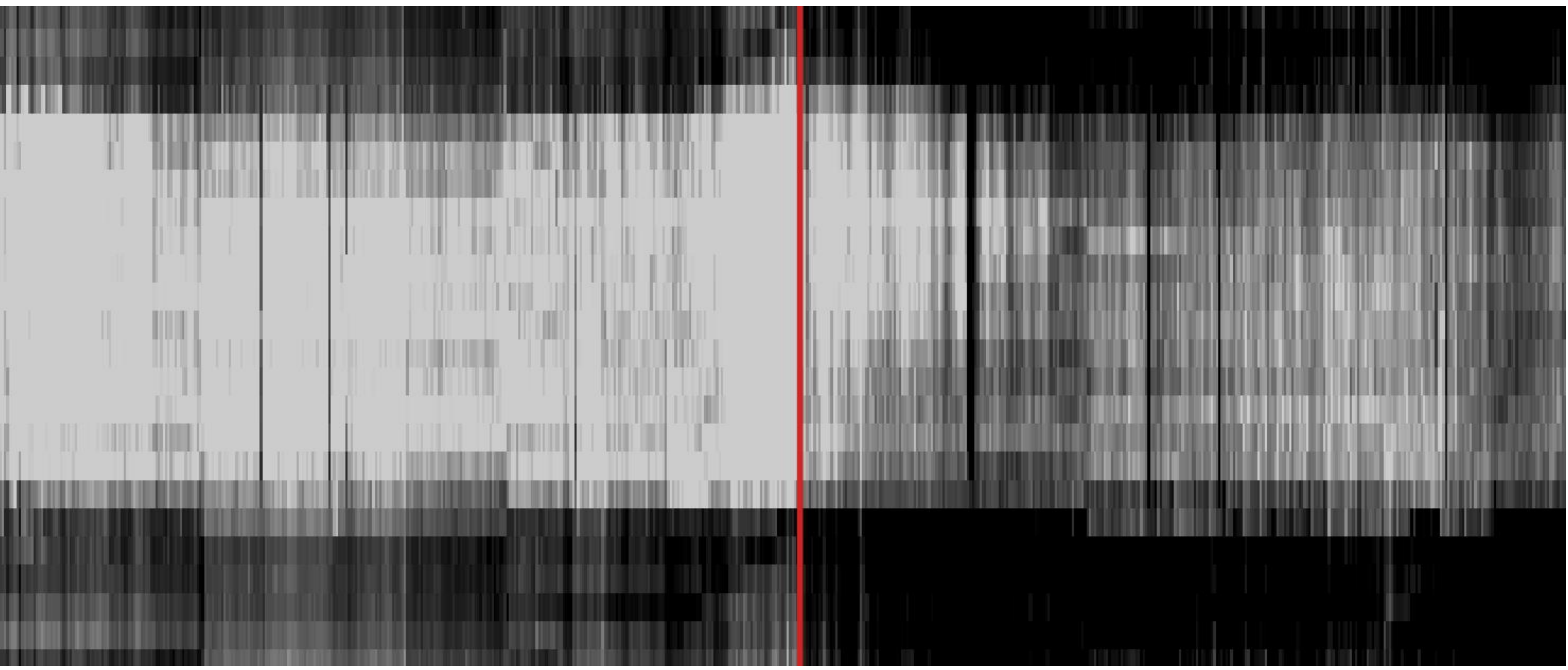
Energy Efficiency Program Targeting:

Using AMI data analysis to improve at-the-meter savings for small and medium businesses

Final Report: Whitepaper

August 10, 2018

CALMAC Study ID PGE0421.01



Sam Borgeson, Partner

Convergence Data Analytics, LLC

Convergence Data Analytics

Adam Scheer

Robert Kasman

Megan Geraci

Francis Dahlquist

Pacific Gas and Electric Corporation

Contents

Note for readers........3
1 Executive summary........4
1.1 Results and conclusions........5
1.2 Summary........8
2 Policy background........8
3 Introduction........8
4 Data........9
5 Key concepts........11
5.1 Considerations of meter-based savings claims vs. traditional approaches........11
5.2 Defining and calculating at-the-meter savings........11
5.3 At-the-meter program savings........12
5.4 Understanding how specific consumption feature filters perform........13
6 Customer targeting: Results and discussion........17
6.1 Control comparison........17
6.2 Filter performance: detailed results........17
6.3 Savings by customer categories........25
7 Discussion........28
7.1 Different ways of evaluating savings........28
7.2 Considerations of targeting power........29
7.3 The relevance of total consumption versus other usage attributes........30
7.4 Program design considerations........30
7.5 The limitations of NMEC savings calculations........31
7.6 Additional considerations........32
8 Conclusions........33

Cover image: What intervention impact look like: heat map of pre and post intervention consumption for an anonymous DI customer, with hours of day along the y-axis and days of the year along the x-axis and the intervention date highlighted with a red line.

Note for readers

This document, written for a general audience, summarizes the background, methods, and results of a longer and more detailed research paper. Readers interested in more technical details on the methods and results should read that report as well. It will be referred to as the "research report" in this summary document. This work also extends and expands upon prior research conducted on residential energy efficiency program results. Interested readers are encouraged to read the whitepaper on that work as well.[1]

[1] https://pda.energydataweb.com/api/view/1945/Customer_Targeting_Final_Whitepaper_ResEE.pdf

1 Executive summary

This research paper explores targeting methods for energy efficiency (EE) programs in the Small and Medium Business (SMB) sector. Meter-based savings[2] are evaluated for recent programs and customer targeting methods with the potential to enhance the savings of future programs are developed and tested. This research addresses several questions:

- To what extent do intervention impacts of current programs “show up” at the meter?
- What metered savings increases are achievable through data-driven targeting?
- What are the objectives of “targeting for savings?” Increasing the magnitude of savings (kWh) and enhancing savings depth (savings as a percentage of pre-program consumption) are explored.
- What metrics best quantify the performance of targeting methods and support comparison across methods?
- How much variation in optimal targeting strategies can be expected across different programs, measures, and customers?

Although this research is based on two specific PG&E energy efficiency (EE) programs in the Small and Medium Business (SMB) sector, the authors believe the methods and results developed here are generally applicable to most other EE programs that are evaluated with meter-based savings.

What is targeting?

Energy efficiency program “targeting” is accomplished through selecting specific customers to be the focus of program marketing efforts.[3] For example, a residential pool pump program would have greater success if targeted at residences known or expected to have pools. This first strategy applies to narrowly focused program offers and focuses on the ability to participate. Other efficiency programs are applicable to most or all customers, and a second type of targeting strategy focuses on improving the propensity, or likelihood, of targeted customers’ participation. Finally, a third type of targeting strategy, which is the subject of this work, identifies customers with the greatest likelihood of saving more than others would - *targeting for savings*.

Why target?

Customer targeting has the potential to increase overall program savings, at lower cost, while delivering greater benefits for participating customers. With these gains, targeting can increase program cost-effectiveness and improve customer satisfaction. Because it allows for marketing to the subset of customers most likely to save, targeting can be enhanced with more specific and individualized marketing messages, which stand to improve conversion rates and limit recruitment costs. In summary, targeting methods are worthy of study because they can enhance the value of EE to participating

---

[2] Meter-based savings are estimated based on changes in customers’ metered energy consumption before and after program interventions. While sub-meters can be installed on customer premises, this research is based entirely on whole building utility meters installed at customer premises.

[3] Energy efficiency programs are generally open to all eligible customers. Targeting does not exclude customer participation but promotes participation by those most likely to achieve greatest savings.

customers, yield higher returns on investments from the ratepayer base, and provide additional benefits to utility program sponsors.

How was this research conducted?

At-the-meter savings estimates and predictive targeting methods were developed for two longstanding PG&E programs for small and medium business (SMB) customers: The Regional Direct Install Program (DI), which focuses on lighting and refrigeration upgrades, and the Commercial HVAC Quality Maintenance Program (HVAC), which services air conditioning equipment based on industry standard maintenance protocols. The first step in the research was to gather *program intervention*, *customer characteristics* (e.g. zip code, NAICS code, etc.) and *interval consumption data* for all program participants from 2013 to 2016.[4] Savings were calculated for each participant based on pre- and post-intervention consumption and compared to a control group of similar non-program participants. Finally, targeting methods based on customer characteristics and energy consumption metrics were developed and then analyzed to see which most significantly improved meter-based savings.

1.1 Results and conclusions

We estimate that well-executed targeting can improve per-customer average savings by a factor of 2-3x by pre-screening potential participants using data-driven targeting methods described here and focusing recruitment efforts on the most attractive 25-50% of potential customers.

Table 1 summarizes the untargeted and targeted savings improvements for the highest performing targeting methods we examined[5] for a) the full Direct Install (DI) program, b) DI participants with only lighting interventions, c) the DI participants with only refrigeration interventions, and d) the HVAC program. For the DI program, targeting is anticipated to double or triple average customer savings. For the HVAC program, targeting stands to increase savings levels from near zero to highly significant.

Table 1: Summary of targeting performance for the DI (including lighting and refrigeration examined separately) and HVAC programs.

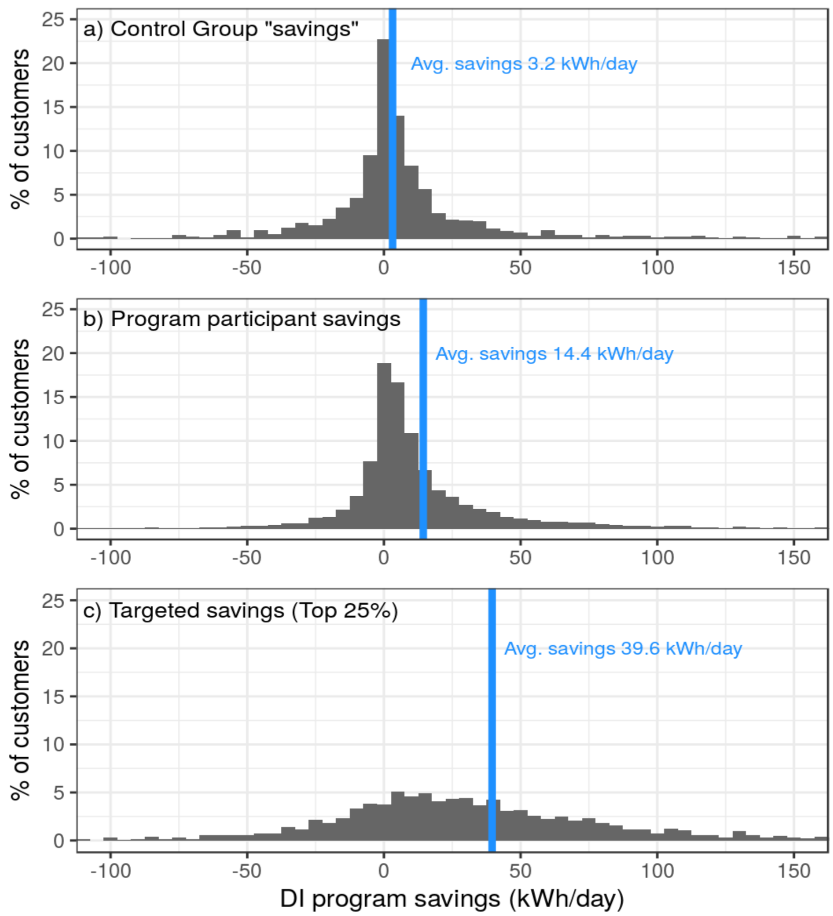
	DI	DI lighting	DI refrigeration	HVAC
Premise count	7,497	5,331	772	1,193
Untargeted savings (kWh/day)	14.4	13.7	10.1	1.0
Matched control savings	3.2	*N/A*	*N/A*	-3.5
Targeted savings for 1 of 2 customers (kWh/day)	25.3	24.5	18.2	13.8
Savings multiple after targeting 1 of 2 customers	1.8x	1.8x	1.8x	13.4x
Targeted savings for 1 of 4 customers (kWh/day)	39.6	39.1	27.3	28.7

[4] Interval data in this study used customer utility meter data averaged at one-hour intervals.
[5] The targeting strategies that performed best in the DI program tend to focus on total consumption, while the best HVAC targeting strategies isolate AC loads.

	DI	DI lighting	DI refrigeration	HVAC
Savings multiple after targeting 1 of 4 customers	2.8x	2.9x	2.7x	27.7x

Where do these average customer savings values come from? Figure ES1 shows estimated "savings" results, in kWh/day, for a) non-participants (comparison group) and estimated saving for b) EE program participants, and c) targeted program participants.

Figure ES1: Savings Distributions of Comparison Group, Program Participants, and Targeted Participants for the DI program



Several key conclusions can be drawn from Figure ES1.

The nature of savings at-the-meter is a distribution. In all three customer groups (Figure ES1) savings vary from one customer to the next and are best understood as a distribution. Even for the comparison group, there are positive savers, zero savers, and negative savers. This is a result of the natural variability of customer energy usage over time. This "background noise" is an inescapable reality when considering meter-based savings.

At-the-meter savings are visible when looking at recent program participants. Even accounting for the "noise" of natural variability, statistically significant program savings show up at the meter and shift and skew savings distributions. Average savings of program participants (Figure ES1b) is 14.4 kWh/day, while the control group (Figure ES1a) shows far lower average savings (3.2kWh/day).

Targeting can significantly increase average savings per participant. After targeting the top 25% of customers (Figure ES1c), average participant savings were nearly a factor of three higher than for general program participants (39.6 kWh/day compared to 14.4kWh/day). The targeted customers still display a distribution of savings, but it is shifted toward higher savings, with a significantly higher fraction of customers achieving positive savings. When targeting in the HVAC program, customers achieved approximately eight times the average participant savings, although this was partly due to low general program participant savings. This is intriguing as it suggests that under meter-based evaluation some programs may only be viable under cost effectiveness rules with effective targeting.

Comparison groups can be used to subtract out "naturally occurring savings." The drivers behind the comparison group's (Figure ES1a) small positive savings are not known with certainty. They could be due to imperfect weather normalization, the impacts of other efficiency efforts, or other market or technology trends toward lower energy usage in this time period. We do know that LED lighting was achieving widespread market adoption and computing efficiency continued to drive down energy consumption during the period evaluated. Whatever the cause, the comparison group "savings" can be subtracted from participant savings for a more accurate estimate of program impacts.

Additional project findings include:

Targeting methods using energy consumption data generally outperform targeting using customer attributes (e.g. zip code, NAICS code, etc.) During this research we tested hundreds of targeting approaches using both consumption data and customer attributes. The best results (greater average savings from more customers) were achieved using consumption data.

Targeting methods need to be tailored for individual programs and measures. Unsurprisingly, this research showed there is no single targeting scheme that works best, or even at all, for the diverse range of programs types and efficiency measures studied. To maximize savings at-the-meter, the best targeting approaches identify both *high usage* <u>*and*</u> *inefficient usage* of the specific end uses addressed by program interventions.

Targeting for absolute savings as compared to percentage savings (depth of savings) can require different targeting approaches. Some of the best targeting metrics for absolute savings were poor performers for depth of savings, and vice versa. This result encourages careful consideration of program goals and implementation strategy prior to selecting targeting methods.

Targeting, by definition, focuses on subsets of populations. When there are large populations, it may be possible to target the top 5-10% of customers and still have sufficient numbers to meet program goals. In other cases, targeting power may need to be reduced to maintain sufficient target population sizes. Targeting can result in lower total participation but yield higher total savings and cost effectiveness.

1.2 Summary

This study demonstrates that targeting can be an effective way to significantly increase savings at the meter and program cost effectiveness. Implementers considering program designs utilizing a meter-based savings approach (including Pay-for-Performance) will benefit from identifying customers at the outset with high savings potential. Especially if overlaid with demographic and/or customer segmentation information, personalized messaging can be developed and targeted at customers most likely to benefit from participation.

2 Policy background

Many energy efficiency (EE) advocates and implementers are able to identify measures that they believe have significant potential to save energy or deliver operational benefits to the grid but are not cost effective under existing policies and rules. With the advent of new evaluation and implementation pathways that are focused on savings at the meter, improved program targeting stands to help unlock some of that untapped potential. Whether through lowering acquisition costs by focusing recruitment on viable participants or improving per-customer outcomes by focusing program recruitment on those with the greatest expectation of savings, data-driven targeting and personalization have a lot to offer EE administrators, policymakers, and implementers.

The passage of Assembly Bill 802[6] in California established existing conditions baselines[7] for many EE programs with the goal of catalyzing the replacement of inefficient equipment. AB802 further specified that the savings should be estimated "…*taking into consideration the overall reduction in normalized metered energy consumption as a measure of energy savings."* The employment of existing conditions baselines directly aligns the savings attributable to a program with the change in a customer's metered energy usage. This has fostered interest in programs whose evaluations are based on Normalized Metered Energy Consumption (NMEC).[8]

"At-the-meter" savings are computed based on changes in customers' whole building NMEC before and after energy EE program-related interventions. **This paper describes practical targeting methods for improving at-the-meter savings outcomes with examples of data-driven targeting applied to real-world small and medium business (SMB) program data.**

3 Introduction

The goals of data-driven targeting are to lower cost and/or increase savings through the identification of customers most likely to achieve savings greater than past participants. This type of targeting is

---

[6]AB 802 instructs the California Public Utilities Commission to authorize EE programs with savings measurement based on *"all estimated energy savings and energy usage reductions, taking into consideration the overall reduction in normalized metered energy consumption as a measure of energy savings"* https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201520160AB802

[7] Before AB 802, the standard California baseline for EE measures was set to equal to current code requirements, or industry standard practice (ISP), in cases where no code exists. EE programs could only incentivize the portion of savings above code. With *existing conditions baselines*, eligible programs can incentivize all program savings.

[8] At the time of this writing, the CPUC is taking comments on its proposed ruling on how NMEC-based evaluation will be handled under the expanded third-party solicitation process.

developed here using data from recent EE programs. Customer characteristics and consumption patterns are evaluated for their correlation with at-the-meter savings. The best performing of the candidate targeting criteria can then be used to identify customers with characteristics and consumption patterns to be targeted for participation in future iterations of similar programs.[9]

This research is motivated by several questions: What magnitude and depth of savings "gains" over current programs are achievable through targeting? What metrics and methods best support the comparisons necessary to evaluate the performance of different strategies? How are these results determined by the methods of estimating at-the-meter savings? How much variation in optimal targeting strategies can be expected across different types of customers, programs, and measure types?

For this work, two longstanding PG&E programs were used to evaluate data-driven targeting schemes. The first is the Regional Direct Install Program (DI), which focuses primarily on lighting and refrigeration upgrades, and the second is the Commercial HVAC Quality Maintenance Program (HVAC), which services air conditioning equipment based on industry standard maintenance protocols. For both programs, we have identified customer characteristics and usage patterns, all available prior to the program start, that predict at-the-meter savings.

We observe that the best predictive targeting strategies depend on the nature of the interventions being made and the end-uses they impact. High total and peak usage are predictive of lighting savings, high total and baseload usage are predictive of refrigeration savings, and high temperature sensitivity and estimated disaggregated AC usage are predictive of air conditioning savings.

When applied as "targeting filters" that select sub-groups of customers based on threshold values, usage metrics related to baseload and total consumption can roughly **double average DI program savings of the remaining customers when targeting 1 out of every 2 customers and roughly triple average DI savings when targeting 1 out of 4 customers**. For the HVAC program, the results are even more dramatic. The program is fairly light touch and the average savings at the meter for the program were nearly indistinguishable from the background noise of other changes in consumption over time. Filters based on usage characteristics that estimate AC loads were able to **elevate average HVAC savings from 1 kWh/day to 13 kWh/day when targeting 1 out of every 2 customers and to 28 kWh/day when targeting 1 out of every 4 customers.** Those performance gains elevate savings well beyond the background noise of other changes.

Although this research applies targeting methods to two specific PG&E energy efficiency (EE) programs, the methods and results developed here are generalizable to a wide variety of EE programs and customer types. The insights gained through this work can be readily employed to guide future interventions toward optimized savings results, both for participating customers and for the EE programs.

4 Data

This work was based on data from all SMB customers (defined as those with a maximum demand during the year prior to the program of less than 200 kW) who completed interventions associated with either

[9] Energy efficiency programs are generally open to all utility customers. Targeting does not prevent customer participation but is intended to encourage those customers most likely to achieve the greatest savings.

the DI or HVAC program during the 2014 and 2015 program years. For this group of participating SMB customers, we worked with four types of data:

1. **Account data** covering service start and stop dates as well as customer characteristics like zip codes, rate plans, business types, and pre-computed customer size code (S, M, L, and N, corresponding to sites with consumption less than 40 MWh/yr, from 40 up to 50 MWh/yr, 50 MWh/yr and larger, and with not enough information to compute, respectively).
2. **Intervention data** covering all known program interventions (i.e. not just DI or HVAC) for the period spanning 2013-2016, with their timing and program and technology type characteristics.
3. **Meter data** for each customer in the form of 24 hourly readings per day for the period spanning 2013-2016.
4. **Weather data** spanning 2013 through 2016 for every zip code for which we had customer data. For the purposes of this project the most important weather variable is the hourly average outside temperature used in NMEC weather normalization.

The same categories of data were used for a 20,000 SMB customers who did not participate in either program between 2013-2016. This pool of "candidate control" customers was used to generate matched control groups for each program.

Quality checks were applied to ensure that all customers examined had sufficient data to support the analyses performed. We required 120 days of meter data prior to the program intervention(s) (pre) and 120 days of meter data after (post). To avoid changes too large to be caused by the program interventions between pre- and post-program periods, we dropped the minority of outlying customers whose energy consumption rose above 200% or dropped to less than 50% of pre-period consumption.

The comparison sample was also subjected to the same data quality requirements. Table 2 below provides the counts of the remaining premises that participated in each program for each data quality test, with the final numbers used for the savings and targeting analysis in bold.

Table 2: Program participating premises passing each cumulative step in our data validation requirements

Criteria	DI	HVAC
all participants	13,428	3,578
and occupied without missing data	10,442	2,004
and no PV	10,363	1,970
and 120 days of pre and post	7,767	1,304
and remove factor of 2 changes	**7,497**	**1,193**
Control group with same criteria	**6,168**	**6,168**

The DI program data includes 5,331 customers with lighting upgrades and 772 with refrigeration upgrades.

5 Key concepts

5.1 Considerations of meter-based savings claims vs. traditional approaches

New program interventions with energy savings not available in deemed measure lists (TRMs), or not easily estimated using simulation software or engineering analysis may be well suited to at-the-meter savings estimation.

To target customers for programs with savings estimated through NMEC changes, it is first useful to review traditional approaches to estimating savings.

Programs that rely on **deemed savings** assumptions (i.e. fixed savings estimates per measure installed) encourage PAs and implementers to pursue efficient customer recruitment and efficient measure installation.

Programs that use **simulated outcomes** to compute savings are often utilized for large commercial, industrial, and agricultural customers with unique usage requirements and patterns. These types of custom programs rely on engineering analysis to facilitate large individual project savings claims. Due to the custom nature of the projects and savings estimates, it is only practical to pursue a limited number of projects of this type.

Deemed and custom programs often have further complexities, including counterfactual projections and the determination of code or industry standard practice baselines, consideration of "early retirement" and corresponding dual-baseline periods, and possible adjustment factors such as realization rates and expected in-service rates. In these cases, assigned savings are several steps removed from any change in metered energy use. New approaches that have not (yet) been included in tables of deemed savings or simulation software, or readily incorporated into engineering analysis cannot be rewarded under traditional program rules.

In contrast, programs that estimate savings from metered consumption changes encourage program designs that drive down observable energy usage. Given these differences across program savings claims approaches, we note that traditional strategies for implementing programs are not optimized for maximizing meter-based savings. Instead, programs utilizing at-the-meter savings estimation will realize deeper savings with program designs and customer recruitment strategies tailored to at-the-meter performance.

5.2 Defining and calculating at-the-meter savings

The at-the-meter **savings** computed for this project are defined as the difference, in kWh/day, between the daily average energy consumption after the interventions (the post-period) and the NMEC baseline derived from data from before the interventions (the pre-period) for each participating customer.[10]

The changes observed across all participants include not only the impacts of the program, but also the non-program changes in energy consumption from business expansion or contraction, control changes, non-program efficiency gains, etc. that happened to occur during the same period. To get a sense of the magnitude of these "naturally occurring" changes, we also compute the at-the-meter "savings" of non-

[10] This computation was done in R, using scripts that heavily leverage the core capabilities of the VISDOM package developed and maintained by Convergence Data Analytics staff. It's open source: https://github.com/ConvergenceDA/visdom

participating customers otherwise similar to the participants. These customers are also known as a "control group." Further details on control group matching, results, and discussion can be found in the research report for this work.

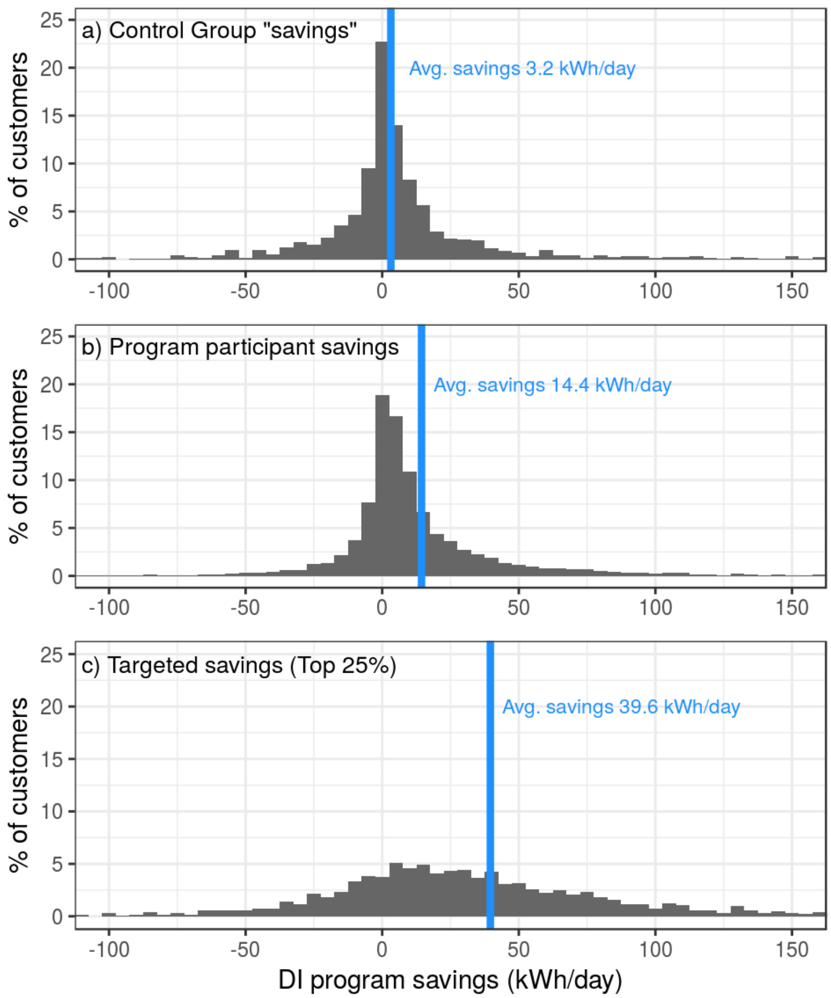
5.3 At-the-meter program savings

The underlying trends and natural variability of energy usage can be illustrated by a histogram "savings"[11] of the matched control group. Similarly, the impact of an energy efficiency program can be assessed by comparing histograms of at-the-meter savings for the control and participant groups respectively. Figure 1a provides a histogram of "savings" (i.e. non-program trends and variability) derived from the control group for the DI program. The x-axis is "savings" in kWh/day and the vertical blue line is the average value for all customers in the distribution. The average of 3.2 kWh/day indicates that energy use naturally dropped slightly for non-program participants during the program period. Figure 1b illustrates the savings for all DI program participants. The average of 14.4 kWh/day is comfortably above the average for the control group, so we conclude that the DI program delivers reliable at-the-meter savings.

Finally, Figure 1c illustrates the savings of the DI customers found in the top 25% of pre-intervention average daily consumption, which is one of many metrics (or "features") that can be derived from pre-program usage data. We define a feature as any metric that can be calculated from a customer's energy consumption data. Examples include average daily consumption (kWh), average consumption on Tuesdays from 2-3am, etc. We define the feature "filter %" as the percentage of customers eliminated using percentiles of the feature value – all customers above the 75th percentile (a 75% filter %) in the case illustrated in Fig. 1c. For this subgroup of "targeted" customers, the average DI program savings is 39.6 kWh/day, or 2.75x the 14.4 kWh/day achieved by all participating customers on average.

[11] "Savings" for the control group are calculated the same was as program participants, but since the control group didn't participate in any energy efficiency programs their expected "savings" are zero.

Figure 1: Computed savings for the direct install (DI) program. a) Control group "savings" for reference; b) All DI program participants; note the positive average and skew. C) Targeted DI program participants using the upper quartile of average pre-program energy consumption; note the skew and the increased mean.



For the DI program, ranking and selecting customers using average daily consumption, which is directly proportional to total consumption, is a viable targeting strategy, but there is much more specific and detailed information about customer energy usage in smart meter data. The remainder of this paper describes and provides targeting results from a systematic evaluation of 100 different consumption-derived features. We have also found that certain categories of customer types and other account data like rate plan and customer size, and details of program implementation, like location and specific EE technologies can reliably amplify savings when compared to the average of all participants.

5.4 Understanding how specific consumption feature filters perform

Consumption feature values (metrics computed using only pre-intervention data) were used to construct "feature filters" that define program participant sub-groups whose average savings

magnitude or depth[12] can be computed and compared to the average performance across all participants. A typical feature filter uses a threshold value to select a subset of customers, for example all customers whose mean daily consumption (a total usage metric) is greater than 100 kWh. The thresholds are selected so they eliminate, or filter out, 10%, 25%, 50%, 75%, and 90% of all customers when applied, keeping 9 in 10, 3 in 4, 1 in 2, 1 in 4, and 1 in 10 of the original customers, respectively. When a consumption feature correlates with program savings, the average savings of the filtered sub-groups are larger than the average for all participants and the feature filter can be said to yield savings gains beyond the average savings of all participants.

Subsequent figures showcase the savings impacts of specific top-performing or otherwise illustrative feature filters. Table 3 lists and defines the features, or consumption metrics, used. Each performs well in predicting either savings magnitude or savings depth (or both) for either the DI or HVAC program (or both).

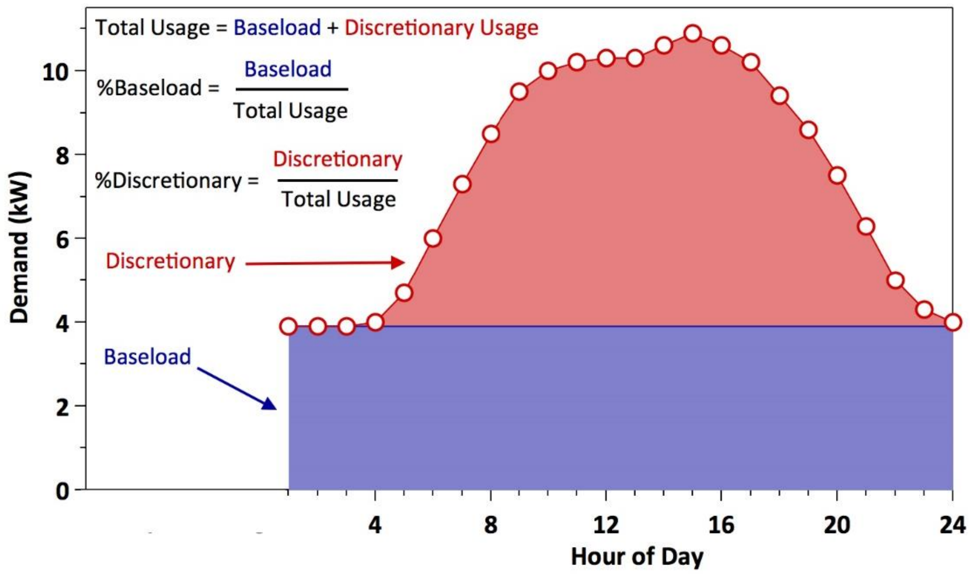
Table 3: The set of features used to illustrate filter performance in this section. These include some of the top performers for both DI and HVAC programs for enhancing savings and savings as a percentage of pre-intervention consumption.

feature	application	definition	units
kw.mean	DI, HVAC	mean demand (all data) – this is mathematically identical to total kWh usage	kWh/hr
discretionary	DI	non-baseload consumption (consumption above the daily minimum)	kWh/day
discretionary_pct	DI	discretionary consumption as a percentage of total consumption	%
baseload	DI	mean of daily minimum hourly demand x 24hrs	kWh/day
baseload_pct	DI	baseload consumption as a percentage of total consumption	%
sum2win	HVAC	ratio of total consumption during summer months to winter months	
pre_CDH	HVAC	pre-period modeled sensitivity of usage to temperature	kWh/day/F
pre_CDH_pct	HVAC	pre-period sensitivity of usage to temperature as a % of daily kWh	%/F
pre_daily.cooling.kwh	HVAC	pre-period modeled daily cooling consumption	kWh/day

Figure 2 provides a visual to help readers understand the total usage (kw.mean feature), discretionary, and baseload features of Table 3.

[12] We use the term "depth" to refer to the percentage of a building's total consumption saved. If a building uses an average of 100 kWh/day, then 90 kWh/day after the program, the savings magnitude is 10 kWh/day and the depth is 10%.

Figure 2: Schematic illustration of the relationship between total, discretionary, and baseload.



'Baseload' is the "always-on" component of usage. For example, refrigeration equipment tends to run 24 hours per day with little variation.[13] The 'baseload_pct' feature defines the percentage of a customer's total usage corresponding to baseload. Similarly, 'discretionary' and 'discretionary_pct' features correspond to the component of total usage that operates during business hours. For example, most office lighting is expected to fall into the discretionary category.

5.4.1 Savings magnitude vs. depth

Energy efficiency potential can be understood in terms of **three distinct drivers that determine savings potential and inform customer targeting strategies**:

1. The **magnitude** of a customer's service demand (i.e. how much lighting does the customer need)
2. The **timing** of the service demand (i.e. when is the lighting in use)
3. The **efficiency of the equipment** used to meet the service demand (i.e. the efficacy of the lights)

Efficiency interventions can yield savings by addressing one or more of these drivers. In some cases, the service being provided (like lighting) is greater than anyone requires. This results in waste that can be corrected by reducing the level service provided. For example, fixtures could be removed from an over-lit room. In other cases, the level of service is appropriate, but it is being provided at time when it is not needed. For example, lighting controls might switch or dim lights based on time of day or occupancy. Finally, whatever the service demand or its timing, there may be new equipment that can meet the demand using less energy. For example, LEDs can replace linear fluorescent lights. Consider three hypothetical lighting program participants in Table 4. Each saves the same amount, but for different reasons that correspond to the three drivers:

[13] Refrigeration equipment actually cycles on/off throughout the day, but if the cycling interval is shorter than the consumption data interval (one hour in this study), then refrigeration indeed looks like baseload.

Table 4: Drivers of efficiency potential for three hypothetical customers

Facility Characteristics			Pre-Program				Post-Program		
Customer	# of Fixtures	Lighting Hours Per Day	Fixture Wattage	Lighting Usage (kWh/day)	Non-Lighting Usage (kWh/day)	Total Usage (kWh/day)	Fixture Wattage	Savings (kWh/day)	% Reduction in Total kWh
A	100	8	64	51	154	205	40	19	9%
B	50	16	64	51	77	128	40	19	15%
C	50	8	88	35	77	112	40	19	17%

Customer A has 100 64 W fixtures that are on 8 hours per day. Compared to Customer A, Customer B has half as many fixtures that are on twice as long, and Customer C has half as many fixtures on the same amount of time, but those fixtures are less efficient. Each customer receives a retrofit to 40 W fixtures.

Customer A saves primarily due to having a large number of moderately efficient products replaced (factor 1). Despite having half the number of fixtures Customer B saves the same amount due to the longer usage time (factor 2). Finally, Customer C achieves the same savings with half the fixtures due to achieving double the wattage reduction (factor 3).

Despite yielding the same savings, the observable usage characteristics that would signal savings potential and inform targeting would be different for each customer, as would the project considerations. With a high total usage, Customer A might be expected to offer a savings opportunity even with modest efficiency gains. Customers B and C allocate a greater percentage of their total usage to lighting. Assuming normal operating hours, their ‘discretionary’ usage during daytime hours will comprise a larger fraction of their total usage. A further practical consideration is project scope. Customers B and C would be more attractive in this example as the implementer would need to replace only half the fixtures as Customer A to achieve equal savings with lower equipment costs and installation time. can impact any or all of these factors.

These considerations point to two main **strategies to increase project savings**:

1. **Focus on larger magnitude consumers.** For example, focus on driver 1 and perform the same LED for linear fluorescent lighting swap for customers with more square footage and lighting fixtures.
2. **Provide deeper improvements for customers with especially wasteful or inefficient systems.** For example, focus on drivers 2 and 3 to identify and swap LEDs for incandescent lights or add controls to high usage fixtures.

If strategy 1 for increasing savings is dominant, targeting customers based on metrics aligned with total usage will tend to yield improved magnitude of savings, but at fixed depth of savings. If strategy 2 is dominant, outcomes will tend to have improved depth of savings as consumption feature filters go deeper. Because project costs, engineering, and equipment profiles are different for each of the two strategies, a given implementer might choose to tune targeting efforts to better support one or the other strategy. **For this reason, our filter results are presented in terms of both savings magnitude and depth. Depending on an implementer’s strategic approach, either or both of these metrics could be incorporated into a customer targeting scheme.**

6 Customer targeting: Results and discussion

6.1 Control comparison

Independent of program interventions, energy consumption fluctuates across customers and over time. To assess the degree to which programs have achieved impacts, we first need to establish a baseline for how energy use would have been expected to change across pre- and post-program time periods for the group of participating customers *if they had not participated*.[14] This baseline serves as a basis for comparison to the observed outcomes of participants when estimating the overall magnitude of program impacts. To make that comparison, we compute the average customer savings value for each participant group and corresponding matched control group. We then compute the difference in savings between the program participant and control groups for each program (aka the net savings), as well as the uncertainty (standard error) in this difference. Table 5 shows the results. There are positive and statistically significant net program savings for both programs, but the HVAC Quality Maintenance net savings are small. It is a light touch program whose impacts are more difficult to differentiate from the noise of natural variability.

Table 5: Average reduction in energy consumption (kWh/day) for the DI and HVAC programs, following outlier rejection, and the net program savings, after subtracting the control-group savings.

Program	Load category	# premises	Gross program reduction (kWh/day)	Avg. control reduction (kWh/day)	Net savings (kWh/day)	uncertainty
DI	total	7497	14.39	3.16	11.23	0.01
HVAC	cooling	1193	1.03	-3.47	4.51	0.05

Note: Control group matching is both an art and science. For simplicity, clarity, and to avoid uncertainties caused by imperfect control group matching,[15] the savings results presented in the remainder of this report are based on gross pre/post reductions in NMEC consumption.

6.2 Filter performance: detailed results

6.2.1 DI program

Figure 3 shows how the average DI program customer savings change upon selecting customers that pass different filter thresholds for five selected features. For both panels, the filter depth is on the x-axis. The left panel places mean daily savings on the y-axis and the right panel places savings as a

[14] One might assume that the baseline for comparison is no change from pre- to post-program periods. This is sufficient for ranking customers (what the feature-based filters are based on), but there are long-term trends and customer-level fluctuations in consumption year over year that suggest a non-zero baseline.

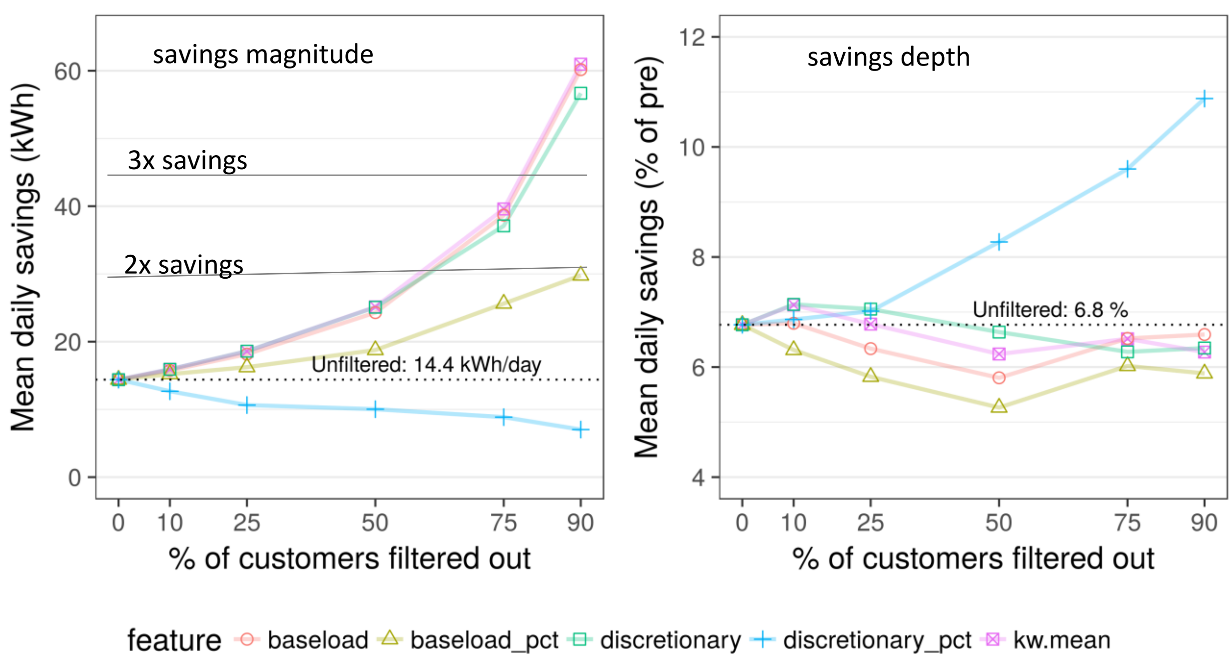
[15] The control matching methods used to establish program level savings rely on the controls being good proxies, on average, for participants. This assumption is defensible for larger samples where the idiosyncratic consumption of a few outliers averages out with others. For smaller samples, on the other hand, matched controls can add more noise to the result than the plain pre/post estimates. Additionally, as we identify targeted subsets of customers, the distributions of usage and customer characteristics may change in ways that require re-matching controls – work that was outside the scope of this research.

percentage of pre-intervention consumption (depth of savings) on the y-axis. The dotted horizontal line is the DI program average savings.

In the left panel it can be seen that three consumption feature filters, average consumption (*kw.mean*), daily non-baseload energy consumption (*discretionary*), and the average daily minimum consumption x 24 hrs (*baseload*) double the magnitude of savings (a gain % of 100 or greater) between 50 and 60% filter depth. These features yield increasing average savings at higher filter depths. They triple unfiltered savings near 80% filter depth.

In the right panel, it can be seen that the consumption features that amplified the magnitude of savings, did not amplify the depth of savings. Instead the percentage of daily consumption represented by non-baseload (*discretionary_pct*) is by far the top performer. It is normalized by total consumption, suggesting that normalized variability metrics can better identify deeper savings opportunities.

Figure 3: Filter performance as mean daily savings magnitude vs. depth for a representative sample of top DI feature filters. Dotted horizontal line is the unfiltered population average savings.

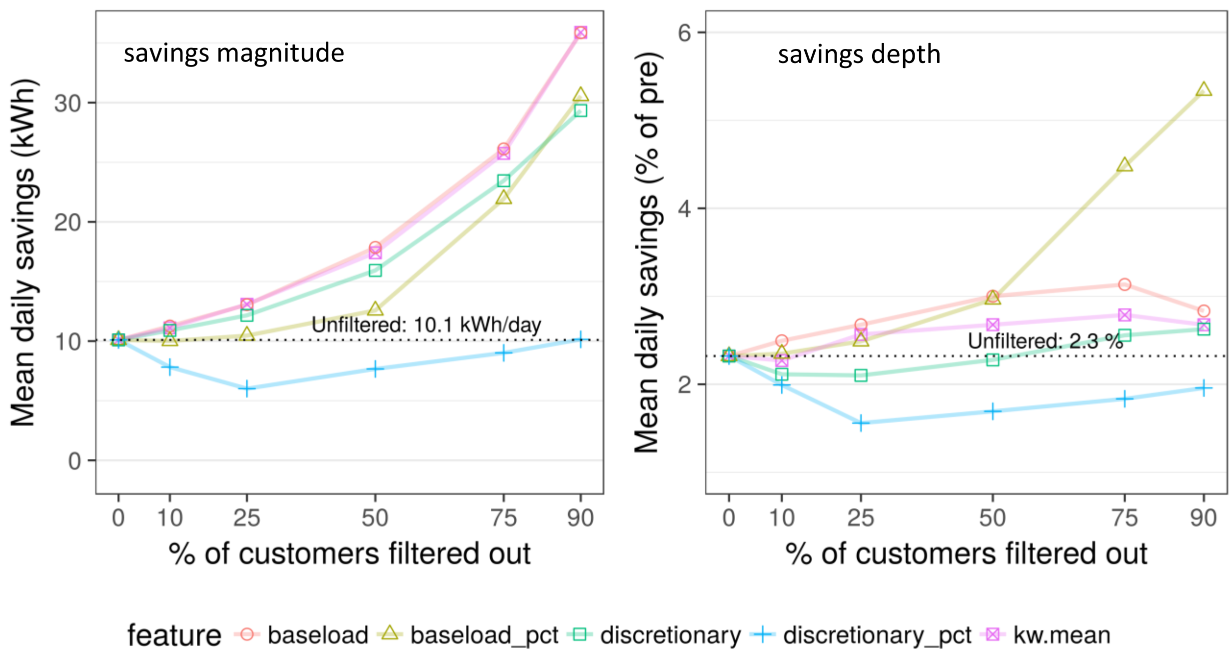


That the percentage of discretionary usage is a good predictor of savings depth for the DI program is reflective of the majority of customers receiving lighting interventions. The DI program data contain sufficient samples to study targeting for lighting and refrigeration end uses independently. The next sections present filter performance analyses for the DI participants that received exclusively refrigeration or exclusively lighting interventions.

6.2.1.1 DI refrigeration outcomes

Refrigeration is an end use that reliably fits into the 24x7 baseload of a building when utilizing hourly interval data, as was done in this study. Some business types, like restaurants and grocery and liquor stores, usually have significant refrigeration loads while others, like office spaces, do not. Figure 4 presents the results of consumption feature filters applied to the subset of 772 DI participants who's only intervention technology family was refrigeration. Among those participants, all of the filters that scale with consumption magnitude amplify savings, with *baseload* and *kw.mean* noticeably out-performing *discretionary*. The only feature filter that significantly improves savings depth is *baseload_pct*, which also performs somewhat well in improving total savings. When *baseload_pct* is elevated, it is more likely to include larger than typical refrigeration loads.

Figure 4: Filter performance as mean daily savings magnitude vs. depth for a representative sample of DI refrigeration feature filters. Dotted horizontal line is the unfiltered population average savings.



The results shown above shed light on why targeting based on specific criteria is effective. The right panel of Figure 4 shows nearly constant savings depth when targeting based on higher total usage (*kw.mean*), variable load (*discretionary*), or baseload (*baseload*). In contrast, when selecting customers with *baseload_pct*, where baseload is a larger than average fraction of total usage, savings depth increases substantially. This is an indication that the *baseload_pct* filter successfully identifies customers with larger than usual refrigeration loads, either due to the nature of the customer business or a higher degree of inefficiency of the legacy equipment.

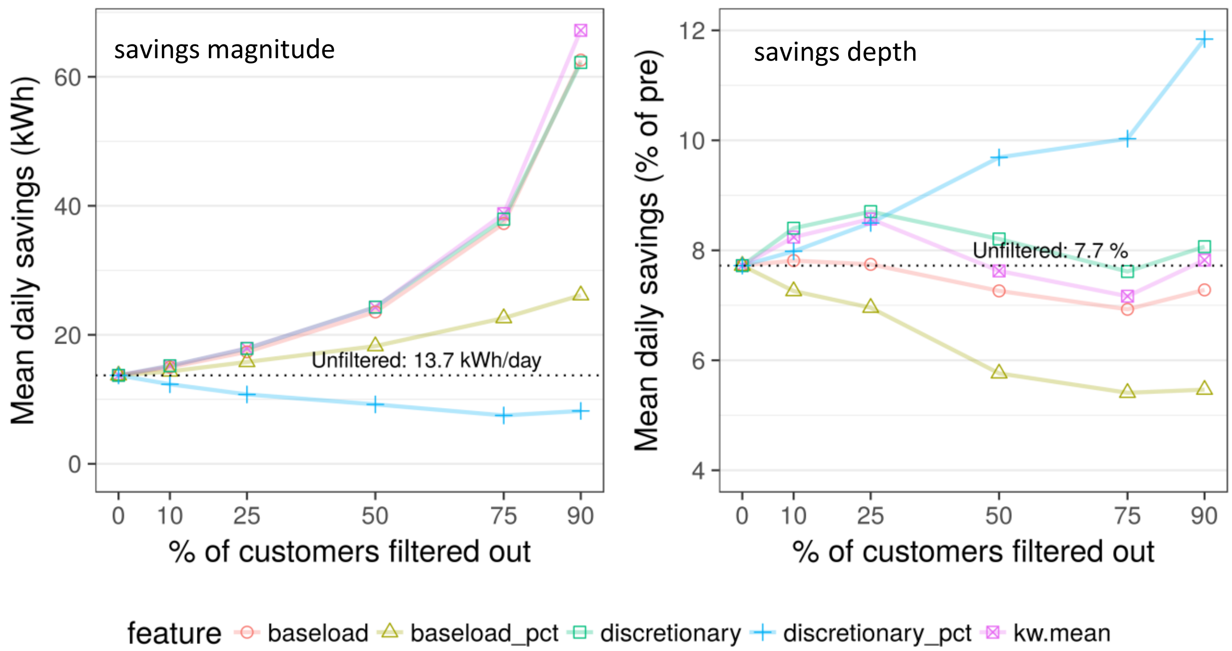
6.2.1.2 DI lighting outcomes

Figure 5 presents the results of consumption feature filters applied to the subset of 5,331 DI participants who's only intervention technology family was lighting.

The left panel shows significant gains in savings from targeting based on features that reflect the magnitude of consumption. While *kw.mean*, or total consumption, breaks ahead at the 90% filter depth (1 in 10 customers preserved), it is in a virtual dead heat with the *baseload* and *discretionary* features at lower depths.

The right panel shows deeper savings can be achieved through concentrating on customers with elevated *discretionary_pct*, or discretionary loads as a percentage of total loads. This feature relates to the fraction of total load that is occupant-controlled, so this finding is consistent with the understanding that most lighting loads (even automated ones) are ultimately occupant driven. Further corroborating this interpretation, *baseload_pct* (which was the best predictor of savings depth for refrigeration interventions) actually decreases the depth of savings compared to the unfiltered average.

Figure 5: DI program filter performance as mean daily savings magnitude (left) vs. depth (right) for a representative sample of top DI lighting feature filters. Dotted horizontal line is the unfiltered population average savings.



Similar to refrigeration, lighting efficiency gains may be achieved for one or more reasons. Changes to service demand (i.e. de-lamping), tighter controls, and/or device efficiency gains would all be expected to yield savings. Therefore some project approaches would be expected to achieve simple efficiency improvements that scale with total consumption, while others will go deeper with improvements to controls as well as luminous efficacy. The targeting metrics and evaluation results can provide insight into which customers offer savings opportunities and why, potentially vital information for implementers.

To design programs to maximize savings one might perform targeting within customer characteristics already known to predict savings. This section presents the results of performing feature filtering within prominent customer and program categories.

6.2.1.3 DI filters within customer size

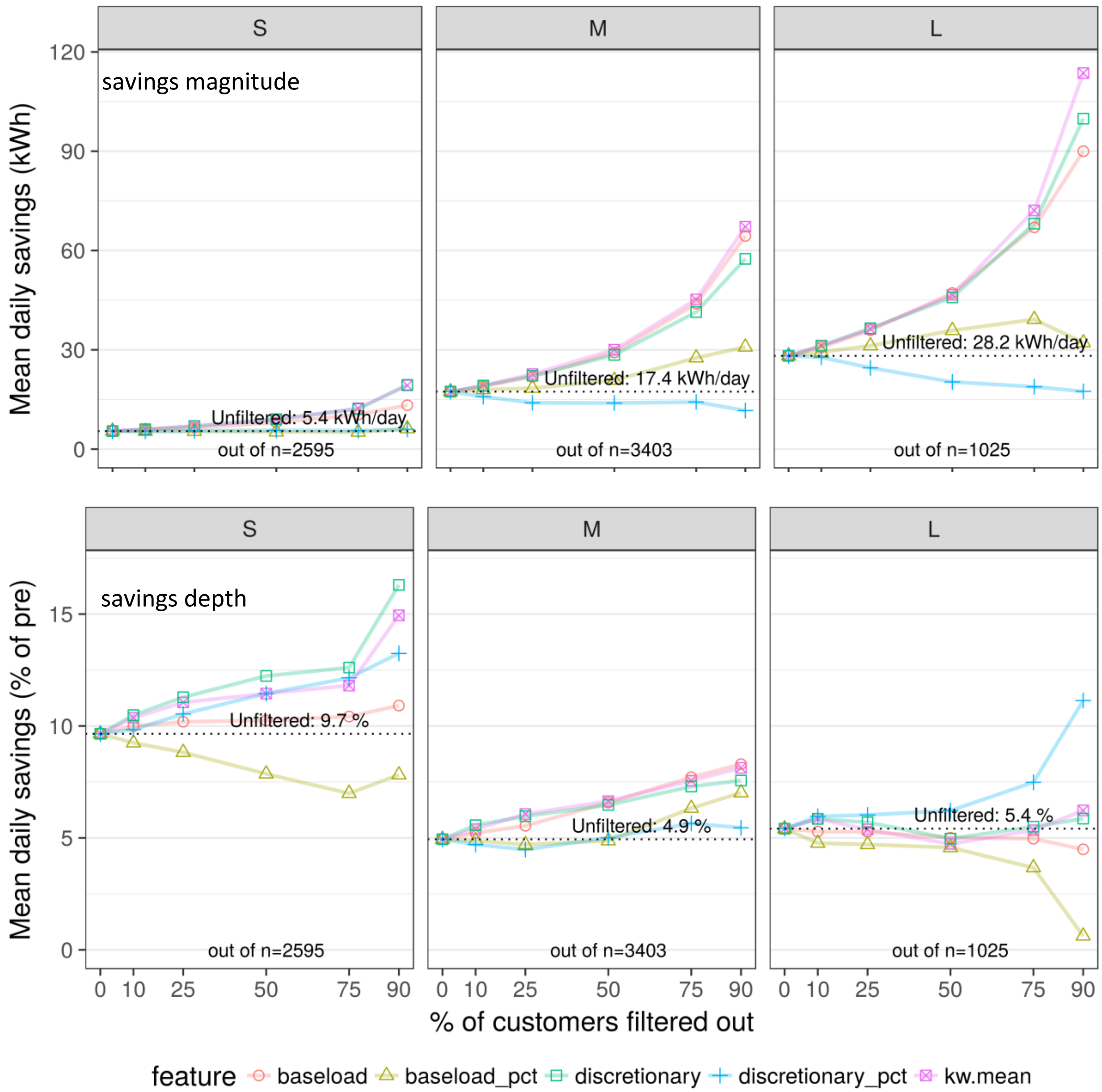
The simplest framing of DI program results is that size, meaning total consumption, is a good predictor of average program savings. If proxies for size are to be the focus of DI program targeting, one might wonder how well features can predict savings within customer size categories. Figure 6 illustrates the results of running feature filtering within the major consumption “size” categories S, M, and L, each in a separate panel. The dashed horizontal lines show the average savings for each size category. Size is defined by the following annual electric and gas usage thresholds:

Large: ≥ 500,000 KWh or ≥ 250,000 Therms
Medium: 40,000 - 500,000 KWh or 10,000 - 250,000 Therms
Small: < 40,000 KWh or < 10,000 Therms
Unknown: Insufficient data (<12 months)

Even within size categories, the best performing filters are related to the magnitude of consumption, with the simple measure of total consumption turning in the best performance. Instead of capturing all of the targeting potential related to total consumption, the **size categories are complimented by the more precise, individual consumption metrics**.

The figure also suggests that targeting cannot squeeze enough program performance out of small customers to match the magnitude of the large ones, but projects for smaller customers tend to have deeper savings. Filter depths of 50% or greater applied to size M customers allow their savings to surpass the average across all size L customers and 3 times as many M customers as L customers participated in the DI program. For depth of savings, it is also noteworthy that each size category has a different top performing consumption feature filter. We hypothesize that this may be the result of different project implementation strategies being employed for customers of different sizes.

Figure 6: DI program filter performance within customer size categories. Dotted horizontal lines are the average savings for each size category.



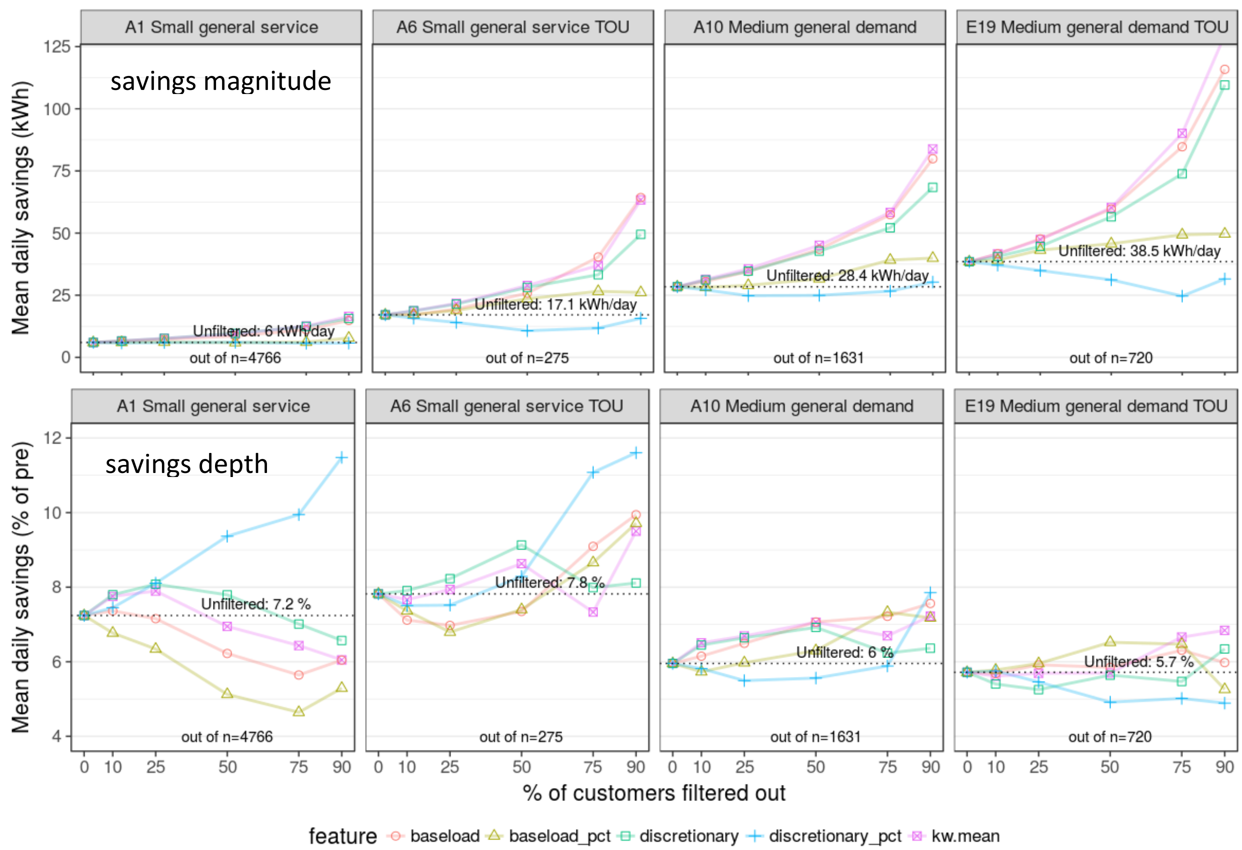
6.2.1.4 DI filters within rate type

A categorical targeting strategy that shows notable DI program performance differences is the utility rates of participating customers. Figure 7 presents the results of feature filtering within the rate categories. As with the size categories, the unfiltered averages are smaller for the small customers (rates A1 and A6) and larger for the medium customers (rates A10 and E19).[16] Customers on time of use (TOU)

[16] https://www.pge.com/tariffs/assets/pdf/tariffbook/ELEC_SCHEDS_A-1.pdf, https://www.pge.com/tariffs/assets/pdf/tariffbook/ELEC_SCHEDS_A-6.pdf, https://www.pge.com/tariffs/assets/pdf/tariffbook/ELEC_SCHEDS_A-10.pdf, https://www.pge.com/tariffs/assets/pdf/tariffbook/ELEC_SCHEDS_E-19.pdf

rates (A6 and E19) easily out-perform their standard rate counterparts (A1 and A10, respectively). It also appears that the smaller A6 customers can match the performance of medium-sized A10 customers at filter depths of about 50%, but the sample of A6 customers is too small to have high confidence in those specific numbers. Due to the strong showing from TOU customers, it is important to keep in mind the difference between the self-selection effect (customers savvy enough to opt-in to TOU rates are possibly more likely to opt into EE programs and do well in them) and the pricing effect of the rates. As customers are steered into TOU rates by default, this distinction will determine the durability of the TOU correlated savings documented here.

Figure 7: DI program filter performance within rate type categories. Dotted horizontal lines are the average savings for each rate category.

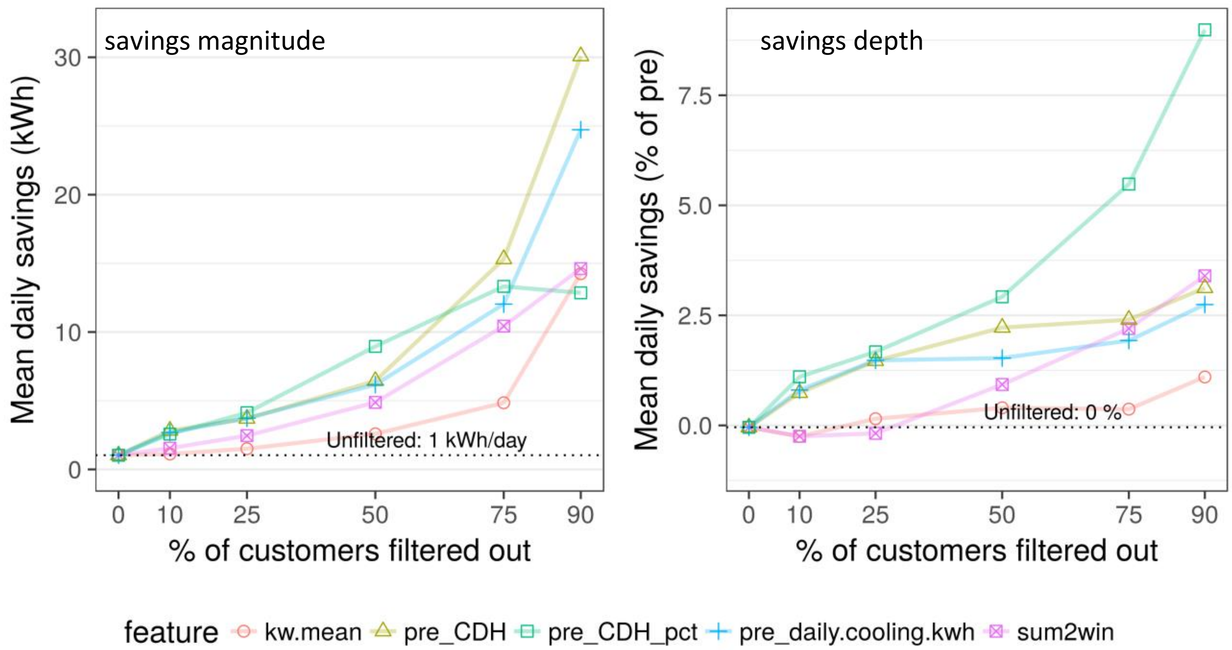


6.2.2 HVAC program

The features whose performance is plotted below for the HVAC program include several that correlate with AC usage, like the degree to which consumption increases in the summer (*sum2win*) and weather normalization outputs for temperature sensitivity (*pre_CDH* and its normalized form *pre_CDH_pct*) and AC consumption (*pre_daily.cooling.kwh*). These results also include *kw.mean* to facilitate comparisons with the DI program results.

The selected HVAC program features are plotted in Figure 8 with filter % from the 10th through the 90th percentile on the x-axis. The left panel places the magnitude of mean daily savings on the y-axis. The right panel places the savings as a percentage of pre-intervention total consumption on the y-axis. The dotted horizontal lines are the HVAC program average savings.

Figure 8: HVAC program filter performance as magnitude of mean daily savings (left) and depth of mean daily savings (right) vs. the % of customers filtered out by HVAC feature filters. Dotted horizontal line is the unfiltered population average savings.



The unfiltered average savings is very small for the HVAC program, but AC-correlated filters are able isolate premises with significant savings – for example, an average savings of 10-15 kWh/day with 75% of customers filtered out. In contrast to the results from DI program filters, *kw.mean* (or total consumption) is only weakly predictive of savings and features that amplify the magnitude of HVAC program savings also amplify the depth of savings. This suggests that HVAC loads and efficiency potential can vary to a significant degree. For example, restaurants have sizeable ventilation requirements in their kitchens, resulting in HVAC loads much larger than another business of similar size. Updates to controls and scheduling, better system zoning, and more efficient equipment can all contribute to savings.

The *pre_CDH* feature based on modeled temperature sensitivity performs best in amplifying the savings magnitude at more aggressive levels of filtering but is out-performed by *pre_CDH_pct* at lower levels. The *pre_CDH_pct* filter performs best in amplifying the depth of savings hands down. The regression approach that produced the '*pre_*' features is a more precise way to isolate the temperature effects from other sources of variability than the other thermal filters.[17]

Even though efficiency potential is often observed to increase with the magnitude of consumption, features that correlate strongly with total consumption, including *kw.mean* are much weaker predictors of savings outcomes than the more specialized thermal features in the HVAC program. In contrast, even

[17] The regression model is also the basis for the evaluation of savings in the first place, so the pre_CDH feature may also have a structural advantage over the others.

though HVAC loads might be considered as discretionary loads that are above and beyond the baseload, the *discretionary* and *discretionary_pct* features are poor performers in the DI program (not shown).

The DI results compared to the HVAC results showcase the need for different targeting strategies for programs with different intervention types. In the case of DI, the end-uses correlate well enough with total consumption that total consumption turns out to be one of the best feature filters in terms of savings magnitude, but the depth of savings was more accurately predicted by end-use appropriate intensity metrics. In the case of the HVAC program, features more finely tuned to pick out cooling loads from meter data are better options.

6.3 Savings by customer categories

6.3.1 DI program customer sub-categories

Table 6 provides highlights of the key findings and best performing sub-groups within the DI program. See the "Tabulation of DI savings by customer characteristics" section of Appendix C of the research report for supporting details. The summary table supports the following conclusions:

- DI savings tend to track with customer size (energy consumption).
- Smaller customers tend to save a greater percentage of their pre-intervention consumption than larger customers, indicating deeper savings are being achieved at such sites.
- Customers on the "Medium" rates (E19 and A10) significantly out-perform the average DI program participant, which is consistent with the correlation between customer size and savings.
- Customers on TOU rates tend to save more energy than their peers. Customers on E19 save an average of 168% more than the average DI program participant! Possibly more striking, the group on the TOU rate amongst "Small" customers (A6) out-saves the general population by 19%, while the standard "Small" rate (A1) saves 58% *less* than the general population.
- Customers who received lighting measures alone out-save customers who received only refrigeration measures, but each of these sub-groups perform worse than the full DI program average. When lighting and refrigeration are done together, the savings are a notable 105% greater than the DI program average. Beyond the additive impacts of addressing both end-uses, this might have to do with the types of customers, like grocery stores, that are eligible and elect to complete both interventions.
    - This general pattern holds across size categories, with average savings from "lighting and refrigeration" interventions for large (L) customers being a remarkable 269% larger than the DI program average.
- Lighting interventions tend to save a little over 7% of pre-intervention total consumption, but refrigeration saves just over 2%.
- Refrigeration savings as a % of pre-intervention total consumption tend to decrease with customer size (refrigeration loads are usually a smaller % of the total load for large customers). The pattern is a little less clear for lighting. At nearly 10%, lighting savings as a % of pre-intervention total consumption are greatest for size S customers, but lowest at 5% for size M.
- Within the lighting category, LEDs (24% higher than average) substantially out-save linear fluorescents (49% less than average), and CFLs (85% less than average). Projects involving both LEDs and linear fluorescents (55% higher than average) are observed to have the best savings performance.

- Savings from walk-in coolers are modest, but savings from walk-in coolers and controls together are nearly 140% greater than all refrigeration projects. Controls appear to be key drivers of walk-in cooler savings.
- Savings from sports, entertainment, and recreation venues, non-department stores, and more technical manufacturing are all around 70% greater than DI projects in general.
- Savings go up and down a bit from year to year, but in expectation, every year will return savings at about the average for all DI. However, customers whose participation spans more than one year save 130% more than typical DI participants. This is likely partially due to the cumulative impact of multiple interventions and partially to the self-selection effect of highly motivated customers participating multiple times in programs.

Table 6: DI program participant savings by customer characteristics.

category	Sub-category group	premise count	daily savings as % of pre-intervention usage (%)	daily savings (kWh)	% gain[18]
All participants		7497	6.77	14.39	0
size	L	1025	5.41	28.17	96
	M	3403	4.94	17.36	21
	S	2595	9.65	5.44	-62
rate	E19 Medium general demand TOU	720	5.72	38.51	168
	A10 Medium general demand	1631	5.95	28.41	97
	A6 Small general service TOU	275	7.82	17.12	19
	A1 Small general service	4766	7.24	6.01	-58
DI tech. family	LIGHTING and REFRIGERATION	1005	7.80	29.46	105
	LIGHTING	5331	7.57	12.77	-11
	REFRIGERATION	772	2.32	10.10	-30
	LIGHTING and REFRIGERATION & size L	112	7.02	53.10	269
lighting	LED and LINEAR FLUORESCENT	645	9.7	19.8	38
	LED	2065	9.0	15.8	10
	LINEAR FLUORESCENT	995	3.3	6.5	-55
	COMPACT FLUORESCENT	40	-0.6	1.9	-87
refrigeration	REFRIGERATION CONTROL and WALK-IN COOLER	42	9.0	24.0	67
	WALK-IN COOLER	687	1.9	9.6	-33
program year	more than one	386	9.32	33.23	131
NAICS code	Arts, Entertainment, and Recreation[19]	218	8.75	24.74	72

[18] % Gain is the percentage savings greater than the average savings across all premises. A % gain of 100 corresponds to a doubling of savings.

[19] Arts, Entertainment, and Recreation is concert halls, sports venues, museums, etc. It includes all NAICS codes starting with 71: https://www.bls.gov/iag/tgs/iag71.htm

category	Sub-category group	premise count	daily savings as % of pre-intervention usage (%)	daily savings (kWh)	% gain[18]
	RETAIL TRADE - 1[20]	1786	9.26	24.27	69
	MANUFACTURING - 3[21]	62	4.60	24.26	69

6.3.2 HVAC program customer sub-categories

Table 7 (below) provides highlights of the key findings and best performing sub-groups within the HVAC program. See the "Tabulation of HVAC savings by customer characteristics" section of Appendix C of the research report for all the supporting details. The summary table supports the following conclusions:

- HVAC program savings are loosely correlated with customer size, but the effect is not nearly as strong as it is for the DI program.
- One might expect HVAC saving to strictly correlate with hotter climate zones. However here we see that the hottest climate zones, cz12, and cz13 are not the strongest performers. The northern Central Valley, cz11, and northern coast including the Bay Area, cz03, perform best. Note that a quality and maintenance program addresses other aspects of air and water distribution in addition to the AC units systems themselves.
- As with DI, the "Medium" rate class, especially the TOU version, out-performs the general population of HVAC program participants.
- Chiller projects dramatically out-perform other types, with unitary AC projects associated with above average savings. Notably, the quality maintenance interventions are associated with below average savings – the actual average is negative, but this is likely just a symptom of the variability being so much higher than the savings so that outliers can dominate average outcomes.
- As with DI, program year 2015 had noticeably better results than others. Unlike DI, premises with interventions spanning more than one year did not perform better than their peers. At least some of the year-over-year variability in outcomes could be due to imperfect weather normalization.
- Accommodation and Food Service has the largest average savings by far, followed by Public Administration and Retail Trade – 2. None of these has a very large premise count, making these results more prone to be impacted by outliers.

[20] Retail Trade - 1 is basically non-department stores. It is composed of all the NAICS codes starting with 44 at this location: https://www.bls.gov/iag/tgs/iag44-45.htm. All code starting with 45, basically department stores, are Retail Trade – 2.

[21] Manufacturing – 3 is more technical manufacturing, including the production of metal products, machinery, and electronics. It is all NAICS starting with 33 here: https://www.bls.gov/iag/tgs/iag31-33.htm

Table 7: HVAC program savings by customer characteristics.

category	Sub-group	premises	daily savings (% of pre)[22]	daily savings (kWh)	% AC gain[23]
All participants		1193	-0.04	1.03	0
size	L	694	0.27	1.52	48
	M	327	0.66	0.96	-7
	S	119	-2.47	-0.36	-135
climate zone	cz11	98	1.51	2.30	122
	cz03	282	0.55	2.06	99
	cz12	345	-0.59	1.78	72
	cz04	202	0.41	1.33	28
	cz13	163	-2.09	-1.75	-269
	cz02	65	1.43	-3.23	-413
rate	E19 Medium general demand TOU	249	0.36	2.76	167
	A10 Medium general demand	354	0.16	1.94	88
technology	CHILLER	45	0.7	7.2	593
	UNITARY AC/HP	721	0.3	1.9	83
	QUALITY MAINTENANCE	90	-0.4	-2.3	-323
program year	2015	235	-0.11	5.03	387
	more than one	189	-0.43	-0.98	-194
	2013	224	-1.22	-3.48	-437
NAICS code	Accommodation and Food Service	49	2.23	11.49	1011
	Public Administration	34	4.20	7.72	647
	RETAIL TRADE - 2	39	-1.39	7.70	645

7 Discussion

7.1 Different ways of evaluating savings

Savings are achieved one project at a time, and they can be tabulated in different ways, depending on the context. There are four primary tabulations of program savings relevant to discussion of targeting and program performance.

7.1.1 Average per-customer savings (magnitude)

These are reported in energy units and are useful when contemplating the energy benefits of data-driven targeting and their customer-level impacts. These are particularly relevant because typical program benefits are computed, and rewarded, in energy terms. This is especially true in the context of pay for performance programs.

[22] This is the mean of each customer's % of pre savings, not the total savings as a % of the total pre-period consumption, thus the average values for some groups can be negative.
[23] Percentage AC savings greater than the average AC savings across all premises. A gain of 100 corresponds to a doubling of savings.

7.1.2 Percentage per-customer savings (depth)

These are the site-level savings as a *percentage* of pre-program energy consumption. All other things being equal, if greater savings depth can be achieved, program administrators and implementers can achieve the same savings across fewer projects and therefore stand to save budget on recruitment and project fixed-costs. Achieving greater savings depth also positions a program better for evaluation. Deeper savings are more readily differentiated from natural variability at-the-meter and therefore easier to attribute to program interventions.

7.1.3 Total program savings

Total savings are a multiple of per-customer savings and the number of customers. The true value of many EE programs is best expressed as their total savings. In the context of targeting, where some fraction of potential participants is ignored, it is important to understand at what point the filter depth is interfering with a program's ability to keep its project pipeline full. See "Considerations of targeting power" for more discussion of this issue.

7.1.4 Savings over time and location

The grid is a dynamic and complex system. The value of load shifting and reduction varies substantially over time and location. Therefore, strategies for grid operations are needed that accommodate rapidly growing sources of renewable energy and other distributed energy resources. It is not hard to imagine a time when EE resources will be valued principally according to the timing and location of their savings. The interval meter data used to derive feature filters for this study will also be integral to the analysis that attributes EE savings accordingly. Moreover, consumption features should be able to predict the timing and location of savings as well as they predict its magnitude and depth. The advent of time and location varying resource valuation in EE program will make the case for data-driven targeting even stronger.

7.2 Considerations of targeting power

Policy makers and implementers looking to keep their EE project pipelines full might be concerned that data-driven targeting filters eliminate potential program candidates. Is there a point where targeting will run into practical limits? The obvious answer is that yes, this is entirely possible and the solution is to ease filter restrictions to the point where there are enough candidates to run programs at full capacity. However, the kinds of EE programs amenable to at-the-meter evaluation typically touch thousands of customers per year out of a pool of hundreds of thousands or millions. Those programs are most often constrained by implementation capacity and the ability to recruit, not by the pool of potential participants.

On the other hand, the more aggressive filtering becomes, the more likely that high and low "savings" outliers (defined as those caused by non-program changes in consumption) will result in misalignment between expectations from this research and program results. At 50% of customers eliminated, the influence of outliers is doubled. At 90% it is multiplied by 10. Although we have taken steps to eliminate outliers within this research,[24] we cannot know for sure what caused the unusual consumption changes

[24] See the "Cleaning data" and sub-section under "Methods" and Appendix B of the research report for more details.

we flagged with the information available to us. For many programs, concerns about the influence of outliers will manifest before there is slack in program capacity.

Here are some real-world numbers: there are approximately 400,000 SMB customers (account + premise) that are eligible to participate in the programs studied. Based on discussion with the program manager, regional DI serves approximately 5,000 customers per year. Implementers report that their conversion rates of approximately 50% for the DI program (higher or lower for particular implementers) so we can assume that 10,000 customers are approached each year.

For HVAC CQM, approximately 2,400 customers participated in 2016 and another 2,400 in 2017. If we assume a more conservative 25% conversion rate, approximately 9,600 customers need to be approached to get 2,400 participants.

Thus, DI requires outreach to 2.5% of all SMB customers a year and HVAC requires outreach to approximately 2.4%. We can also assume there are practical limitations on customer outreach - geography, business type, etc. It is probable that meeting those requirements would require a larger pool of candidates to draw upon. Nevertheless, it appears that programs like those examined here could be fairly aggressive with targeting.

A level of 50% targeting would be a reasonable starting point, but it may be possible to go to 75% filtered out or higher, again cautioning that at very aggressive levels of targeting the influence of outliers increases.

7.3 The relevance of total consumption versus other usage attributes

As a rule of thumb, EE program savings are often assumed to scale with total energy consumption (i.e. customer size), but there are many cases of high energy consumption with efficient and well operated systems and there are many cases of smaller customers with significant waste and/or inefficiency or who are dominated by a program-relevant end-use. A more accurate rule of thumb would be that savings tend to scale with the magnitude of the relevant end-use(s).

In reality, beyond rules of thumb, savings scale with the magnitude of correctable waste and inefficiency in the relevant end-uses and the interactions between the premise occupants and those end uses. But that presents a practical problem: end uses, waste, inefficiency, and occupants are not directly observed. The best resource widely available for estimating these core drivers of efficiency potential is interval meter data and the question of how well that data can support program improvements is an open question – and the subject of this research.

Furthermore, savings are not the only, or at times even the most important metric of EE program success. Cost effectiveness calculations require understanding the costs associated with implementation in opposition to the savings achieved. Social, policy, or corporate goals will often dictate that a broad mix of customers should be served by programs. The most well-considered and effective real-world targeting will not be one dimensional and even the most aggressive profit maximizing approach to efficiency would benefit from targeting the least expensive savings, not just the largest savings projects.

7.4 Program design considerations

The features used here to develop customer targeting strategies shed light on the ways in which customers use energy and their opportunity for savings. On one hand, an implementer may have a prescriptive program and wish to find customers who can benefit most from that specific set of interventions. On the other hand, an implementer may be able to take a more flexible, individualized approach to treating customers. Using lighting as an example, very high mid-day discretionary usage may signal a potential opportunity for de-lamping over-lit rooms in addition to swapping out inefficient lamps. In contrast, high discretionary usage outside of normal business hours may signal an opportunity for savings via lighting controls that switch off or dim lights based on time of day or occupancy. In other words, some project approaches may be best left to simple efficiency improvements that scale with total consumption, while others would benefit from the incremental investment to go deeper with the addition of controls. Often the answer to which strategy will serve the customer and program best is written in the feature data of individual customers.

7.5 The limitations of NMEC savings calculations

Confidently attributing savings to program actions as opposed to other factors is a perennial challenge in energy efficiency, both in traditional and NMEC programs. As we describe above, standard pre/post at-the-meter NMEC "savings" calculations reflect all changes that occurred between the pre and post periods regardless of cause. As we have seen in our control groups, there can be very broad distributions of "natural variability," presumably caused by a combination of mean-zero fluctuations in consumption from year to year, site-specific changes in energy use intensity (i.e. business expansion or contraction), and long-term trends in consumption (i.e. LED lighting adoption or more efficient computing). On top of those actual fluctuations, NMEC also adds the uncertainties associated with weather normalization methods.

Potential sources of NMEC savings bias:

1. In large samples, mean-zero fluctuations and site-specific changes in consumption are often assumed to cancel out across premises (for every site with an increase, there is a corresponding site with a decrease). However, shared factors like droughts, prevailing economic conditions, etc. can cause shifts in consumption that do not cancel out. Further, these exogenous factors can impact certain customer segments more than others.
2. Similarly, a weather normalization model that is overly temperature sensitive or was trained using relative cool weather data, could create systematic biases when trying to normalize consumption for a relatively hot year.
3. Actual trends in energy consumption (i.e. LED adoption or plug load growth) can also undermine the assumption that models trained on pre-period data can provide unbiased estimates of the counterfactual conditions for the post-period.

The method used in this first analysis was among the simpler methods that could be used to estimate what usage *would have been* in absence of an energy efficiency intervention. For example, it is not a time-series method that can account for seasonality (beyond temperature) or trend.

With the advent of machine learning, additional powerful methods may prove capable of better predicting usage. These methods should be explored including their ability to isolate and incorporate trends, seasonality, and the effect of group membership as well as exploiting techniques such as using multiple models to create more reliable prediction.

7.5.1 Synthetic controls for NMEC savings

Another potential solution to the biases listed above is to identify control groups that experienced the same fluctuations, weather, and energy trends as the participants. To serve as controls, customers would also need to respond to all those conditions in the same manner as participants would have. For example, if participants are all higher than average AC users, their controls would need to be as well.

The best controls come from randomly controlled trials (RCT) that randomly assign participants and controls out of a single, large pool of customers. In those cases, there is no risk that some unobserved characteristic (politics, energy awareness, past participation, etc.) has influenced the participants. In the real world, there are limited opportunities to run RCTs.

When RCTs are impractical or impossible, there are several promising methods for generating synthetic control groups using methods that match between participants and controls based on various metrics of similarity. This matching can be done at any time and can therefore be applied to a program in retrospect. Unfortunately, there is no way to rigorously *prove* that a matched control is going to behave as the participants would have, but there are some practical tests that can be run to evaluate both the matching methods and the groups they select.

With all this as context, it does appear based on this and related research that computing NMEC savings without control groups runs a significant risk of counting savings that were not caused by the program interventions or ignoring savings that were. More work should be done on methods for forming synthetic control groups for at-the-meter savings calculations and methods for evaluating the resulting matched groups.

As a corollary, NMEC savings should only be expressed as a point estimate with error bars. Those error bars should shrink with larger samples sizes and controls, but will never be zero, and the errors on individual projects will likely be so large as to often make individual assessments extremely difficult.

7.6 Additional considerations

Targeting logic selects some people and leaves others out. As a matter of policy, we should be concerned about ensuring that the inclusions and exclusions continue to serve the greater public purpose of efficiency programs. At first blush, one might assume that targeting will result in programs that just focus on the biggest customers, who represent a modest fraction of the full rate base. However, targeting based on individual consumption spans a variety of customers far better than targeting based on broad categories. For example, if your HVAC program targets hot climate zones, you are likely to pick up some serious AC users, but you will also be including some customers without much AC at all and excluding the higher AC users from other climates.

It is important to note that untargeted programs are not evenly applied. Customers typically self-select to enroll, introducing a bias towards the profile of the most motivated participants. To the extent that targeting is merit based and results in contact with atypical program participants, it would naturally improve the diversity of characteristics of participants.

One category of customers that can't be well served by data-driven methods is anyone who just started a new account. Without the data to analyze, data-driven programs could have a blind spot for customers that move frequently. However, many aspects of efficiency are more accurately tied to the premise than the customer, so it may be possible to overcome this blind spot to some extent by looking at the location's consumption history rather than the specific customer.

In the case of targeting based on retrospective analysis, it is also important to recognize cases where targeting could amplify one or more aspects of current practice while ignoring the potential gains from a strategy that hasn't yet been tried. This means that it is important that data-driven programs are designed to ensure that they meet the varied needs and savings potential of all customers. In that setting, the targeting logic that excludes customers from one program would help to identify the other program(s) that are the best fit for each customer.

Whatever biases and blind spots targeting rules have, the best response is to improve the rules. Targeting can be used to recruit a population of program participants with any number of constraints, including income, business type, etc. and, using the methods of retrospective analysis described in this report, the mix of program participants can be monitored and adjusted as needed with more precision that untargeted programs are currently able to provide.

8 Conclusions

This research demonstrates that significant gains in at-the-meter EE program savings can be achieved using targeting based on energy consumption features and customer characteristics. We estimate that well-executed targeting can improve per-customer average savings by a factor of 2-3x by pre-screening potential participants using data-driven targeting methods described here and focusing recruitment efforts on the most attractive 25-50% of potential customers.

These results are based on practical calculations that can be undertaken by a wide variety of program planners and implementers, provided they have access to meter data, and the lessons already learned can be applied even without the benefit of meter data, for example by using coarser metrics derived from billing data.

Because this research was performed retrospectively using data from actual programs, we can state with confidence that the targeting gains we have observed are additional to real world best practices already in place.

Because the savings calculations were based on NMEC pre/post comparisons, our methods and findings are particularly relevant to the planning and evaluation of pay-for-performance programs.

Based on the strength of these findings, we recommend follow-up in two areas:

First, every program is different, so **more program data should be retrospectively analyzed** using these or similar methods to expand the body of knowledge on generalizable and program-specific findings on the drivers of program savings. These methods are well suited to **applications in both research and evaluation settings**.

Second, program designers, planners, and implementers should **incorporate data-driven targeting into their program recruitment strategies**. There are many details, potential pitfalls, and synergies that can only be worked out in the field. For example, is data-driven targeting best suited to help improve the performance of existing programs or should programs be designed with individualized information in mind from the ground up? However, based on the savings improvements we've documented, efforts along the lines of this work should out-perform current practice, with improvements to savings and cost effectiveness that make the effort well worth it.

More broadly, the end goal of data-driven programs should be the **tailoring of efficiency services to the specific needs, waste, and inefficiency of individual customers**. Properly done, individualized targeting, diagnosis, education, and support has the potential to significantly increase realized benefits of program interventions, unlock program specialties that are not cost effective when prescriptively applied, and improve customer satisfaction. Further, the insights and implications derived from customer targeting research have the potential to drive the frontier of our collective understanding of how customers use energy and what can be done to make that use more flexible at a time when grid operators are looking to demand side resources for greater flexibility in time and location.

There is valuable information about the drivers of program savings and the wide range of "energy behaviors" of customers locked away in the customer and intervention data of past programs. We now have the tools to unlock that information, with a focus on increasing benefits for both programs and customers.