Using Smart Meter Data to Identify Non-Performing Load Control Devices

Colin Kerrigan, Pacific Gas and Electric, San Francisco, CA
Wendy Brummer, Pacific Gas and Electric, San Francisco, CA
Christine Hartmann, Freeman, Sullivan & Co., San Francisco, CA
Josh L. Bode, Freeman, Sullivan and Co., San Francisco, CA

ABSTRACT

A common question among utilities has been, to what degree we can understand the components of household loads by observing patterns in the aggregate household hourly load. In particular, for utilities operating direct load control programs with one-way communication, the question is if we can use smart meter data to identify premises with unresponsive control devices? Unresponsive devices weaken the load impacts of load control programs, often times without the utility really having a grasp on the magnitude of issues. This paper describes methods to identify missing or non-performing load control devices using smart meter hourly household data. It also present the results of field study to assess the accuracy of diagnostic tests using hourly household data.

Introduction

A significant problem in load control programs is non-performing devices. These can be due to broken or disconnected control devices or because some devices fail to receive control event paging signals. For utilities using devices with one-way communication, there is no easy answer to these questions; due to the significant cost of direct verification of device operation, often utilities just assume a customer remains a part of the program without any ongoing verification. It is not financially feasible to blindly send service technicians to every property to check device operation. Up until recently, with no way to identify broken devices, it has just been easier, and more cost effective, to recruit new customers.

If utilities were able to remotely identify broken and missing devices, it could increase the aggregate impacts of the program without as much cost as new customer acquisition. In this paper we discuss the challenges and accuracy of identifying unresponsive or missing devices using hourly whole-building meter data.

Using hourly interval data from Pacific Gas & Electric's (PG&E's) air conditioning cycling load control program, SmartAC, we undertook the task of creating methods to identify probable broken or missing devices and assessing their accuracy. Our effort involved three main steps:

- A field survey to identify non-responsive devices among a recruited random sample of 416 devices. The survey allowed us to quantify the incidence rate of non-responsive devices. As we discuss later, the incidence rate is one of the critical components that affects the accuracy of efforts to identify broken or missing devices.
- Development and application of a method for identifying non-responsive devices using interval data. A device that is not functional does not reduce air conditioner demand over multiple events.
- A verification test to determine the accuracy of the diagnosis and whether these devices could be brought back into service. To test the accuracy, PG&E randomly chose to visit and service 100 of the 1,210 devices identified as non-performing or missing.

Our conclusion is that using whole building smart meter data to identify non-performing or missing devices leads to substantial improvements over blindly sending technicians to assess if devices are performing. These efforts are most accurate if they are restricted to households whose electricity use clearly spikes with hotter weather conditions. These customers also provide the most value since they use air conditioners during peaking conditions. Diagnosis of non-performing devices is less accurate when it is applied to the general population. Prior to discussing the results from each of the three main steps, we provide context by comparing whole building data to air conditioner end use data and by discussing fundamentals of diagnosis.

Whole Building Data Versus Air Conditioner End Use Data

Most utility smart meters collect residential whole building data for each hourly interval and business data for 15-minute intervals. To identify non-performing or missing load control devices, it is necessary to assess whether air conditioner units are on when load control events are called and whether or not the devices lead to reduction in the air conditioner load control demand. There are two related challenges for doing so with whole building data: air conditioner use varies substantially across households and the footprint of air conditioner use is often not clearly identifiable with hourly data on individual days.

Figure 1 shows the distribution of air conditioner use across the 15 highest system load days in 2009 over the period of 2-6 pm, when the PG&E system typically peaks. It is based on a sample of 500 air conditioners for which PG&E collected directly-metered air conditioner demand. These units did not experience load control events.

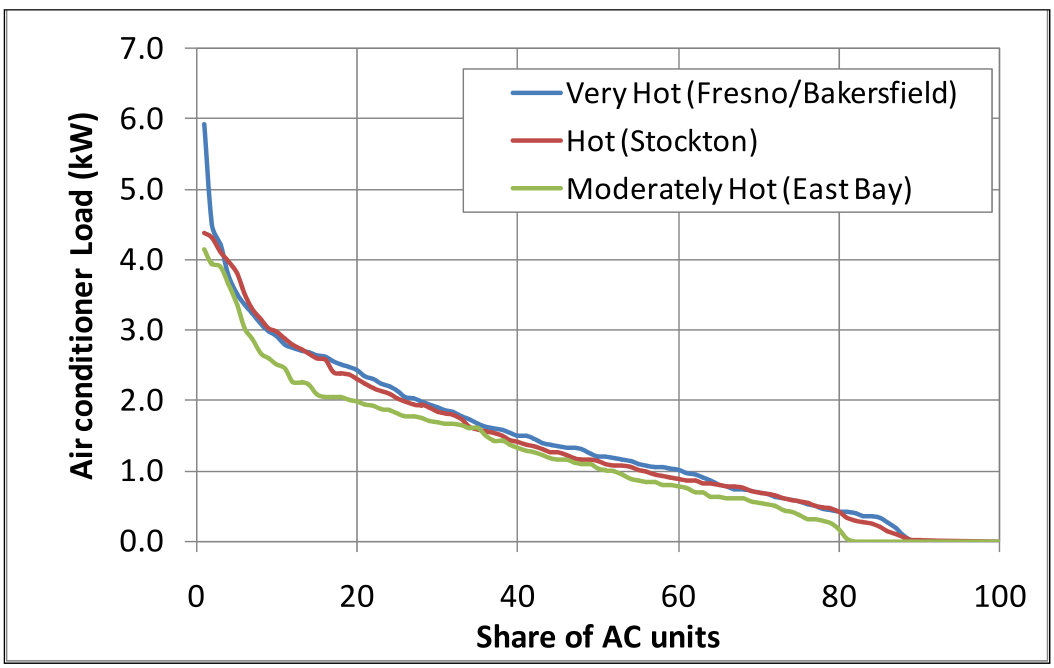

Figure 1. Distribution of Air Conditioner Demand Varies Substantially for Individual Households

The amount of variation in air conditioner use is striking, even in hotter areas such as Fresno/Bakersfield where temperatures during summer months commonly exceed 100°F. Many households use little or no air conditioning during peak hours on weekdays. They may not turn on their air conditioner until the evening hours either because they are not home during the day or because of the significant overnight cooling common in the Western U.S. The amount of variation in air conditioner use has significant implications for the ability to identify non-performing or missing devices using whole building data. If an air conditioner unit is not on during peak hours or if the compressor only needs to operate for a few minutes during an hour to cool the home, it is more difficult to determine whether a device is or is not performing.

The second challenge is identifying the demand signature of air conditioner units on an hourly basis using whole building data. With more granular data, the signature of air conditioner data is more easily identifiable, since the compressor (which accounts for most of the demand) is either on or off. However, over the course of an hour, the air conditioner signature is not as distinct. Figure 2 shows whole building and air conditioner end use data for three randomly selected households on PG&E's system peak day and its third highest system load day in 2009, when PG&E collected air conditioner end use data for a sample of households.

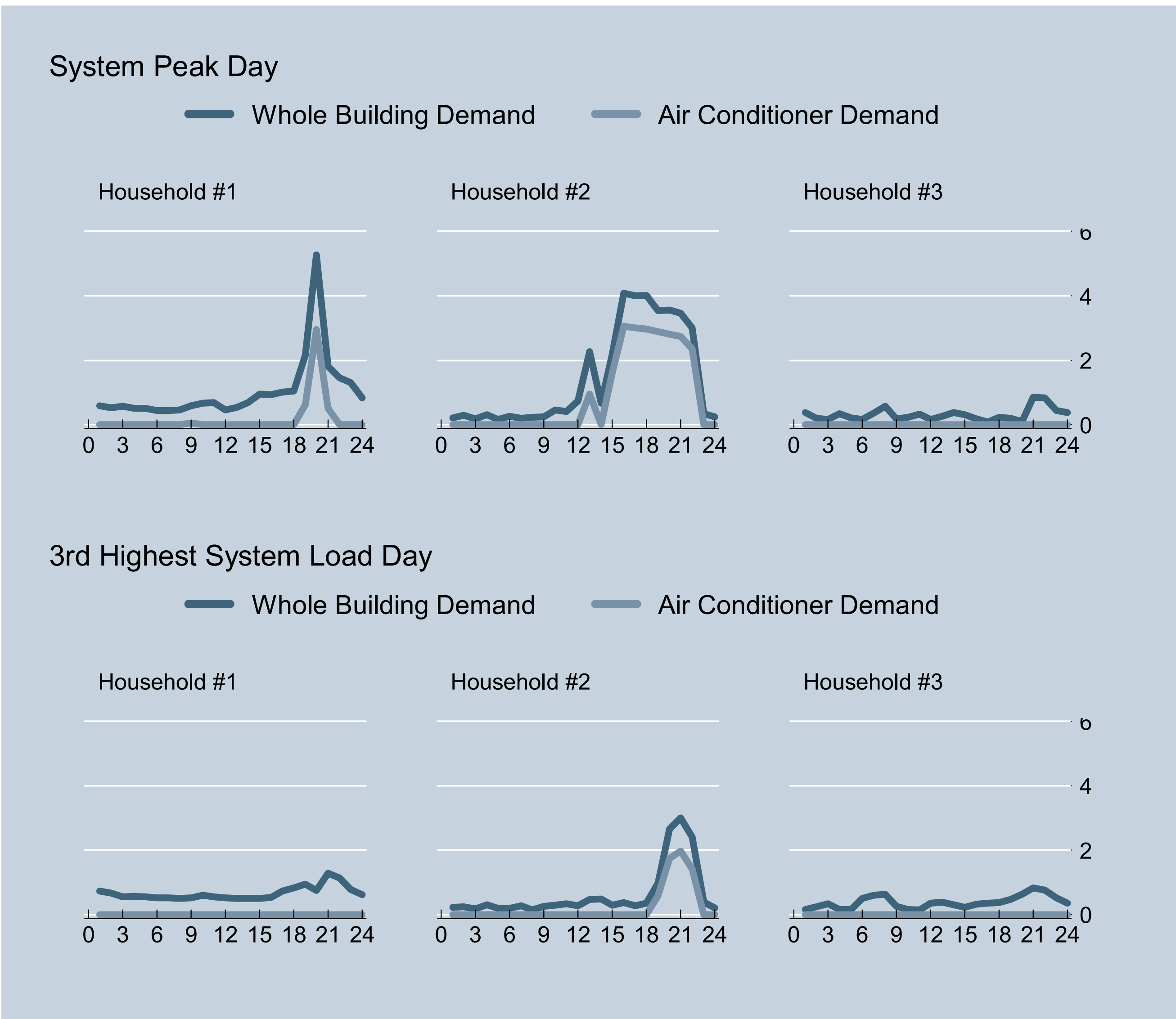

Figure 2. Comparison of Whole Building and Air Conditioner End Use Data

Even on two of the hottest days of the year, air conditioner use is highly diverse, with different patterns across customers and within individual sites. The first household used air conditioning for a short evening period on the peak day but did not turn on the air conditioner at all on the third highest system load day. The second household had nearly twice as much air conditioning demand on the system peak day than on the third highest system load day; and also had the air conditioner unit in operation for a longer period. The difference in air conditioner demand is likely due to a higher duty cycle – the share of an hour an air conditioner compressor has to be on to cool the home – due to hotter conditions. The third household did not use their air conditioner on either day.

The illustration shows that detecting whether a load control device works for any single day is difficult, particularly since most load control operations reduce air conditioner demand, but do not eliminate it entirely. Air conditioners may not be on at all when a control event starts, so a drop in load is not always observed. Users may also coincidentally turn on an air conditioner unit at the time a load control event ends, which may be confused with snapback, when in fact the unit was never controlled because it was not on during the event. Another scenario is air conditioners that are never turned on at all, in which case it is difficult to assess if a device failed because the air conditioner load would not have changed irrespective of whether the device was performing properly.

There are two main conclusions from the above discussion. It is extremely difficult to determine whether load control devices are or are not working on an event by event basis; to successfully identify non-performing devices it is necessary to rely on data from multiple events. The second conclusion is that it is difficult to identify non-performing devices for the subset of the population that rarely uses air conditioners during peak hours. In fact, it is easy to mix up limited operation of air conditioners with non-performing devices.

Fundamentals of Diagnosis

A few fundamentals of diagnosis are useful for understanding the accuracy of tests designed to identify non-performing events. The accuracy of any test depends on the answer to three questions:

- Are failures common? Technically, this is the incidence rate of failures.
- How well does the test identify failures when there is indeed a failure? The question can also be reversed – how often does a test miss a failure? – In that case, the answer describes the rate of false negatives.
- How often does the test incorrectly diagnose a failure when none occurred? Technically, this is the false positive rate.

In describing diagnostic tests, it is common to focus on how well the test identifies failures when there are indeed failures. However, the failure rate and frequency of false positives typically play a larger role in the accuracy of a diagnostic test.

Diagnosis is inherently difficult when failures are not common. When that occurs, most tests to diagnose the failure perform poorly when applied to the full population. This is best illustrated through an example. Suppose that we know 5% of devices do not perform or are missing. Out of every 1,000 devices, 50 do not work properly. A test that identifies actual failures 90% of the time will properly classify 45 of the non-performing devices and miss 5 of them. However, the bigger issue is the frequency of incorrectly diagnosed failures. Suppose the test incorrectly classifies 10% (95) of the 950 performing devices as non-performing. In total, the test will identify 140 devices (45+95) as non-performing when in fact only 45 (32%) of those devices are actually not performing or missing.

When failures are more common, diagnostic tests perform better. Table 1 illustrates this point. It shows how the accuracy of the diagnostic test improves as failures become more common. It assumes the test correctly identifies failures when there are indeed failures 90% of the time and that it incorrectly diagnoses a failure when there is none 10% of the time. If the failure rate is 5%, as in the earlier example, 32% of devices classified as non-performing will indeed be non-performing. If the failure rate is higher, e.g., 15%, 61% of the devices diagnosed as non-performing will be correctly identified even though the diagnostic test is identical. The only difference is the frequency of failures (the incidence rate).

Table 1. Incidence Rates and Accuracy of Diagnostic Tests – Example

Failure incidence rate	Non-performing devices			Performing devices			Test accuracy (among those classified as non-performing)		
	Properly identified as failing	Incorrectly diagnosed as performing	Total	Properly identified as performing	Incorrectly diagnosed as failing	Total	Actual failures	Total classified as failing	% Properly identified
2.5%	22.5	2.5	25.0	877.5	97.5	975.0	22.5	120.0	19%
5.0%	45.0	5.0	50.0	855.0	95.0	950.0	45.0	140.0	32%
7.5%	67.5	7.5	75.0	832.5	92.5	925.0	67.5	160.0	42%
10.0%	90.0	10.0	100.0	810.0	90.0	900.0	90.0	180.0	50%
15.0%	135.0	15.0	150.0	765.0	85.0	850.0	135.0	220.0	61%
20.0%	180.0	20.0	200.0	720.0	80.0	800.0	180.0	260.0	69%
25.0%	225.0	25.0	250.0	675.0	75.0	750.0	225.0	300.0	75%
30.0%	270.0	30.0	300.0	630.0	70.0	700.0	270.0	340.0	79%

Table 1 assumes the test correctly identifies 90% of failures when devices are indeed non-performing (10% false negative rate) and that the test incorrectly classifies 10% of devices that are performing as non-performing (10% false positive rate)

For many newer programs, the rate of failure will be relatively low. A well-constructed diagnostic test will perform better than sending technicians at random to inspect sites. However, even well designed screens for identifying non-performing or missing devices will perform poorly if applied to the entire population. Many of the sites identified as having non-performing devices will have functional devices.

One way to improve accuracy is to target the diagnostic tests rather than apply them to the entire population. This an approach often applied to medical diagnosis. Tests are only conducted once the patient exhibits characteristics – such as age or a particular symptom – that place them in a segment with a higher incidence rate. Another related alternative is to apply it only to populations where the diagnostic test is known to perform well. In the case of direct load control devices, it is more difficult to assess whether a device is functional for customers that rarely or never use air conditioning during peak hours since a clear change in load shape is not evident. In other words, diagnostic tests should perform better among customers that are more likely to use air conditioning on a regular basis.

Frequency of Non-Responsive Devices in Population

As a first step, site visits were conducted on 416 selected load control devices. The sample was stratified between customers who were dually enrolled in both PG&E's load control program, SmartAC and critical peak pricing rate, SmartRate and customers enrolled only in SmartAC. The main goal of the field study was to quantify the failure rate of devices and understand if and how they varied between dually-enrolled customers and those enrolled on SmartAC alone. At the time, dually-enrolled customers accounted for roughly 15% of the SmartAC participants, but experienced events more frequently since their air conditioners were also curtailed for critical peak pricing event days.

A sample of customers was randomly recruited into the study. However, the sample of devices visited cannot be considered a completely representative sample because only about 1 in 10 customers who were phoned agreed to allow the technician access to their device. Due to low response rates, there is the potential for non-response bias based on unobservable characteristics, although it's hard to know what effect, if any, that has on the results below. It could overstate or understate the failure rate of devices.

During the site visits, technicians downloaded data from the internal logs of the load control devices. These logs identify whether or not the devices were functioning properly. Each PG&E device is capable of recording 90 days of various operating data, which includes the amount of minutes the device was running and the minutes of shed during a control event. Most of the air conditioning units in question have load control switches installed, though nearly 20% have programmable communicating thermostats (PCTs), which can be and are operated like load control switches. A device is determined to be non-responsive if the device did not receive the control event paging throughout the summer and there are no load shed minutes.

Table 2 summarizes the results from the field study. Out of the 416 devices, 23 of them were non-responsive. The failure rates were very similar for customers who were and were not dually enrolled, 5.3% and 5.9%, respectively. The estimated failure rate across all customers with devices was 5.8% with a 95% confidence interval of ± 2.2%.

Table 2. Population Failure Rates Based on Field Study

	Type of Device	Total devices downloaded data	Devices that received signal and/or shed load	Non-performing devices	Failure rate	95% Confidence Interval	
						Lower	Upper
Dually enrolled	Thermostats	194	193	1	0.5%	0.0%	1.5%
	Switches	34	23	11	32.4%	19.2%	51.4%
	Total	228	216	12	5.3%	2.4%	8.2%
SmartAC Only	Thermostats	165	165	0	0.0%	-	-
	Switches	23	12	11	47.8%	27.4%	68.2%
	Total	188	177	11	5.9%	2.5%	9.2%

Non-performing devices were more frequent on sites that had PCTs. While less than 1% of air conditioners with switch devices were not performing, based on the field study, an estimated 45% (±12.9%) of sites with thermostats were not functional, after applying the population weights. There are several potential reasons for the higher failure rates for thermostats. Customers who had a load control device at an earlier date are more likely to have thermostats. Because thermostats have been in the field longer, they are more likely to have failed or to have had the control device removed. Thermostats also experience higher communication failure rates, particularly in areas where the paging network is weaker, because the thermostats are inside the home while load control switches tend to be outside.

The main implication from the field study is that any diagnostic test will be more accurate among sites with thermostats. In contrast, the same test is likely to incorrectly diagnose failures for a large number of sites with functional load control switches if it is applied to full population.

Methods for Identifying Non-Responsive Devices Using Interval Data

In order to determine which were non-responsive, we focused on the whole building load shape over multiple events. Devices that are functional reduce demand or notch the load shape during the load control events. Devices that are non-performing do not alter the load shape. There were three main components:

- Load drop in the first hour of the event
- Snapback immediately after control of the air conditioner
- A high correlation between temperature and loads

Figure 3 demonstrates the first two features. The green line reflects the first hour of the control event, from 2-3 pm, and shows a significant drop for the treatment group, which is not experienced by the control group. Similarly, shown in orange, is a rebound effect experienced by the treatment group when the control of air conditioner units is released, referred to as snapback.

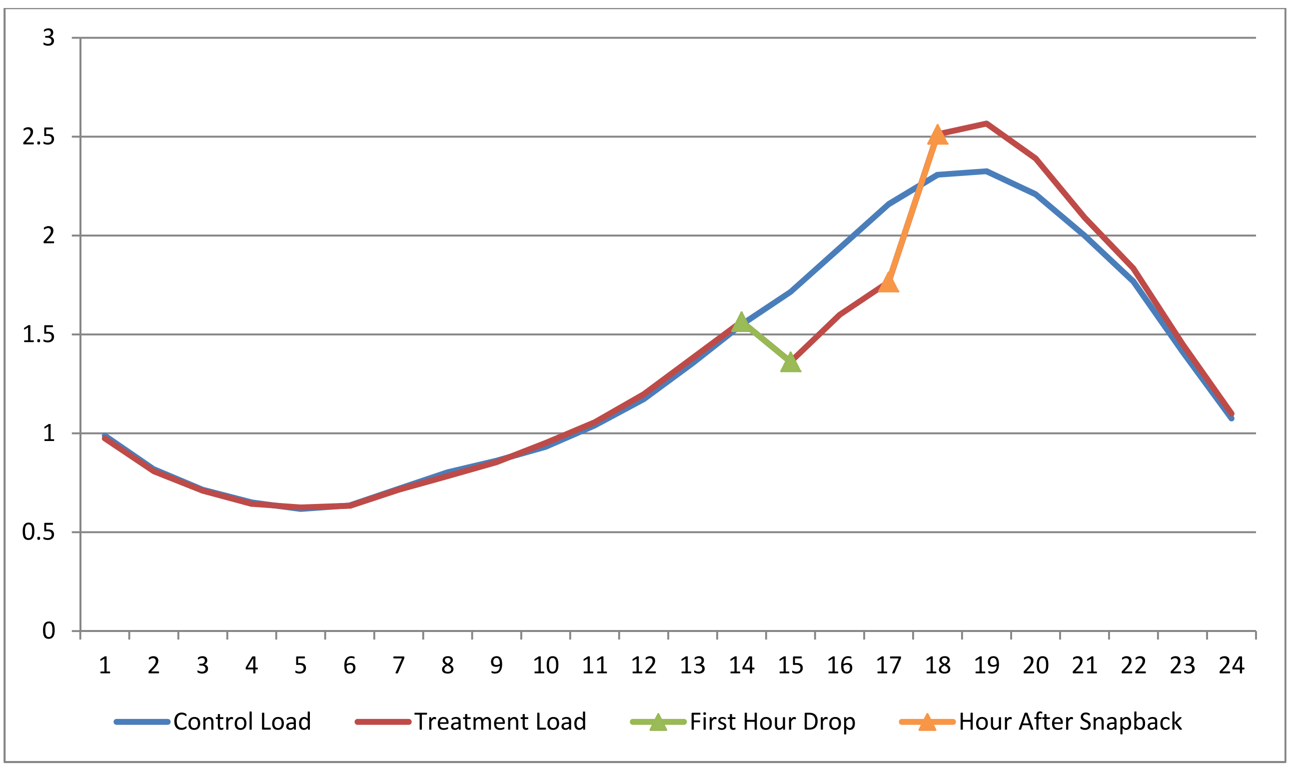

Figure 3. Load for Dually-Enrolled Customers on July 12, 2012

The third component is a high correlation between temperature and customer loads. Since such participants are inherently more likely to use air conditioner during SmartAC control event hours, it is easier to detect if they reduce load during a control event. In contrast, customers with a low correlation between electricity use and temperature are more likely to be incorrectly diagnosed as failing when in fact they simply may not have had any air conditioning load to drop.

The diagnostic test was based on the percentage difference between the average load of the first and last control event hours and the average of the two hours immediately surrounding the control event. This approach combines information about load drop in the first event hour and snapback after control of the air conditioner unit is released. For simplicity, we refer to it as the combined metric for load drop and snapback. Customers who dropped load because of a functional control device should have a negative percentage value. Customers who did not experience a demand reduction are more likely to have a non-performing device and should have small or no differences.

Figure 4 shows the load patterns when all customers are ranked based on the combined metric for load drop and snapback. It shows the load patterns when this process is applied to the full population. The load drop and snapback pattern is clearly evident in roughly 50%-60% of customers. However, for a substantial share of customers, their load shape is such that it is hard to determine whether the lack of load drop and snapback is because the air conditioner unit was not in use or because the load control device was not functional. The potential for misdiagnosis for these devices is high.

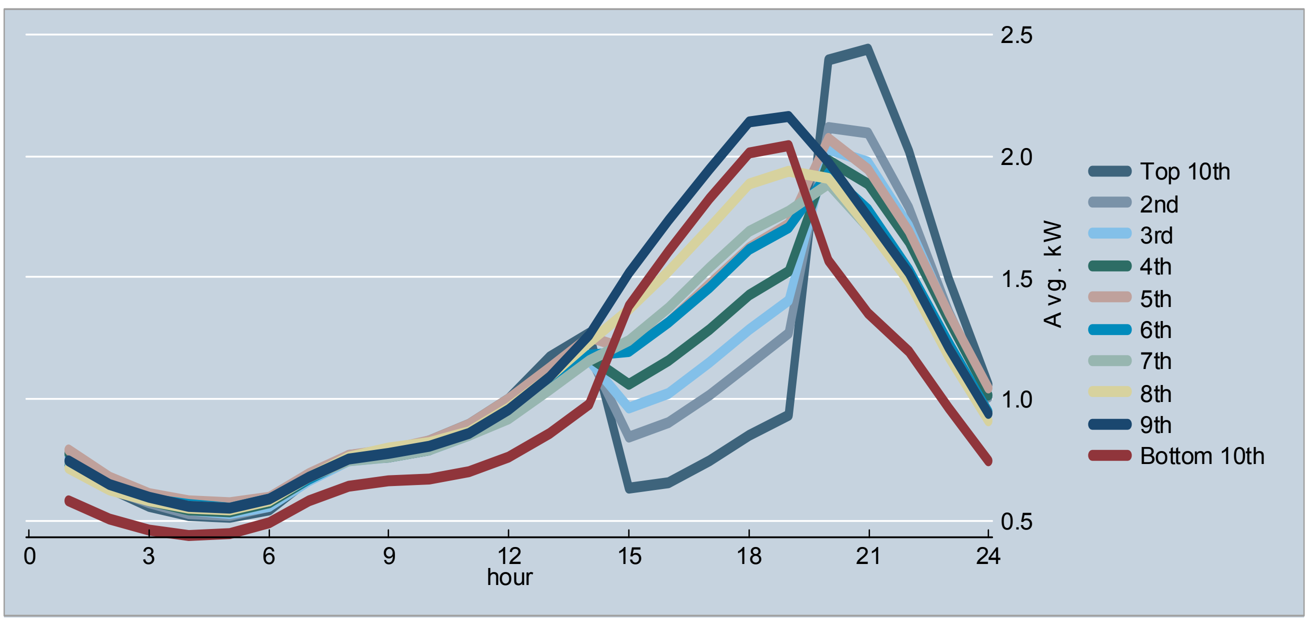

Figure 4. Load Shapes Based on Performance Diagnostic – Full Population

To reduce the risk of misdiagnosis, we only relied on the results of the diagnosis test for customers who experienced a minimum of two events and had peakier load shapes during non-event days. First, load shapes were normalized to a percent of the mean load for each customer on hot (over 90°F) weekdays when the air conditioner units were not controlled. Each customer was assigned to 1 of 10 load shapes through cluster analysis. The main purpose of this initial step was to classify each customer into natural groupings (or clusters) based on their load shape. This allowed us to identify customers with load shapes associated with air conditioner loads (loads correlated with temperature). It also allowed us to avoid applying the test to customers who rarely used their air conditioner and had the highest risk of misdiagnosis. Figure 5 shows the average customer load for each shape cluster.

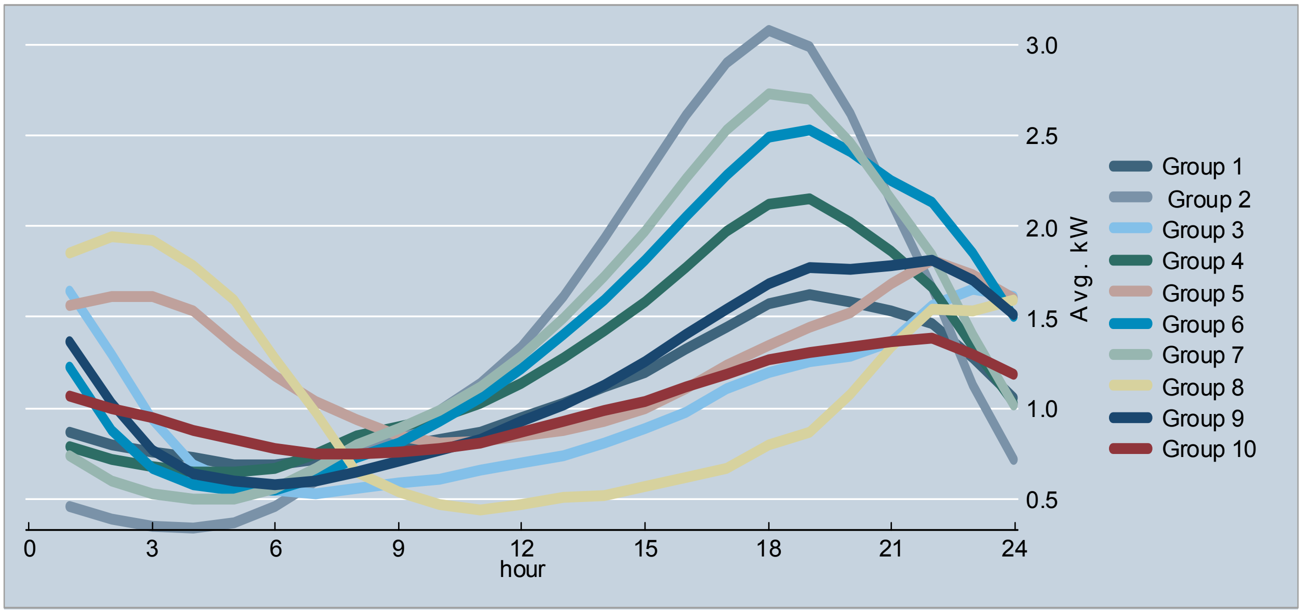

Figure 5. Load Shape Clusters – Full Population

Results and Accuracy of Tests

In total, 135,000 residential households had load control devices and whole building hourly interval data. The diagnostic tests were only considered reliable for approximately 30,000 customers who had experienced a minimum of two events and were classified into the two peakiest load shape clusters. Out of the 30,000 sites screened, 1,210 (4.0%) were identified as having potentially disabled devices.

To test the accuracy of this identification, PG&E randomly chose 100 of the 1,210 devices classified as non-performing to visit and service the devices, if needed. Our technicians were able to visit 95 customers after calling to confirm a visit. Shown in Table 3, 100% of the programmable communicating thermostats (PCT) were missing or had internal errors, while 76% of the switches visited were not functioning for some reason. The targeting was very successful with an overall 83% success rate. The higher amount of misdiagnosis among switches is not surprising given that the initial field survey indicated a low rate of failure.

Based on the verification tests, out of the 1,210 devices classified as non-performing, approximately 1,000 (83%) are likely not performing. This represents 3.3% of the target sites and 57% (3.3%/5.8%) of the estimated failures based on the initial field survey. Even if the failure rate were higher, at the upper bound of the 95 percentile, the effort would have identified 40% of non-performing devices in the targeted segments.

While the diagnosis of non-performing devices was relatively successful, only 56% of the missing or broken devices were able to be brought back into the program. One of the most common problems found was that the AC unit had been replaced and either the device was missing or not connected again. The SmartAC program incentive is currently associated with the install of the DLC device, and an ongoing incentive is not offered. Additionally, a portion of the devices passed visual inspection but were found to be broken only by examining the internal device logs.

Table 3. Number of Failures Found in the Devices Visited

Status	Switches	PCTs
Physically Damaged	3	0
Missing	13	6
Internal Errors	6	12
No Connection	8	6
Disconnected	24	0
Functional	17	0
Total	**71**	**24**

Conclusion

PG&E's experience shows that using whole building interval data to identify non-performing devices can be very successful under the right settings. The accuracy of the diagnosis is higher if failure rates are high or efforts are directed at devices that are known to have a higher failure rates. The accuracy also improves when it focuses on peakier customers who are less prone to misdiagnosis. Finally, the diagnosis works better for customers who have experienced multiple events.

Importantly, the diagnostic tests perform better among customers who use their air conditioners during peaking conditions. These customers are more valuable and cost-effective to reactivate. The risk of misdiagnosis is highest among customers who rarely use their air conditioner during peaking conditions and who should not be targeted for re-activation in the first place.

While these results are encouraging, there are limitations to diagnosis of non-performing devices through whole building interval data. The results were accurate because we relied heavily on selecting customers who were both valuable and had a lower risk of misdiagnosis. As a result, the initiative targeted approximately a quarter of customers with load control devices. The frequency of misdiagnosis would likely increase if the diagnostic tests are applied to a larger share of customers with load control devices.