Memo to:

Massachusetts Program Administrators Research Team and Energy Efficiency Advisory Council EM&V Consultants

From: Michelle Marean and Ken Agnew, DNV GL

Date: July 22, 2019

Copy to: Wendy Todd, DNV GL

Statistical Memo: Methods and Evaluation of Control Measures Phase 2 Consumption Data Analysis

1 EXECUTIVE SUMMARY

Building automation systems (BAS) have become a popular measure in the food service segment in Massachusetts. One large coffee chain franchisor has accelerated installations in the quick-serve food service segment by marketing an energy management program to their franchisees. This program features BAS and a suite of additional energy efficiency measures including lighting, heating, ventilation, air conditioning (HVAC), and process end uses. The Massachusetts PAs are interested in quantifying savings from this important measure.

The Evaluation of Control Measure study was driven by the search for a cost-effective way to estimate savings in a sample of BAS participants using readily available utility consumption data. BAS measures, installed alone, are generally considered challenging to evaluate given the diversity, complexity, and interactivity of the controls that produce savings. BAS measures installed with the additional energy efficiency measures, with the additional level of interactivity, represent a degree of complexity for estimating savings beyond typical energy consumption data analysis (ECDA) and engineering analyses.[1]

The first phase of the Evaluation of Control Measures study assessed the feasibility of using monthly utility consumption data for estimating savings for BAS installed as a standalone measure. An unresolved question was whether monthly utility consumption data could be used to effectively estimate BAS savings in targeted small commercial offerings. A simulation exercise was used to calculate the level of precision and the degree of bias with which monthly data could estimate BAS and associated measure savings.

Phase 1 of the Evaluation of Control Measures study established that an aggregate, ECDA or billing analysis approach including a comparison group was both:

- Feasible, with respect to precision, given the number of BAS being installed and the energy usage characteristics of these kinds of sites, and
- An approach that addressed the potential for biased savings estimates due to non-stable consumption, year over year.

In Phase 2, site- and measure-level savings estimates were developed using consumption data for all available BAS-enabled sites. Phase 2 attempted to produce BAS-specific savings estimates at the measure level despite the presence of additional EE measures. In phase 2 we found that:

- CDA using an aggregate whole facility approach is a technically feasible way to estimate savings for a challenging measure such as BAS, if a BAS has been installed without additional energy efficiency measures. However, given the likelihood that multiple measures are installed with BAS, site-level

[1] The term “billing analysis” is typically used in a residential context. Although this analysis resembles a billing analysis in some ways, the term “energy consumption data analysis” is used here to distinguish its application in a commercial context.

estimates are not particularly useful to the PAs for estimating savings from BAS in the absence of other measures.

- For measure-level estimates, the combination of increased standard errors and low estimated BAS savings produced results that were not satisfactory with respect to precision.

1.1 Findings

- Two different ECDA approaches produced whole facility (site-level savings) estimates of similar magnitude. These results include the combined effects of BAS and all other installed measures. The forecast approach indicated annual savings of between 17 MWh and 18 MWh, calculated with 12-month rolling averages. The pre-post approach estimates savings for 124 sites of just over 18 MWh.
- The forecast approach illustrates the seasonality of the savings from the full set of installed measures. Summer savings are double winter savings and savings vary more from summer to winter than site-level consumption itself. At all sites, other than the two with HVAC measures, BAS is the only measure with a direct impact on the HVAC system. All non-HVAC measures, including BAS, could lower internal heat gains thus lowering cooling consumption.
- Using the pre-post statistically-adjusted engineering (SAE) approach for the measure-level regression indicates BAS measure savings of just over 5 MWh per site, about a quarter of the expected (tracked) gross savings. The savings could range from 1 MWh to 9 MWh, and considering the low precision, do not support the savings claims of the measure.
- If BAS installations had delivered the expected magnitude of tracked gross savings, an ECDA based on a participant population of just over 100 sites would have delivered high-precision savings estimates. Under the observed conditions, with BAS installed along with a range of other energy efficiency measures, the ECDA approach could still deliver a result with reasonable confidence if savings were in the expected range.
- Despite the shortcomings of the measure-level regression result for BAS with respect to precision, it clearly indicates that BAS savings are below expected gross savings.

1.2 Next steps

The BAS requires extensive end-use level monitoring to support the controls. It captures high-frequency kWh data for all major end uses along with temperature and setpoint information, and appears to include an overall site-level load. However, to our knowledge, the implementer collects no pre-installation consumption data. If the goal is to establish a high-precision estimate of BAS savings, potential next steps include a baseline study using BAS-generated data, and an engineering analysis also leveraging BAS-generated data.

A baseline study should collect metered data on all HVAC and non-HVAC end uses for a limited number of sites for a minimum of six months before controls are implemented. Special arrangements would need to be made with the facility owners so that implementation could be delayed. An engineering analysis should include on-site metering and data collection to validate the extensive BAS-generated data. The validation of BAS-generated data would be essential to assess whether any BAS data from the pre-implementation period is reliable enough to be used to develop savings estimates, or in building simulations. An engineering analysis would include on-site metering and data collection to validate the extensive BAS-generated data. The validation of BAS generated data would be essential to assess whether any BAS from the pre-implementation period is reliable enough to be used to develop savings estimates, or in building simulations.

For example, BAS-generated end-use data, if valid, should support reasonable estimates of control-specific savings for non-seasonally-correlated process measures (such as toasters and impingement ovens).[2] For more complicated systems such as HVAC, or systems with multiple measures such as controls paired with new and more efficient equipment, engineering analytical techniques can leverage the presence of some pre-control interval data to provide separate estimates of end-use efficiency and control-related savings estimates for the full set of end uses.

2 INTRODUCTION

Building automation systems (BAS) have become a popular measure in the food service segment in Massachusetts. One large coffee chain franchisor has accelerated installations in the quick-serve food service sector by marketing an energy management program to their franchisees. This program features BAS and a suite of additional energy efficiency measures including lighting, heating, ventilation, air conditioning (HVAC), and process end uses.

The Evaluation of Control Measure study was driven by the Massachusetts PAs interest in quantifying savings from this important measure and identifying a cost-effective way to estimate savings in a sample of BAS participants using readily available utility consumption data.

The varied new measures and the interactive interventions of BAS combined represent a degree of complexity beyond the typical ECDA or engineering analysis. This difficulty is due to a combination of changes in efficiency and usage characteristics and the potential thermal interaction of systems controlled by BAS. Our initial concern was focused on interactive effects caused only by BAS. We hypothesized that, for example, reductions in hours of use for food production process and lighting end uses would lower internal heat gains, affecting HVAC consumption separate from BAS-related changes in setpoint. Because of these complexities, it made sense to measure consumption reduction at the site level. The site-level approach would account for the interactive effect inherent in BAS projects and offer a valid estimate of the overall savings taking place at these sites.

During the initial data collection phase, we discovered that most of the participating sites installed additional energy efficiency measures at the same time as BAS. This introduced additional interactive effects, because BAS was now controlling new, more efficient equipment in addition to the pre-existing standard efficiency equipment. The combined savings of these overlapping efforts–controls applied to new, higher efficiency lighting for example—would be lower than a combination of the two effects applied independently.

The first phase of the study assessed the feasibility of using utility consumption data for estimating savings for BAS systems installed without other energy efficiency measures. It used a simulation exercise to calculate the level of precision and the degree of bias with which monthly data could estimate BAS and associated measure savings. In the study's second phase, site- and measure-level savings estimates were developed using consumption data for all available BAS-enabled sites. Phase 2 attempted to produce BAS-specific savings estimates at the measure level despite the presence of additional EE measures. This memo describes the results of the Phase 2 analysis.

[2] Impingement ovens work on the same principle as convection ovens except that an impingement oven uses jet nozzles, instead of fans, to force hot air on to the surface of the food. Impingement ovens heat food much faster and use more energy than traditional convection ovens. "Turbo Chef" is a common brand of impingement ovens.

3 METHODS

The methods section describes the source data used in the analysis, the number of sites in the analysis, and attrition rates, and the development of comparison groups.

3.1 SOURCE DATA

Billing data

We extracted billing data from January 2014 through December 2017 from the Massachusetts Evaluation Database. National Grid and Eversource provided additional consumption data for January through June 2018 to extend the analysis period and increase the number of eligible BAS enabled sites in the analysis.

Tracking data

Tracking data from the Massachusetts Evaluation Database provided the official record of BAS installations and additional energy efficiency measures installed at the same time as BAS. In addition, we extracted all instances of program activity for participants and potential comparison group sites for measures that were not installed concurrently with the BAS. The data indicates whether participants took part in any additional program activity other than a BAS project, and what program activity of any kind occurred at non-participant sites.

3.2 Available sites and attrition analysis

An attrition analysis tracks the preparation of the analysis datasets from the source billing and tracking data. The first row of Table 1 presents the counts of the coffee chain's sites identified in the Massachusetts Evaluation Database for National Grid and Eversource. The second row indicates the number of customer sites that remained in the analysis after data cleaning or removal of participants with program activity not related to a BAS project. The last two rows indicate the number of sites included in two different analysis datasets. The first dataset (row 3) includes any participant or comparison group site that has 12 months of pre-installation period data and at least one month of post-installation period data. The last row indicates the counts of sites with 12 months of pre-installation data and at least 11 months of post-installation data.

Table 1. Counts of available customer-level data and site attrition

Analysis step	Count of participants			Count of comparison		
	National Grid	Eversource	Total	National Grid	Eversource	Total
Projects in the Massachusetts Evaluation database, tracking data	121	72	193	0	0	0
Subset with good billing data	116	67	183	153	118	271
Matched analysis subset: pre-period forecast savings analysis	116	67	183	78	104	182
Matched analysis subset: pre-post modeled savings analysis	70	54	124	78	46	124

3.3 Comparison group development and outcomes

In preparation for the CDA, DNV GL developed two comparison groups for this analysis; a matched comparison group and a cluster-weighted comparison group. We used non-participating sites from the same coffee chain brand as the eligible sites for these comparison groups.

Matched comparison group–method

Comparison group development used a minimum distance algorithm to find optimal non-participant matches for each participating site among eligible comparison group sites. We matched comparison group sites to participants based on pre-installation electrical consumption. The comparison group sites were then assigned the same installation date as the participants with which they matched so that we could compare the same pre-post periods.

The minimum distance algorithm minimized the square root of the sum of squared differences across the 12 pre-installation months. We matched on modeled consumption for the same calendar period and fit to typical weather conditions. This approach leveraged the structural information underlying the model specification, while still comparing across models based on the same underlying actual weather. We randomized treatment sites before starting the process and selected comparison sites without replacement. A one-to-one ratio comparison group was all that was possible for this analysis. A larger set of eligible comparison group sites would improve the analysis.

Cluster-weighted comparison group–method

We also attempted a cluster-weighted comparison group approach that used all available eligible comparison group sites. The hypothesis behind the cluster-weighted approach was that the remaining sites do not vary widely from those that were included in the matched comparison group, and that the additional available population counteracts any negative effects from increased variation if the remaining sites do differ from the participant sites. To develop the cluster-weighted comparison group, we grouped all sites according to clusters based on monthly consumption patterns. We then weighted comparison group sites in each cluster to match the number of treatment group sites in that cluster.

Comparison group outcomes

This cluster-weighted approach proved to be much less effective than the matched comparison group approach. For this analysis, although we intended the full set of eligible comparison group sites to be similar in building construction, operations schedules, and load profiles, consumption variability within the overall eligible comparison group was substantially higher than within the participant group. The additional variability in the comparison group undermined the explanatory power of the model using the cluster method, while the additional sites appeared to introduce bias into the comparison group estimate of exogenous change.

The error statistics for the one-to-one match process supported this conclusion. Across multiple replications of the matching process, there was a slight but consistent degradation of the quality of matches as the matching algorithm moved through the randomly-ordered participants. The population of eligible matches with which to perform a one-to-one match was relatively small, and the additional sites beyond the one-to-one matches were increasingly different from the participant population with respect to consumption characteristics.

4 CONSUMPTION DATA ANALYSIS METHODS

For the ECDA we estimate savings using two regression approaches that highlight different aspects of the data; average site-level savings estimates from a forecast approach and a pre-post approach. The pre-post approach also supports a statistically adjusted engineering (SAE) model for measure-level results.

4.1 CDA protocols, methods, and approaches

As with Phase 1, we followed the Uniform Methods Project two-stage modeling protocols for whole facility consumption data analysis.[3] This is consistent with the original PriSM billing analysis method[4] as well as more recent efforts such as CalTRACK.[5] The approach used customer-level model specifications that, in addition to an intercept term representing baseload, included heating degree-days (HDD) and cooling degree-days (CDD) to represent a site with both heating and cooling end uses; just HDD (heating only) or CDD (cooling only); or no weather variables. For all specifications, HDD and CDD were modeled across a wide range of degree-day bases (a variable degree-day model) to facilitate a more tailored model fit. Ultimately, we chose a single model across different DD bases of all specifications that represented the highest adjusted R-square. Though these monthly models were relatively simple and based on only 12 data points, they were effective for putting pre- and post-installation consumption on the same weather basis.

Forecast approach. The forecast approach allows the use of all post-installation data by including sites with less than a year of post-installation data. It is conducted at the customer level (for all measures), then aggregated. The pre-installation period model is used to forecast an estimate of post-installation baseline consumption under post-installation weather conditions. We applied the forecast approach to both treatment and comparison groups, with the participant change of consumption (baseline minus actual load) adjusted by the comparison group change of consumption in the standard difference-in-difference structure. The change in consumption in the comparison group served as a proxy estimate of non-program related change across the analysis period.

The savings estimates from the forecast approach reflected the weather conditions during those post-installation months. This kind of actual-weather savings estimation is more common for estimating savings ex post, such as for home energy report programs, than as a basis for expected, or an ex ante savings value. Actual-weather savings estimates are usually close to weather-normalized savings estimates, but can vary from them if the weather during the post-installation period is mild or extreme.

Pre-post site-level and SAE model. For sites with sufficient post-installation data, a second-stage regression model compares normal-weather consumption from the pre-and post-installation customer-specific models for both participants, and comparison group to develop an estimate of savings under any weather condition. These results can be aggregated to yield an average site-level estimate of savings across the population, or broken out to measure-level estimates. The site-level savings approach estimates savings related to all measures, addressing all interactive effects between BAS and other measures. Given the installation of other measures, the measure-level model was the only way to isolate savings related to BAS

[3] Ken Agnew and Mimi Goldberg, Chapter 8: Whole-Building Retrofit with Consumption Data Analysis Evaluation Protocol. (2018). The Uniform Methods Project: Methods for Determining Energy-Efficiency Savings for Specific Measures, 8–13. Golden, CO; National Renewable Energy Laboratory. NREL/SR-7A40-70472. http://www.nrel.gov/docs/fy18osti/70472.pdf

[4] Fels, Margaret F, PRISM: An Introduction, Energy and Buildings 9 (1986): 5–18.

[5] CalTRACK Methods, Version 2.0, "CalTRACK is considered a best-practices guide to calculating meter-based changes in energy consumption." http://docs.caltrack.org/en/latest/methods.html

alone. Measure-level results are less reliable, as they are sensitive to the distribution of measure bundles and may not distribute interactive effects appropriately.

For measure-level results, if the expected savings for a particular measure vary across sites, the second stage model may be improved by including expected savings as an additional variable in the second-stage regression.[6] This approach is referred to as an SAE model. For the measures installed by BAS participants for this program, only lighting varied across sites sufficiently to justify the inclusion of the additional variable in the regression. We estimated a partial SAE model by including expected gross savings for lighting measures along with the full set of measure indicator variables. The inclusion of gross lighting savings in the model substantially improved the explanatory power of the model.

4.2 Installed measures at BAS enabled sites

Only 15% of the sites installed BAS without any additional measures. Table 2 shows the installed measure mix for BAS enabled sites. Most sites installed some combination of at least two of the measure groups, primarily lighting, refrigeration, and process measures. There were eight measure mixes among 124 participants. Table 1 shows the eight combinations and the number of sites in each combination. For example, there were 18 sites that installed BAS only, and six that installed BAS and process measures.

Table 2. Measure combinations for full-year-post participants

BAS	HVAC	Lighting	Refrigeration	Process	# of Site combinations
√	–	–	–	–	18
√	–	–	–	√	6
√	–	–	√	–	2
√	–	√	–	–	11
√	–	√	–	√	54
√	–	√	√	-	20
√	–	√	√	√	11
√	√	–	–	√	2
124	2	96	33	73	**124**

[6] In addition to the measure-specific dummy, measure-level expected savings are included in the regression where installed and with a zero where not installed. See the Universal Methods Project (NREL 2018) for a discussion on why expected savings should be included in addition to measure dummies rather than replacing them.

5 RESULTS

This section summarizes the results produced by the forecast and pre-post approaches using the one-to-one matched comparison group.

5.1 Forecast approach: predicted post vs actual

Figure 1 and Figure 2 provide monthly savings estimates using the forecast approach. Figure 1 provides the monthly savings estimates in kWh (left axis) while Figure 2 provides monthly saving estimates as a percentage of consumption (left y-axis). Both figures show results from April 2017 to July 2018. The solid red line is the monthly savings point estimate (kWh or percentage) surrounded by 90% confidence intervals. In both figures, the height of the bars (right y-axis) is the cumulative count of sites included in the average savings estimates.

The analysis includes 183 participants and starts in April 2017 when approximately 40 sites installed BAS. Relatively few participants installed BAS prior to January 2017. More than 100 BAS were installed by July 2017, and consumption data is available through June 2018. The reduction in site counts in July 2018 is an artifact of monthly billing periods around the cut-off of available data.

Figure 1. Monthly results for all participants in kWh

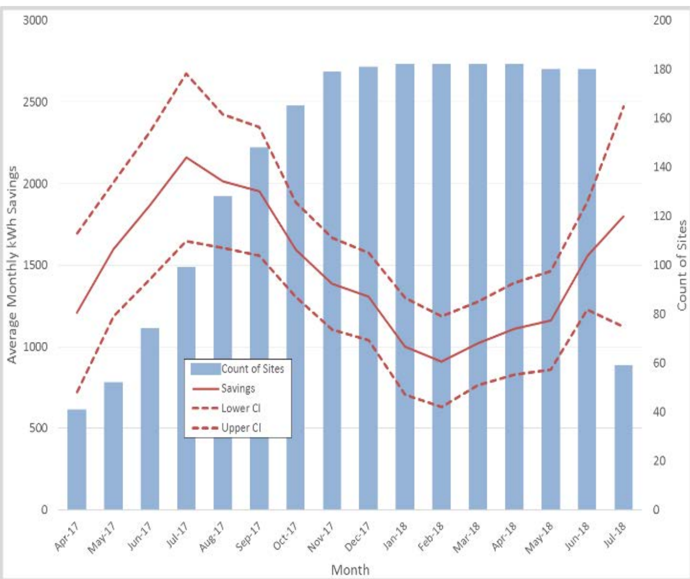

Figure 2. Monthly results for all participants as a % of total consumption

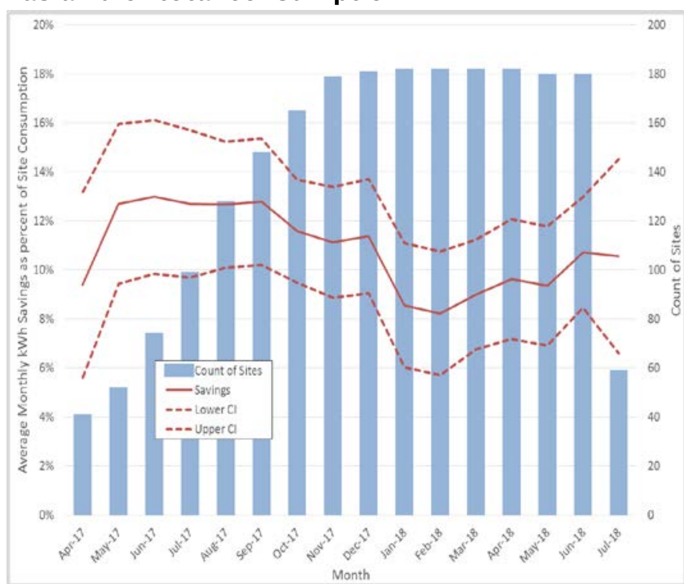

Figure 1 indicates all-measure, site-level, monthly savings of over 2,000 kWh per site for July 2017. Winter monthly savings decreased below 1,000 kWh from January 2018 to March 2018. The monthly kWh savings indicate seasonally correlated savings consistent with reductions in HVAC-related consumption. The results for the summer 2017 have wider confidence intervals because the number of installed participants was smaller at the beginning, when the program was ramping up.

Figure 2 shows savings of over 12% of total consumption during summer months, May 2017 through September 2017, and as low as 8% during the winter (February 2018). These results further highlight the importance of summer season savings. The increase in the percentage of savings during the summer season indicates that the cooling savings represent a greater portion of cooling consumption than do baseload savings of baseload consumption.

The extent of cooling-related savings is striking. Other than two sites with HVAC measures, BAS was the only measure with a direct impact on the HVAC system. All non-HVAC measures, including BAS, could lower internal heat gains, thus lowering cooling consumption. This is a prime example of the difficulty of correctly attributing savings to individual measures, which may require a more informed engineering approach.

The monthly results support annual estimates of savings of between 17 MWh and 18 MWh, calculated as 12-month rolling averages. The forecast approach, on which these monthly estimates are based, can be extended as an ongoing analysis to assess the persistence of savings.

5.2 Pre-post site- and measure-level savings estimates

There were 124 participating sites with at least 11 months of post-installation data. For those sites, it was possible to compare modeled annual consumption between the pre- and post-installation periods. The second-stage models applied to these data (site-level savings and SAE) provided estimates of site-level or measure-level, weather-normalized savings. Table 3 provides the savings estimates for the site-level and measure-level models. The measure-level version of the second-stage regression in Table 4 distributes the site-level expected gross savings estimates across five measure groups, including BAS.

Average site-level savings, which include savings from all installed measures, were estimated at 18,421 kWh. This site-level savings estimate has a low standard error of 1,589 and 90/15 relative precision. This is a strong result from an ECDA with so few sites and reflects the homogeneity of site consumption and savings across the sample of coffee chain sites.

These all-measure savings were approximately the same magnitude as the estimated ex-ante gross savings for BAS alone. These results, then, support an important part of the original hypothesis: that a high-precision estimate is possible with a small but homogeneous group of small commercial sites installing similar measures. If the savings at each site had been generated by BAS only, this would have been considered a successful analysis result. Savings across the sites would also likely have been more homogeneous and the savings estimate would have had even higher precision.

Table 3. Site-level and measure-level results

Savings Result	Savings (kWh)	Confidence Interval (+/-, kWh)	90/10 Relative Precision	Standard Error (kWh)
Site-level	18,421	(15,807, 21,035)	14%	1,589
Measure-level				
BAS	5,344	(1,274, 9,414)	76%	2,474
Lighting	8,490	(3,803, 13,178)	55%	2,850
Refrigeration	10,673	(6,359, 14,986)	40%	2,622
Process	5,298	(1,299, 9,297)	75%	2,431

Savings Result	Savings (kWh)	Confidence Interval (+/-, kWh)	90/10 Relative Precision	Standard Error (kWh)
HVAC	24,403	(10,747, 38,059)	56%	8,302

Measure-level savings estimates

To estimate the BAS savings when multiple measures are installed concurrently with the BAS required a measure-level second-stage regression to distribute savings to the measure level. The measure-level version of the second-stage regression in Table 4 distributes the site-level expected gross savings estimates across 5 measure groups, including BAS. As expected, the measure-level results demonstrated lower precision than the site-level results. However, except for the HVAC measure group (installed at only two sites), the standard errors were between 2,500 kWh and 3,000 kWh. This is greater than the site-level standard errors but still less than 2% of average site-level consumption. The degradation and variation in precision for the measure-level estimates was primarily driven by the smaller magnitude of the measure-level savings estimates.

The measure-level regression found average per-site BAS measure savings of just over 5 MWh, about a quarter of the expected BAS savings. The result is statistically significantly different that zero, with a 90% probability of falling between 1,274 kWh and 9,414 kWh. While this result does not demonstrate the desired magnitude of savings or level of precision, it indicates that BAS produced savings, although at a level below expected gross savings.

Table 4 Provides the same results in the context of average consumption and expected measure-level savings. The site-level savings represent 12% of average site-level consumption which is below expected gross savings. The BAS savings estimate represents only 3% of average site-level consumption and also below expected gross savings. Other than HVAC, which was installed in just two sites, the other measures ranged from 3% to 7% of average consumption. The widely varying realization rates may be a further indication that the measure-level results were less reliable.

Table 4. BAS savings from measure-level regression

Savings Result	Average Evaluated Savings (kWh)	Savings as a % of Average Consumption	Average Tracking Savings (kWh)	Realization Rate
Site-level	18,421	12%	37,908	49%
Measure-level				
BAS	5,344	3%	19,943	27%
Lighting	8,490	6%	15,933	53%
Refrigeration	10,673	7%	6,211	172%
Process	5,298	3%	6,357	83%
HVAC	24,403	16%	10,024	243%

Challenges to isolating BAS-estimated savings

There are two possible ways that the installation of BAS and non-BAS measures could work against achieving expected savings levels. As discussed previously in section 5.2, the interactive effects between BAS and many of these measures means that the marginal additional savings for the next measure, whether it is BAS or the other measure, is lower than expected. The site-level savings on which the regression breaks out to measures reflects the decreased marginal savings of the additional measures. The decreased combined savings are distributed out to the measures at lower levels than each measure would likely generate if installed in isolation.

Furthermore, the additional measures may have been installed at targeted stores that would have shown the greatest potential savings from BAS without those additional measures. The BAS implementer recruited the stores for BAS installation, but then also offered the additional measure installations. A store with existing lower-efficiency equipment would represent the greatest opportunity for BAS, while a store that was updated more recently would realize lower control-related savings because BAS would be controlling more efficient equipment. If the BAS implementer were to install only BAS for six to twelve months, it would be more straightforward to estimate the savings potential for BAS under those conditions.

Under the scenario of multiple measure installations in combination with BAS, an evaluation that applies engineering approaches to the available end-use metering data available from the BAS could be promising. BAS represent the beginning of a trend of using sensor data for many or most end uses.