Improvements in SEM Program Impact Evaluation Methods: Lessons Learned from Several Recent Projects

Heidi Ochsner, Cadmus, Portland, OR
Jim Stewart, Cadmus, Portland, OR
Lauren Gage, Bonneville Power Administration, Spokane, WA
Erika Kociolek, Energy Trust of Oregon, Portland, OR

ABSTRACT

Strategic Energy Management (SEM) programs differ from traditional energy-efficiency programs in that the focus is on integrating a holistic energy-management strategy. These programs provide long-term consulting services that educate and train commercial or industrial energy users to: (1) develop and execute a long-term energy-planning strategy, and (2) integrate energy management into their business planning permanently. Participants implement efficient behaviors, operational and maintenance improvements, and equipment which lead to energy savings. Due to the holistic approach and wide range of activities implemented, the evaluation approach is a whole-building analysis following the IPMVP Option C.

Three challenging evaluation areas are discussed: (1) statistically detecting energy savings, (2) designing a sampling strategy, and (3) accounting for equipment and custom measures that received rebates through other programs. Statistically detecting savings is largely dependent upon the length of the data series, whether savings are large enough relative to noise in the energy consumption data, and whether the available data explain the majority of variance in energy consumption. Sampling approaches are not widely employed due to small population sizes and the uniqueness of participating sites, particularly within the industrial sector. Desk reviews, interviews, and site visits can verify equipment and custom measures incented through other programs, each with their own benefits and costs.

Lastly, we end with recommendations for evaluators to standardize how results are reported, include confidence intervals around the savings, and better explain how evaluated savings are determined so that other utilities considering SEM programs can review successes and challenges in this area.

Introduction

In 2014, there were at least 12 Strategic Energy Management (SEM) programs (CEE, 2014a). SEM programs differ from traditional energy-efficiency programs in that the focus is on integrating a holistic energy-management strategy that extends beyond replacing inefficient equipment. These programs provide long-term energy-management consulting services that educate and train commercial or industrial energy users to: (1) develop and execute a long-term energy-planning strategy, and (2) integrate energy management into their business planning permanently. As a result of these activities, participants implement efficient behaviors, operational and maintenance improvements, and equipment which lead to energy savings. Due to the holistic approach and wide range of measures implemented, evaluating SEM programs and quantifying energy savings can be challenging and intensive. Though there are 12 SEM programs, only a handful of these have quantified and reported their energy savings and even fewer have undergone evaluation.

There are several challenging areas in designing the impact evaluation for SEM programs. The first area is quantifying energy savings, the second is with sample design, and the third is accounting for savings from equipment or custom measures that received rebates through other programs. In the following sections, we discuss these challenging areas in more detail, and then present potential solutions.

Energy Savings Quantification

Program implementers and program evaluators use different approaches to quantify energy savings, though both rely on a regression analysis. The evaluators begin with the regression model specification chosen by the program for predicting the baseline consumption, however, the final model used for evaluation may or may not have the same specification. Evaluators review energy savings several months to years after the initial baseline model was developed by the program, and typically, evaluators can incorporate a longer timeframe of engagement data than program staff included. When revisiting the model, the initial baseline model may no longer be valid at some sites due to additional data collected from facility or due to changes at the facility. In other cases, the evaluators may find that including other variables such as weather can improve the explanatory power of the model.

The methods used by program implementers and evaluators are described below, followed by a discussion of the challenges associated with these approaches.

Program Methodology for Quantifying Energy Savings

The programs that quantify savings typically use the cumulative sums (CUSUM) method, where a regression model is developed using the baseline energy consumption as the dependent variable and independent variables include facility production, weather, and other documented drivers of energy use. The model specification and resulting coefficients are then used to predict what energy consumption would have been during the period of participation, in the absence of the program. The participant's actual consumption data is then subtracted from the predicted baseline consumption to determine the energy savings. The result is a CUSUM plot such as the one shown in Figure 1, showing the participant's cumulative change in energy usage over time as compared to the baseline. In this example, the participant increased energy consumption compared to the baseline during their first year in the program (May 2009 through May 2010) and then began seeing savings around June 2010. The savings reflected in this plot are cumulative.

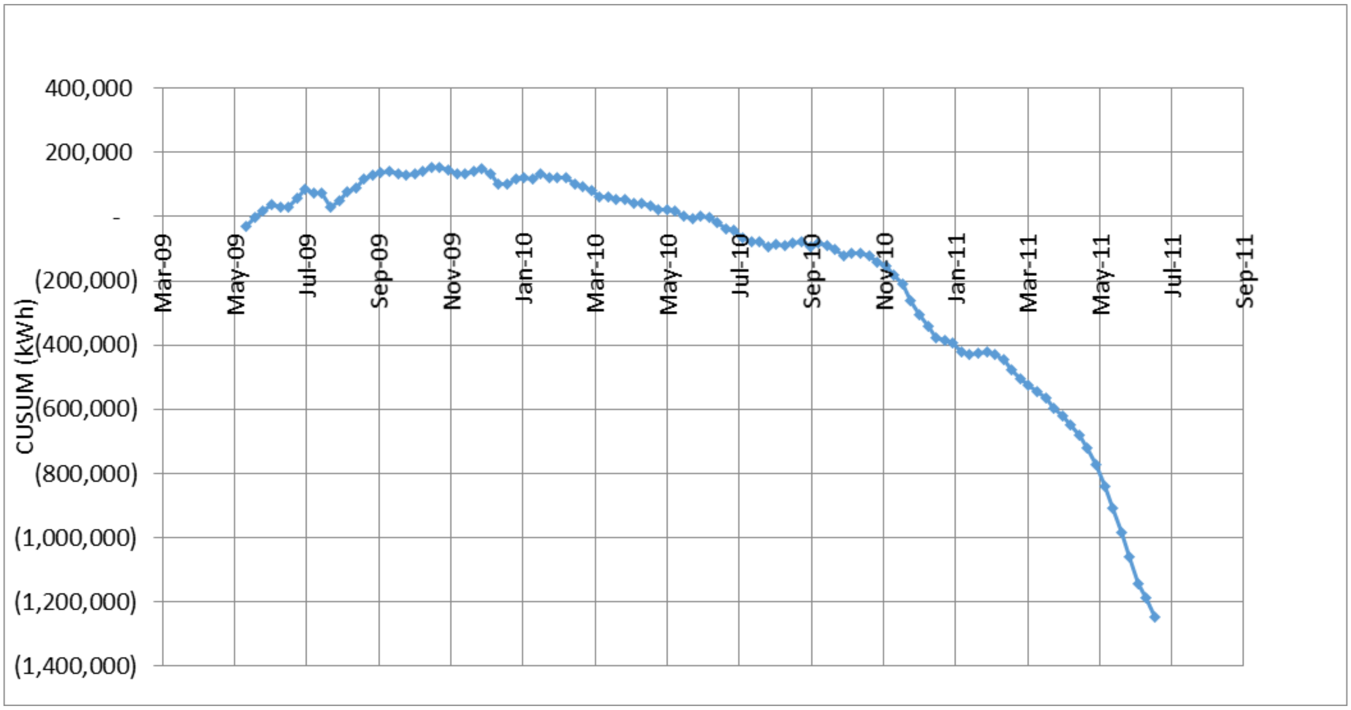

Figure 1. Example CUSUM plot

As shown by the plot, energy savings increase with time, as the participants implement more energy saving actions. Different programs use these results differently, depending on the program reporting objectives. For example, programs with a multi-year engagement period quantify the savings by summing the differences between the projected baseline data and the actual consumption for each interval within the engagement period. However, a program with a one-year engagement may exclude a ramp-up period and only consider the CUSUM results for the last three to six months of the engagement period and extrapolate these to determine annual energy savings.

Lastly, if participants received rebates for equipment or custom measures during the same time period as their participation in SEM, the savings associated with those measures are subtracted from the whole-facility savings to avoid double-counting those savings in both programs.

The advantage to the CUSUM approach is that it is easy for the participants to understand and continue using on their own. Once the baseline model is established, the CUSUM can be updated as frequently as billing and production data are collected and savings results can be tracked on an ongoing basis.

Evaluation Methodology for Quantifying Energy Savings

The most common evaluation method to quantify energy savings for SEM participants is by conducting a regression analysis to calculate whole-facility savings, following the guidance in the International Performance Measurement and Verification Protocol (IPMVP) Option C: Whole Building. Similar to the CUSUM approach, the regression model independent variables include facility production, weather, and other documented drivers of energy use. Where the approaches differ is that the regression model includes both pre-engagement period data and post-engagement period data and incorporates an indicator variable defining the participation period. The coefficient on the participation period indicator variable represents the average savings per data interval (e.g., monthly savings if data are monthly). The coefficient is then multiplied by the number of data intervals in a year to determine the annual energy savings. Also similar to the program method, savings associated with projects rebated through a different program are subtracted from the whole-facility savings to avoid double-counting.

The advantage of the evaluation approach is that all pre- and post-engagement period data are available and taken into account in the regression model, and this can provide more flexibility in the model. For example, if production efficiency improved as part of SEM, the model can take this into account by interacting the engagement period variable with the production variable allowing the relationship between production and energy consumption to change. Other advantages include that this approach follows the widely accepted IPMVP Option C and the confidence intervals are simple to calculate. Lastly, if the model specification chosen for the evaluation is similar or matches the baseline regression model used in the CUSUM approach that uses all of the engagement period data, the results will be similar.[1]

Energy Savings Estimation Challenges

There are several challenges and considerations when using a regression analysis approach:

- **Timing of analysis.** What is the best timing to quantify energy savings that are representative of all activities implemented during the program? If data are only collected through the end of a participant's engagement, savings from activities implemented just before the end of the engagement period may be underestimated.
- **Statistically detecting savings.** Savings must be large enough to detect amidst the noise in the energy consumption data. Unexplained variability in energy consumption can lead to large confidence intervals around the savings estimates. Additionally, if the collected data do not include all important drivers of energy consumption, it will be more difficult to detect energy savings. In some cases, the model may result in biased program results if unexplained changes in energy consumption coincide with the SEM engagement period.
- **Risk of participants making changes that invalidate the baseline.** If a participant's building undergoes renovations, changes in product mix or occupancy type, or other significant changes affecting energy use during the engagement period that are unobserved or unmeasurable and not due to the program, the SEM savings estimates may be biased.
- **Risk of finding negative savings (increase in consumption).** Negative savings can occur if there are unobserved changes at the facility that caused consumption to increase, or if the savings from the incented measures were overestimated.

Potential Solutions to Improve the Likelihood of Detecting Savings

Based on our experience evaluating SEM programs, the following recommendations improve the likelihood of detecting savings.

Evaluability Assessment. Anticipating the evaluation's data needs early on will help ensure the required data are collected to improve the chance that energy savings can be detected. Program administrators who are new to implementing SEM programs should collaborate with an experienced SEM program implementer or evaluator early in the program design phase to understand and identify specific data needs for quantifying energy savings. Beyond identifying the data to collect, an implementer or evaluator with experience quantifying SEM savings can provide insight as to the types of scenarios where savings may be difficult or impossible to determine and scenarios where the program administrator may need additional customer data or information.

[1] The evaluation of BPA's Energy Management Pilot program found that the savings estimated by the program were within the confidence interval of the evaluation savings (Cadmus, 2013).

Fractional Savings Uncertainty. Fractional savings uncertainty (FSU) analysis indicates whether the time series data—in particular, the frequency and series length—are sufficient to detect the expected (*ex ante*) savings at a particular significance level. A site's FSU is the ratio of the expected uncertainty about the savings estimate to the total expected savings. It depends positively on the coefficient of variation of the regression root mean square error and the expected savings as a percentage of total consumption, and it depends negatively on the number of observations in the baseline and test periods. A lower FSU indicates that savings are more likely to be detected.[2]

Program implementers could conduct an FSU analysis of new program participants to test whether the currently collected data will be sufficient to detect savings. If the analysis shows that data are not sufficient, the implementer can work with the participant to improve the likelihood of detecting savings by collecting data more frequently, incorporating submetering on specific processes that drive energy consumption, or through other methods. Alternatively, the program administrator and customer could determine it is not worth the additional effort or expense to collect these data, or not estimate the savings at this site, or terminate participation in the SEM program. This is a valuable process since it reduces the risk of investing in a site where it is unlikely that savings can be quantified. From the customer's perspective, since the program incentive is based on achieved energy savings that must be quantifiable, it is best to set their expectations around data needs early on.

Bonneville Power Administration (BPA) uses FSU analysis with new program participants. The evaluation of BPA's Energy Management Pilot shows the FSU analysis result compared to the energy savings result. The evaluation confirmed that facilities with statistically significant and positive energy savings tended to have a smaller FSU. Facilities with lower frequency data (e.g., monthly or bi-monthly) tended to have higher FSU than facilities with higher frequency data (e.g., daily or weekly). Also, the results showed it is possible to detect savings at sites with a high FSU if the savings are larger than expected (Cadmus, 2013).

Timing of Analysis and Length of Analysis Period. Both program implementers and evaluators should carefully consider the best timing for the energy savings analysis to occur, whether the purpose of the analysis is to report program savings or to evaluate program savings. We recommend collecting data for an additional three to six months after a participant's engagement in order for the savings analysis to fully capture the impacts of any activities implemented at the end of a participant's engagement. Post-engagement data ideally covers at least one full year in order to capture weather-sensitive energy savings.

Energy Management Information Systems. The evaluation of BPA's Energy Management Pilot showed that the regression analysis was more successful for participants who collected data more frequently than monthly (Cadmus, 2013). Higher frequency data could also provide increased certainty in energy savings and decrease the confidence interval range. Energy management information systems (EMIS) could play a future role in SEM programs, being used to collect higher frequency data and to provide more immediate feedback that improves measurement, tracking and reporting, and savings verification. Some EMIS have the functionality to automate a user-specified regression model that calculates and reports savings to the customers and program administrators. EMIS also offers additional benefits to customers, such as the ability to manage energy by sending alerts when usage is outside a defined range.

Communication between Evaluators and Program Staff. It is important for evaluators to communicate with the program staff when considering a different model specification which incorporates new variables or incorporates existing variables in a new way. Program staff have visited the participant sites and have a greater knowledge of the activities and events that may not always be fully documented in the site reports. When considering a different model specification, evaluators should discuss the program

[2] American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) Guideline 14 for measurement of energy, demand, and water savings provides guidance for FSU analysis (ASHRAE, 2014).

approach and the new approach with program staff to understand the reasoning behind the program's chosen model specification and to gather feedback about the new approach.

Sample Design for Industrial SEM Programs

The number of participants in industrial SEM programs has been increasing and soon it will no longer be cost-effective to evaluate a census of participants. However, current program participation is still small enough that the realization rate from sampled sites is highly dependent upon which sites are sampled. We performed a simulation study to test the impact that sampling from a small SEM program population would have on overall program results, and the accuracy of the resulting savings estimate.

We performed a simulation study using the reported and verified electricity savings observed in a previous program evaluation for BPA's Industrial Energy Management Pilot, where a census of the 15 High Performance Energy Management (HPEM) projects were evaluated. For the simulation study, we defined the certainty stratum to include projects that contributed to the top 65% of annualized reported savings. This resulted in a sample size of six projects in the certainty stratum. From the remaining nine projects, we randomly sampled five in the sample stratum to reach a target sample size of 11. We performed the random sampling procedure 10 times and calculated the resulting realization rates, verified total savings, and precision at the study level. The results are provided in Table 1. The simulation results can be compared to the results from verifying a census of the participants, which gave a realization rate of 94% with a total verified savings of 9.9 MWh.

Table 1. Simulation of Sampling Results Compared to Actual Results

Simulation	Realization Rate	Study Level Estimated Verified Savings (kWh)	Relative Precision	Does Confidence Interval Contain Census RR result?
1	91%	9,650,107	13%	Yes
2	83%	8,820,294	14%	Yes
3	90%	9,483,861	12%	Yes
4	86%	9,070,269	12%	Yes
5	100%	10,567,872	1%	No
6	86%	9,115,166	13%	Yes
7	88%	9,356,357	12%	Yes
8	86%	9,129,230	14%	Yes
9	86%	9,116,314	13%	Yes
10	108%	11,402,619	3%	No
Census	**94%**	**9,922,931**	**N/A**	**N/A**

The resulting realization rates and total verified savings estimates vary widely, with realization rates between 83% and 108%, and savings results between 8.8 and 11.4 MWh. Eight of the 10 simulations yielded a confidence interval range that included the realization rate result from the evaluation of a census of participants. Additionally, the relative precision ranges between 1% and 14%; however, with such a wide range of estimated savings, tight precision may create the false impression that there is little uncertainty about the true savings. In fact, there is tight precision around results that are not within the confidence interval of the true savings.

Sample Design Recommendations

Industrial facilities are each unique, making it difficult to select a representative sample. Unless there is a large sample frame or the facilities have similar energy use and expected savings, we recommend continuing to evaluate a census of participants to eliminate the risk of calculating inaccurate savings. Further evaluations of industrial SEM programs may be able to identify characteristics, such as data frequency, that could be used to stratify the sample frame.

Accounting for Measures Rebated Through Other Programs

SEM program staff often identify opportunities for equipment and custom upgrades which are eligible for incentives under other programs and encourage participants to leverage the available incentives for these measures. The program that provided the incentive for these measures claims the savings. To avoid double-counting, these savings are subtracted from the SEM total facility savings. If the savings from these measures are overestimated, the SEM savings will be underestimated. Likewise, if the savings from the equipment and custom measures are underestimated, the SEM savings will be overestimated. Verifying the savings for equipment and custom upgrades (rather than just subtracting savings claimed by the program) will therefore improve the SEM savings estimate.

The evaluation activities to verify these savings are desk review, phone interviews, and site visits. Table 2 compares the costs, benefits, and risks of these activities. Desk reviews are the lowest cost, however the documentation may not be accurate or may be missing some information. Phone interviews are medium cost and can achieve other purposes such as measuring progress in SEM adoption and participant satisfaction. However the information gained through a phone interview about equipment and custom measures is limited by the respondent's familiarity and memory about the measure. Site visits are the most expensive, but can potentially provide the most benefits including verifying both SEM measures and measures rebated by other programs, and whether regression model results make sense. For sites where the regression model was determined to be invalid due to missing variables, a site visit could inform a bottom-up savings approach using engineering algorithms for certain SEM activities.

Table 2. Evaluation Activities for Verifying Measures Rebated Through Other Programs

Evaluation Activity	Relative Cost	Benefits	Risks
Desk Review	Low	• Low cost	• Documentation may not accurately reflect actual implementation; • Documentation may be missing important information about equipment specifications or operating characteristics
Phone Interviews	Medium	• Can be combined with a process evaluation; • Can be combined with measuring progress in adopting SEM; • Can verify SEM activities and measures rebated by other programs	• Not practical to ask about all measures or measure details; • Respondent may not be familiar with the measure or may not remember equipment specifications/operating characteristics during the interview

Evaluation Activity	Relative Cost	Benefits	Risks
Site Visits	High	• Can verify both SEM activities and measures rebated by other programs; • Provides the most data and detail and can fill in any gaps in the program's documentation; • Can verify regression model results make sense; • If regression model is invalid, a site visit could inform a bottom-up savings method	• High cost to verify savings claimed by other programs; • Sampling from a small heterogeneous sample is a challenge; • Unclear if data collected through site visits can improve the regression analysis

Evaluators weigh the cost and benefits of the above activities when designing the SEM evaluation plan. As previously mentioned, sampling from a small heterogeneous sample of participants is challenging, however conducting phone interviews or site visits for a census of participants may not be practical. Evaluators and program staff should consider the importance of accurately measuring the SEM savings separately from the measures rebated through other programs. It could be important when both programs are measured independently, and must meet independent savings and cost-effectiveness targets.

Recommendations for Accounting for Measures Rebated Through Other Programs

To aid in deciding on the best evaluation approach, we recommend compiling a list of the measures rebated through other programs, including the date implemented and their claimed savings. Review the list to determine which measures are best verified though desk review, phone interviews, or site visits. For example, most measures with deemed savings can be verified through a phone interview as the required information is simply confirmation that the measure was implemented. Some O&M measures can be verified through a phone interview if a list of the measures and operating characteristics is provided to the respondent in advance so they can gather the needed information in advance of the interview. Equipment specifications and their operating characteristics are typically best verified during a site visit.

In this way, the evaluator can then review the measures at each site and identify sites where the evaluation would not benefit from the site visit. For example, it would not be necessary to visit a site which did not implement any measures rebated through other programs or a site that only implemented measures with deemed savings. The evaluator can then prioritize which sites to visit, thereby lowering the evaluation costs.

Conclusions

Evaluating SEM program savings has its challenges, but it is possible, particularly when program staff have collected the necessary data. The current evaluation methods are defensible, relying on best practice approaches outlined in IPMVP and evaluator experience with other types of behavioral programs. Past experience evaluating SEM program savings has led to the following findings:

- Evaluability assessments conducted just after the program design phase can ensure required data are collected at an appropriate frequency to improve the likelihood of detecting energy savings.
- Ongoing communication between evaluators and program staff during the evaluation is important because program staff have visited the participant sites and have a greater knowledge of the activities and events that may not always be fully documented in the site reports.

- Industrial facilities are each unique and most sample populations are small, making it difficult to select a representative sample for the evaluation. More evaluations of SEM programs may provide insights into the best sampling approach.
- Evaluators must account for measures rebated through other programs so as to not double-count savings. Reviewing the measures at each site to identify where site visits may be necessary can lower evaluation costs.

Evaluators will continue to learn more about best practices as additional SEM programs are implemented, as participant numbers increase, and as the years of participation increase. With more data points, evaluators will be better able to recommend the best timing for the savings analysis and the required time period for data collection, assess when sampling can be successful, when site visits are valuable, and when a bottom-up engineering approach may be more successful than a regression analysis approach.

Other Evaluation Recommendations

We recommend evaluators standardize their reported results so that other utilities considering SEM programs can easily review other programs' successes and challenges. Currently, there are not enough publicly available evaluation results, and those that exist are reported in different units. In some cases the reports do not adequately explain the methodology used to calculate savings. These make it difficult for others to make an informed decision on the value of SEM.

In particular, we recommend that savings are not only reported in MWh, but also as a percent of baseline consumption. This is the most comparable metric across programs, as this normalizes for the number and size (in terms of energy use) of program participants. Confidence intervals should be included so readers understand the uncertainty around the savings and whether savings are significant. The methodology used to evaluate energy savings should be adequately discussed so that readers can understand whether a higher or lower savings result than other programs can be explained by the savings methodology, rather than the program design.

References

American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE). ASHRAE Guideline 14: 2014 Edition. December 2014.

Cadmus. Energy Smart Industrial - Energy Management Pilot Impact Evaluation. February 2013. Prepared for Bonneville Power Administration. Available online: http://www.bpa.gov/EE/Utility/research-archive/Documents/BPA_Energy_Management_Impact_Evaluation_Final_Report_with_Cover.pdf

Consortium for Energy Efficiency (CEE). Industrial Strategic Energy Management Initiative. January 2014. Available online: http://library.cee1.org/sites/default/files/library/11282/Industrial_SEM_Initiative.pdf

Consortium for Energy Efficiency (CEE). Strategic Energy Management Minimum Elements. February 2014. Available online: http://library.cee1.org/content/cee-strategic-energy-management-minimum-elements/

Efficiency Valuation Organization. January 2012. International Performance Measurement and Verification Protocol: Concepts and Options for Determining Energy and Water Savings Volume 1.

Available online: http://www.evo-world.org/index.php?option=com_content&view=article&id=272&Itemid=379&lang=en