LBNL-63728

Ernest Orlando Lawrence Berkeley National Laboratory

Estimating Demand Response Load Impacts: Evaluation of Baseline Load Models for Non-Residential Buildings in California

Katie Coughlin, Mary Ann Piette, Charles Goldman and Sila Kiliccote

Demand Response Research Center

Ernest Orlando Lawrence Berkeley National Laboratory

1 Cyclotron Road, MS 90R4000

Berkeley CA 94720-8136

Environmental Energy Technologies Division

January 2008

http://eetd.lbl.gov/ea/EMS/EMS_pubs.html

The work described in this report was funded by California Energy Commission (Energy Commission), Public Interest Energy Research (PIER) Program, under Work for Others Contract No.150-99-003, Am #1, and the Office of Electricity Delivery and Energy Reliability, Permitting, Siting and Analysis of the U.S. Department of Energy, under contract No. DE-AC02-05CH11231. The authors are solely responsible for any omissions or errors contained herein.

Table of Contents

List of Tables

List of Figures

Acknowledgements

The work described in this report was funded by the Demand Response Research Center which is funded by the California Energy Commission (Energy Commission), Public Interest Energy Research (PIER) Program, under Work for Others Contract No.150-99-003, Am #1 and the Office of Electricity Delivery and Energy Reliability, Permitting, Siting and Analysis of the U.S. Department of Energy under contract No. DE-AC02-05CH11231.

Researchers who contributed to this paper include Ranjit Bharvirkar, and June Han at Lawrence Berkeley National Laboratory (LBNL). The authors are grateful for the support from Mike Gravely, Martha Brook, and Kristy Chew (California Energy Commission). The authors would like to thank the following for review comments on draft of this report: GlennPerez and John Goodin (CAISO), Carmen Hendrickson (Enernoc), and Mark Martinez (SCE).

Abstract

Both Federal and California state policymakers are increasingly interested in developing more standardized and consistent approaches to estimate and verify the load impacts of demand response programs and dynamic pricing tariffs. This study describes a statistical analysis of the performance of different models used to calculate the baseline electric load for commercial buildings participating in a demand-response (DR) program, with emphasis on the importance of weather effects. During a DR event, a variety of adjustments may be made to building operation, with the goal of reducing the building peak electric load. In order to determine the actual peak load reduction, an estimate of what the load would have been on the day of the event without any DR actions is needed. This *baseline load profile (*BLP) is key to accurately assessing the load impacts from event-based DR programs and may also impact payment settlements for certain types of DR programs. We tested seven baseline models on a sample of 33 buildings located in California. These models can be loosely categorized into two groups: (1) *averaging methods*, which use some linear combination of hourly load values from previous days to predict the load on the event, and (2) *explicit weather models*, which use a formula based on local hourly temperature to predict the load. The models were tested both with and without *morning adjustments,* which use data from the day of the event to adjust the estimated BLP up or down.

Key findings from this study are:

- The accuracy of the BLP model currently used by California utilities to estimate load reductions in several DR programs (i.e., hourly usage in highest 3 out of 10 previous days) could be improved substantially if a morning adjustment factor were applied for weather-sensitive commercial and institutional buildings.
- Applying a morning adjustment factor significantly reduces the bias and improves the accuracy of *all* BLP models examined in our sample of buildings.
- For buildings with low load variability, all BLP models perform reasonably well in accuracy.
- For customer accounts with highly variable loads, we found that no BLP model produced satisfactory results, although averaging methods perform best in accuracy (but not bias). These types of customers are difficult to characterize with standard BLP models that rely on historic loads and weather data.

Implications of these results for DR program administrators and policymakers are:

- Most DR programs apply similar DR BLP methods to commercial and industrial sector customers. The results of our study when combined with other recent studies (Quantum 2004 and 2006, Buege et al., 2006) suggests that DR program administrators should have flexibility and multiple options for suggesting the most appropriate BLP method for specific types of customers.

- Customers that are highly weather sensitive, should be given the option of using BLP models that explicitly incorporate temperature in assessing their performance during DR events.
- For customers with more variable loads, it may make more sense to direct these facilities to enroll in DR programs with rules that require customers to reduce load to a firm service level or guaranteed load drop (e.g. which is a common feature of interruptible/curtailable tariffs) because DR performance is difficult to predict and evaluate with BLP models.
- DR program administrators should consider using weather-sensitivity and variability of loads as screening criteria for appropriate default BLP models to be used by enrolling customers, which could improve the accuracy of DR load reduction estimates.

1. Introduction

Both Federal and California state policymakers are increasingly interested in developing more standardized and consistent approaches to estimate and verify the load impacts of demand response programs and dynamic pricing tariffs (e.g. critical peak pricing) [FERC Staff Report 2006; CPUC 2007].[1] For example, the California Public Utility Commission is overseeing a regulatory process to develop methods to estimate the load impacts of demand response (DR) programs. These methods will be useful for measuring the cost-effectiveness of programs, assist in resource planning and long-term forecasting exercises, and allow the California Independent System Operator (CAISO) to be able to more effectively utilize DR as a resource.

Policymakers are concerned that the methods used to estimate load reductions and compensate customers and load aggregators are fair and accurate, and that protocols for estimating load impacts can be used by resource planners and system operators to incorporate demand-side resources effectively into wholesale (and retail) markets. One of the challenges to developing protocols for estimating load impacts is the diversity of customers (and their loads) and the heterogeneity in types of DR programs and dynamic pricing tariffs. In its Order Instituting Rulemaking on DR load impact protocols, the CPUC [2007) acknowledged that calculating the load impacts of DR programs is not easy given the diversity in curtailment strategies, customer characteristics, and DR event characteristics (e.g., timing, duration, frequency, and location).

This paper describes a statistical analysis of the performance of different models used to calculate the baseline electric load for buildings participating in an event-driven demand-response (DR) program, with emphasis on the importance of weather effects. During a DR event, a variety of adjustments may be made to building operation, with the goal of reducing the building peak electric load. In order to determine the actual peak load reduction, an estimate of what the load would have been without any DR actions is needed. This is referred to as the *baseline load profile* or BLP and is key to accurately assessing the load impacts from certain types of demand response programs that pay for load reductions.[2] The impacts estimate uses the BLP calculated for a specific

---

[1] In their report to Congress on Demand Response and Advanced Metering, FERC Staff identified the need for consistent and accurate measurement and verification of demand response as a key regulatory issue in order to provide system operators with accurate forecasts and assessments of demand response, to support just and reasonable rates for the delivery of DR in wholesale markets, and to accurately measure and verify demand resources that participate in capacity markets.

[2] Note that an explicit customer baseline calculation is not as important if the DR program design requires customers to reduce usage to a "firm load" level (e.g. an interruptible/curtailable tariff) [KEMA 2007].

time period on the event-day. This calculation should ideally account for all those factors which are known to systematically impact the building load at any given moment, such as weather, occupancy, and operation schedules.

The sample of buildings included in this study is mainly commercial (e.g., office and retail) and institutional (e.g. schools, universities, government) buildings. There are a few industrial facilities including a bakery, electronics manufacturing, laboratories and large mixed-use office/data center. Historically, many utilities have marketed emergency DR programs and interruptible/curtailable tariffs to large industrial facilities with process loads or onsite generation. The mix and type of industries has changed in California and other states due to the growth in light industry, high technology (e.g. computer electronics, bio-technology), commercial office space, the institutional sector, and retail services. As DR programs continue to evolve, it is important that the program rules and protocols for determining load impacts take into account the increasingly diverse types of customers that can participate in DR programs.

The BLP methods discussed in this study are most relevant for non-residential buildings and have not been broadly evaluated for relevance to industrial facilities. DR events are called during times of system stress, which are also typically related to weather. For California, DR may be used in the summer to deal with high peak loads on weekdays, which are often driven by space cooling in buildings. This study looks at results for buildings participating in an Automated Demand Response pilot sponsored by the PIER Demand Response Research Center[3] [Piette et al 2007; Piette et al 2005] and who face a critical peak price. In these cases DR events are only called on normal working days, during the period 12 pm. to 6pm. Weather-sensitivity is likely to be especially important during DR events.

Accurate BLP estimates help ensure that individual participants in DR programs are fairly compensated as part of settlement procedures for their actual load reductions, and that the contribution of demand response resources in aggregate is properly accounted for in resource planning and benefit cost screening analysis. In both cases it is important to avoid systematic bias in estimating the load reductions. Given the correlation between temperature and increased building energy use for space conditioning, non-weather corrected models may under-predict the baseline and therefore systematically underestimate the response. This can be true even for buildings with large non-weather responsive loads, if the weather-dependent load is significant relative to the estimated DR reduction. On the other hand, many customers, load aggregators and DR program administrators have a strong preference for simpler calculation methods with limited data requirements that can be used for customer settlement processes. It is useful therefore to establish how much quantitative improvement is gained by introducing more complicated calculation methods.

---

[3] The California Energy Commission's Public Interest Energy Research (PIER) Program sponsors the DRRC, which is managed by LBNL.

Table 1: Sites included in this study

Site name	Description	Location	Weather Station
Office1	Office	Fremont	Hayward Airport
Office2	Office	Martinez	Buchanan Field
Office3	Office	Martinez	Buchanan Field
Detention Facility	Detention Facility	Martinez	Buchanan Field
School1	Jr. High School	Fremont	Hayward Airport
Museum	Museum	Oakland	Oakland Foothills
Office4	Office	San Jose	San Jose Airport
Office/Lab1	Office/Lab	Foster City	San Francisco Airport
Office/Lab2	Office/Lab	Foster City	San Francisco Airport
Office/Lab3	Office/Lab	Foster City	San Francisco Airport
Retail1	Big Box Retail	Emeryville	Oakland Airport
Retail2	Big Box Retail	Palo Alto	Palo Alto Airport
School2	High School	Fremont	Hayward Airport
Office/DC1	Office/Data Center	Concord	Buchanan Field
Office5	Office	Rocklin	Fair Oaks
Supermarket	Supermarket	Stockton	Stockton Airport
Office/LM1	Office/Light Manufacturing	Milpitas	San Jose Airport
Office/LM2	Office/Light Manufacturing	Milpitas	San Jose Airport
Office/LM3	Office/Light Manufacturing	Milpitas	San Jose Airport
Office/LM4	Office/Light Manufacturing	Milpitas	San Jose Airport
Office/LM5	Office/Light Manufacturing	Milpitas	San Jose Airport
Office/LM6	Office/Light Manufacturing	Milpitas	San Jose Airport
Office/LM7	Office/Light Manufacturing	Milpitas	San Jose Airport
Office/LM8	Office/Light Manufacturing	Milpitas	San Jose Airport
Office/LM9	Office/Light Manufacturing	Milpitas	San Jose Airport
Bakery	Bakery	Oakland	Oakland Airport
Office/DC2	Office/Data Center	Dublin	Pleasanton
Office/DC3	Office/Data Center	Dublin	Pleasanton
Retail3	Big Box Retail	Antioch	Buchanan Field
Retail4	Big Box Retail	Bakersfield	Meadows Field
Retail5	Big Box Retail	Hayward	Hayward Airport
Retail6	Big Box Retail	Fresno	Fresno Airport

1.1. Project Objectives and Analytical Approach

In this study we evaluate seven BLP models, for a sample of 32 sites in California incorporating 33 separately metered facilities. In some cases the meter may include electricity use for multiple buildings at one location. Such is the case, for example with the High School and the Office/Data Center. For each BLP model, we tested two implementations: models without and with a morning adjustment (which incorporates site usage data from the morning of the DR event prior to load curtailment). The site locations, building types and associated weather data sites are listed in Table 1. The majority of the sites in the dataset are commercial buildings, but the analytical methods we develop here can be applied to any building type. For each site, 15-minute electric interval load data are available through the web-based customer energy metering site maintained by Pacific Gas and Electric (PG&E). While the models differ in the details, each uses electric load data from a period before the event to predict the electric load on an event day.

Our main objective in this work is to provide a statistically valid evaluation of how well each BLP model performs, and to relate the performance to more general building characteristics. To do so, we need to define both the sampling procedure and the evaluation metrics. Building loads always have a random component, so the baseline estimation problem is inherently statistical in the sense that to properly assess the performance of a method, a sufficiently large sample of applications must be considered. Because our building sample is small, to develop a large enough data set, we define a set of *proxy event days* (days on which no curtailment occurs and the load is known, but which are similar in terms of weather to actual event days). For these days, we use the historical data and the BLP model to predict the load, and compare the prediction to the actual load for that day. If the proxy event set is large enough, we can evaluate each model for each site separately. We focus on metrics that quantify the *bias* and the *accuracy* of the model at the building level.

1.2. Prior Work

Several recent studies have reviewed and analyzed alternative methods for calculating DR peak load reductions, either as part of working groups or evaluations of California DR Programs using customer load data [KEMA 2003, Quantum 2004, Quantum 2006). The most extensive review of BLP methods is provided in the KEMA (2003) study *Protocol Development for Demand Response Calculation—Findings and Recommendations*. This study examined a number of methods in use by utilities and ISO's across the country, and evaluated them in terms of accuracy and bias. As noted there, a BLP method is defined by specifying three component steps:

- A set of data selection criteria,
- An estimation method,
- An adjustment method.

The difference between the estimation and the adjustment step is that estimation uses data prior to the event day to predict the BLP during the event period, while adjustment uses data from the event day, before the beginning of the curtailment period, to align and shift the predicted load shape by some constant factor to account for characteristics that may affect load on the day of the event.

The KEMA 2003 report, while quite comprehensive, included only three accounts from California in their total sample of 646 accounts. There are 32 accounts from the Northwest and 24 from the Southwest, so the sample is dominated by data from the eastern U. S. Given significant climatic and demographic variation across the country, with corresponding differences in building practices, occupancy, etc., it is unclear how well results really generalize across different regions. In particular, the KEMA study found that explicitly weather–dependent models did not generally outperform models that did not include weather. One of the goals of this work is to determine whether this hypothesis also holds true for California.

Quantum Consulting (2004) conducted an analysis of methods to estimate customer baselines as part of its broader evaluation of California's 2004 DR programs targeted at industrial and commercial customers. The baseline assessment had billing data for a large sample (450 customers) of non-participants that were eligible for the DR programs; customers' peak demand ranged from 200 kW to greater than 5 MW. The sample was weighted appropriately to represent the population of eligible customers. Eight proxy event days were selected for each utility from the period July 1, 2003 to August 31, 2003. These event days were classified in to three categories: high load (potential event days), low load (as potential "test" days), and consecutive high load days (series of three high load days that occurred back-to-back). This study coupled with subsequent analysis of load impacts in the Quantum (2006) evaluation provides a more detailed analysis of the bias and accuracy of BLP methods for large industrial and commercial buildings located in California.

In developing the statistical sample of test profiles, KEMA (2003) and Quantum (2004) used a large number of accounts, but a relatively small number of calendar days, comprised of only those days where an actual curtailment was called in the region (as in KEMA) or proxy event days (as in Quantum). Our statistical approach is different, using a much larger selection of proxy event days. This allows us to create a statistical picture for each building, which is useful both because our building sample is smaller, and because we can then evaluate whether different methods perform equally well for different building types.

The methods investigated in this study overlap with the KEMA (2003) and Quantum (2004) reports, with a somewhat different approach to adjustment for weather effects. We have also developed a different method for estimating the degree of weather-sensitivity of a building, and different diagnostics to quantify the predictive accuracy of the BLP, and the estimated peak load savings values that are used in bill settlement. The metric used for measuring the bias of the BLP is similar to that used by Quantum (2004). We also provide detailed results for the baseline model that is currently in wide use in California, based on a simple average of the hourly load over the highest 3 of the previous 10 days in the sample. Some of the baseline models tested in Quantum (2004) are the same as those included in this study (e.g. 10-day unadjusted and 10-day adjusted). Our approach to testing BLP models that include an Adjustment Factor is similar to the Quantum (2004) study, although the number of hours and time period (e.g. day of vs. day ahead) used for calculating the adjustment factors is different.

The Quantum evaluation reports (2004 and 2006) and a subsequent article based on those reports by Buege et al. (2006) conclude that the 10-day adjusted BLP is significantly better than the currently used 3-day unadjusted BLP in California. Specifically, the authors assert that the 3-day unadjusted BLP method is biased high by two to four times. They also find that the presence of large customers with highly variable load can add considerable uncertainty to the estimation of baselines.

The remainder of the paper is organized as follows: In Section 2 we present an overview of the technical steps involved in preparing the data sets, defining the sample of proxy event days, running the models and developing the diagnostics. In Section 3 we describe our weather sensitivity metrics, and in Section 4 we define each of the methods investigated in this paper. Section 5 presents the results for our building sample. In Section 6 we provide a discussion of the limitations of the analytical approach used here, and outline some suggestions for future work.

2. Data Processing and Evaluation Metrics

In this section we describe the preparation of the data, the mechanics of implementing different models, and the diagnostic metrics used in this report.

2.1. Data Sources

The building load data used in this project consists of 15-minute electric interval load data for each metered building, which we convert to hourly by averaging the values in each hour. We use data from May through October of 2005 and 2006 to define the sample days and test the methods. Only the warm-weather months are included here, as these are the periods when (to date) events are more likely to be called in California's DR Programs. The amount of data available depends on how long the account has participated in the DR program (in some cases interval meters were installed because the site was willing to go onto a DR program), and whether there is any missing data during the sample period.

The explicit weather models require hourly temperature data for each site. The data were obtained by assigning each site to a weather station that is currently active and maintained by either a state or a federal agency. A website developed at the University of California at Davis (www.ipm.ucdavis.edu/WEATHER) provides maps of the weather monitoring stations maintained by various entities for each county in California. These are used to select the weather station closest (both geographically and in elevation) to each site. The sites were chosen from those maintained by NOAA (available by subscription) or by the California Irrigation Management Information System (CIMIS), which is a program of the state Department of Water Resources. Only outdoor dry bulb air temperature data are used currently in developing the weather-dependent models.

2.2. Proxy Event Days

The goal of using proxy event days is to have a large sample set for which (i) the actual loads are known and (ii) the days are similar in some sense to the actual DR event days that were called by the CAISO and California utilities in 2005 and 2006. Before selecting the proxy set, we first need to define the set of what we call *admissible days*, which is the set of days that can be used as input to the BLP model calculations. We define admissible days as normal working days, i.e. eliminating weekends, holidays and past curtailment events, which follows standard procedures.

The proxy event days are selected as a subset of the admissible days. DR events are typically called on the hottest days, and can be called independently in each of several climate zones defined by the CEC (all the sites available for this study are located in either zone 1 or zone 2, as indicated in Table 1). To define the weather characteristics associated with an event day, we first construct a spatially-averaged zonal hourly temperature time series, using a simple average over the weather stations located in the zone. The hourly zonal temperatures are then used to construct three daily metrics: the maximum daily temperature, the average daily temperature, and the daily cooling degree hours (using 65 °F as the base temperature).

Sorting the weather data on the value of the daily metric provides a list of the hottest to coolest days in the sample period. We defined the proxy event days as the top 25 percent of the admissible days sorted in this manner. The three metrics give consistent results for the hottest days, but select slightly different samples. A little over ¾ of the actual event days in each year are included in the top 25 percent selected.[4] The results presented here use the sample associated with cooling-degree hours.[5] For each building, a proxy event day is included in the analysis only if there is sufficient load data for that day. Hence, the proxy event sets vary somewhat from building to building. On average, this procedure leads to about 60 proxy days for each site.

2.3. Model Runs and Diagnostics

In our procedure, model results are calculated for all the admissible days, but diagnostics are calculated only for the set of proxy event days. For each model and each building site, the BLP for each hour from 9 am-6 pm is calculated. While the event period is limited to 12 pm-6 pm, the adjustment factors may require model and actual data from the early morning period. Our notation is as follows:

- the admissible day is labeled **d**
- the hour is labeled **h;** our convention is h = time at the beginning of the hour
- the predicted load is **pl(d,h)**
- the actual load is **al(d,h)**
- the adjustment factor for day d is **c(d)**

---

[4] It is possible that a metric based on the deviation of the daily value from a monthly average would capture the rest of the event days, however it is also the case that event days may not be entirely determined by the daily temperature.

[5] The results do not appear to be sensitive to the daily metric used to define the sample. Note that the proxy event days are defined purely from temperature data, so there is one set for each zone.

- the absolute difference between the actual load and the predicted load is defined as $\mathbf{x(d,h) = al(d,h) - pl(d,h)}$
- the relative difference between the actual load and the predicted load (or percent error) is defined as $\mathbf{e(d,h) = x(d,h)/al(d,h)}$

For each combination of a model and a site we calculate the absolute and relative difference between predicted and actual loads, x(d,h) and e(d,h), for each proxy event day and each hour in the event period, which gives us about 360 observations for each building site. Our statistical metrics are defined for these sets of numbers.

Often, utilities or ISOs settle payments for performance during DR events based on the average hourly load reduction during the hours of the event. It is therefore useful to compare the prediction of the average hourly load to the actual value. To do so we define:

- $\mathbf{A(d) = \langle al(d,h) \rangle}$ the actual hourly load averaged over the event period
- $\mathbf{P(d) = \langle pl(d,h) \rangle}$ the predicted hourly load averaged over the event period
- $\mathbf{X(d) = A(d) - P(d)}$ the absolute difference in average event-period hourly load
- $\mathbf{E(d) = X(d)/A(d)}$ the percent difference in average event-period hourly load

2.3.1. Adjustment Factors

As noted in the KEMA 2003 report, the algorithm for predicting a customer's load shape includes a modeling estimation step and an adjustment step. In our analysis, we evaluate each model both with and without a *morning adjustment* factor applied. The KEMA report reviews several methods for calculating the adjustment factor. Most are based on some comparison of the actual to the predicted load in the hours immediately preceding an event. In this study, we use a multiplicative factor defined as the ratio of the actual to the predicted load in the two hours prior to the event period:[6]

$$\mathbf{c(d) = [\ al(d,h{=}10) + al(d,h{=}11)\ ] / [\ pl(d,h{=}10) + pl(d,h{=}11)\ ]}.$$

To adjust the BLP, we multiply the predicted value in each hour by the daily adjustment factor:

$$\mathbf{pl'(d,h) = c(d) * pl(d,h)}.$$

The Adjustment Factor essentially scales the customer's baseline from admissible days to the customer's operating level on the actual day of a DR event.[7]

[6] Quantum (2004 and 2006) studies use the three hours preceding the event period.

[7] Deciding on the period to use for the Adjustment Factor can be more problematic for DR programs or tariffs where the event is announced on prior days (e.g. Critical Peak Pricing), as there may be some concern about customers "gaming" their baseline by intentionally increasing consumption during the hours just prior to the event. Quantum [2004] addressed this issue by

We also tested an alternative adjustment approach that used the two hours preceding the event to define an additive, rather than multiplicative, correction factor. In our sample, there is no significant difference in the results.

2.3.2. Diagnostic Measures

For each BLP model, both with and without adjustment, and each site, we calculate the set of absolute and percentage errors x(d,h) and e(d,h). Our evaluation of the performance of a model is based on the statistical properties of these errors. To measure any bias in the model, we calculate the median of the distribution of errors.[8] If the method is unbiased the median will be zero. If the median is positive (negative) it means that the model has a tendency to predict values smaller (larger) than the actual values. To quantify the accuracy of the model, we calculate the average of the absolute value of the error terms (|e(d,h)| or |x(d,h)|). These metrics can also be applied to the average event-period values X(d) or E(d).

3. Weather Sensitivity

Weather sensitivity is a measure of the degree to which building loads are driven directly by local weather. By far the most important weather variable is temperature. Physically, space-conditioning loads are affected by the total heat transfer to the building from the environment, which is affected by such details as the orientation and shading of the building, shell characteristics, thermal mass, cooling and ventilation strategies, and occupant behavior. In modeling baseline energy consumption, the cooling load in a given hour is related to some kind of weighted integral of the temperature over an earlier set of hours, with the weighting and the number of hours depending on the specific building. Practically, weather dependence is often represented by using regression models relating hourly load to hourly temperature, possibly including lagged variables or more complex functions of temperature. The KEMA 2003 report investigated a number of weather regression models, some fairly complicated, but it is not clear from that study that including additional variables leads to a consistent improvement in the accuracy of the models tested. In some climates humidity may be an important factor in weather sensitivity, but for sites in California, weather behavior is likely to be dominated by dry bulb outdoor air temperature (OAT). The models tested here are based on straightforward correlation of hourly load with

---

selecting the three hours prior to the utility notifying customers of an event on the prior weekday. For purposes of our analysis, we have included the two preceding hours prior to a CPP event on the same day for the Adjustment Factor.

[8] The median of a set of numbers is the value such that one half of the set is greater than the median, and one half of the set is less than the median. The average value of the error could also be used as a bias measure, however the median tends to be more robust as it is not sensitive to outliers.

hourly OAT. This approach effectively rolls all other building-specific factors into the regression coefficients.

To develop an a priori sense of whether a building is likely to be weather sensitive, we use a simple and robust correlation function known as Spearman Rank Order Correlation (ROC) (Press et al. 2007). Given two time series (X(t), Y(t)) of equal length M, the ROC is obtained by (1) replacing each variable with its rank relative to the rest of the set and (2) calculating the linear correlation coefficient between the two sets of ranks. While the distributions of the X and Y variables may be unknown, the ranks are distributed uniformly on the interval [1,M]. Thus, the ROC can be calculated explicitly without approximation, along with the associated statistical significance. The ROC coefficient is insensitive to the size of hourly variation in X and Y, and measures only the degree to which they tend to rise and fall together. This makes it more straightforward to compare correlation magnitudes across different types of buildings. The ROC should also provide a more robust measure of weather sensitivity for buildings with highly variable loads.

For each site, we calculate the ROC between load and temperature for each hour separately for all the admissible days. We calculate an ROC coefficient in each hour separately to avoid spurious correlations driven by the daily work schedule. The average of these calculated values during event period hours is shown in Table 2. These have been color-coded to indicate high (>= 0.8), medium (0.65-0.8), low (0.5-0.65) and very low (<0.5) degrees of correlation. We also calculate an average coefficient over all the hours, which is used as an overall indicator for the building. In all cases except two the significance is greater than 95%. The two exceptions are the School1 and School2 sites, which also show negative correlation coefficients. These schools are closed from mid-June to September. The algorithm works correctly for these sites, but what it picks out is an anti-correlation between load and temperature and a strong random component.

Table 2: Hourly rank order correlation (ROC) coefficients

Site Name	Avg	h=10am	h=11am	h=12pm	h=1pm	h=2pm	h=3pm	h=4pm	h=5pm	h=6pm
Retail6	0.97	0.96	0.96	0.96	0.97	0.97	0.97	0.97	0.97	0.97
Supermarket	0.93	0.87	0.89	0.91	0.94	0.94	0.95	0.97	0.96	0.96
Retail4	0.91	0.91	0.90	0.89	0.92	0.89	0.90	0.92	0.93	0.94
Office/LM5	0.88	0.82	0.87	0.85	0.88	0.88	0.91	0.92	0.91	0.89
Retail3	0.83	0.80	0.73	0.78	0.80	0.85	0.87	0.88	0.87	0.87
Retail5	0.83	0.71	0.79	0.82	0.82	0.85	0.85	0.85	0.85	0.90
Office2	0.82	0.78	0.77	0.77	0.83	0.87	0.85	0.88	0.87	0.81
Office3	0.82	0.82	0.83	0.82	0.80	0.82	0.85	0.88	0.89	0.69
Office4	0.82	0.89	0.90	0.88	0.87	0.84	0.68	0.73	0.77	0.84
Office/DC3	0.79	0.73	0.80	0.82	0.85	0.83	0.85	0.82	0.83	0.54
Office/Lab2	0.78	0.78	0.71	0.73	0.79	0.83	0.80	0.80	0.79	0.82
Retail1	0.77	0.75	0.82	0.72	0.79	0.76	0.75	0.77	0.78	0.79
Office/LM7	0.77	0.82	0.80	0.79	0.77	0.74	0.74	0.74	0.75	0.75
Office1	0.75	0.71	0.70	0.70	0.72	0.76	0.76	0.77	0.81	0.78
Office/DC1	0.75	0.87	0.79	0.74	0.68	0.65	0.66	0.71	0.81	0.87
Office/DC2	0.74	0.72	0.73	0.72	0.73	0.75	0.79	0.82	0.75	0.67
Detention Facility	0.71	0.63	0.64	0.70	0.72	0.71	0.67	0.66	0.80	0.83
Retail2	0.71	0.66	0.67	0.67	0.70	0.70	0.73	0.78	0.76	0.73
Office/LM1	0.65	0.56	0.62	0.66	0.68	0.72	0.69	0.69	0.59	0.61
Office/LM2	0.64	0.63	0.70	0.66	0.63	0.60	0.61	0.61	0.69	0.62
Office/LM4	0.63	0.57	0.58	0.61	0.61	0.63	0.66	0.68	0.68	0.67
Office/Lab1	0.61	0.51	0.30	0.51	0.62	0.69	0.74	0.71	0.69	0.72
Office/LM8	0.60	0.62	0.59	0.62	0.61	0.63	0.61	0.62	0.62	0.47
Office/Lab3	0.49	0.34	0.39	0.43	0.52	0.52	0.53	0.50	0.57	0.56
Museum	0.48	0.47	0.50	0.56	0.58	0.54	0.56	0.55	0.49	0.10
Office/LM3	0.45	0.43	0.45	0.43	0.47	0.45	0.47	0.43	0.43	0.46
Office5	0.40	0.39	0.42	0.42	0.43	0.42	0.40	0.37	0.38	0.34
Office/LM9	0.36	0.30	0.34	0.35	0.37	0.35	0.38	0.39	0.38	0.39
Office/LM6	0.17	0.16	0.16	0.21	0.18	0.18	0.18	0.16	0.16	0.15
Bakery	0.01	0.07	0.10	0.07	0.02	-0.01	-0.06	-0.05	-0.06	-0.01
School1	-0.05	-0.12	-0.04	0.00	0.01	0.01	0.03	-0.07	-0.09	-0.13
School2	-0.23	-0.24	-0.12	-0.12	-0.19	-0.17	-0.24	-0.33	-0.34	-0.35

4. Baseline Profile (BLP) Models

We tested seven baseline models for our sample of buildings, with and without the morning adjustment factor applied. These models can be loosely categorized into two groups: (1) *averaging methods*, which use some linear combination of hourly load values from previous days to predict the load on the event day (models 1 through 4), and (2) *explicit weather models*, which use a formula based on local hourly temperature to predict the load (models 5 through 7). The methods are summarized in Table 3, and described in more detail below. To improve the readability of the results tables, we have given each model a code (BLP1 through BLP7). For the version of the model with no morning adjustment factor applied we append an *n* to the code. For example, *BLP1* refers to the simple average model with morning adjustment, and *BLP1n* refers to the simple average with no adjustment.

Table 3: Summary of BLP models evaluated

Code	Description
BLP1	10-Day simple average baseline with morning adjustment
BLP2	Weighted average formula using previous 20 admissible days with morning adjustment
BLP3	Simple average over the highest 3 out of 10 previous admissible days with morning adjustment
BLP3n	Simple average over the highest 3 out of 10 previous admissible days without morning adjustment
BLP4	Simple average over the highest 5 out of 10 previous admissible days with morning adjustment
BLP5	Seasonal regression baseline with morning adjustment
BLP6	10-day regression baseline with morning adjustment
BLP7	Limited seasonal regression baseline with morning adjustment

4.1. 10-Day Simple Average Baseline with Morning Adjustment (BLP1)

In simple averaging, the average of the hourly load over the N most recent admissible days before the event is used to predict the load on the event day. Typically, N is set equal to 10, which is the value used in our analysis. Note that averaging will tend to under-predict the load by definition. Both BLP1 and BLP1n (without morning adjustment) were also tested in the Quantum (2004) study.

4.2. Weighted Average Baseline with Morning Adjustment (BLP2)

In recent regulatory discussions on load impact estimation protocols, EnerNOC has proposed a recursive formula to predict the load on day d from predictions over a set of N previous days (EnerNOC 2006). This is equivalent to a weighted average of actual loads over the previous N days, with weights defined by:

pl(d,h) = 0.1*[sum(m=0,N-1) $(0.9)^{m}$ * al(d-m,h)] + $(0.9)^{N}$ * al(d-N,h)

We applied EnerNOC's proposed BLP using 20 previous days.

4.3. Simple Average over the Highest 3 out of 10 Admissible Days with Morning Adjustment (BLP3)

In this model, the 3 days with the highest average load during the event period 12pm-6pm are selected from the previous 10 days, and the simple average of the load over these three days is calculated for each hour. The unadjusted version

(BLP3n), is the baseline method currently used in California's Demand Bidding and Critical Peak Pricing programs[9] and was also tested in Quantum (2004).

4.4. Simple Average over the Highest 5 out of 10 Admissible Days with Morning Adjustment (BLP4)

This method is similar to BLP3, except the highest five days are used.

4.5. Seasonal Regression Baseline with Morning Adjustment (BLP5)

In this method, we use a year's worth of data to calculate the coefficients of a linear model : pl(d,h) = C1(h) + C2(h)*temperature(d,h). The coefficients are calculated using linear regression. We have calculated two separate sets of coefficients, using 2005 data and 2006 data. The coefficients differ slightly, and the 2006 values are used here. All the admissible days from May through October are used. This is the Linear Regression with Seasonal Coefficient method.

4.6. 10-Day Regression Baseline with Morning Adjustment (BLP6)

This method uses a linear regression model as defined for BLP5, but the coefficients are calculated using only data from the N most recent admissible days prior to the event period. In this analysis we set at N equal to 10.

4.7. Limited Seasonal Regression with Morning Adjustment (BLP7)

This method is a variation of BLP5. Here, in calculating the regression coefficients, instead of using all the admissible days from May through October, we use only those hours for which the temperature is greater than or equal to 60°F. The results for this model do not differ significantly from those for model BLP5, and are not included in the tables.

5. Results

As an illustration of how the actual and estimated load profiles look, Figure 1 shows data for an office building in Fremont, for a summer day in 2006. The plot shows the

---

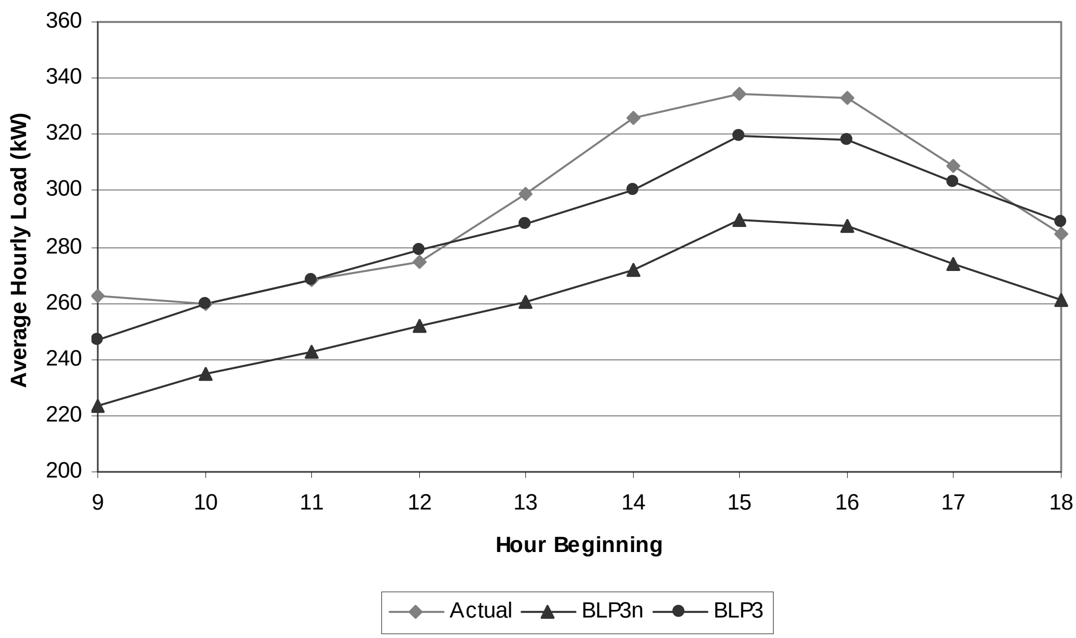
[9] The California Capacity Bidding Program uses a different version of the BLP3n model, in which the selection of the highest 3 of 10 days is based on analysis of the total load for all the sites included in the portfolio of a load aggregator. For the methods evaluated in this study, the 3 highest days are chosen separately for each individual site/facility. This approach is commonly used by other U.S. ISO/RTOs (e.g., NYISO, PJM, ISO-NE) in their DR programs. We believe that calculating customer baseline load profiles from individual participating facilities is likely to enhance customer acceptance and transparency because individual customers can determine and verify their baseline load profile and load reductions (compared to the aggregator portfolio approach in which the CBL depends on the usage patterns of all other customers).

estimated BLP for method BLP3 with adjustment and method BLP3n with no adjustment, and the actual load. The model values are calculated for all the hours from 9 am to 6 pm, and the values for the beginning at 10 am and 11 am are used to calculate the morning adjustment factor. In this particular case, the unadjusted prediction is below the actual load, so the adjustment boosts the load profile upward.

The percent error in the estimate is the ratio of the difference between the actual and estimated load, divided by the actual load. For this example, the actual hourly load is on the order of 300 kwh. The BLP3n prediction is roughly 30 kwh below the actual, so the percent error in model BLP3n is about +10%. This error is slightly larger during the afternoon, high load period. The difference from the adjusted BLP3 profile is roughly 5-15 kwh during the event period, so the adjustment reduces the error to roughly 5%.

Our statistical analysis is based on calculations of profiles like the one illustrated in Figure 1 for all sites, all proxy event days and all models. For a given site and model, the performance of the model is characterized by the average size of the absolute percent error over all proxy event days, and whether there is a bias towards predominantly positive or negative errors.

Figure 5-1: Example results for models BLP3n and BLP3

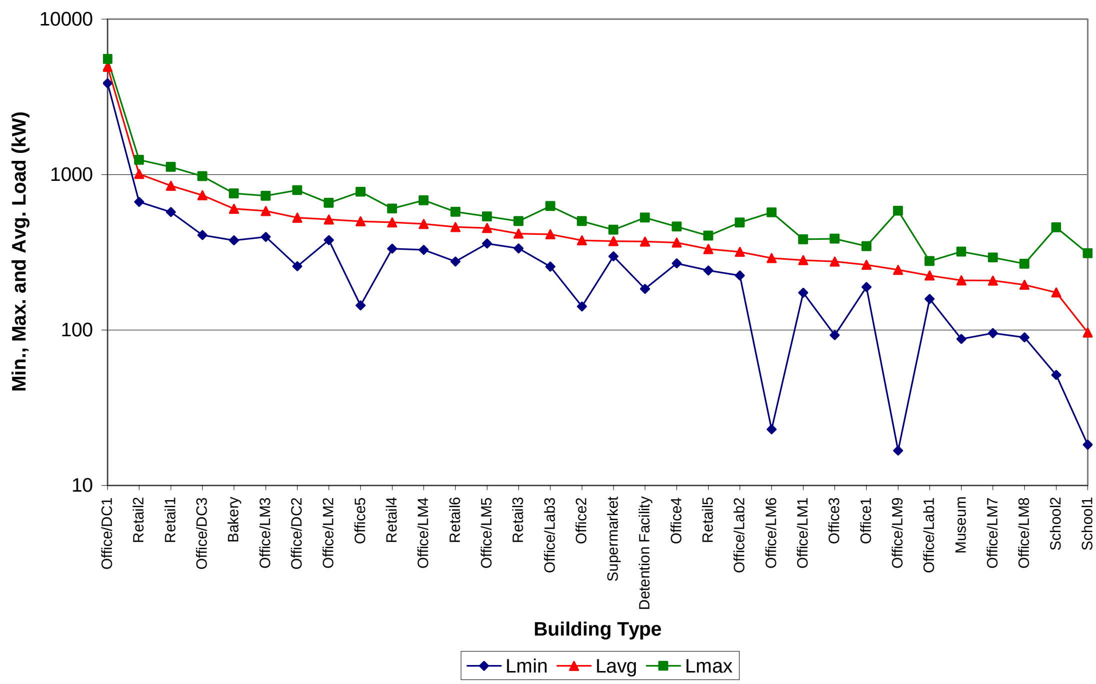


5.1. Building Characteristics

An examination of some general characteristics of our sample of buildings is very helpful in interpreting the results of our analysis of BLP models. The characteristics we use are the weather sensitivity (discussed above) and the load variability of each building in the sample. In this context, load variability refers to how different the load profiles are from one day to another, which will affect the degree to which the loads on a given day can be predicted from previous data.

There are a variety of ways of measuring the load variability. In Figure 2, we show one approach, where for each building site the minimum, maximum and average hourly load are plotted. The sites are labeled in Figure 2 by building type, and the order on the horizontal axis is determined by sorting the average loads from largest to smallest. Note that the vertical axis uses a logarithmic scale. This plot shows that while for most sites variability is moderate, for several sites the variability exceeds two orders of magnitude. In these cases, the building was essentially "turned off" for some part of the sample period (for example, the Museum is closed on Mondays).

Figure 5-2: Maximum, minimum and average hourly load at each site



To quantify the variability, we use a simple measure based on the deviation of the load in each hour from an average calculated over all the admissible days. The deviation is defined as the average value of the difference between the load in a given hour and the period average load for that hour. This is converted to a percent deviation by dividing by the period average. This variability coefficient can take on any value greater than zero, with low values indicating low variability. In order to derive a single value for each facility in our sample, we average the values calculated for each hour. Facilities are classified as either high or low variability. The cutoff is chosen at 15 percent. We also classify building weather sensitivity as either high or low, with the cutoff set at an ROC coefficient of 0.7. Using this segmentation scheme, we disaggregate our sample of facilities into four categories, as shown in Table 4.

In our sample there are three buildings with non-standard schedules, shown in the table in italics. Two are schools that are closed during the summer as noted above. The third is a museum that is closed on Mondays and most Tuesdays. Although these schedules are perfectly predictable, they deviate from the assumption that normal operating days are Monday through Friday year-round. This results in an artificially high level of variability in load (and corresponding reduced estimate of weather sensitivity) for these sites.

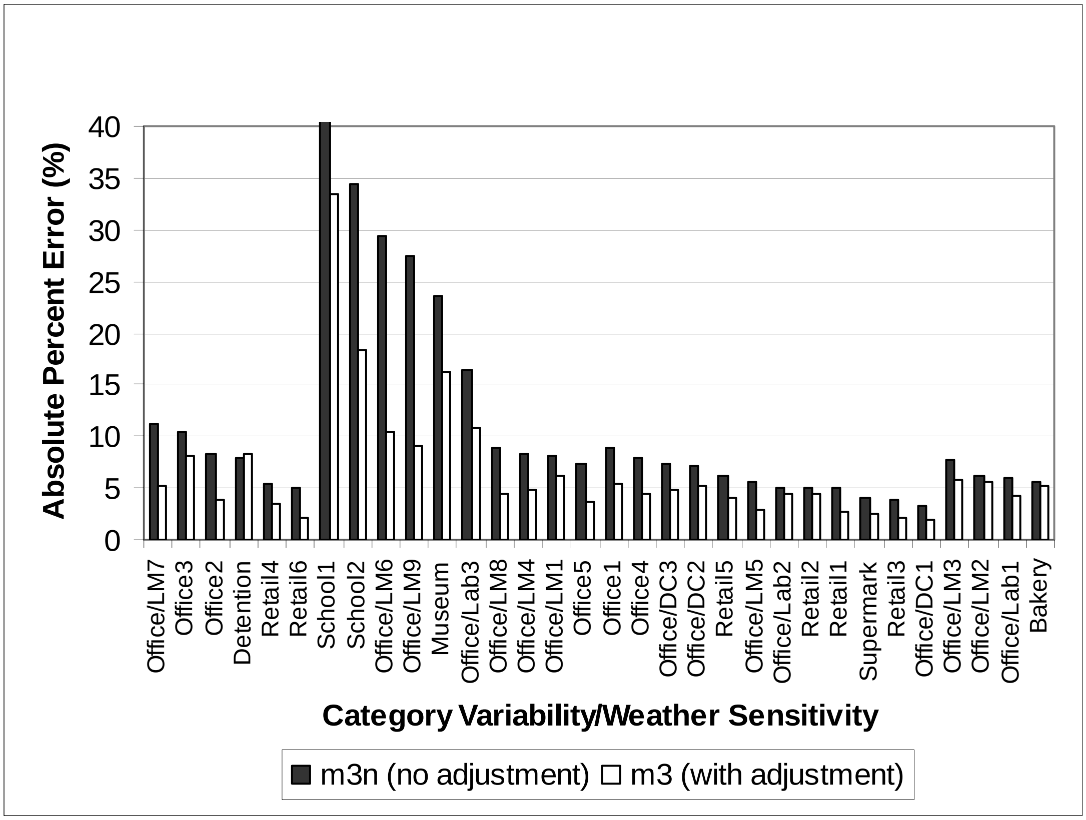
Table 4: Classification by load variability (var) and weather sensitivity (ws)

Site Name	ROC	VAR	ws	var	Site Name	ROC	VAR	ws	var
Retail6	***0.97***	***0.20***	h	h	Supermarket	***0.93***	0.10	h	l
Retail4	***0.91***	***0.19***	h	h	Office/LM5	***0.88***	0.11	h	l
Office2	***0.83***	***0.22***	h	h	Retail3	***0.83***	0.13	h	l
Office3	***0.82***	***0.27***	h	h	Retail5	***0.83***	0.10	h	l
Office/LM7	***0.77***	***0.19***	h	h	Office4	***0.82***	0.14	h	l
Detention Facility	***0.71***	***0.24***	h	h	Office/DC3	***0.79***	0.11	h	l
Office/LM1	0.65	***0.17***	l	h	Office/Lab2	***0.79***	0.15	h	l
Office/LM4	0.63	***0.15***	l	h	Retail2	***0.77***	0.10	h	l
Office/LM8	0.60	***0.32***	l	h	Office/DC1	***0.75***	0.10	h	l
**Museum*	*0.49*	***0.29***	*l*	*h*	Office1	***0.75***	0.15	h	l
Office/Lab3	0.49	***0.18***	l	h	Office/DC2	***0.74***	0.14	h	l
Office5	0.40	***0.29***	l	h	Retail1	***0.71***	0.12	h	l
Office/LM9	0.36	***0.63***	l	h	Office/LM2	0.64	0.11	l	l
Office/LM6	0.17	***0.96***	l	h	Office/Lab1	0.61	0.13	l	l
**School1*	*-0.05*	***0.41***	*l*	*h*	Office/LM3	0.45	0.14	l	l
**School2*	*-0.23*	***0.34***	*l*	*h*	Bakery	0.01	0.11	l	l

5.2. Morning Adjustment

Overall, we find that the morning adjustment factor substantially improves the performance of each baseline model; both in terms of reduced bias and improved accuracy (see Figures 3 and 4). In Figure 3, we show the average of the absolute errors between predicted and actual load, which is our accuracy measure, for each site using the BLP3/BLP3n (highest 3 of 10) model. The sites are labeled by name, and have been ordered along the x-axis according to the category they belong to with respect to variability and weather sensitivity. The category order is high-high, high-low, low-high, and lastly low-low. The shaded bars are for the model with no morning adjustment applied, and the white bars with the morning adjustment. The vertical axis limits are chosen to ensure that all the data are visible, and as a result one of the unadjusted values is off the chart. This plot shows that for almost all the sites, and in particular for the high variability sites, the morning adjustment leads to a large improvement in the accuracy of the model prediction. For cases where the adjustment does not improve the result (for example, Detention Facility) use of the adjustment does not substantially degrade the model performance.

Figure 5-3: Error magnitude for model BLP3 without and with adjustment



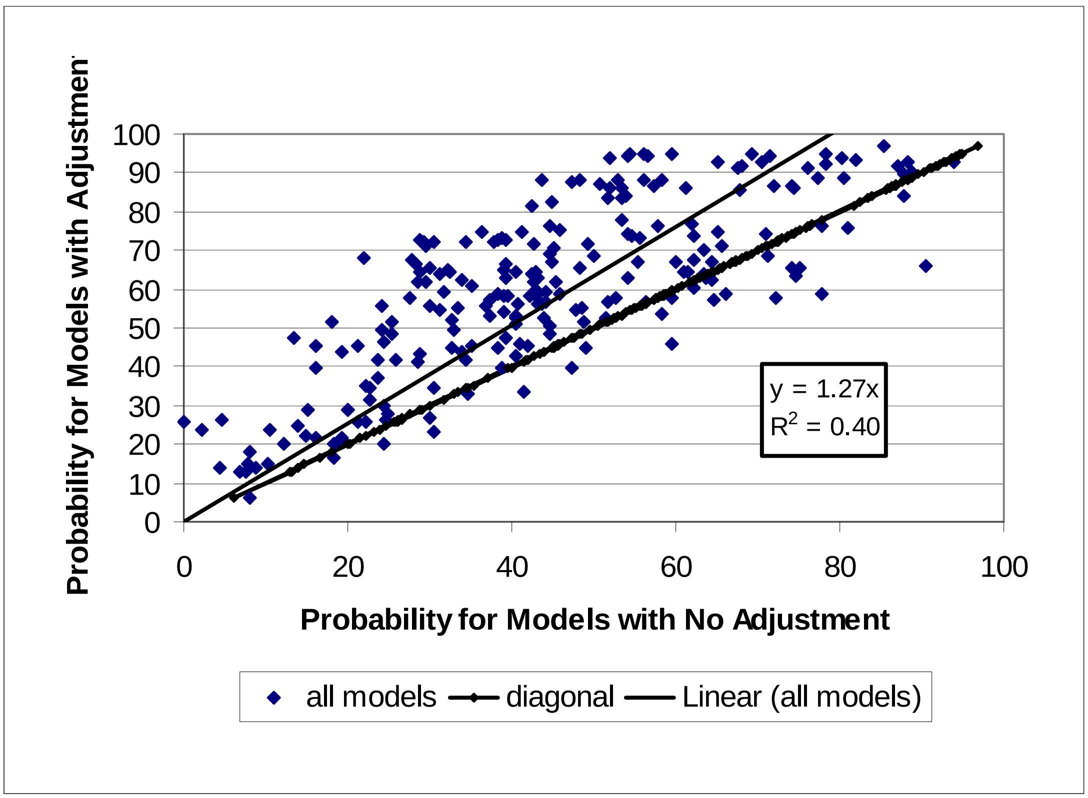
The results in Figure 3 illustrate the decrease in magnitude of errors between predicted and actual load when the morning adjustment is applied. In Figure 4, we provide a slightly more complicated representation of the effect of applying the morning adjustment factor, using data from all BLP models and all sites. It illustrates the impact of the adjustment on the likelihood that the model will have a small (less than 5%) error. Each point on the chart represents a single building-model pair. It is constructed as follows:

1. For a site and a model with ***no*** adjustment applied we calculate the probability that the absolute value of the error |e(d,h)| is less than 5%.
2. For a site and a model *with* the adjustment applied we calculate the probability that the absolute value of the error |e(d,h)| is less than 5%.

3. The probability calculated in case (2) is plotted against the probability in case (1).
4. The diagonal is shown on the plot as a heavy dark line.
5. A linear trend line passing through (0,0) is also plotted in black.

The diagonal corresponds to a situation where the morning adjustment has no effect on the likelihood of a small error. If a point lies above the diagonal it means that the probability of a small error is larger when the adjustment is used. The fact that most points are above the diagonal means that in most cases the morning adjustment increases the probability that the error will be small. The linear fit shows that on average, for a given model-site pairing, the probability of small error is increased by about 25 percent when the morning adjustment is applied. There is broad scatter in the plot, indicating that some cases are improved a great deal, where as others are improved only slightly. Below the diagonal, there are a few cases where the adjustment factor produces worse results, but in general these differences are small.

Figure 5-4: Comparison: probability of error less than 5% with or without morning adjustment



We have observed two situations where building or facility operating issues are likely to be misrepresented with morning adjustments. These are related to demand response end-use strategies that begin prior to the start of the DR event, and are important for day-ahead or other pre-notification DR programs. The first situation is when pre-cooling is done only on DR event days, and not on normal days. If the chiller load is higher than normal on the morning of a DR event day, the baseline load will be adjusted to a higher value than if the pre-cooling had not occurred. The adjustment reflects a demand response strategy, not the fact that the day is hotter than normal. In the second situation, we have observed industrial demand response strategies that involve reducing the end-use loads one to two hours prior to the beginning of the DR event. This is done because some industrial loads take time to "unload". In this case the morning load is lower than it would have been in the absence of a DR event, so the morning adjustment will scale the baseline down more than is appropriate. These issues suggest that some information about the building DR strategies would be very useful in assessing whether and how a morning adjustment should be applied to a baseline model.

5.3. Bias and Accuracy

The next two tables present our analysis of the relative bias and accuracy among the various BLP models that we tested in our sample of buildings. Table 5 provides results for the distribution of hourly percent errors e(d,h) between predicted and actual load, while Table 6 shows the same metrics for the distribution of daily values of the percent error in the average event-period hourly load E(d). The bias is measured using the median of the sample of values, and the accuracy is measured by the average of the absolute value of the error. We present only the percent error data as these are easiest to compare across buildings. In Tables 5 and 6, the best and worst performing models for each building, are highlighted in blue and grey shading respectively. The table rows are sorted on the categories for variability (var) and weather sensitivity (ws). The three sites with anomalous schedules (the two schools and the museum) are noted in italics.

In the table of results for the hourly values e(d,h) we present both model BLP3 and model BLP3n (highest 3 of 10 with and without the adjustment applied), as the current practice in California is to use the BLP3n method (no adjustment).

Table 5: Metrics for the percent hourly error e(d,h) by site and model

site	var	ws	Median of e(d,h) (Bias measure) m1	m2	m3	m3n	m4	m5	m6	Average of |e(d,h)| (Accuracy measure) m1	m2	m3	m3n	m4	m5	m6
Office2	h	h	0.0	0.1	-0.8	2.4	-0.5	4.4	1.6	3.9	4.0	3.9	8.3	3.8	5.9	4.8
Office3	h	h	0.7	0.5	-1.0	3.6	-0.7	7.5	1.1	7.5	7.5	8.2	10.5	8.0	11.2	8.6
Detention Facility	h	h	-0.6	-0.8	0.5	1.9	0.2	-0.6	0.0	7.9	7.7	8.3	8.0	8.6	7.2	8.2
Office/LM7	h	h	-2.3	-2.4	1.0	1.8	0.1	-4.7	0.2	5.3	5.4	5.2	11.2	5.3	6.8	5.1
Retail4	h	h	-0.9	-0.5	-0.5	2.0	-0.5	-1.0	-0.2	3.0	2.9	3.5	5.4	3.4	3.0	3.5
Retail6	h	h	-0.3	-0.4	-0.7	2.2	-0.5	-1.1	-0.3	1.9	2.0	2.1	5.0	2.0	2.0	2.1
**School1*	*h*	*l*	*-7.1*	*-7.2*	*-3.8*	*7.3*	*-7.8*	*0.2*	*0.0*	*31.0*	*31.6*	*33.5*	*55.1*	*32.3*	*44.5*	*34.6*
**Museum*	*h*	*l*	*1.2*	*3.4*	*1.6*	*3.5*	*1.6*	*4.4*	*1.6*	*15.0*	*15.8*	*16.2*	*23.6*	*15.4*	*14.9*	*18.2*
**School2*	*h*	*l*	*-0.2*	*0.1*	*-1.2*	*7.0*	*-3.4*	*1.6*	*2.6*	*18.9*	*20.7*	*18.3*	*34.4*	*18.1*	*27.5*	*22.7*
Office/Lab3	h	l	-4.7	-4.9	0.3	5.1	-3.5	-1.9	-0.7	10.6	10.6	10.9	16.5	11.1	8.1	11.4
Office5	h	l	-1.4	-2.0	0.1	2.1	-0.2	-2.1	0.2	3.6	3.7	3.6	7.4	3.5	4.6	3.6
Office/LM1	h	l	-1.4	-1.1	1.8	2.0	0.0	-0.7	-0.2	5.8	5.7	6.1	8.1	5.8	6.1	6.0
Office/LM4	h	l	-2.7	-2.9	0.0	3.4	-1.4	-4.8	-1.6	5.1	5.1	4.9	8.3	4.9	6.1	4.4
Office/LM6	h	l	-1.0	-1.3	2.4	4.0	0.7	8.3	0.9	7.7	7.8	10.5	29.3	9.1	12.1	12.0
Office/LM8	h	l	-0.4	-0.8	0.1	0.5	-0.4	6.7	-1.2	4.7	4.8	4.5	8.9	4.8	9.3	5.1
Office/LM9	h	l	-2.9	-3.1	-1.0	8.9	-1.6	-11.1	1.0	7.2	7.0	9.0	27.4	8.0	13.9	10.8
Office1	l	h	-2.4	-2.7	0.2	2.3	-0.5	1.1	0.1	5.3	5.2	5.4	8.9	5.3	4.2	4.9
Office4	l	h	-1.9	-2.0	-0.8	-1.5	-0.9	0.2	-0.6	4.3	4.3	4.5	8.0	4.3	3.6	4.5
Office/Lab2	l	h	0.7	0.6	0.5	-0.4	0.8	-0.4	0.4	4.4	4.1	4.5	5.1	4.2	4.8	4.9
Retail1	l	h	1.0	1.4	-0.2	-0.9	0.4	1.2	0.4	2.5	2.5	2.7	5.0	2.6	2.6	2.5
Retail2	l	h	-0.7	-0.9	-0.3	2.8	-0.4	0.6	0.0	4.7	4.7	4.5	5.1	4.9	4.1	5.2
Office/DC1	l	h	1.7	1.3	0.7	0.6	0.7	3.3	0.7	2.4	2.1	1.9	3.2	2.1	4.1	2.8
Supermarket	l	h	-1.6	-1.6	-0.4	1.0	-0.5	0.3	-0.3	2.7	2.5	2.5	4.0	2.3	2.1	2.0
Office/LM5	l	h	-1.0	-1.3	0.7	0.5	0.1	0.3	0.2	2.6	2.7	2.9	5.6	2.7	1.9	2.4
Office/DC2	l	h	-4.0	-5.3	-1.7	-1.6	-2.4	-0.1	-3.2	5.8	6.7	5.2	7.1	5.1	4.3	5.1
Office/DC3	l	h	-3.4	-3.9	-0.5	0.4	-2.1	-1.0	-1.1	5.1	5.4	4.8	7.3	4.8	3.3	3.9
Retail3	l	h	-0.7	-0.8	-0.1	1.3	-0.2	-0.2	0.2	2.0	2.1	2.1	3.8	2.0	2.2	2.3
Retail5	l	h	-2.0	-2.2	0.0	0.5	-0.6	0.0	0.4	4.2	4.2	4.1	6.1	4.1	2.7	3.5
Office/Lab1	l	l	-2.1	-1.9	1.6	0.3	-0.7	0.7	-0.3	4.4	4.4	4.2	6.0	4.3	4.2	5.1
Office/LM2	l	l	0.2	-0.5	0.6	1.4	0.5	0.9	-0.8	5.2	5.0	5.6	6.2	5.3	5.3	5.7
Office/LM3	l	l	-0.9	-1.1	0.8	2.7	-0.6	1.4	-0.8	5.4	5.3	5.8	7.8	5.7	5.3	6.4
Bakery	l	l	0.6	0.8	0.0	3.7	0.0	0.2	-0.1	4.4	4.3	5.2	5.6	4.6	6.6	5.4

blue/white = best performance | grey/black = worst performance

With respect to the bias indicator, both the BLP3 and the BLP6 models perform well (BLP6 is the load-temperature model based on the 10 previous days of data). The weather-dependent BLP6 model is distinguished by the fact that it is the only model that consistently avoids bias in our sample of buildings. For the accuracy metric it is clear that the unadjusted 3-in10 model BLP3n is the least accurate. Table 5 also shows that, for buildings with low variability, all models (except BLP3n) perform reasonably well, which is not surprising. For buildings with high weather sensitivity, overall the explicit weather models (BLP5 and BLP6) either improve the performance for that building or do not affect it much.

Table 6: Metrics for the average hourly load percent error E(d)

site	var	ws	Median of E(d) (Bias measure) m1	m2	m3	m4	m5	m6	Average of |E(d)| (Accuracy measure) m1	m2	m3	m4	m5	m6
Office2	h	h	0.0	0.3	-1.0	-0.2	4.6	1.6	3.4	3.5	3.3	3.1	5.3	4.0
Office3	h	h	2.1	1.1	-0.5	0.5	8.4	1.3	6.8	6.7	7.2	6.9	10.6	7.8
Detention Facility	h	h	-0.6	-0.4	0.3	0.1	-0.9	0.6	7.0	6.8	7.3	7.6	5.9	6.7
Office/LM7	h	h	-2.5	-3.0	1.0	0.5	-5.4	-0.2	4.4	4.7	4.1	4.2	6.1	3.9
Retail4	h	h	-1.0	-0.6	-0.6	-0.7	-0.9	0.4	2.5	2.3	2.9	2.9	2.4	2.9
Retail6	h	h	-0.3	-0.3	-0.9	-0.6	-1.2	-0.1	1.8	1.9	2.0	1.9	1.9	1.9
*School1	h	l	-10.4	-10.3	-3.6	-9.0	2.2	1.2	22.2	22.0	25.2	25.5	28.2	24.6
*Museum	h	l	1.7	5.0	1.6	3.5	5.4	0.1	14.1	14.7	15.1	14.6	14.0	16.5
Office/Lab3	h	l	-6.9	-6.7	0.8	-4.4	-2.3	0.0	9.3	9.3	8.7	9.9	6.2	8.1
*School2	h	l	0.4	1.3	-3.1	-3.4	-0.4	3.4	14.0	15.5	15.0	14.8	21.2	17.6
Office5	h	l	-1.8	-2.1	-0.4	-0.6	-3.0	-0.2	3.1	3.2	3.1	2.9	3.6	3.0
Office/LM1	h	l	-1.3	-1.0	1.0	-0.3	-0.5	-1.2	4.8	4.7	4.9	4.8	4.3	5.1
Office/LM4	h	l	-3.6	-3.6	-0.6	-1.4	-5.2	-1.0	4.6	4.5	3.8	3.9	5.7	3.2
Office/LM6	h	l	-1.2	-1.6	2.2	0.5	8.7	0.3	6.7	6.9	9.3	8.0	11.5	10.7
Office/LM8	h	l	-0.6	-0.7	0.0	-0.5	7.1	-1.3	4.3	4.3	4.0	4.4	9.1	4.5
Office/LM9	h	l	-2.4	-2.5	0.6	-0.3	-15.2	0.6	6.1	5.9	7.9	6.8	13.9	9.8
Office1	l	h	-2.5	-2.5	0.2	-0.2	1.4	0.0	4.7	4.6	4.8	4.6	3.0	3.6
Office4	l	h	-2.0	-2.1	-1.1	-1.3	0.2	-1.2	3.7	3.8	3.7	3.5	2.9	4.0
Office/Lab2	l	h	0.3	0.7	0.7	0.8	0.0	0.5	4.1	4.0	4.3	4.0	4.3	4.3
Retail1	l	h	1.0	1.2	-0.4	0.9	1.3	1.3	4.5	4.4	4.3	4.6	3.5	4.6
Retail2	l	h	-1.1	-1.2	-0.6	-0.3	0.5	-0.1	2.3	2.4	2.4	2.3	2.0	1.7
Office/DC1	l	h	1.7	1.4	0.9	0.6	3.5	0.6	2.2	2.0	1.9	2.0	4.0	2.6
Supermarket	l	h	-1.2	-1.5	-0.4	-0.4	0.1	-0.2	2.4	2.2	2.1	2.0	1.6	1.6
Office/LM5	l	h	-1.1	-1.5	0.8	0.1	0.2	-0.1	2.3	2.4	2.5	2.3	1.5	1.9
Office/DC2	l	h	-4.4	-5.5	-2.1	-2.2	0.0	-3.4	5.2	6.3	4.6	4.3	3.8	4.0
Office/DC3	l	h	-3.7	-4.8	-1.1	-2.4	-1.0	-2.0	4.8	5.1	4.4	4.3	2.8	3.6
Retail3	l	h	-0.7	-1.1	-0.1	0.0	-0.3	0.4	1.5	1.6	1.5	1.5	1.8	1.7
Retail5	l	h	-2.1	-2.1	-0.4	-1.0	-0.5	0.3	3.9	3.9	3.6	3.7	2.1	2.7
Office/Lab1	l	l	-2.5	-1.9	1.8	-0.2	1.2	0.2	4.1	4.0	3.7	3.7	3.1	4.0
Office/LM2	l	l	0.9	-0.6	0.2	0.7	1.1	-1.1	4.3	4.2	4.6	4.5	4.6	4.5
Office/LM3	l	l	-1.4	-1.0	0.9	-1.0	1.2	-0.4	4.0	4.0	4.4	4.2	4.2	5.0
Bakery	l	l	1.0	1.6	-0.1	0.3	1.2	1.2	3.8	3.7	4.3	3.8	4.6	4.6
			blue/white = best performance						grey/black = worst performance					

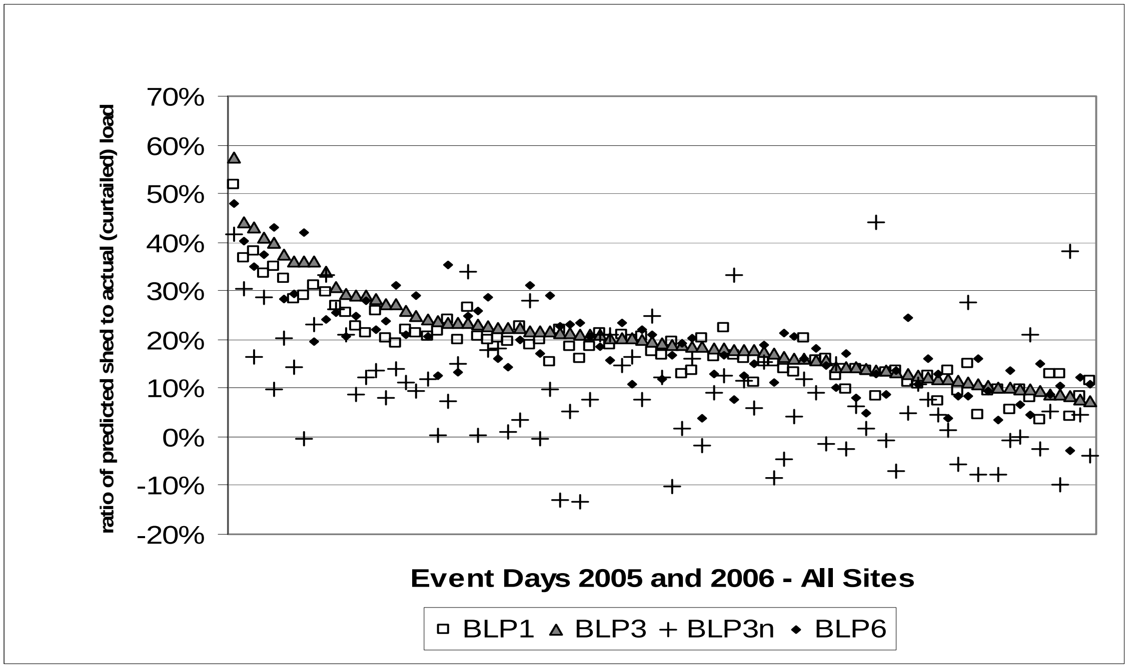
Table 6 is similar to Table 5, except that the error metrics are derived for the sample of event-period average hourly load differences E(d). We have also removed the BLP3n column from this table. The BLP3n model is clearly the least accurate, and by removing it we can get a sense of which of the BLP models is best/worst when all the models include the morning adjustment factor. For the bias measure, the results are similar to Table 5. This is to be expected, as averaging is a linear operation, which is unlikely to strongly affect the median results. The BLP5 model (seasonal load-temperature) tends to be the most biased in our sample of buildings. In the accuracy metric, no model stands out as clearly worse or better than the others. It is interesting to note that the BLP5 model is frequently both the best and the worst. The building load categorizations are reasonably good at predicting performance, with BLP5 performing poorly for "h-l" buildings (high load variability and low weather sensitivity) and well for "l-h" (low variability and high weather sensitivity facilities). The "h-h" and "l-l" sample sizes are small, so one should be careful in drawing conclusions from these data. They do suggest that, as noted above, for buildings in the “l-l” category all models perform reasonably well. For the “h-h” category, the BLP6 model (load temperature based on 10 days of data) consistently avoids bias. It is not clear from this data if explicit weather models out-perform averaging models in this category.

5.4. Event Day Shed Load Estimates

ISOs or utilities with DR programs use BLP models to estimate the customer load reduction achieved from changes to building operation during DR events. The reduction is defined as the estimated baseline value minus the actual (presumably curtailed) value. For this analysis, we have used models to predict electric loads on DR event days for sites that showed some significant demand reductions (these are itemized in Piette et al 2007 and Piette et al 2005). Figure 5 shows the estimated load reductions for each site and event day in the data set. For clarity, only a few representative BLP models are shown. We include three models: BLP3n represents current practice in California’s Demand Bidding program, BLP6 is an example of an explicit weather model, and BLP3 is the preferred model for most of the facilities in our sample, which includes a representative day approach with a same-day morning adjustment.

Load shed estimates are defined as the difference between the estimated average event period hourly load and the measured (curtailed) event-period average hourly load. The results are expressed in percentage terms (i.e., estimated shed load during an event divided by the actual average hourly load). The data in Fig. 5 are sorted on the value of the predicted shed for the BLP3 model (highest 3 of 10 with morning adjustment). We exclude sites for which *no* BLP model predicts a shed of greater than 10%. Note that in some cases the BLP model baseline values for a site are lower than the actual load; these negative values are included in Figure 5. This leads to about 85 building-event day records in the data set. From Figure 5, it is clear that the BLP3n model (no morning adjustment) generally predicts lower values for the sheds than BLP3, i.e. the morning adjustment raises the value of the predicted baseline and hence of the load reduction. The load-temperature (BLP6) model results are scattered around the line defined by the BLP3 results.

Figure 5-5: Predictions of the shed load for event days in California 2005 and 2006



It is also useful to compare the aggregate load reductions for our sample of buildings predicted by the different baseline load profile models. The aggregate is defined as the total over all buildings participating in the DR event on a given day. Figure 6 shows the estimated total load reduction from buildings that participated in DR events in June and July 2006. Eight to ten sites participated in these eleven events as part of PG&E's critical peak pricing tariff; events covered six hours (noon - 6 pm). Not all sites participated in every event, but Figure 6 shows the sum of the participant load reductions for the facilities listed in Tables 5 and 6. The average of the maximum hourly outside dry bulb temperatures for each site is also shown; average peak temperatures ranged from the mid-80's °F to about 100 °F.

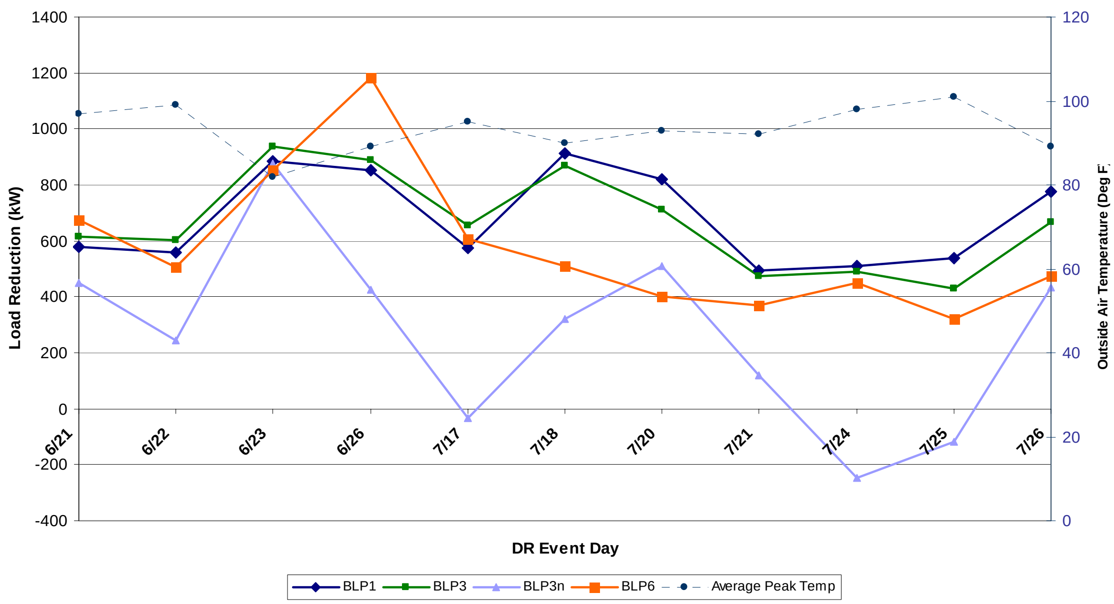
Our analysis of this sample of buildings that actually reduced load during DR events suggests the following key results.

- First, for each DR event, the BLP3n model (highest 3 of previous 10 days with no morning adjustment) estimates the lowest level of demand response and actually shows a net negative response in 3 of the 11 events.
- Second, the negative load reductions with the BLP3n model often occur on the hottest days.

The lowest negative aggregated load reduction took place on July 24th during a severe heat wave in California where DR events were called for several days during a second week of record high temperatures. While there may be some "participant fatigue" in the load reductions from these 8 to 10 sites on those hot days, the other three baseline models show 400 to 500 kW of reduction. Interviews with the facility managers at these sites indicated that they continued to implement their DR strategies during this heat wave but their load reductions were not revealed by the existing BLP approach used in the critical peak pricing tariff (i.e. BLP3n approach).

This also illustrates a problem that occurs with all averaging methods during multi-day events. Because event days are excluded from the set of admissible days, an averaging method will calculate the same unadjusted baseline for every event day if there are events on consecutive days. The adjustment factors will differ on each day during the event, but because of alterations to the building operation induced by the event, the morning loads used to calculate the adjustment may no longer be representative of the normal correlation of that building's load with that day's weather. Explicit weather models do not have this problem.

Figure 6: Aggregate estimated load reduction by baseline model



- Third, in general the three models that include a morning adjustment (BLP1, BLP3, and BLP6) show load reductions that average at three to five times larger than the CBL model without the morning adjustment (BLP3n).

This result illustrates the problem of using the BLP3n model for commercial buildings during heat waves when the previous days in the baseline were not as hot as the DR event days, and the morning adjustment factor is no longer representative of typical load-weather correlations.

- Fourth, the results from this research on baseline models to assess DR load impacts in commercial buildings stands in sharp contrast to previous work in California by Buege et al (2006). Buege et al found that the 3 in 10 day baseline model with no morning adjustment (BLP3n) produced the highest estimates of customer baseline and the largest savings estimates for the California demand bidding and CPP tariffs. However, the load impacts from the sample of sites that Buege et al evaluated were dominated by a relatively small number of large industrial customers.[10] In contrast, our results suggest that for weather-sensitive commercial/institutional customers in California, the 3 in 10 day baseline model (BLP3n) produces estimates of the customer's baseline that are biased on the low side, which results in estimated load curtailments that are biased on the low side.

6. Conclusions and Suggestions for Further Work

We believe that the methods used in this study provide a statistically sound approach to evaluating the performance of different BLP models for a building or set of buildings, provided sufficient historical data are available. The results indicate in general that:

1. The BLP3n model currently used by California utilities to estimate load reductions in several of their DR programs could be improved substantially if a morning adjustment factor were applied for commercial and institutional buildings.[11]
2. Applying a morning adjustment factor significantly reduces the bias and improves the accuracy of *all* BLP models examined in our sample of buildings.
3. Characterization of building loads by variability and weather sensitivity is a useful screening indicator that can be used to predict which types of BLP models will perform well. We believe that DR program administrators can use the analytic techniques described in this study to characterize and possibly screen participating customer's loads.
4. In our sample, BLP models that incorporate temperature (e.g. explicit weather models) improve accuracy of the estimated baseline loads, and in cases where it doesn't improve the accuracy it has relatively little impact.
5. Explicit weather models (in particular, the 10-day version BLP6) are the only model type that consistently avoids bias in the predicted loads in our sample of buildings.

---

[10] We believe that large industrial customers account for most of the load impacts in the California Demand Bidding and CPP evaluation study conducted by Buege et al, because of their load shapes (i.e., high nighttime loads) [Buege et al 2006]. Industrial facilities may have nighttime electric loads that are twenty to thirty percent lower, or even greater than daytime peak loads. By contrast, the primarily commercial and institutional sector participants in our sample of California buildings all have night time loads that are typically a factor of three lower than peak hour electric loads.

[11] DR baselines are used to estimate load reductions in both the California Demand Bidding program and CPP tariff for resource planning and B/C screening analysis. The Demand Bidding Program also uses a BPL method to determine payments to customers for their load reductions as part of a settlement process.

6. For customer accounts with highly variable loads, we found that no BLP model produced satisfactory results, although averaging methods perform best in accuracy (but not bias). These types of customers are difficult to characterize with standard baseline load profile models that rely on historic loads and weather data. Because the DR potential and performance in actual DR events for facilities with more variable loads is harder to predict, measure, and evaluate, it may make more sense to direct these facilities to enroll in DR programs with rules that require customers to reduce load to a firm service level or guaranteed load drop (e.g. interruptible/curtailable tariffs).
7. For buildings with low load variability all BLP models perform reasonably well in accuracy.
8. Similarly, customers that are highly weather sensitive, should be given the option of using BLP models that explicitly incorporate temperature in assessing their performance during DR events.
9. Many DR programs apply similar DR BLP methods to both commercial and industrial sector (C&I). The results of our study when combined with results of other recent studies (Quantum 2004 and 2006, Buege et al., 2006) suggests that DR program administrators should have flexibility and multiple options for suggesting the most appropriate BLP method for specific types of customers. Key load characteristics to be considered in BLP methods are weather-sensitivity (which is an issue for many commercial and institution buildings but not common in industrial process loads) and variability of loads.

Suggestions for Future Work

From our detailed examination of both the data and the model predictions, we can also suggest some new approaches that are reasonably straightforward and could improve the utility of a given model. Below is a list of specific suggestions for future work.

1. For many sites the seasonal load-temperature model (BLP5) is either the best or worst performer. From the data, it is fairly clear that a linear load-temperature relationship is crude, and simply changing to a quadratic fitting function may substantially improve the model performance.
2. Application of the methods developed here to a larger sample of buildings, covering a wider geographical area, would be very useful in determining the robustness of the results. The calculation methodologies are fully automated, so larger data sets could be handled without significant additional effort.
3. The weather data provided by NOAA and CIMIS may occasionally contain erroneous values, which produce outliers (large errors) in the model predictions. We have not screened for weather data errors in our analysis, as we wanted to evaluate the methods as they are currently used by DR program administrators in California. To screen for consistency in the weather data is technically straightforward, but burdensome if each program participant has to do it on their

own. Given the large number of state agencies that use weather data, and the extensive infrastructure that already exists for collecting and maintaining it, it should be feasible to provide DR program participants with access to weather information that is periodically screened and updated. This would greatly facilitate the use of explicit weather models.

4. Some buildings have predictable but non-standard schedules (for example, closed Mondays, closed in summer etc.) Including this scheduling information in the selection of the admissible set would reduce the variability in the load data, and therefore improve BLP model performance. Technically, because the admissible day selection process used by utilities and ISOs typically screens for weekends etc., it should be simple to add additional building-specific criteria.

5. Our data set of proxy events is similar to but not the same as the actual event day set in California, and in particular contains milder weather days than is typical for real events. It may also be useful to investigate whether using a more restricted proxy event set (e.g., the highest 10% of days in temperature instead of the highest 25%) would significantly impact the results.[12]

References

[Buege 2006] Buege, A, M. Rufo, M. Ozog, D. Violette, and S. McNicoll 2006. "Prepare for Impact: Measuring Large C/I Customer Response to DR Programs," 2006 ACEEE Summer Study on Energy Efficiency in Buildings, Monterey, CA August.

[CPUC 2007] California Public Utilities Commission 2007. "Order Instituting Rulemaking Regarding Policies and Protocols for Demand Response Load Impacts Estimates, Cost-Effectiveness Methodologies, Megawatt Goals and Alignment with California Independent System Operator Market Design Protocols," OIR-07-01-041, January.

[EnerNOC 2006] Kozikowski, D., A. Breidenbaugh and M. Potter 2006. *The Demand Response Baseline, v.1.75*. EnerNOC OPS Publication.

[FERC 2006] FERC Staff Report 2006. *Assessment of Demand Response and Advanced Metering*, Docket Number AD-06-2-000, Aug.

[KEMA 2003] Goldberg M.L and G. Kennedy Agnew 2003. *Protocol Development for Demand-Response calculations: Findings and Recommendations*. Prepared for the California Energy Commission by KEMA-Xenergy. CEC 400-02-017F.

[12] If sufficient time periods of data are available, this is easy to implement. For example, if 2007 data were added to our data set, the top 10% of days with high temperatures would provide a large enough sample to do this type of analysis.

[KEMA 2007] Goldberg M.L., *Customer Baselines for Demand Response Programs.* Presentation to Midwest Demand Response Initiative DR Program Design Sub-group, July 9.

[Piette et al, 2007] Piette, M.A. Piette, D. S. Watson, N. Motegi and S. Kilicotte 2007. *Automated Critical Peak Pricing Field Tests: 2006 Program Description and Results.* Draft Report.

[Piette et al, 2005] Piette, M. A., D. S. Watson, N. Motegi and N. Bourassa 2005. *Findings from the 2004 Fully Automated Demand Response Tests in Large Facilities*. LBNL Report Number 58178.

[PJM 2007] Amended and Restated Operating Agreement of PJM Interconnection LLC, effective June 12 2007, Section 3.3.A.

[Quantum 2004] *Working Group 2 Demand Response Program Evaluation – Program Year 2004 Final Report*. Prepared for the Working Group 2 Measurement and Evaluation Committee, by Quantum Consulting Inc. and Summit Blue Consulting, LLC, 2004.

[Quantum 2006] *Evaluation of 2005 Statewide Large Nonresidential Day-ahead and Reliability Demand Response Programs*. Prepared for Southern California Edison and the Working Group 2 Measurement and Evaluation Committee, by Quantum Consulting Inc. and Summit Blue Consulting, LLC, 2006.