Title
Incorporating residential AC load control into ancillary service markets: Measurement and settlement

Authors
Bode, Josh L.
Sullivan, Michael J.
Berghman, Dries
et al.

Publication Date
2013-05-01

LBNL-XXXX

ERNEST ORLANDO LAWRENCE
BERKELEY NATIONAL LABORATORY

Incorporating residential AC load control into ancillary service markets: Measurement and settlement

Josh L. Bode[1]
Michael J. Sullivan[1]
Dries Berghman[1]
Joseph H. Eto[2]

[1]The FSC Group
[2]Lawrence Berkeley National Laboratory

Environmental Energy Technologies Division

May 2013

This report is a post-print version of an article published in *Energy Policy*, Volume 56, May 2013, Pages 175–185. http://dx.doi.org/10.1016/j.enpol.2012.12.024

The work described in this report was coordinated by the Consortium for Electric Reliability Technology Solutions (CERTS) and funded by the California Energy Commission (CEC), Public Interest Research (PIER) Program, under Work for Others Contract No. 500-05-001 and by the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

LBNL-XXXX

Incorporating residential AC load control into ancillary service markets: Measurement and settlement

Principal Authors:

Josh L. Bode[1]
Michael J. Sullivan[1]
Dries Berghman[1]
Joseph H. Eto[2]

[1]The FSC Group
[2]Lawrence Berkeley National Laboratory

May 2013

This report is a post-print version of an article published in *Energy Policy*, Volume 56, May 2013, Pages 175–185. http://dx.doi.org/10.1016/j.enpol.2012.12.024

The work described in this report was coordinated by the Consortium for Electric Reliability Technology Solutions (CERTS) and funded by the California Energy Commission (CEC), Public Interest Research (PIER) Program, under Work for Others Contract No. 500-05-001 and by the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Acknowledgements

This project was sponsored by the California Energy Commission Public Interest Energy Research (PIER) Program under contract #500-05-001. The project team gratefully acknowledges the support of Ron Hofmann, PIER Advisor, and Kristy Chew, PIER Contract Manager.

We also thank John Hernandez, Michael Alexander, Ulric Kwan and Greg Mandelman from PG&E for their support throughout this project and for permission to use the data for the analysis. In addition, we extend thanks to Steve Braithwait and Johanna Mathieu for helpful feedback on early drafts of this study.

Abstract

Many pre-existing air conditioner load control programs can provide valuable operational flexibility but have not been incorporated into electricity ancillary service markets or grid operations. Multiple demonstrations have shown that residential air conditioner (AC) response can deliver resources quickly and can provide contingency reserves. A key policy hurdle to be overcome before AC load control can be fully incorporated into markets is how to balance the accuracy, cost, and complexity of methods available for the settlement of load curtailment. Overcoming this hurdle requires a means for assessing the accuracy of shorter-term AC load control demand reduction estimation approaches in an unbiased manner. This paper applies such a method to compare the accuracy of approaches varying in cost and complexity – including regression analysis, load matching and control group approaches – using feeder data, household data and AC end-use data. We recommend a practical approach for settlement, relying on an annually updated set of tables, with pre-calculated reduction estimates. These tables allow users to look up the demand reduction per device based on daily maximum temperature, geographic region and hour of day, simplifying settlement and providing a solution to the policy problem presented in this paper.

Keywords: Measurement; Load management; Ancillary services

1. Introduction

Adding variable generation such as wind and solar to an existing power system increases the need for flexible resources to respond to system changes and uncertainties, including wind ramps, demand ramps and forced transmission or generation outages. Flexible resources are defined by the speed in which they can increase (ramp up) or decrease (ramp down) production. Traditionally, much of the system flexibility required to maintain reliability is obtained from peaking generation units. However, using generators to provide operational flexibility can impose significant costs and lead to extra wear and tear on the generating equipment. In 2010, the North American Electric Reliability Council (NERC), which sets reliability standards for operation of the electric grid, investigated emerging flexible resources, including demand response, battery storage and electric vehicles (North American Electric Reliability Council (NERC), 2010). It identified residential air conditioner (AC) response as an existing technology that is particularly valuable because it is typically available during peak load times when energy and ancillary services are expensive and when generation is typically in short supply. The study recommended adjusting regional and federal reliability standards that might limit the deployment of these resources, developing operation infrastructure, and modifying market rules or non-market rules/procedures that limit technically capable resources from providing flexibility. Several markets in North America, including those in Texas, New York, Ontario, and California, are currently in the process for developing standards to incorporate demand response and other emerging flexible resources.

Residential AC response is particularly well suited for providing operators flexibility and, more specifically, contingency response, which requires fast deployments to stabilize the grid, but are used infrequently (<30 times per year) and for short periods (usually less than 10 min). Recent advances in communications technology allow for more precise control of AC units and operator visibility. Residential AC response is a disseminate resource that is not subject to transmission constraints and can be used to deliver specific incremental load reductions at specific locations. In addition, as we detail below, the operational capability of residential AC load control program and their ability to be used for grid operation has been tested extensively in recent years. It is also a large pre-existing resource that can be incorporated into grid operations through adjustments in reliability standards, market rules, and load control dispatch practices. Based on the Federal Energy Regulatory Commission 2010 Demand Response (DR) survey, there are over 4.8 million households and over 200,000 businesses currently enrolled in AC load control programs (Federal Regulatory Energy Commission, 2011). Historically, these resources have been used for emergency operations and to offset the need to build additional peak generation, but they can also provide operators' significant flexibility if incorporated into ancillary service electricity markets.

To date, there have been several studies that have tested the potential of controlling residential AC loads in order to provide flexible operating reserves and assessed the ability of integrating control of AC loads into operations. The conceptual framework and the policy reasons for using AC as spinning reserves were detailed in a series of reports by the Oakridge and Lawrence Berkeley National Laboratories (Eto et al., 2010 and Kueck et al., 2001). In addition, Lawrence Berkeley National Laboratory, Pacific Gas and Electric (PG&E) and Southern California Edison (SCE) sponsored a series of demonstration studies testing the ability to use AC load control to provide operating reserves (Kirby, 2003, Eto et al., 2007, Eto et al., 2009, Sullivan et al., 2009 and Gifford et al., 2010).

Combined, the demonstration studies showed that:

- Residential AC load control reduces demand quickly. AC units begin to noticeably shut down or cycle compressors within 60 s of when the load control signal is sent out and reach 80% of capacity within 3 min.
- The effect of short-duration residential AC curtailments on customer comfort is negligible.
- AC load drops can be observed on near real time basis using samples.
- The demand reductions observed in the samples were also observed in the distribution feeder circuits.

A key policy hurdle to be overcome before AC load control can be fully incorporated into markets is how to quickly and accurately measure shorter-term (e.g., ten's of minutes to a couple of hours) demand reductions from residential AC curtailments for settlement. This is an important policy question that will affect the ability of residential AC load control programs to participate in electric markets. The challenge is that measurements for settlement and operations need to be conducted in real time or on a monthly basis—much faster than traditional program evaluations, which are conducted on an annual basis. In addition, measuring demand reductions, sometimes referred to as “negawatts,” is an entirely different task than measuring power production. While power production is metered and thus is measured directly, demand reductions cannot be metered. They must be estimated by indirect approaches. In principle, the reduction is simply the difference between electricity use with and without the AC curtailment. However, it is not possible to directly observe or meter what electricity use would have been in the absence of curtailment. Instead, the electricity that would have been used in the absence of the curtailment – the counterfactual, sometimes referred to as the baseline – must be estimated. In doing so, it is important to systematically eliminate or control for alternative explanations for the change in electricity consumption.

Much of the existing research on estimating demand reductions for settlement has focused on large industrial and commercial customers because electricity markets operated by Independent System Operators (ISO) have allowed these customers to participate in energy and capacity markets for well over a decade. The accuracy of many day-matching baselines for settlement of large commercial and industrial customers has been studied on several occasions. In 2003, KEMA compared the accuracy of 6 settlement baselines in 2003 using 646 accounts from multiple regions across the U.S (Coughlin et al., 2008). In 2004, Quantum Consulting and Summit Blue Consulting (2004) estimated the accuracy of 4 settlement baselines using data from 450 accounts in California, none of which were enrolled in DR programs. In 2008, Lawrence Berkeley National Laboratory (2008), (Kema, 2003) compared accuracy of 7 alternate settlement baselines using data from 32 sites in California. It was the first study to assess accuracy by comparing actual and predicted baseline load for demand response program participants. All prior studies had drawn conclusions based on results from non-participants or comparisons of one estimate to another. Since then, assessments of baseline accuracy have relied on the use of proxy event days because this allows a comparison of estimated values to actual known values.

Several additional studies have been conducted since, all of which focused on large commercial and industrial customers (Braithwait and Armstrong, 2009, Braithwait and Armstrong, 2010, KEMA, Inc., 2011 and Bode et al., 2010).

This paper presents a method for assessing the accuracy of shorter-term residential AC load control demand reduction estimation approaches and compares the accuracy of various alternatives for measuring AC reductions using three data sources: feeder data, household data and AC end-use data. The method relies on inserting pre-determined values measured in prior studies into naturally occurring electricity use. It then measures how well each approach estimates (or "predicts") the known demand reductions under different conditions. In total, we evaluate 10 different demand reduction estimation approaches using feeder data, household data and end-use AC data. The approaches tested include both within- and between-subject estimators. Within-subject estimators use customer's electricity use patterns during days when AC units are not curtailed to estimate AC load absent curtailment operations during actual event days, while between-subject estimators rely on an external control group of AC units that is not curtailed to provide information about electricity use absent curtailment.

While highly accurate results are desirable, there is often a tradeoff between simplicity and incremental accuracy. In order to help gauge the benefit of more complex and costly approaches, each of the estimation approaches are compared with one of the simplest and least technical approaches—a set of tables with pre-calculated load reduction estimates. These tables are based on annual evaluations and allow users to look up the demand reduction per device based on the daily maximum temperature, geographic region and hour of day. They facilitate quick settlement when resources are dispatched and provide operators a quick estimate of the DR resources available for operations.

The study presented in this article differs from the studies cited above because it focuses explicitly on the policy problem of how to measure demand reductions from residential AC response. In addition, it compares a wider range of approaches for estimating demand reductions, including day-matching baselines, weather-matching baselines, regression models, and approaches that rely on control groups. Finally, it also assesses how the accuracy of the demand reduction estimation approaches varies as a function of the data source employed. Decisions about whether to rely on end-use, household, or feeder data directly affect the ability to accurately measure demand reductions. This is because the data source affects the amount of background noise from which the signal – the demand reduction – must be identified.

The remainder of this paper is structured as follows. Section 2 documents the methodology, including the data sources used, estimation approaches tested, and metrics used to assess accuracy. Section 3 presents the results. Section 4 concludes by discussing the implications of the findings.

2. Methodology

To assess the accuracy of different approaches for estimating AC demand reductions, we introduced pre-determined AC load curtailments on actual feeder, household, and AC end-use data from customers enrolled in Pacific Gas and Electric's SmartAC program (George et al., 2012, KEMA, 2009). That is, while the estimates of AC demand have been predetermined for purposes of this evaluation, they were developed from a prior analysis of recorded data collected on AC usage patterns and demand reductions by temperature conditions and geographic location. This process is used in order to ensure that the true demand reductions are known in advance. This enables us to determine exactly how well different approaches estimate these known demand reductions and whether or not they exhibit tendencies to over or underestimate them.

A. Simulation framework

Fig. 1 summarizes the general framework used for assessing the accuracy of the demand reduction estimation approaches. To implement the assessment framework, we

1. Calculate the magnitude of controllable AC loads. This step estimates how much of the load from the data source can be controlled for each date and hour. This estimate is based on a sample of unperturbed AC end use data for 547 Pacific Gas and Electric residential customers. The load data was used to create 24 daily load profiles, based on temperature conditions, for each of 3 distinct climate regions.
2. Select proxy curtailment events. In total, 15 curtailment days were randomly selected in 2009 and 2010 from the set of weekdays where daily maximum temperature exceeded 85 °F (29.5 °C). Days with a daily maximum temperature below 85 °F generally have little AC load in California because overnight temperatures cool off substantially compared to more humid regions in North America. In the hottest climate region – the Fresno/Bakersfield region of the California Central Valley – daily maximum temperatures exceeded 90 °F (32 °C) on 76% of summer days. In the second hottest region – the Northern part of the Central Valley near Stockton and Sacramento – daily maximum temperatures exceeded 90 °F on 43% of the days. Curtailment event start times were randomly selected between 12 PM and 10 PM with durations of one hour. One-hour events were simulated because residential smart meter data were only available on an hourly basis. In practice, most contingency reserves operations are much shorter, typically less than 10 min, though contingency reserves must be able to deliver resources for 90 min or up to 2 h, based on market rules.
3. Calculate demand reductions. The demand reductions were estimated based on the variation observed in historical reductions of AC load from 34 curtailment events documented in annual program impact evaluations. The resulting relative (percent) demand reductions incorporate the effect of weather, plus a random variation component that reflects the variation in the historical evaluation results. The percent reductions were then multiplied by controllable AC loads to produce a simulated (estimated) demand reduction. With this process, the demand reductions for each curtailment event are known, making it possible to test how accurately each of the different settlement alternatives measures the load drop.

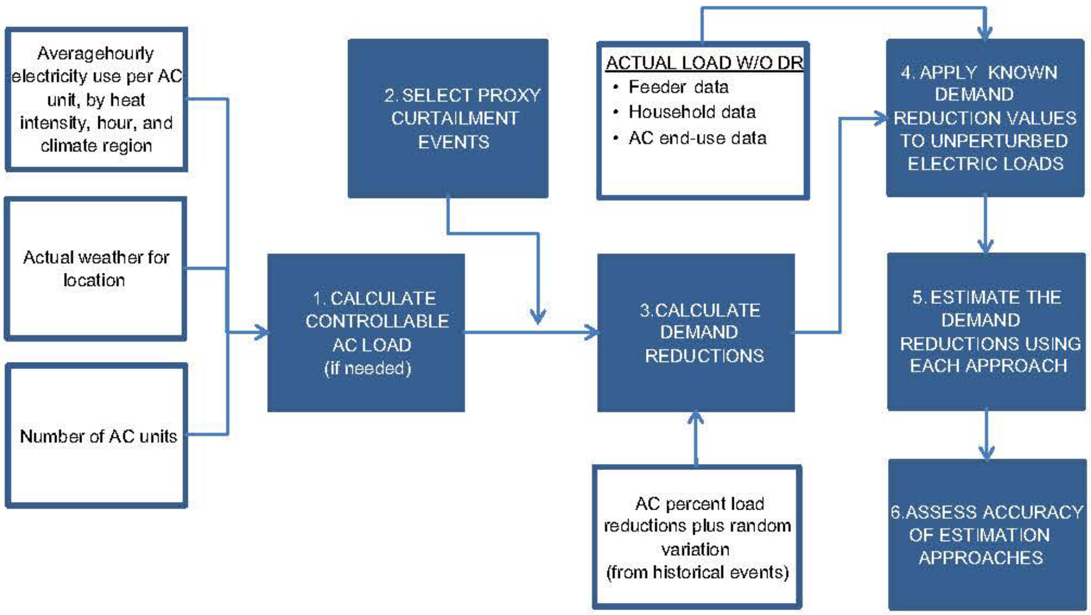

Fig. 1. Methodological framework.

4. Apply demand reductions to unperturbed loads. During each of the proxy curtailment event periods, simulated demand reductions were subtracted from the unperturbed loads. In other words, we knew the actual demand with and without the simulated curtailments, as well as the magnitude of the demand reductions. For AC end-use data, the unperturbed load consisted of the sample of 547 AC units with directly metered end use data. The simulated percent demand reductions were applied directly to the actual loads. For household-level data, the unperturbed load consisted of smart meter data from 6000 households located in 204 randomly sampled feeders. For each feeder, 100 households participating in PG&E's AC load control program were randomly selected. For feeders with less than 100 participating households, we included data from all participants. The household data was aggregated to the feeder level prior to applying the demand reductions. Unperturbed feeder load was based on the same 204 randomly selected feeders; however, hourly data was only available for 85 of those feeders. By applying the simulated demand reduction to unperturbed electric loads, we not only knew the actual curtailments, but were also able to realistically simulate the background noise and effect size.
5. Estimate the demand reductions using each data source and 10 estimation approaches. The demand reductions were calculated using feeder, household, and AC end use data. The estimation approaches tested included day- and weather-matching methods, regressions, and approaches that relied on control groups.
6. Assess the accuracy of each of the estimation approaches. For each of the curtailment events, we knew the true load patterns without curtailment and the true demand reductions. As a result, we were able to assess the accuracy of each estimation approach. To standardize the comparison, we used metrics designed to assess if the estimation approaches systematically over or under-reported demand reductions (bias) and metrics that summarized how close the estimates were to the true demand reductions (goodness-of-fit).

B. Demand reduction estimation approaches

Table 1 summarizes the estimation approaches we evaluated and provides greater detail for the fifth step in the simulation framework described above. A total of 10 different demand reduction estimation approaches were applied to feeder, whole household, and individual customer and aggregated AC end-use data. Individual AC load is the AC load for a single customer, while aggregated AC load is the combined AC load for multiple customers. With AC load, the order of aggregation and calculation can lead to different results. Calculating baselines for individual AC units and then aggregating program results produce a different answer than if the data is aggregated before calculating the baselines. This occurs because individual AC units tend to be either off or on while aggregated AC loads show a more continuous pattern.

The least technical approach – a set of tables that provides estimates of the load curtailment based on daily maximum temperature, region, and hour of day – is used as a benchmark to assess the extent to which more complex demand reduction estimation approaches improve accuracy. The approaches can be classified into 2 broad categories: within- and between-subject estimators.

Within-subject estimators use customer's electricity use patterns during days when AC units are not curtailed to estimate AC load absent curtailment operations during actual event days. They include demand reduction calculation methods such as individual customer regressions and day- and weather-matching baselines. They work because the AC curtailment is introduced on some days and not on others, making it possible to observe behavior with and without the load control in effect.

Within-subject approaches can be less reliable when curtailment events lack comparable non-event periods. For example, if an AC load control program is utilized on all of the hottest days, much like AC programs have normally been operated, there may not be any similarly hot days left over for comparison. However, contingency reserve operations are typically triggered by random generation or transmission outages. They tend to be short in duration and do not always affect the same hours. As a result, there are typically a large number of similar non-curtailment periods.

Between-subject estimators rely on an external control group of AC units that are not curtailed to provide information about AC units that were curtailed and would have used electricity if they were not instructed to shed load. We considered two simple options that rely on random assignment to load control operations: a simple comparison of means and a difference-in-differences calculation.

1) Impact estimate tables.

Impact estimate tables are the least technical demand reduction estimation approach and are typically constructed at the AC unit level. They are essentially a set of tables with pre-calculated load reduction estimates based on annual impact evaluations of historical curtailment. They allow the user to look up estimates of the reduction per AC unit based on the hour of day, temperature condition category, and climate region. These estimates

Table 1
Demand reduction estimation methods tested.

Type of estimator	Method	No.	Calculation	Data source				Summary description
				Individual AC	Aggregated AC	Feeder	House data	
Within- subject estimators	Day-matching baseline	1	10-in-10 with a 20% in-day adjustment cap	X	X	X	X	A subset of weekdays when units were not cycled is identified and the average is calculated for each hour to produce a baseline. The days are selected from the 10 non-event weekdays closest to the load curtailment day. The baseline is calibrated or adjusted using information about demand patterns in the hours preceding the curtailment (in-day adjustment). Demand reductions are calculated as the difference between the adjusted baseline and metered load. The process for weather-matching baselines is similar except that the baseline load profile is selected from non-event days with similar daily maximum temperatures and then calibrated with an in-day adjustment.
		2	10-in-10 without an in-day adjustment cap	X	X	X	X	
		3	Top 3 in 10 without an in-day adjustment cap	X	X	X	X	
	Weather-matching baseline	4	Profile selected based on daily maximum temperature without an in-day adjustment cap	X	X	X	X	
	Regression	5	Treatment variables and no day or hourly lags or leads	X	X	X	X	Regression analysis quantifies how different, observable factors such as weather, hour of day, day of week, location, and load curtailments affect AC electricity use patterns. With regressions, the impacts are usually directly estimated through the model parameters that reflect the effect of load control operations known as treatment variables. With treatment variables, the impacts are the difference between the regression estimates of AC use with and without load control. Regression models can be informed by electricity use patterns in the day prior (day lags) and in the hours before or after an event (lags or leads).
		6	Treatment variables with a day lag	X	X	X	X	
		7	Treatment variables with hourly lags and leads	X	X	X	X	
		8	No treatment variables but use of hourly lags and leads	X	X	X	X	
Between- subject estimators	Random assignment of load control operations	9	Comparison of means		X		X	AC load control program participants are randomly assigned to groups that do and do not have their AC unit instructed to reduce or shed AC load. Any differences between the two groups are random, not systematic. The group that is not subject to the load curtailment is typically referred to as the control group and provides information about normal electricity use patterns in the absence of AC curtailment. Impacts are calculated as the difference in average demand between the group that is and is not dispatched (comparison of means). The estimate can be refined by assessing inherent differences between the two groups in hot non-event days and netting them out of the demand reduction calculation (difference-in-differences).
		10	Difference-in-differences		X		X	
Pre-calculated load reduction estimate tables		11	Multiply the number of AC units in each geographic location by the corresponding estimate of demand reductions per AC unit for the corresponding area, hour of day, and temperature bin.		X	X		Empirical data on AC end use (absent curtailments) is used to produce estimates of average load by geographic location, hour of day, and temperate bins (based on cooling degree days). The historical evaluation data is used to estimate how the percent load reductions vary with temperature conditions and geographic locations. The percent load reductions are also estimated by geographic location, hour of day, and temperate bins. The percent load reductions are then applied to estimates of AC end-use (absent curtailment) to produce tables. The tables contain estimates of the per AC unit demand reduction for specific geographic locations, hour of day, and 10 temperature profiles.

per AC unit can be multiplied by the number of AC units dispatched in each climate region and aggregated to obtain an estimate of the aggregate demand reduction. While not complex, the approach is practical and of low cost. It serves as a useful baseline for assessing how much value is added by using more complex demand reduction estimation approaches or requiring placing more extensive data requirements for settlement.

The accuracy of the impact estimate tables was tested using a sub-sampling approach. Approximately one eighth of the data was sampled and used to develop the impact estimate tables. To assess accuracy, we then compared the estimates from the table to the known impacts for population, which were artificially introduced. The process is repeated 100 times to reflect both sampling and estimation error.

2) **Within subject estimators**

   a. *Day- and weather-matching baseline estimation approaches*

   Day-matching baselines are a widely used technique for developing an estimate of what electricity use would have been in the absence of load control. The approach was developed and tested for large commercial and industrial (C&I) customers and have been used by Independent System Operators (ISOs) for settlement of DR products targeting large C&I customers. This approach relies solely on electricity use patterns when the AC unit is not controlled. A subset of weekdays when units were not cycled in close proximity to the event day is identified. The electricity use in each hour of the identified days is averaged to produce a baseline. While more accurate approaches are available, baselines are useful because they allow settlement to be conducted quickly and are relatively intuitive and easy to understand. They are also relatively accurate for commercial and industrial customers, since many of these customers are not particularly weather sensitive; it is unlikely that their load on hot event days will be much higher than on the days used to calculate the baseline.

   Many options exist for calculating day-matching baselines. They are often supplemented with corrections to incorporate information about usage patterns in the hours preceding an event—usually referred to as in-day or same-day adjustments. In-day adjustments are common and typically reduce the error between the unadjusted baseline and actual loads.

   Weather-matching baselines are a variation of day-matching approaches. The main difference is that the comparable days are based on average hourly load patterns during non-event days with similar weather conditions, as defined by temperature bins. These days may or may not be immediately prior to the curtailment event. For example, to produce a baseline for a weekday with a daily maximum temperature between 90 °F (32 °C) and 95 °F (35 °C), the first step would be to identify weekdays with similar temperatures and without AC curtailments. If there were six such days, the electricity use for each time period would be averaged for those six days to produce a baseline. As with the day-matching baseline, the weather-matching baseline can be calibrated or adjusted using actual usage patterns in the hours preceding an event. Given the weather sensitivity of AC load, this approach is preferable to using a simple baseline.

b. *Regression analysis estimation approaches*

Regression analysis quantifies how different, observable factors such as weather, hour of day, day of week, location, and cycling affect AC electricity use patterns. With regressions, the impacts are directly estimated through the regression model parameters. In other words, the impacts are the difference between the regression estimates of AC use with and without load control.

The analysis consists of applying regression models separately at the unit of analysis. The regression specification is common over all units but estimated coefficients vary for each unit. The variables in the regression specifications model time-based and weather-based impacts. The fact that each feeder has its own specification automatically accounts for variables that are relatively constant for each unit, such as geographic location, mix of load control switches versus smart thermostats, and the strength of the communication network. Because the coefficients are specific to the unit, they can better explain the variation in weather sensitivity and load patterns.

Regression analysis estimation approaches work because AC load control naturally produces an alternating or repeated treatment design. The primary intervention – AC load control – is present on some days and not on others, making it possible to observe AC use with and without cycling under similar conditions. A repeated introduction and removal of curtailment events allows for an assessment of whether the outcome – electricity consumption – rises or falls with the presence or absence of AC cycling.

We tested four regression analysis estimation approaches; the appendix contains the mathematical expression of the regression models tested.

3) Between-subject estimation approaches (control groups)

Another way to estimate demand reductions is by using a control group of customers that does not participate in the event. In essence, the electricity demand patterns by the group that did not participate are used to infer what the usage patterns of the curtailment group would have been in the absence of curtailment.

However, on its own, using a control group does not guarantee more accurate results. To eliminate alternative explanations for differences in electricity use, it is critical that the only systematic difference between the two groups is the fact that one group had their AC units curtailed while the other group did not.

The best way to ensure there are no systematic differences between the two groups is to randomly assign customers to the curtailment and control groups and use large sample sizes. Because of random assignment, on average, both groups can be expected to have similar characteristics such as household size and to experience the same weather, economic conditions, and occupancy patterns. The only systematic difference between the two groups is whether or not they were curtailed.

The demand reductions were calculated in one of two methods:

- A simple comparison of means: with this approach, for each time period, demand reductions are estimated as the difference between the group that did not have their AC loads curtailed and one that did.
- A weather-matched difference-in-differences calculation: this approach is useful when sample sizes are smaller. The demand reduction is calculated as the difference between the two groups, but then adjusted with one additional step. We subtracted out differences between the two groups observed during days without curtailments and similar weather. This nets out differences that are irrelevant and mainly due to sampling variation.

To simulate the effect of sampling accuracy, we (1) randomly selected a sample of customers, (2) randomly assigned half of them to receive the curtailment and half to act as the control group, (3) simulated the impacts for the group assigned curtailment, (4) calculated the demand reduction using the control group, and (5) recorded the degree of error in the estimate. This process was repeated 100 times to reflect the distribution of errors in the estimation approach.

C. Metrics for assessing accuracy

Since AC load control programs aggregate tens and sometimes hundreds of thousands of the AC units, the focus is less on the accuracy of estimates for individual AC units or feeders and more on the overall accuracy of the results for the program and for larger zones in the electric grid.

To standardize the comparison between various estimation approaches, we used metrics designed to assess to what extent the estimates systematically over or under-estimated the known, true demand reductions (bias) as well as metrics that summarized how close the estimates were to the known, true demand reductions (goodness-of-fit). An accurate estimator produces results that are on average unbiased and minimize amount of error for individual periods (i.e., it has a high goodness-of-fit).

In comparing various demand reduction estimation approaches, it is important to understand whether an approach is unbiased on average and accurate for individual curtailment hours. An approach that produces correct measurements on average can perform poorly for individual events. This occurs if the errors cancel each other out.

Table 2
Metrics for bias and goodness of fit.

Type of metric	Metric	Description	Mathematical expression
Bias	Mean Percentage Error (MPE)	The mean percentage error (MPE) indicates the percentage by which the measurement, on average, tends to over or underestimate the true demand reduction.	$MPE = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{y}_i - y_i\right)}{\overline{y}}$
Goodness-of-fit	Mean Absolute Percentage Error (MAPE)	The mean absolute percentage error (MAPE) is a measure of the relative magnitude of errors across event days, regardless of positive or negative direction. It is normalized allowing comparison of results across different data sources.	$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left\lvert\frac{\hat{y}_i - y_i}{y_i}\right\rvert$
	CV(RMSE)	This metric normalizes the RMSE by dividing it by the average of the actual demand reduction.	$CV(RMSE) = \frac{RMSE}{\overline{y}}$

Table 2 summarizes the metrics for bias and goodness-of-fit used to assess the different estimation approaches. It includes a brief description and the corresponding mathematical equations.

3. Results

This section summarizes the accuracy of the estimation approaches tested. In total, we evaluated 10 estimation approaches that relied on either feeder data, household data, or end-use AC data. Each combination of data source and estimation approach is considered as a separate alternative. Results are shown for bias, using the mean percent error (MPE), and for goodness-of-fit, using the mean absolute percent error (MAPE), and normalized root mean square error (CV RMSE). To illustrate, a bias statistic of 5% indicates that the approach tends to overestimate demand reductions by 5%. In contrast, the goodness-of-fit metrics selected indicate the magnitude of the errors for individual curtailment periods, with lower values indicating less error. MPE can be positive or negative, while MAPE and CV RMSE can only be positive.

Table 3 shows results for the within-subjects estimation approaches and compares those results with the impact estimate tables for the average event day. The within-subjects' approaches include three different day-matching baseline methodologies, a weather-baseline methodology, and four regression models. As a reference, the table also shows impact estimate tables, which are the least technical approach available, and serve as a benchmark to test more sophisticated approaches. The impact estimate results shown here are for a sample of 500 customers drawn, 100 times, and show both the median result and a 90% confidence band around that result.

Table 3
Within-subjects methods results.

No.	Result type				Bias (MPE)				Goodness-of-fit (MAPE)				Goodness-of-fit (CV RMSE)			
	Data source				Ind. AC	Agg. AC	Household	Feeder	Ind. AC	Agg. AC	Household	Feeder	Ind. AC	Agg. AC	Household	Feeder
1	Within-subjects	Baseline methods	Day-MATCHING	1	−136%	−105%	−123%	−1810%	141%	111%	120%	2114%	1.82	1.26	1.88	29.84
2				2	94%	14%	−36%	−526%	129%	32%	70%	1663%	2.24	0.50	0.92	20.10
3				3	960%	9%	−11%	−203%	146%	31%	49%	932%	2.72	0.43	0.57	12.27
4			Weather-matching	4	55%	−3%	−6%	−463%	99%	29%	25%	1165%	1.64	0.35	0.40	15.30
5		Percentiles		5	80%	3%	0%	−89%	23%	24%	29%	304%	0.26	0.26	0.25	3.17
6				6	2%	5%	12%	−140%	26%	25%	32%	367%	0.29	0.29	0.27	3.62
7				7	6%	15%	14%	−98%	30%	30%	36%	331%	0.30	0.31	0.28	4.01
8				8	−3%	−2%	−8%	−72%	31%	24%	24%	549%	0.35	0.22	0.30	7.26
11	Impact estimate tables	Percentiles		5%		−6%				33%				0.39		
				50%		0%				36%				0.41		
				95%		7%				40%				0.44		

Feeder data provides the worst results across the board, regardless of the estimation approach employed. Simply put, feeder data includes a lot of irrelevant load variation that dilutes the signal and makes it harder to detect. It includes load variation due to customers that are not enrolled in the AC load curtailment program (including commercial and industrial businesses), as well as end-uses that are not AC related. It is difficult to pinpoint the amount of load being curtailed by an AC program because the program signal is very small, while the noise of other loads is large. For the average feeder, the curtailment events led to an average reduction of 0.2% for the feeder loads. Even for the feeders with the highest penetration of load control devices, the curtailments rarely exceed more than 1% or 2% of the feeder loads. While demand reductions can be observed in feeders with high AC load control penetration on very hot afternoons, they are not a viable option for settlement. Not only does it lead to inaccurate demand reduction estimates, but many utilities such as PG&E cannot readily access sub-hourly data for a large share of their feeders.

Table 3 also shows that baseline approaches are inferior to regression approaches. The day-matching baselines that are typically used for large C&I customers produced the least accurate estimates of residential AC demand reductions. They both exhibited larger bias and more error for individual curtailment periods (goodness-of-fit). This is likely because residential AC loads are far more weather-sensitive than large C&I loads. Weather-matching baselines tend to provide results that are lower in bias and have better goodness-of-fit than day-matching approaches because they better account for residential AC weather sensitivity. They work well with aggregated AC end-use data, less so with household data. The regressions are much better at providing accurate estimates of load curtailments than day- or weather-matching baselines. They produce the most accurate results and perform with both AC end-use and household data. Regression methods 1 and 4 do particularly well. (Regression specifications are included in the Appendix to this article.)

As the table shows, alternatives that rely on AC end-use data tend to do the best job of estimating the true demand reductions. Individual AC loads show a very clear usage pattern – they are either on or off – those patterns are very difficult to predict, as any individual AC unit's load can be rather volatile. Aggregated AC data can be more accurate because it is easier to predict the aggregate behavior of many customers than to accurately predict the individual behavior of one customer.

Even though it provides the most accurate results, collecting large amounts of AC end-use data is an expensive proposition. Generally, data loggers must be installed on individual AC units and retrieved at the end of a study period or have data transmittal capability. Data collection of AC end-use data requires large expenditures in both labor and capital. On the other hand, household-level data is much easier to collect, especially as smart meters become more and more common.

Several evaluations have recently relied on smart meter data from tens or hundreds of thousands of households with very little incremental cost (George et al., 2012 and Hartmann et al., 2012). While household load data is "noisier" than AC end-use data because it includes the load of many other household devices, the AC load is still quite easy to detect, especially on hotter days. This makes it an affordable and very useful data source.

The impact estimate tables provide fairly good results. In terms of bias, they consistently do better than baseline approaches, and the median result only shows bias of 0.1%, which is better than even the regression approaches. Their goodness-of-fit statistics are not quite as good, indicating that while they do a good job of estimating demand reductions for the average event day, there is considerable variation across individual event days. In addition, goodness-of-fit does not improve as sample size increases; the results shown in the table are for a sample of 500 customers, but our results for a sample of 2000 customers are very similar. Importantly, the quality of results using this approach depends on the amount of historical event data incorporated, the quality of the underlying evaluations, and the granularity of the cell tables.

However, considering the simplicity of this very low-cost approach, impact estimate tables provide a good method of achieving a relatively accurate settlement. Regression approaches are preferable for accurate ex-post measurement and verification, but impact estimate tables are quite capable of providing quick, unbiased results for settlement purposes.

Table 4 shows results for the two between-subjects methods. The first approach is a simple comparison of means, while the second approach is a difference-in-differences calculation. With the first approach, demand reductions are estimated as the difference between the group that did not have their AC loads curtailed and one that did. With the second approach, the difference between the two groups is also calculated for the curtailment day. However, differences between the two groups observed during days without curtailments and similar weather are then subtracted out. This additional step nets out differences that are irrelevant and mainly due to sampling variation. It improves precision of the estimates, particularly if smaller samples are employed. The table also shows results for the impact estimate table approach using a sample size of 500 customers.

Table 4
Between-subjects methods results.

No.	Result type			Bias (MPE)				Goodness-of-fit (MAPE)				Goodness-of-fit (CV RMSE)			
	Data source			Agg. AC	Household			Agg. AC	Household			Agg. AC	Household		
	Sample size			500	500	1000	2000	500	500	1000	2000	500	500	1000	2000
9	Between-subjects	Comparison of means	5%	−32%	−72%	−57%	−41%	20%	36%	25%	17%	0.23	0.33	0.24	0.17
			Median	−2%	−7%	4%	−2%	30%	58%	40%	27%	0.34	0.54	0.37	0.24
			95%	39%	78%	56%	32%	50%	106%	92%	64%	0.55	0.97	0.73	0.52
10		Diff-in-diff	5%	−16%	−21%	−13%	−9%	18%	28%	21%	15%	0.20	0.28	0.21	0.15
			Median	0%	0%	0%	−1%	26%	42%	30%	21%	0.30	0.41	0.29	0.20
			95%	14%	22%	13%	10%	36%	58%	42%	28%	0.43	0.58	0.39	0.27
11	Impact estimate tables	Percentiles	5%		−6%				33%				0.39		
			Median		0%				36%				0.41		
			95%		7%				40%				0.44		

By definition, between-subjects approaches require aggregating multiple customers into two groups to make a comparison. Thus, individual AC data does not lend itself to doing this type of comparison. In addition, it is not possible to carry out a meaningful comparison between randomly assigned groups of feeders. Thus, the table only shows results for aggregated AC data and household data. In comparing the results of AC end-use and household data, it is important to keep in mind that collecting AC end-use data is prohibitively expensive in comparison to extracting household data from smart meters that have or will be deployed. Collecting AC data for an entire summer for a sample of 500 customers can cost from $300,000 to $600,000.

Despite the fact that the aggregated AC data source only has 500 customers, it exhibits less bias and better goodness-of-fit than household data with a sample size of 500 or 1000 customers. This echoes the results shown in the within-subjects comparison; aggregated AC data includes only the “signal” of AC usage with none of the “noise” of other end-uses found in household data. With household sample sizes of 2000 or more, household data does do better than 500 AC units. Increasing the sample size tightens up the confidence bands for both bias and goodness-of-fit statistics considerably.

The difference-in-differences approach is more accurate than the simple comparison of means. The additional step of netting out random differences that are mainly due to sampling variation improves measurement precision considerably, especially for smaller sample sizes.

Both impact estimate tables and between-subjects approaches do not tend to over or underestimate impacts, provided sample sizes are large enough. However, goodness-of-fit is considerably improved when using the between-subjects approaches, indicating that these approaches do much better for individual event days. Some other considerations regarding the use of impact estimate tables have already been described above.

Comparing Table 1 and Table 2, it is clear that a between-subjects design using household data and a sample size of 2000 customers does a better job of estimating curtailment on both average and individual event days than any within-subjects approach; in addition, it also does a better job of estimating curtailment on individual days than the impact estimate tables.

Residential AC load control devices are well suited for between-subject approaches that rely on random assignment. It is possible to randomly assign and/or rotate curtailment operations rather than have to deny or delay an intervention for a subset of customers. For example, with many systems, it is possible to instruct the load control device of a house to shed load and to instruct the load control devices at an adjacent house not to do so. This approach was successfully executed in the 2011 evaluation of Pacific Gas & Electric's SmartAC program, where approximately 140,000 AC units were randomly assigned to 10 different groups and test operations were systematically called for research purposes (George et al., 2012). For each curtailment event, one or two groups were curtailed and the remaining groups served as controls.

4. Conclusions and policy implications

Residential AC load control programs are substantial, existing resources that can be deployed quickly. If integrated into grid operations and ancillary service markets, they can provide operators with a resource that can be deployed quickly in response to system shocks or unexpected changes in demand or supply that requires a fast response. In order to fully incorporate AC load control into grid operations and markets, it is necessary to define the rules and requirements for settlement.

Much of the debate to date regarding settlement methods for demand reduction has focused on day-matching baselines, metering requirements, and telemetry. Our research shows that day-matching baselines, which, as shown in previous research, can be accurate to measure reductions in commercial and industrial DR programs, are not well suited for measuring AC demand reductions. Moreover, more granular meters do not necessarily increase the accuracy of demand reduction measurement because measuring demand reduction is fundamentally different than measuring the output from generation resources.

The fact that relatively accurate estimates can be cheaply and quickly obtained using pre-calculated tables of demand reduction estimates raises several interesting policy questions. Is it really necessary to use more complex and more expensive estimation approaches for each individual AC curtailment event? How much value does the incremental accuracy of more complex estimation approaches and metering provide for settlement? How much value is gained by increasingly granular measurement (1 min versus 15 min versus 1 min data)?

As a solution to the policy problem identified at the outset of this paper – how to quickly, cheaply, and accurately measure residential AC load curtailments for settlement – we recommend a practical approach, using tables with pre-calculated load reductions per AC unit to estimate demand reductions during the summer and conducting a more detailed evaluation at the end of the summer to reconcile settlements. The demand reduction tables should be updated on an annual basis using a transparent process that allows for independent verification by a third party. As the measurement uncertainty in annual evaluations improves and the number of AC load operations increases, the accuracy of the tables is expected to increase. The use of such tables allows for quick settlement when resources are dispatched and provide operators a quick estimate of the DR resources available for operations. This approach will go a long way toward solving the policy problem of how to quickly, cheaply, and accurately measure short-term reductions for settlement.

The accuracy of pre-calculated tables depends in part on the amount of historical curtailment data incorporated, the quality of the evaluations, and the granularity of the tables. When possible, it is highly recommended that direct load control program administrators systematically execute test operations to better define the performance of the programs and that they rely on large sample sizes, with random assignment of devices to curtailment operations, and a difference-in-differences method.

5. References

Bode et al., 2010

Bode, J., M. Perry, Morgan, S., 2010. Assessment of Settlement Baseline Methods for Ontario Power Authority's Commercial a Industrial Event Based Demand Response Programs. Prepared for Ontario Power Authority.

Braithwait and Armstrong, 2009

Braithwait, S., Armstrong, D., 2009. Load Impact Evaluation of California Statewide Aggregator Demand Response Programs, vol. 2. Baseline Analysis of AMP Aggregator Demand Response Program. Prepared for the California Demand Response Measurement and Evaluation Committee.

Braithwait and Armstrong, 2010

Braithwait, S., Armstrong, D., 2010. Highly Volatile-Load Customer Study. Prepared for Southern California Edison, Pacific Gas and Electric, and San Diego Gas and Electric.

Coughlin et al., 2008

Coughlin, K., Piette, M.A., Goldman, C., Kiliccote, S., 2008. Estimating Demand Response Load Impacts: Evaluation of Baseline Load Models for Non-Residential Buildings in California. Ernest Orlando Lawrence Berkeley National Laboratory. LBNL-63728. 〈http://eetd.lbl.gov/ea/EMS/EMS_pubs.html〉.

Eto et al., 2010

Eto, J., Marnay, C., Goldman, C. Lawrence Berkeley National Laboratory; Kueck, J., Kirby B.Oak Ridge National Laboratory; Dagle, J., Pacific Northwest National Laboratory; Alvarado, F., Mount, T., Oren, S., PSERC; Martinez, C., Southern California Edison, 2010. An R&D Agenda to Enhance Electricity System Reliability by Increasing Customer Participation in Emerging Competitive Markets. IEEE Power Engineering Society Winter Meeting, January 28–31, 2001.

Eto et al., 2007

Eto, J., Nelson-Hoffman, C., Torres, S., Hirth, B., Yinger, J., Kueck, B.Kirby (Oak Ridge National Laboratory), Bernier, C., Wright, R., Barat, A., Watson, D., 2007. Demand Response Spinning Reserve Demonstration. Ernest Orlando Lawrence Berkeley National Laboratory. LBNL-62761.

Eto et al., 2009

Eto, J., Bernier, C., Young, P., Sheehan, D., Kueck, J., Kirby, B., Nelson-Hoffman, J., Parker, E., 2009. Demand Response Spinning Reserve Demonstration—Phase 2 Findings from the Summer of 2008. Ernest Orlando Lawrence Berkeley National Laboratory. LBNL-2490E.

Federal Regulatory Energy Commission, 2011

Federal Regulatory Energy Commission, 2011. 2010 Assessment of Demand Response and Advanced Metering. Available at 〈http://www.ferc.gov/industries/electric/indus-act/demand-response/2010/survey.asp〉.

George et al., 2012

George, S.G., Bode, J., Berghman, D., 2012. 2011 Statewide Evaluation of California Aggregator Demand Response Programs, vol. II. Baseline Calculation Rules and Accuracy. Prepared for Pacific Gas & Electric, Sand Diego Gas & Electric and Southern California Edison. Submitted to the California Public Utilities Commission on June 1, 2012. Available at: 〈http://apps.pge.com/regulation/search.aspx〉

George et al., 2012

George, S.G., Hartmann, L., Perry, M., 2012. 2011 Load Impact Evaluation for Pacific Gas and Electric Company's SmartAC Program. Prepared for Pacific Gas & Electric and submitted to the California Public Utilities Commission on June 1, 2012. Available at: 〈http://apps.pge.com/regulation/search.aspx〉

Gifford et al., 2010

Gifford, W., Bodmann, S., Young, P., Eto, J., Laundergan, J., 2010. Customer Impact Evaluation for the 2009 Southern California Edison Participating Load Pilot. Lawrence Berkeley National Laboratory LBNL-3550E.

Hartmann et al., 2012

Hartmann, L. Berghman, D., Bode, J., Perry, M., George, S., 2012. 2011 Load Impact Evaluation of PG&E's Residential Time-Based Programs. Prepared for Pacific Gas & Electric and submitted to the California Public Utilities Commission on June 1, 2012. Available at: 〈http://apps.pge.com/regulation/search.aspx〉

Kema, 2003

KEMA, Inc., 2003. Protocol Development For Demand Response Calculation—Findings and Recommendations. Prepared for the California Energy Commission. 〈http://www.energy.ca.gov/reports/2003-03-10_400-02-017F.PDF〉.

KEMA, Inc., 2009

KEMA, Inc., 2009. Pacific Gas & Electric SmartAC™ 2008 Residential Ex Post Load Impact Evaluation and Ex Ante Load Impact Estimates. Prepared for Pacific Gas & Electric. 〈http://www.calmac.org/publications/FINAL_SmartAC_Load_Impact_2009_03_31_-_CALMAC_Study_Id_PGE0278.01.pdf〉.

KEMA, Inc., 2011

KEMA, Inc., 2011. PJM Empirical Analysis of Demand Response Baseline Methods. Prepared for the PJM Markets Implementation Committee. 〈http://pjm.com /markets-and-operations/demand-response/~/media/markets-ops/dsr/pjm- analysis-of-dr-baseline-methods-full-report.ashx〉.

Kirby, 2003
Kirby, B., 2003. Spinning Reserve From Responsive Loads. Oak Ridge National Laboratory. ORNL/TM-2003/19.

Kueck et al., 2001
Kueck, J., Kirby, B., Staunton, R., Eto, J., Marnay, C., Goldman, C., Martinez, C.A., 2001. Load As a Reliability Resource in Restructured Electricity Markets. Prepared for the U.S Department of Energy. 〈http://certs.lbl.gov/pdf/load-reliability.pdf〉.

North American Electric Reliability Council (NERC), 2010
North American Electric Reliability Council (NERC), 2010. Special Report: Potential Reliability Impacts of Emerging Flexible Resources.

Quantum Consulting and Summit Blue Consulting, 2004
Quantum Consulting and Summit Blue Consulting, 2004. Working Group 2 Demand Response Program Evaluation—Program Year 2004 Final Report. Prepared for the California Public Utilities Commission.

Sullivan et al., 2009
Sullivan, M., Bode, J., Mangasarian, P., 2009. 2009 Pacific Gas and Electric Company SmartAC Ancillary Services Pilot. Prepared for Pacific Gas and Electric and submitted to the California Public Utilities Commission on December 31, 2009

Appendix A1.

Table A1
Definition of regression terms and symbols.

Term	Definition
i,h,t	The subscripts *i*, *h*, and *t* are attached to various regression terms and apply to a given person/device (i), hour of day (*h*), and day (*t*)
α	Regression constant
β	Regression coefficient
ϵ	Error term
kw	Measured kilowatts (kW)
$kw_{i,h-2,t}$	Subtraction signs on subscripts indicate a lagged variable; in this case, the term signifies kW for a given customer, on a given day, two hours prior to the current hour
weekend	Dummy variable indicating whether the day is a weekend
CDH	Cooling degree hours; a cooling degree hour is the maximum value of either the temperature in Fahrenheit minus 70 or 0
CDH^2	Cooling degree hours, squared; the squared symbol applies to other variables in a similar fashion
past24hrCDH	Cooling degree hours experienced during the previous 24 h, summed
noschool	Dummy variable indicating whether the day is a school vacation day
event	Dummy variable indicating whether there is a load control event during that hour

Table A2
Mathematical expression of regression models.

#	Model description	Mathematical expression
1	*Treatment variables and no day or hourly lags or leads.* This model estimates demand reductions as a function of the temperature during the hour, as measured by cooling degree hours (CDH), and total heat intensity in the day prior to curtailment (past 24 h CDH). Several additional variables are included to explain variation in electricity use so the demand reduction signal can be better detected, including: • The effect of hour, weekend, and school vacations. • Total heat intensity in the day prior. • Heat intensity on during each time period. • The interaction between school vacation periods and the use of cooling. The model can be applied for both ex post estimation and to forecast available AC loads and demand reduction potential. It uses data from prior curtailments to inform the estimate of the demand reduction for the curtailment in question.	$kw_{i,h,t} = \alpha_{i,h} + \beta_{i,h} \cdot weekend_t + \beta_{i,h} \cdot noschool_t + \beta_{i,h} \cdot past\ 24\ hr\ CDH_{i,t} + past\ 24hrCDH^2_{i,t} + \beta_{i,h} \cdot CDH_{i,h,t} + \beta_{i,h} \cdot CDH^2_{i,t} + \beta_{i,h} \cdot (noschool_t \cdot CDH_{i,h,t}) + \beta_{i,h} \cdot (noschool_t \cdot CDH_{i,h,t})^2 + \beta_{i,h} \cdot (event_{i,h,t} \cdot CDH_{i,h,t}) + \beta_{i,h} \cdot (event_{i,h,t} \cdot past24CDH_{i,t}) + \varepsilon_{i,\ h,\ t}$
2	*Treatment variables with a hourly lag.* This model estimates demand reductions as a function of the temperature during the hour, as measured by cooling degree hours (CDH), and total heat intensity in the day prior to curtailment (past 24 h CDH). The same variables as those in Model 1 are included to explain variation in electricity use so the demand reduction signal can be better detected. In addition, this model includes the electricity use two hours prior to the time period in question to explain variation in electricity use. The model can be applied for ex post estimation; however, it requires continuous data uploading to use it for forecasting available AC loads and demand reduction potential for operations.	$kw_{i,h,t} = \alpha_{i,h} + \beta_{i,h} \cdot weekend_t + \beta_{i,h} \cdot noschool_t + \beta_{i,h} \cdot past\ 24\ h\ CDH_{i,t} + \beta past\ 24\ h\ CDH^2_{i,t} + \beta_{i,h} \cdot CDH_{i,h,t} + \beta_{i,h} \cdot CDH^2_{i,t} + \beta_{i,h} \cdot (noschool_t \cdot CDH_{i,h,t}) + \beta_{i,h} \cdot (noschool_t \cdot CDH_{i,h,t})^2 + \beta_{i,h} \cdot (event_{i,h,t} \cdot CDH_{i,h,t}) + \beta_{i,h} \cdot (event_{i,h,t} \cdot past24CDH_{i,t}) + \varepsilon_{i,\ h,\ t} \cdot kw_{i,h\ \ 2,t} + \varepsilon_{i,h,t}$
3	*Treatment variables with hourly lags and leads.* Like Models 1 and 2, this model estimates demand reductions as a function of the temperature during the hour, as measured by cooling degree hours (CDH), and total heat intensity in the day prior to curtailment (past 24 h CDH). The same variables as those in Model 1 are included to explain variation in electricity use. In addition, this model includes the electricity use in both hours preceding and after the time period in question to explain variation in electricity use. The model can be applied for ex post estimation, but cannot be used to forecast available AC loads and demand reduction potential for operations.	$kw_{i,h,t} = \alpha_{i,h} + \beta_{i,h} \cdot weekend_t + \beta_{i,h} \cdot noschool_t + \beta_{i,h} \cdot past\ 24\ h\ CDH_{i,t} + \beta past\ 24\ h\ CDH^2_{i,t} + \beta_{i,h} \cdot CDH_{i,h,t} + \beta_{i,h} \cdot CDH^2_{i,t} + \beta_{i,h} \cdot (noschool_t \cdot CDH_{i,h,t}) + \beta_{i,h} \cdot (noschool_t \cdot CDH_{i,h,t})^2 + \beta_{i,h} \cdot (event_{i,h,t} \cdot CDH_{i,h,t}) + \beta_{i,h} \cdot (event_{i,h,t} \cdot past24CDH_{i,t}) + \varepsilon_{i,\ h,\ t} + \sum_{j=\ 3,\ \ 2,2,3} \beta_{i,h} \cdot kw_{i,h+j,t} + \varepsilon_{i,h,t}$
4	*No treatment variables but use of hourly lags and leads.* This model differs from the prior three models in that the demand reductions is not explicitly calculated using regression coefficient. Rather the regression is used to provide an estimate of load without curtailment and the impacts are calculated as the difference between the reference load and the metered load during the curtailment. Except for the lack of treatment variables, the explanatory variables are the same as in Model 3.	$kw_{i,h,t} = \alpha_{i,h} + \beta_{i,h} \cdot weekend_t + \beta_{i,h} \cdot noschool_t + \beta_{i,h} \cdot past\ 24\ h\ CDH_{i,t} + \beta past\ 24\ h\ CDH^2_{i,t} + \beta_{i,h} \cdot CDH_{i,h,t} + \beta_{i,h} \cdot CDH^2_{i,t} + \beta_{i,h} \cdot (noschool_t \cdot CDH_{i,h,t}) + \beta_{i,h} \cdot (noschool_t \cdot CDH_{i,h,t})^2 + \beta_{i,h} \cdot (event_{i,h,t} \cdot CDH_{i,h,t}) + \beta_{i,h} \cdot (event_{i,h,t} \cdot past24CDH_{i,t}) + \varepsilon_{i,\ h,\ t} \cdot (noschool_t \cdot CDH_{i,h,t})^2 + \sum_{j=\ 2,\ \ 1,2,3} \beta_{i,h} \cdot kw_{i,h+j,t} + \varepsilon_{i,h,t}$

All four regression models have program level R^2 values greater than 0.95 for each data source, meaning that they explain more than 95% of all variation in customer load.