widely, and, given occasional missing data, calculating the mean of similar values (ignoring missing data) is more robust than calculating the sum (ignoring missing values). However, we implemented Option #1 and encountered a significant problem. On a certain day in 2008, for several hours, data was missing from all weather stations except one, Bakersfield, which is in the hottest climate. Averaging, ignoring missing data, produced the highest average temperature seen all summer. However, in actuality, by looking at average temperatures before and after the period of missing data, it is clear that the average temperature was likely mid-range. In addition, the aggregate load, much lower than that seen on the hottest days, corresponds to a mid-range average temperature. Therefore, the 'best fit' temperature dependance was drastically affected: while at low/mid-range average temperatures load increases as temperature increases, as we would expect, at the highest average temperatures load *decreases*, which is counter-intuitive. Since most DR day temperatures are near the highest average temperatures, this 'best fit' temperature dependence significantly affected our prediction of DR day demand. As a result of this problem, we implemented Option #2. The compromise is that we have more missing average temperature data so the baseline models are made with less data, and, in some cases, we are unable to predict load on DR days because of missing average temperature data.

After preparing the load and temperature data, we performed load prediction and error analysis using the same methods as those used for each facility-year, with one exception: we redefined a 'power outage day.' In the aggregate, outages are far less obvious so we need

[10]Note that we did not use the raw data but instead used temperature data that had been interpolated and, if spotty, filled in with data from a back-up weather station, as described in Section 2.2.2.

to use a stricter rule or else no 'power outage days' are removed, despite obvious outliers. Fortunately, the daily minimum load for the aggregate system is less volatile than that for most facility-years so we are able to use a stricter rule. For the aggregate system, we define 'power outage days' as days in which the minimum power consumption during that day is less than 80% of the average minimum daily power consumption during the summer. One 'power outage day' (non-DR days) was removed from 2006 Zone 1, and two were removed from 2006 Zone 2. No other days were removed.

Appendix B

Extensions to the Variability Metric Analysis

B.1 Possible Distributions of Variance and Covariance

We can use the empirical SVMs distributions in Section 3.5 to generate possible distributions of Var(RS) and Corr(RS, UL) that lead to distributions like those in Figure 3.5. Knowing that about 55 facility-years have Var(RS) ≈ 0, we experiment with various values of Var(RS) and Corr(RS, UL) for the remaining facility-years. Some combinations that produce variability metric distributions that are similar to those in Figure 3.5 are shown in Figure B.1. Note that the variances and covariances in the plots do not sum to the SVM in the plots (Figure B.1) because plotted is the real Var(OS) whereas the SVM is computed with the estimated Var(OS).

As a results of this analysis, we can comment on the likely distribution of the variances and covariances. Moreover, we find that for most facilities Corr(RS, UL) is likely negative.

B.2 Disaggregation of the Variability Metric Results

We attempted to determine if SVM correlates to various aspects of the facilities. Therefore, we disaggregated the results presented in the previous section in several ways:

1. **By facility type.** We disaggregated the facility-years by facility type: retail stores (n=8), retail stores with solar PV (n=7), manufacturing facilities (n=17), prisons/jails (n=6), offices (n=45), and other (n=4).[1] Using the 80% confidence interval, we then sorted facility-years into three bins (SVM<nb, nb<SVM<pb, and SVM>pb) and the plotted the results both versus facility type and versus SVM, for each shed period. Results are shown in Figure B.2.

[1] In this analysis, all facility-years with building type 'other' are museums.

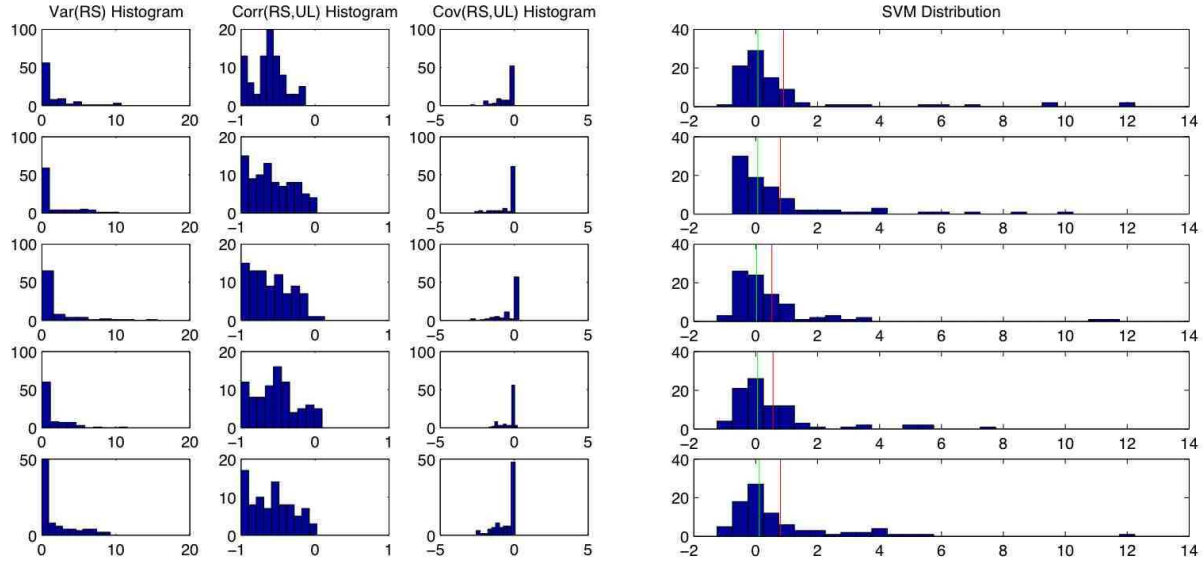


Figure B.1: Five possible Var(RS) and Corr(RS, UL) histograms that produce SVM distributions similar (in terms of mean, median, scatter, etc.) to those in Figure 3.5. We also include Cov(RS, UL) histograms (computed from Var(RS) and Corr(RS, UL)). The green line is the median and the red line is the mean.

2. **By DR strategy.** We also disaggregated the facility-years by DR strategy. Strategies in use by the facilities in the data set are as follows (most definitions from [98]):

   (a) HVAC

      i. Global temperature adjustment (GTA): increase zone temperature setpoints for an entire facility.

      ii. Duct static pressure decrease: decrease duct static pressure setpoints to reduce fan power.

      iii. Supply air temperature (SAT) increase: increase SAT setpoints to reduce cooling load.

      iv. Fan variable frequency drive (VFD) limit: limit or decrease fan VFD speeds or inlet guide vane positions to reduce fan power.

      v. Chilled water temperature increase: increase chilled water temperature to improve chiller efficiency and reduce cooling load.

      vi. Fan quality reduction: shut off some of multiple fans or package units to reduce fan and cooling loads.

      vii. Pre-cooling: decrease zone temperature setpoints prior to DR operations to store cooling energy in the building mass.

      viii. Cooling valve limit: limit or reduce cooling valve position to reduce cooling loads.

      ix. Lockout rooftop unit (RTU) compressors

      x. RTU shutdown: turn off a portion of RTUs, perhaps cycling which ones are off.

   (b) Lighting

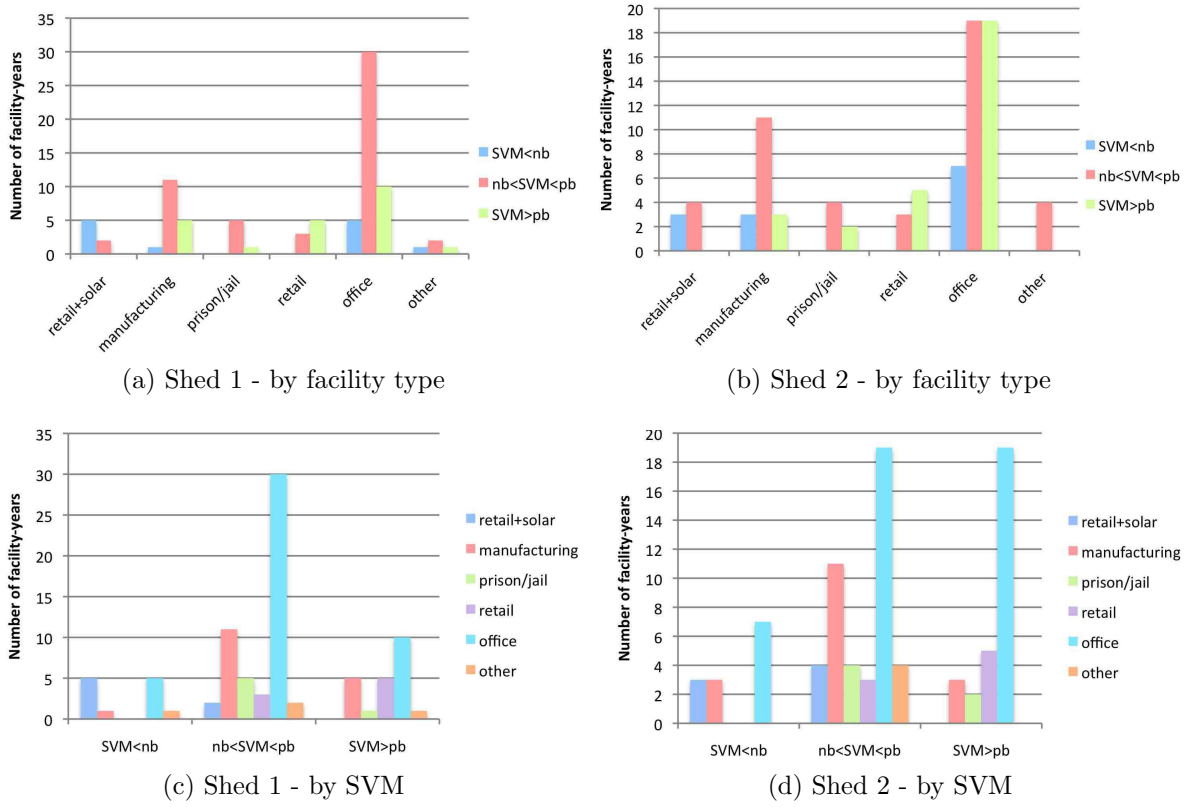
      i. Light switching: switch off lights.

      ii. Light dimming: dim lights with dimmable ballasts or bi-level switches.

   (c) Other

      i. Manufacturing process shed: turn off manufacturing process.

   It is difficult to disaggregate the facility-years by DR strategy because most of the facilities employ multiple DR strategies. Ideally, we would at least like to disaggregate HVAC from lighting, but all of the sites that employing lighting strategies also employ HVAC strategies. Therefore, we disaggregate into the follow groups: GTA only (n=28), GTA + other HVAC strategies (n=37), HVAC + lighting strategies (n=20), and process shed strategies (n=2). Again, using the 80% confidence interval, we sort facility-years into a bins (SVM<nb, nb<SVM<pb, and SVM>pb) and the plot the results both versus DR strategy and versus SVM, for each shed period. Results are shown in Figure B.3.



(a) Shed 1 - by facility type

(b) Shed 2 - by facility type

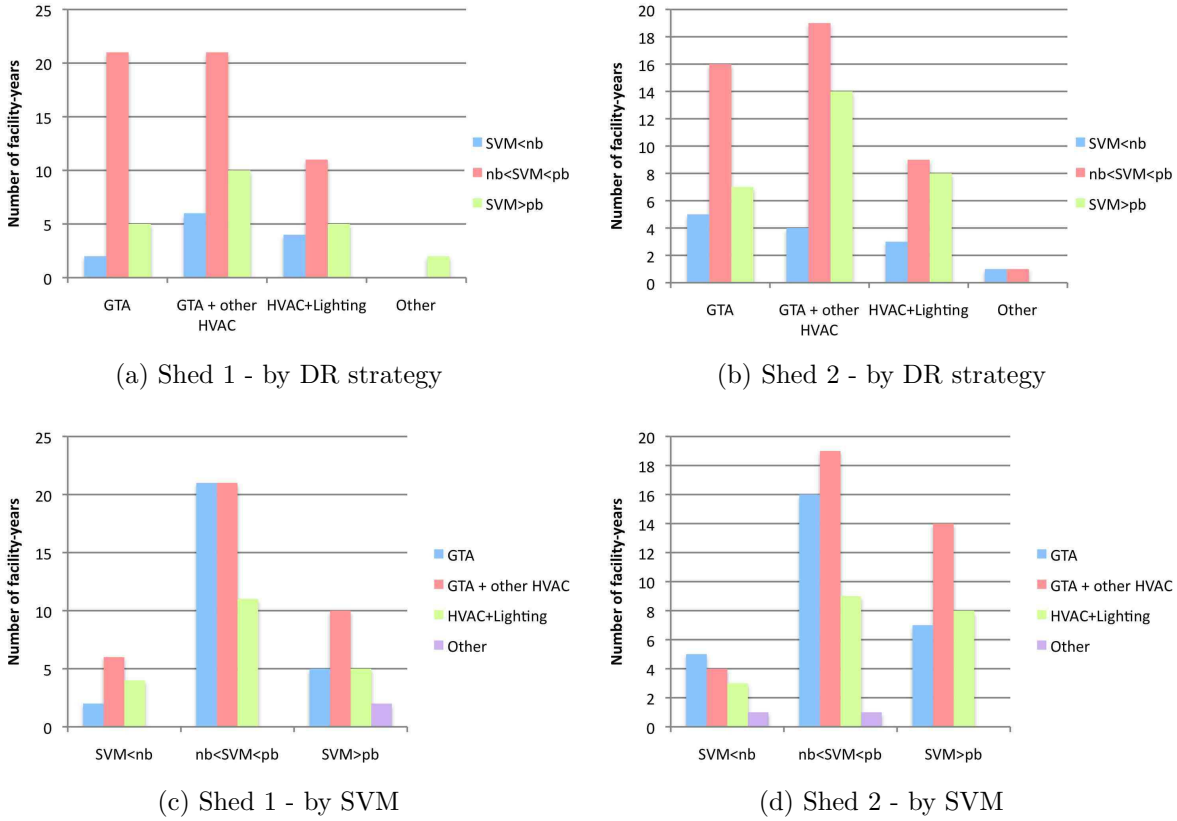
(c) Shed 1 - by SVM

(d) Shed 2 - by SVM

Figure B.2: SVM by facility type.

3. **By shed size.** Lastly, we disaggregated the facility-years by shed size, both absolute and relative (i.e., shed size divided by average predicted load during the shed period). Results for absolute shed size are shown in Figure B.4 and results for relative shed size are shown in Figure B.5.

Results of these analyses were inconclusive. In the future, it would be worth repeating this analysis with more data.



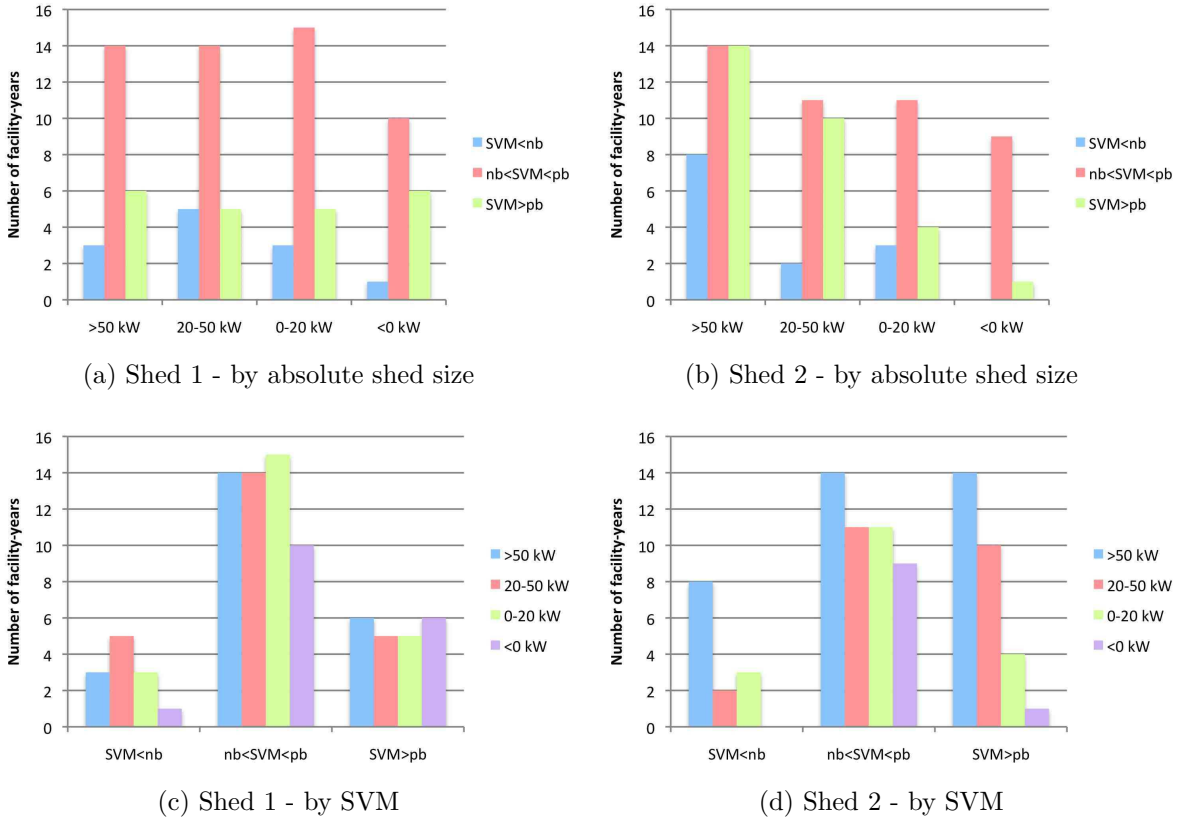
(a) Shed 1 - by DR strategy

(b) Shed 2 - by DR strategy

(c) Shed 1 - by SVM

(d) Shed 2 - by SVM

Figure B.3: SVM by DR strategy.



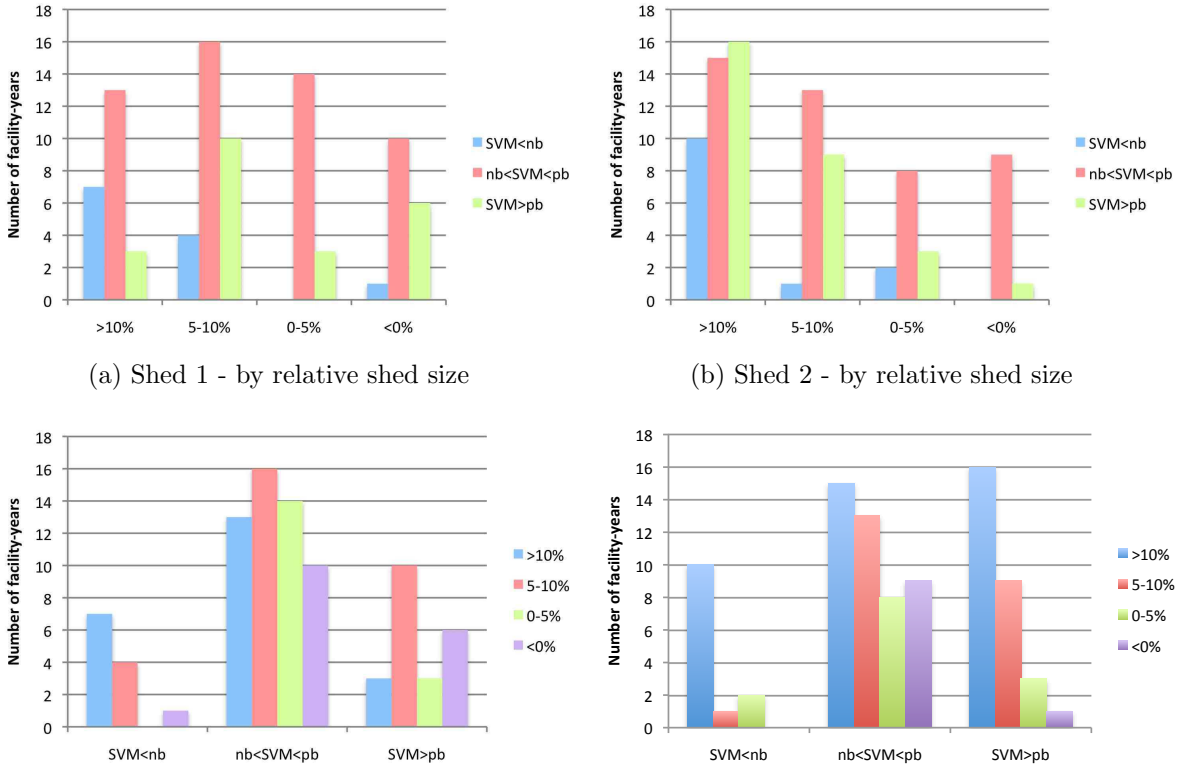
(a) Shed 1 - by absolute shed size

(b) Shed 2 - by absolute shed size

(c) Shed 1 - by SVM

(d) Shed 2 - by SVM

Figure B.4: SVM by absolute shed size.



(a) Shed 1 - by relative shed size

(b) Shed 2 - by relative shed size

(c) Shed 1 - by SVM

(d) Shed 2 - by SVM

Figure B.5: SVM by relative shed size.

Appendix C

Attempts to Derive the $\boldsymbol{A}$-Matrix

This appendix details our attempts to derive the $\boldsymbol{A}$-matrix used in the reduced for TCL population model in Chapter 4.

C.1 Uniformly Distributed a

If a is uniformly distributed between a_{min} and a_{max} then:

$$p(a) = \begin{cases} c & \text{if } a_{min} \leqslant a \leqslant a_{max} \\ 0 & \text{otherwise} \end{cases} \tag{C.1}$$

where,

$$c = \frac{1}{a_{max} - a_{min}}. \tag{C.2}$$

We consider the probability of appliances going from θ_{start} to some θ_{end}, which is within a temperature bin ($\theta_n < \theta_{end} < \theta_{n+1}$). Here, we have:

$$P(\theta_n < \theta_{end} < \theta_{n+1} | \theta_{start}) = \int_{a_1}^{a_2} p(a)\,\mathrm{d}a = \begin{cases} \int_{a_1}^{a_2} c\,\mathrm{d}a & \text{if } a_{min} < a_1 < a_2 < a_{max} \\ \int_{a_{min}}^{a_2} c\,\mathrm{d}a & \text{if } a_1 \leqslant a_{min} < a_2 < a_{max} \\ \int_{a_1}^{a_{max}} c\,\mathrm{d}a & \text{if } a_{min} < a_1 < a_{max} \leqslant a_2 \\ \int_{a_{min}}^{a_{max}} c\,\mathrm{d}a & \text{if } a_1 \leqslant a_{min} < a_{max} \leqslant a_2 \\ 0 & \text{if } a_1 > a_{max} \text{ or } a_2 < a_{min} \end{cases} \tag{C.3}$$

We can compute the range of reachable temperatures $[\theta_{min}, \theta_{max}]$ in one time step as a function of θ_{start}, a_{min}, and a_{max}. For appliances traversing from low temperatures to high temperatures:

$$\theta_{min} = a_{max}\theta_{start} + (1 - a_{max})(\theta_a - u_t\theta_g) \tag{C.4}$$

$$\theta_{max} = a_{min}\theta_{start} + (1 - a_{min})(\theta_a - u_t\theta_g). \tag{C.5}$$

For appliances traversing from high temperatures to low temperatures:

$$\theta_{min} = a_{min}\theta_{start} + (1 - a_{min})(\theta_a - u_t\theta_g) \tag{C.6}$$
$$\theta_{max} = a_{max}\theta_{start} + (1 - a_{max})(\theta_a - u_t\theta_g). \tag{C.7}$$

Now, we can simplify (C.3). For brevity, consider only appliances traversing from low temperatures to high temperatures:

$$\int_{a_1}^{a_2} p(a)\,\mathrm{d}a = \frac{cx}{\theta_a - \theta_{start} - u_t\theta_g}, \tag{C.8}$$

with

$$x = \begin{cases} \theta_n - \theta_{n+1} & \text{if } a_{min} < a_1 < a_2 < a_{max} \\ \theta_{max} - \theta_{n+1} & \text{if } a_1 \leqslant a_{min} < a_2 < a_{max} \\ \theta_n - \theta_{min} & \text{if } a_{min} < a_1 < a_{max} \leqslant a_2 \\ \theta_{max} - \theta_{min} & \text{if } a_1 \leqslant a_{min} < a_{max} \leqslant a_2 \\ 0 & \text{if } a_1 > a_{max} \text{ or } a_2 < a_{min} \end{cases} \tag{C.9}$$

Therefore, we find:

$$P(\theta_n < \theta_{end} < \theta_{n+1} | \theta_m < \theta_{start} < \theta_{m+1}) = \int_{\theta_m}^{\theta_{m+1}} \frac{cx}{\theta_a - \theta_{start} - u_t\theta_g}\,\mathrm{d}\theta_{start}. \tag{C.10}$$

However, since both the case choice and value of x is a function of θ_{start}, it is impossible to evaluate this integral analytically. We are able to evaluate the integral numerically by (1) discretizing the starting temperature bin $[\theta_m, \theta_{m+1}]$ into 1,000 'starting temperatures'; (2) for each starting temperature, computing the range of reachable temperatures; (3) discretizing the total range of reachable temperatures into 1,000 'ending temperatures'; (4) summing the number of starting temperatures that could end up in each ending temperature; (5) normalizing the sums so that the total probability of going from all starting temperatures to all ending temperatures is 1; and (6) mapping the ending temperatures to the ending temperature bin $[\theta_n, \theta_{n+1}]$. This method is repeated for each combination of starting and ending temperature bins.

C.2 Normally Distributed a

If a is normally distributed with mean μ and standard deviation σ, then:

$$p(a) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(a-\mu)^2}{2\sigma^2}}. \tag{C.11}$$

Again, we begin by considering the probability of appliances going from θ_{start} to some θ_{end}, which is within a temperature bin ($\theta_n < \theta_{end} < \theta_{n+1}$). Here, we have:

$$P(\theta_n < \theta_{end} < \theta_{n+1}|\theta_{start}) = \int_{a_1}^{a_2} p(a)\,\mathrm{d}a = \frac{1}{2}\mathrm{erf}\left(\frac{a_2-\mu}{\sigma\sqrt{2}}\right) - \frac{1}{2}\mathrm{erf}\left(\frac{a_1-\mu}{\sigma\sqrt{2}}\right), \tag{C.12}$$

where a_1 and a_2 are defined as in (C.18) and (C.19). Set $f = \theta_a - u_t\theta_g$. Then, for appliances traversing from low temperatures to high temperatures, we find:

$$a_1 = \frac{f-\theta_{n+1}}{f-\theta_{start}} \tag{C.13}$$

$$a_2 = \frac{f-\theta_n}{f-\theta_{start}}. \tag{C.14}$$

Therefore, we find:

$$P(\theta_n < \theta_{end} < \theta_{n+1}|\theta_m < \theta_{start} < \theta_{m+1}) = \tag{C.15}$$

$$\frac{1}{2}\int_{\theta_m}^{\theta_{m+1}} \mathrm{erf}\left(\frac{\alpha+\mu\theta_{start}}{\beta-\delta\theta_{start}}\right) - \mathrm{erf}\left(\frac{\gamma+\mu\theta_{start}}{\beta-\delta\theta_{start}}\right)\,\mathrm{d}\theta_{start}, \tag{C.16}$$

where $\alpha = (1-\mu)f - \theta_n$, $\beta = \sigma\sqrt{2}f$, $\delta = \sigma\sqrt{2}$, and $\gamma = (1-\mu)f - \theta_{n+1}$. There is no closed form solution to this integral.

We attempted the problem again by considering the probability of appliances going from θ_{start}, which is within a temperature bin ($\theta_m < \theta_{start} < \theta_{m+1}$) to some θ_{end}:

$$P(\theta_{end}|\theta_m < \theta_{start} < \theta_{m+1}) = \int_{a_1}^{a_2} p(a)\,\mathrm{d}a, \tag{C.17}$$

Now a_1 and a_2 are defined as:

$$a_1 = \frac{\theta_a - \theta_{end} - u_t\theta_g}{\theta_a - \theta_1 - u_t\theta_g} \tag{C.18}$$

$$a_2 = \frac{\theta_a - \theta_{end} - u_t\theta_g}{\theta_a - \theta_2 - u_t\theta_g}. \tag{C.19}$$

with $\theta_1 = \theta_m$ or θ_{m+1}, and $\theta_2 = \theta_{m+1}$ or θ_m, depending on which direction the appliances are traversing.

Then, the probability of appliances going from θ_{start}, which is within a temperature bin ($\theta_m < \theta_{start} < \theta_{m+1}$) to some θ_{end}, which is within a temperature bin ($\theta_n < \theta_{end} < \theta_{n+1}$) is:

$$P(\theta_n < \theta_{end} < \theta_{n+1}|\theta_m < \theta_{start} < \theta_{m+1}) = \int_{\theta_n}^{\theta_{n+1}}\int_{a_1}^{a_2} p(a)\,\mathrm{d}a\,\mathrm{d}\theta_{end}, \tag{C.20}$$

which also has no closed form solution because, again, we find ourselves trying to integrate error functions.

We tried to solve the integral by approximating the error function. The function $\text{erf}(z)$ can be approximated as:

$$\text{erf}(z) = \frac{z}{|z|}\sqrt{1 - \exp\left(-z^2 \frac{4/\pi + az^2}{1 + az^2}\right)} \tag{C.21}$$

where $a \approx 0.40012$. However, due to the absolute value sign, integration of this function must be done with cases (i.e., $z \geqslant 0$ and $z < 0$). If z were always negative or always positive we could simplify (C.21) and integrate without the use of cases. From (C.12), we have:

$$z_1 = \frac{a_2 - \mu}{\sigma\sqrt{2}} \tag{C.22}$$

$$z_2 = \frac{a_1 - \mu}{\sigma\sqrt{2}}, \tag{C.23}$$

so the sign of z is a function of the relationship between the specific appliance parameter, a_1 or a_2, and the mean appliance parameter, μ. Therefore, we can not simplify (C.21).

It is also be possible to approximate the error function as a piecewise linear and continuous function, such as a saturation function. This approximation is undesirable because it again requires the use of cases. Note that taking the derivative of the saturation function leads to a uniform distribution of a, which is equivalent to the case discussed in Section C.1.

The Taylor Series expansion of $\text{erf}(z)$ is:

$$\text{erf}(z) = \frac{2}{\sqrt{\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n+1}}{n!(2n+1)} \tag{C.24}$$

$$= \frac{2}{\sqrt{\pi}} \left(z - \frac{z^3}{3} + \frac{z^5}{10} - \frac{z^7}{42} + \frac{z^9}{216} - ...\right) \tag{C.25}$$

The error function along with several Taylor Series approximations are shown in Figure C.1. The relevant range of z for this problem can be computed using (C.22) or (C.23). Because a is normally distributed, approximately 95% of appliances have an appliance parameter, a_i, that is within two standard deviations of the mean value. Therefore, set $a_{min,95\%} = \mu - 2\sigma$ and $a_{max,95\%} = \mu + 2\sigma$, which results in $z_{min,95\%} = -2/\sqrt{2}$ and $z_{max,95\%} = 2/\sqrt{2}$, meaning that for approximately 95% of the appliances z will be between $-2/\sqrt{2}$ and $2/\sqrt{2}$. Similarly, $a_{min,99\%} = \mu - 3\sigma$ and $a_{max,99\%} = \mu + 3\sigma$, which results in $z_{min,99\%} = -3/\sqrt{2}$ and $z_{max,99\%} = 3/\sqrt{2}$, meaning that for approximately 99% of the appliances z will be between $-3/\sqrt{2}$ and $3/\sqrt{2}$. Note these results are independent of the particular choice of μ and σ. Dashed lines are drawn on the plot in Figure C.1 to indicate the relevant range of z. In this range, the 3^{rd} term approximation may be sufficient. While this approximation may suffice in predicting

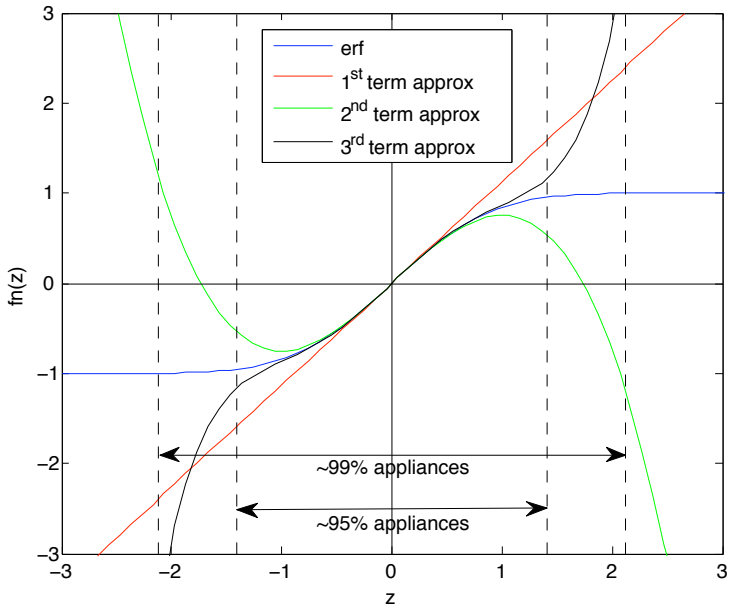


Figure C.1: Error function and Taylor series approximations.

the probably mass in ending bins where the bulk of the appliances from a given starting bin end up, it will do a poor job of predicting the probably mass in ending bins where few appliances (those with a very high or very low a_i) end up.

Appendix D

EKF Setup for the 4-bin System

This appendix details the EKF setup, local observability analysis, and state/parameter convergence results for the 4-bin system. The EKF has trouble converging to the actual states/parameters, and so we have not used this setup in Chapter 4.

D.1 Four-bin System Setup

For a 4-bin system, $\boldsymbol{A}$ and $\boldsymbol{B}$ are defined as follows:

$$\boldsymbol{A} = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \\ a_5 & a_6 & a_7 & a_8 \\ a_9 & a_{10} & a_{11} & a_{12} \\ 1-a_1-a_5-a_9 & a-a_2-a_6-a_{10} & 1-a_3-a_7-a_{11} & 1-a_4-a_8-a_{12} \end{bmatrix} \tag{D.1}$$

$$\boldsymbol{B} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{D.2}$$

Treating the parameters as states leads to a system of 15 nonlinear equations (3 states, 12 parameters).

For simplicity, we assume that the system is structured:

$$\boldsymbol{A} = \begin{bmatrix} a_1 & 0 & 0 & 1-a_4 \\ 1-a_1 & a_2 & 0 & 0 \\ 0 & 1-a_2 & a_3 & 0 \\ 0 & 0 & 1-a_3 & a_4 \end{bmatrix} \tag{D.3}$$

This leads to a system of 7 nonlinear equations (3 states, 4 parameters):

$$x_2(k+1) = (1-a_1)(1-x_2(k)-x_3(k)-x_4(k)) + a_2x_2(k) - u_2(k) \tag{D.4}$$
$$x_3(k+1) = (1-a_2)x_2(k) + a_3x_3(k) + u_2(k) \tag{D.5}$$
$$x_4(k+1) = (1-a_3)x_3(k) + a_4x_4(k) + u_1(k) \tag{D.6}$$
$$a_1(k+1) = a_1(k) \tag{D.7}$$
$$a_2(k+1) = a_2(k) \tag{D.8}$$
$$a_3(k+1) = a_3(k) \tag{D.9}$$
$$a_4(k+1) = a_4(k) \tag{D.10}$$

We also examine the highly structured system:

$$\boldsymbol{A} = \begin{bmatrix} a_1 & 0 & 0 & 1-a_2 \\ 1-a_1 & a_1 & 0 & 0 \\ 0 & 1-a_1 & a_2 & 0 \\ 0 & 0 & 1-a_2 & a_2 \end{bmatrix} \tag{D.11}$$

This leads to a system of 5 nonlinear equations (3 states, 2 parameters):

$$x_2(k+1) = (1-a_1)(1-x_2(k)-x_3(k)-x_4(k)) + a_1x_2(k) - u_2(k) \tag{D.12}$$
$$x_3(k+1) = (1-a_1)x_2(k) + a_2x_3(k) + u_2(k) \tag{D.13}$$
$$x_4(k+1) = (1-a_2)x_3(k) + a_2x_4(k) + u_1(k) \tag{D.14}$$
$$a_1(k+1) = a_1(k) \tag{D.15}$$
$$a_2(k+1) = a_2(k) \tag{D.16}$$

D.2 Local Observability Analysis

As in Section 4.4.3, we use the method in [3] to compute local observability. We begin by constructing the $\boldsymbol{\Theta}$-matrix:

$$\boldsymbol{\Theta} = \begin{bmatrix} dh(x) \\ dh(f_{u_1}(x)) \\ \vdots \\ dh(f_{u_{n-1}} \circ \ldots \circ f_{u_1}(x)) \end{bmatrix} \tag{D.17}$$

where $dh(x)$ is the gradient of the output equation, $f_{u_m}(x)$ are the state equations given inputs u_m, and n is the order of the system.

Using the highly structured system (D.12)–(D.16), we computed the 5×5 $\boldsymbol{\Theta}$-matrix shown in Table D.1.

Table D.1: $\mathbf{\Theta}$-matrix for the highly structured system (D.12)–(D.16).

0	1	1	0	0
$1-a_1$	1	a_2	$-x_2$	x_4
$2a_1-2a_1^2$	$2a_1-a_1^2-(a_2-1)^2$	$-1+2a_1-a_1^2+a_2^2$	$u_2(1)+2(-1+x_2+x_3+x_4)-2a_1(-1+2x_2+x_3+x_4)$	$u_1(1)+2(x_3-a_2x_3+a_2x_4)$
$-(-1+a_1)(3a_1^2-(-1+a_2)^2)$	$-1+3a_1^2-2a_1^3+3a_2^2-2a_2^3$	$-2a_1^3-a_1^2(-4+a_2)+2a_1(-1+a_2)+a_2(-1+a_2^2)$	$2u_1(1)+u_2(2)+x_2-2a_2x_2+a_2^2x_2-2x_4+2a_2x_4-2a_1(3+u_1(1)-u_2(1)-3x_2-3x_3-4x_4+a_2x_4)-3a_1^2(3x_2+2(-1+x_3+x_4))$	$u_1(2)+2u_2(1)+2x_2-2a_1x_2+2a_2(u_1(1)-u_2(1)-x_2+a_1x_2+3x_3)-x_4+2a_1x_4-a_1^2x_4+a_2^2(-6x_3+3x_4)$
$-2(-1+a_1)(2a_1^3-a_1(-1+a_2)^2-(-1+a_2)^2a_2)$	$4a_1^3-3a_1^4-4a_1(-1+a_2)^2+2a_1^2(-1+a_2)^2-(-1+a_2)^2(-1+2a_2+3a_2^2)$	$1-3a_1^4-2a_1^3(-3+a_2)+2a_1(-1+a_2)+2a_1^2(-1+a_2)-2a_2+a_2^4$	$2-2u_1(1)+2u_1(2)-u_2(1)+u_2(3)-2x_2+2a_2^3x_2-4x_3-a_2^2(-2+u_2(1)+6x_2+4x_3)+2a_1(-1+4u_1(1)-u_1(2)+u_2(2)+2x_2+2x_3+a_2^2(-1+2x_2+2x_3)-a_2(-2+u_1(1)+4x_2+4x_3-2x_4)-2x_4)-2x_4+2a_2(-2+u_1(1)+u_2(1)+3x_2+4x_3+x_4)-3a_1^2(4+2u_1(1)-u_2(1)-4x_2-4x_3-6x_4+2a_2x_4)-4a_1^3(4x_2+3(-1+x_3+x_4))$	$2-u_1(1)+u_1(3)-2u_2(1)+2u_2(2)-2x_2+3a_2^2(u_1(1)-2(u_2(1)+x_2-2x_3))-4x_3+2a_2(-1+u_1(2)+4u_2(1)-u_2(2)+4x_2+2x_3)-2x_4-2a_1^3x_4+4a_2^3(-3x_3+x_4)+2a_1(-2+u_1(1)+u_2(1)+3x_2+3a_2^2x_2+4x_3-a_2(-2+u_2(1)+6x_2+4x_3)+x_4)+a_1^2(2-u_1(1)-4x_2-4x_3+a_2(-2+4x_2+4x_3)+2x_4)$

The determinant is a complicated function of all of the variables. When $x_1 = 0$, $x_2 = 0$, $x_3 = 0$ (implying $x_1 = 1$) and $u_m(k) = 0 \,\forall\, k, m$ then:

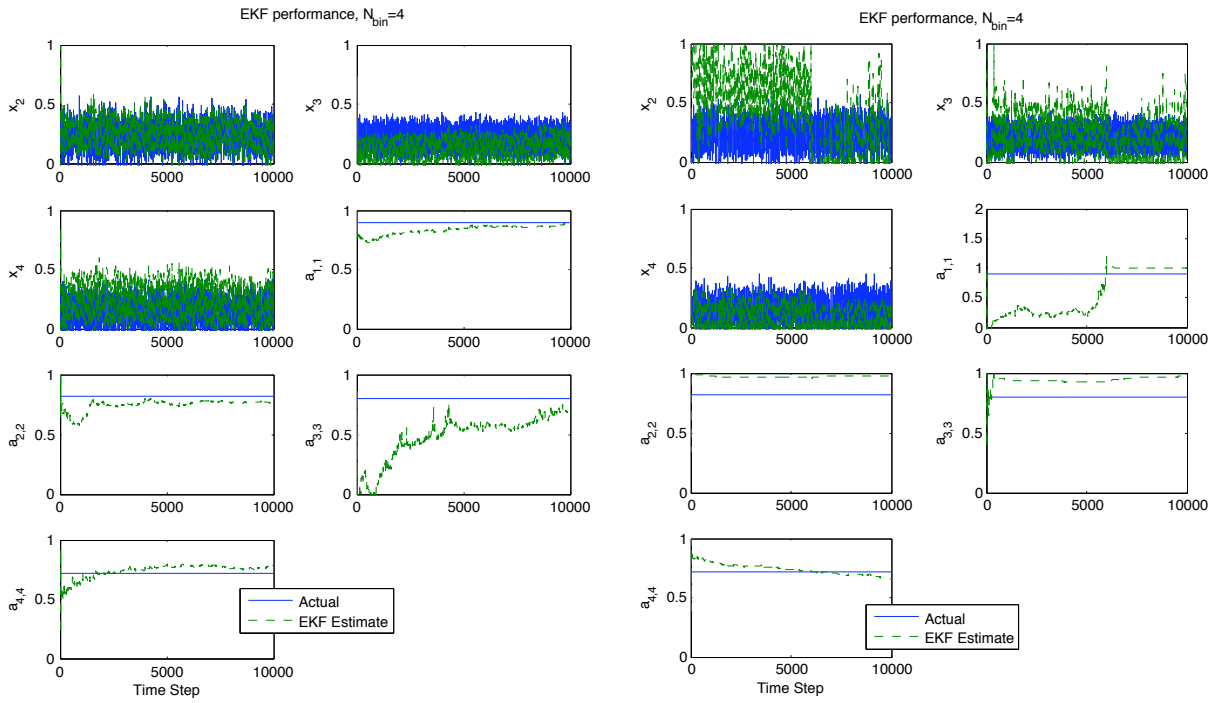
$$\det(\mathbf{\Theta}) = 8(-1+a_1)^4(-1+a_2)^2(a_1^2 + a_1(1-3a_2) + a_2(-1+2a_2)) \tag{D.18}$$

The determinant is zero if $a_1 = a_2$ or $a_2 = (1+a_1)/2$. In the general case (with non-zero states/inputs), if $a_1 = a_2$, $\det(\mathbf{\Theta})$=0, which implies that $\mathbf{\Theta}$ is not full rank and the system is not locally observable. However, in the general case, when $a_1 \neq a_2$ (including when $a_2 = (1+a_1)/2$), $\mathbf{\Theta}$ is full rank (except possibly for some numerical degeneracies) and the system is locally observable.

A similar analysis could be down for the less-structured 4-bin systems, but the $\mathbf{\Theta}$-matrix becomes exceedingly complicated and it is hard to gain insights about local observability.

D.3 State/Parameter Convergence

Two runs of the EKF estimating/identifying the structured system (D.4)–(D.10) are shown in Figure D.1. Two runs are shown to demonstrate that the results are inconsistent from run to run. Also, many more time steps are needed to identify the 4-bin system than the 2-bin system (in this analysis, 10,000 time steps corresponds to 100,000 seconds, or $\sim$ 28 hours). We find that the EKF is better able to consistently identify the highly structured



(a) Run 1.

(b) Run 2.

Figure D.1: EKF convergence results for the structured 4-bin system. Only the states/parameters identified by the EKF are shown. The other states/parameters are computed with the constraints.

system (D.12)–(D.16). Results are shown in Figure D.2. The parameters are identified after about 4,000 time steps, which corresponds to 40,000 seconds or $\sim$ 11 hours.

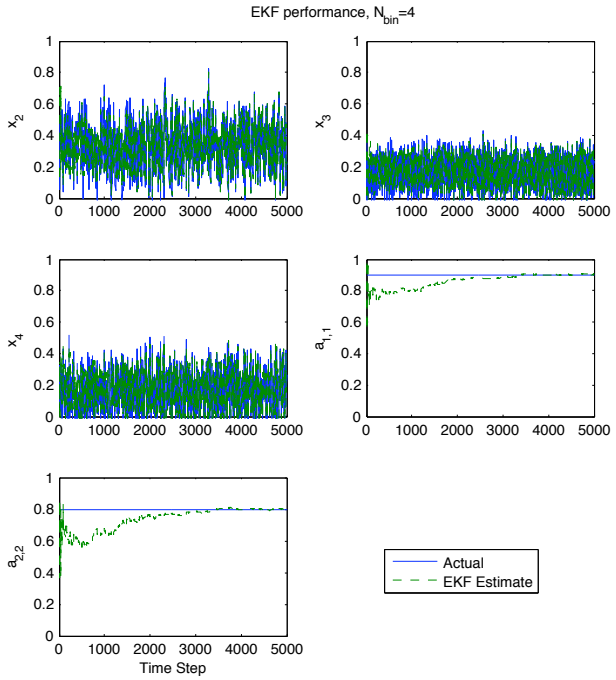


Figure D.2: EKF convergence results for the highly structured 4-bin system. Only the states/parameters identified by the EKF are shown. The other states/parameters are computed with the constraints.

Appendix E

Source of the TCL parameters

This appendix details the source of the TCL parameters used for the resource potential, revenue potential, and cost analysis in Chapter 5.

E.1 ACs

- δ: We used a range expanded around the value in [19].
- θ_{set}: We developed a range based on EIA survey data [36].
- R: We used a range expanded around the value in [19].
- C: We tuned this parameter until the model-predicted mean annual energy consumption per appliance matched data from [35].
- COP: We used the value from [19].
- P_{trans}: We used a range expanded around the value in [19].

E.2 Heat pumps

- δ: We assumed this to be the same as for ACs.
- θ_{set}: We developed a range based on EIA survey data [36].
- R: We assumed this to be the same as for ACs.
- C: We assumed this to be the same as for ACs.
- COP: We assumed this is greater than that of ACs.
- P: We assumed this to be the same as for ACs.

E.3 Water heaters

- δ: A standard water heater dead-band is 5°F [128]. We assumed a range around this value.
- θ_{set}: Different organizations recommend different set points (e.g, [33]. We chose a range of 110-130°F.
- R: We tuned this parameter until the model-predicted mean annual energy consumption per appliance matched data from [35].
- C: We computed the amount of energy required to heat the water contained in a 4 ft tall, 22 in diameter water heater by 1°C. We use a range around the computed value.
- COP: For electric resistance devices, the COP is 1.
- P: We use a range around the standard value of 15,354 BTU/hr [34].

E.4 Refrigerators

- δ: We used a range based on data collected by a UC Berkeley student [113].
- θ_{set}: We used the range given in [121].
- R: We tuned this parameter until the model-predicted mean annual energy consumption per appliance matched data from [35].
- C: Assuming half of a 5.5 ft $\times$ 3 ft $\times$ 2.5 ft refrigerator is filled with water (i.e., food), we computed the amount of energy required to heat the water by 1°C. We used a range around the computed value.
- COP: We assumed this is less than that of ACs.
- P: We compute this value from the mean annual energy consumption per appliance [35], assuming a 33% duty cycle [32].