Growing Pains: Lessons from the Edge of SEM Program Evaluation

Lauren Gage, Apex Analytics, LLC
Todd Amundson, Bonneville Power Administration
Jim Stewart, Cadmus
Andrew Bernath, Cadmus
Michael Baker, SBW

ABSTRACT

This paper provides insights into complex issues discovered during Bonneville Power Administration's (BPA) Energy Smart Industrial Strategic Energy Management (SEM) evaluation. BPA's SEM program began in 2010 and was one of the nation's first large-scale industrial SEM deployments. By 2015, the program had engaged 65 facilities and had modeled site savings for up to four years. The program involves multi-year engagement with facilities and conducts site-specific savings modeling each year using high-quality data sets and well-defined protocols. The evaluation industry lacks experience evaluating multi-year industrial SEM programs. Although BPA completed an evaluation of first-year SEM participation in 2012, this evaluation of multi-year SEM participation had unexpected technical complexities, contained multiple innovative exploratory analyses, and underwent a methodology change mid-study. This paper will provide high-level findings from the evaluation, which found statistically significant SEM savings that were stable across multiple years. More importantly, this paper will dive into the evaluation challenges that arose, many of which have implications for utility planning and evaluation activities, such as: variation in savings among sites and across years, the tradeoffs of accounting for capital projects, frequency of negative savings estimates, differences in program and evaluation models, and the complexity of reporting policies and multi-year engagements. Although this evaluation represents innovative work in evaluation, there remains a significant need for more research. We recommend multiple areas for evaluators nationally to pursue, including: the feasibility of sampling, use of pre-post models, program design impacts on persistence, how to minimize bias in savings reporting, persistence of SEM savings, and impacts of re-baselining.

Introduction

The goal of this paper is to share evaluation insights learned during an impact evaluation of Bonneville Power Administration's (BPA) Industrial Strategic Energy Management (SEM) program, which was completed in 2017 and covered program years 2010-2014 (SBW 2017). This paper builds on and contributes to the evaluation literature that exists for SEM programs and whole building modeling.

Strategic Energy Management has been defined by the Consortium for Energy Efficiency as follows: *Strategic Energy Management can be defined simply as taking a holistic approach to managing energy use in order to continuously improve energy performance, by achieving persistent energy and cost savings over the long term. It focuses on business practice change from senior management through shop floor staff, affecting organizational culture to reduce energy waste and improve energy intensity. SEM emphasizes equipping and enabling plant management and staff to impact energy consumption through behavior and operational change. While SEM does not emphasize a technical or project centric approach, SEM principles and objectives may support capital project implementation.* (CEE 2014).

In general, evaluators of industrial SEM programs have followed International Performance Measurement and Verification Protocol (IPMVP) Option C – Whole Facility guidelines, which recommends regression analysis of facility consumption data to estimate savings. Evaluators use a regression model to forecast baseline energy consumption during the reporting period and estimate savings as the difference between baseline and metered consumption. IPMVP recommends the application of Option C when the expected energy savings are 10% or more of facility consumption.

There are, however, a number of issues relevant to industrial SEM program evaluation that IPMVP Option C and other evaluation protocols have not addressed or not in insufficient detail, including when (a) expected savings are less than 10% of consumption, (b) SEM program facilities implement capital projects, and (c) savings estimates require non-routine adjustments to baseline consumption. Furthermore, Option C does not address methods for quantifying the uncertainty of SEM savings estimates. Recently, in an effort to fill these gaps, evaluators have updated the IPMVP and ASHRAE protocols or issued new protocols such as DOE Uniform Methods Project (UMP) SEM Protocol (US DOE, 2016) and the DOE Superior Energy Performance Protocol (2016).

This paper draws from the SBW/Cadmus 2010-2014 impact evaluation and the authors' learnings from the three-year evaluation process (SBW 2017). The evaluation methods were built upon the evaluation approaches developed as part of the first-year evaluation of BPA's SEM program (Cadmus 2013). During the 2010-2014 evaluation, the UMP was nearing completion of its Strategic Energy Management protocol. The UMP is relevant because it draws upon the expertise and experiences of SEM evaluator and SEM program implementation staff, while providing guidance for evaluation of utility SEM programs, and clarifying and distinguishing from conformance with ISO 50001 and the US DOE's Superior Energy Performance Program.

This paper also builds on other papers, including Ochsner (ACEEE 2015) which summarized evaluation results, and Ochsner (IEPEC 2015) which summarized challenges in evaluation and the high-level guidelines Energy Trust of Oregon developed for evaluation.

It also builds on Amundson (2013), which illustrated that annual increases in both capital and behavior-based energy savings were achieved from developing an industrial facilities' organizational capability to manage energy. Finally, this paper builds on Koran (2017) and Bernath (2017), which provide deeper insight into the issues of uncertainty estimation and comparison of forecast and pre-post methods.

Program Background

The Bonneville Power Administration began offering its Energy Smart Industrial Energy Management Program to industrial facilities in 2010. Through the program, BPA provides long-term energy management consulting services to educate and train industrial energy users for two primary purposes: (1) to develop and execute a long-term strategy for energy planning and (2) to permanently integrate energy management into their business planning. BPA's program was one of the nation's first large-scale deployments of a strategic energy management program in the industrial sector, which had engaged 65 facilities by the end of 2014.

Through the Energy Smart Industrial program, BPA assists their utility customers'[1] industrial facilities in achieving electrical energy performance improvement goals. Industrial facilities may request (1) to focus on developing their organizational capability to manage energy, typically program-provided via a cohort-based, SEM training delivery model[2]; (2) program assistance to provide individualized, focused technical support in developing new low-to-no cost operations and maintenance (O&M) strategies within their existing systems[3]; or (3) a combination of the two. Co-funding for enhancements to energy performance tracking systems were also provided on a case-by-case basis. The program also offers co-funding for an energy project manager in conjunction with these components to enable a facility to dedicate staff time to energy management.

BPA's Energy Performance Tracking (EPT) program team developed monitoring, targeting, and reporting (MT&R) guidelines that include the methodology for measurement and verification of energy savings for program participants (BPA, 2015). The methodology aligns with best practices from IPMVP Option C. The EPT team analyzed facility meter data, production data, and other relevant data to estimate annual energy savings for each facility, and BPA recorded savings in its reporting system.

The program team estimated two types of savings: (1) facility savings and (2) SEM savings. The team estimated facility electrical energy savings during the performance period using the forecast method. This method compares electricity usage at the utility meter level to the facility baseline consumption forecasted for the same

[1] BPA is a wholesale power provider with over 140 customer utilities in the Northwest.
[2] Historically known as High Performance Energy Management (HPEM).
[3] Historically known as Track and Tune (T&T)

period using key independent variables (e.g., production). Facility savings included SEM savings and savings from capital equipment projects that received rebates through other energy efficiency programs.

To avoid double counting, the program calculated SEM savings, equal to the difference between the facility savings and the reported savings from prorated capital equipment projects. BPA also recorded SEM savings in its reporting system. Reported savings equaled the SEM savings, except in cases when SEM savings were negative. In these cases, BPA recorded zero SEM savings.

Methodology

For this impact evaluation, the evaluation team (SBW and Cadmus) focused on the performance of 31 facilities that had the longest history of participation. The evaluation team estimated savings for these facilities and did not extrapolate to the program population. The evaluation included the following objectives:

- Estimate SEM energy savings and characterize year-to-year SEM savings trends for sampled facilities
- Verify program's estimated SEM savings and BPA's reported SEM savings
- Survey participants about their adoption of SEM practices and assess whether differences in adoption can explain the energy savings results

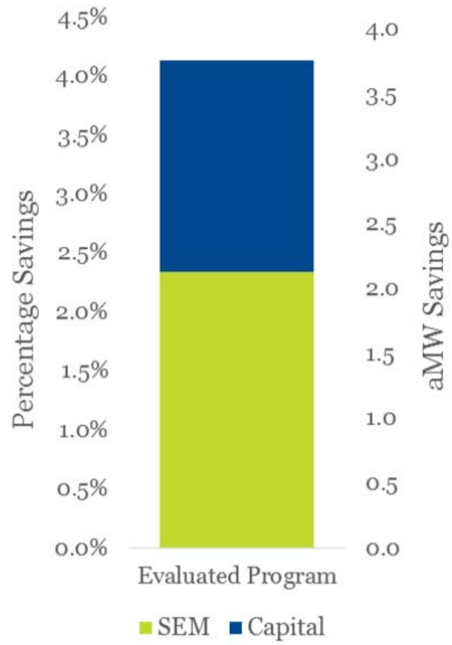
The evaluation team independently estimated annual energy savings for each facility using regression analysis. Similar to the program's MT&R process, the evaluation estimated annual facility savings by comparing metered consumption during program engagement to an adjusted baseline. The evaluation estimated SEM savings as the difference between total facility energy savings and energy savings from any capital projects incentivized by other energy efficiency programs. BPA provided the data used to estimate savings, which it collected by working closely with each participating customer.

The evaluation team used two methods for estimating SEM savings: (1) the forecast method (as the default method) and the (2) pre-post method. The pre-post method differs from the forecast method in that it uses facility data from both the baseline and performance periods. This method employs an indicator variable that represents the average per-period savings attributed to SEM program activity. The evaluation team developed decision logic to determine which method was expected to produce a more accurate savings estimate for a given facility.

High-Level Results

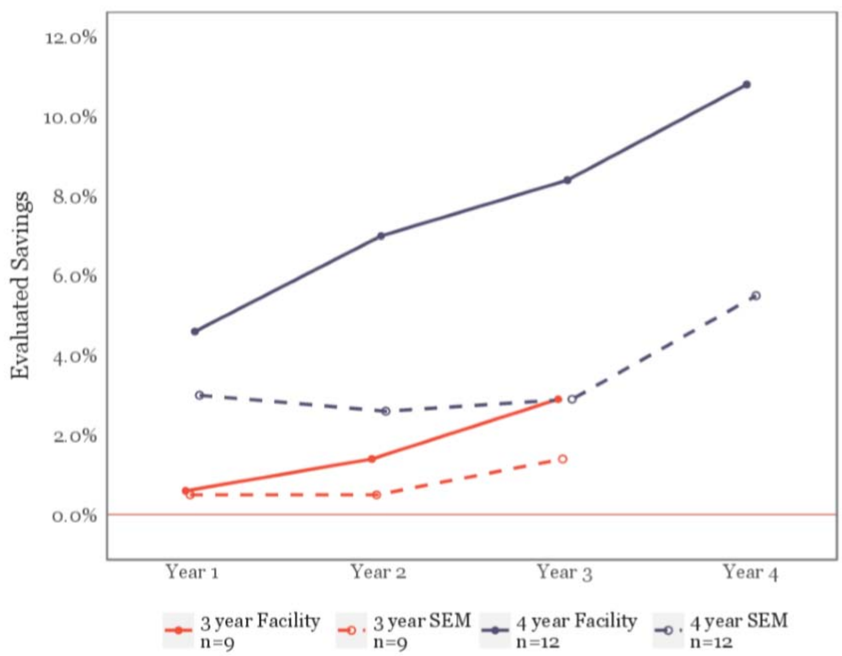
At the highest level, the evaluation found that the facilities participating in SEM saved an average of 4.1% of electricity consumption from the combination of SEM and capital projects. Facilities participated for an average of 3.4 years. These savings were significant at the 10% significance level and were estimated across all facilities and program years. Capital projects were less than half of facility savings and SEM savings were 2.3%, as shown in Figure 1. Statistical significance could not be determined for the capital savings as they were estimated using physical models of equipment energy use.

Figure 1. Evaluated Percentage Savings



The evaluation also found that the SEM savings persisted during the participation period. As shown in Figure 2, facility savings (solid lines) increased throughout the participation period and SEM savings (dashed lines) persisted after the first year and increased slightly in the final year.

Figure 2. Annual Percentage Savings by Year in Program



The evaluation verified the MT&R SEM savings estimated by the program. The realization rate (ratio of MT&R SEM savings to evaluated savings) was above 106%. In addition, the MT&R SEM and evaluated savings estimates for individual facilities and years were also similar: in 73% of cases (i.e., the combination of a facility and program years), the savings estimates were not statistically different.

Finally, the evaluation found that adoption of SEM elements was not correlated with SEM percentage savings. The Consortium for Energy Efficiency identified 13 management practices, called “elements,” for facilities to continuously improve their energy performance. The evaluation team surveyed 24 EM Program participants to assess their adoption of these elements and found no pattern of specific SEM elements with respect to SEM savings.[4] A study that tracks SEM adoption in individual facilities over time and correlates changes in adoption with changes in facility savings may be more likely to detect a relationship.

Evaluation Lessons Learned

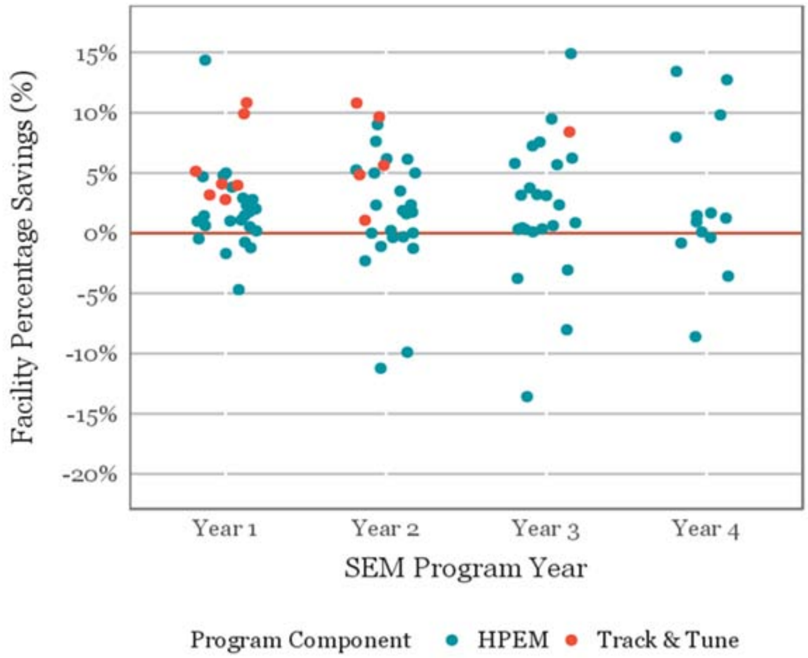
Through the process of the SEM evaluation, the team learned lessons on many topics, which are relevant to other whole-facility program evaluations in non-residential sectors. The following sections outline the learnings and implications for future evaluations.

Engaged Programs Allow for Effective Evaluation. The evaluation found that the duration and depth of program engagement has an impact on the cost, effort and accuracy of the of impact evaluation. For this evaluation, the program had carefully documented multi-year program implementation and collected the data required for evaluation. This enabled evaluation to avoid expensive data collection and end-user contact efforts. For industrial projects, this can be made especially difficult by the sensitivity of industrial facilities to releasing what they consider to be proprietary data. For each facility and year, BPA’s program team prepared a project completion report, which described the facility operations, energy consumption, production activities, documented implemented SEM activities, and provided an estimate of the facility and SEM energy savings. The evaluation stated that this approach met the needs of evaluation and can serve as an industry standard for SEM programs and had essentially no recommendations for improvement in data collection and MT&R modeling.

Be Comfortable with Variation. Figure 2 above shows that trends in program savings on average are smooth and increasing. Yet, program implementers and evaluators need to be prepared that on a site-by-site basis, the results are much less likely to have that smooth result. The BPA evaluation found significant variation in savings between facilities and from year-to-year for individual facilities, as shown in Figure 3, below. The coefficient of variation on percent savings (the ratio of the sample standard deviation to the sample mean) was 201%. Also, out of 29 facilities with more than one year of evaluated savings, 38% increased percent savings each year, 21% decreased percent savings each year, and 41% had seesawing percent savings that increased in some years and decreased in others. This variation in annual savings likely reflected differences in SEM implementation, changes in electricity consumption, uncertainty of the savings estimates, variation due to differences between facilities (e.g., type of production, weather sensitivity), and variation within facilities across program years (e.g., changes to staff, changes to production lines).

[4] This may be due to the small sample size, unexplained variation in percentage savings between facilities, or because savings depended on factors outside this survey (such as how well participants implemented the SEM practices).

Figure 3. Summary of Variability of Annual SEM Percentage Savings Estimates[5]



Accounting for Capital Projects has Tradeoffs. The ways that programs account for savings from capital projects had implications on evaluation methods, as well as utility planning and reporting. There are two primary approaches[6] to accounting for capital projects, which we describe below. We also discuss the implications of these approaches.

1) *Estimate facility savings - combining capital project and SEM savings*. Regression models for savings are typically conducted at the whole facility level, meaning that all efficiency actions included at the facility during program performance periods are captured in the estimate of facility savings. Therefore, combining capital projects and SEM savings is most consistent with the facility regression model and therefore the most straightforward for programs and evaluations.

   One drawback to this approach is that the persistence of SEM savings (i.e., expected duration of savings, both during and after program engagement) is less known than capital projects and is expected to be shorter. This can cause difficulties for paying customer incentives and reporting savings. Measures with shorter lifetimes have less economic value to utilities and therefore incentives are often less for those measures. Also, reporting policies may be different for measures with shorter and longer lifetimes, which may separate SEM and capital projects. For example, when reporting multiple years, BPA adds up first-year savings of capital projects, but makes adjustments for some short duration measures where adding each year's savings would over-count achievements. One solution is to create a blended measure life for projects, although this could add complexity to program efforts and is made difficult by lack of information on SEM savings persistence.

2) *Net out capital project savings from facility savings*. The alternative to using a whole facility savings estimate is to separately allocate capital and SEM savings. As described above, BPA's program reporting and evaluation approach was to estimate total facility savings using a regression model and then

[5] HPEM and Track and Tune are the major program components offered by BPA during the evaluation period.
[6] It is also possible to use indicator variables in a regression model.

subtract the capital project savings from that estimate to determine the portion attributable to SEM. Capital project savings are estimated using standard custom project measurement and verification protocols and the project is reimbursed with a payment rate that reflects the expected measure life of the equipment. Due to the relative certainty of longer measure lives of capital projects and the increased value to the power system, the program can pay a higher incentive. This also allows SEM savings to be tracked annually, which allows for greater understanding of the persistence of SEM savings.

The first drawback of this approach is that any errors in the savings estimation of capital projects affects the SEM savings. For example, if a facility regression model estimates 100 MWh in the first year, but the capital project measurement and verification result is 125 MWh, the resulting SEM savings estimate is -25 MWh. Although this could be a reflection of actual events (i.e., SEM efforts led to an increase in consumption), it is likely that the capital project was overestimated and the SEM program "takes the hit" for that error. Another scenario is where the first-year savings of the capital project is accurate, but the project savings degrade over time. Therefore, as the SEM program tracks annual savings, degradations in capital project savings would reduce the SEM savings in later years and it would appear that the SEM savings are falling. Another drawback of netting out capital projects is that the uncertainty for SEM savings cannot be determined because it is not practical to estimate uncertainty for the physical models the program uses to estimate capital measure savings. Finally, this approach may underrepresent the importance of the SEM program activities in capital project installation. It is likely that facilities engaged in SEM activities are more likely to plan and implement capital projects to save energy. But, the process of netting out capital projects from facility savings may make it appear that the SEM program didn't influence the capital project savings. For this reason, the BPA SEM impact evaluation reports total facility savings as well as SEM results.

In summary, there are tradeoffs for utility planning, reporting and evaluation of these two different approaches. Understanding their implications will allow organizations and evaluators to make informed decisions about how to approach capital projects.

Be Prepared for Negative Savings Estimates. One of the most difficult aspects of the BPA SEM impact evaluation was the concern about negative savings estimates, or increases in consumption. BPA's program practice was to report zero savings for facilities with negative M&TR SEM savings estimates. BPA reasoned that an increase in facility electrical consumption was not likely to have been caused by SEM implementation. Also, because incentives are based on savings, this convention mitigates a change in payment policies.

This evaluation found that for the majority (78%) of cases, which refer to the combination of a facility and program years, evaluated SEM savings estimates were positive. However, in 22% of cases, the MT&R SEM savings estimate was negative. This includes 10% of cases where both facility and SEM savings were negative, as well as 12% of cases when the facility savings estimate was positive but the SEM savings estimate was negative after subtracting capital project savings.

Estimated increases in consumption likely reflect difficulties in the measurement of savings because of omitted variables, degradation in capital equipment performance, or unaccounted for non-programmatic effects—not that the program caused consumption to increase. However, an increase in facility consumption due to the program cannot be ruled out. As there is no accepted method for differentiating between omitted variables and a program causal effect, the evaluation results included estimated consumption increases and the realization rate for program-reported savings was 88%[7].

[7] The realization rate for reported savings was lower than the realization rate for the MT&R SEM savings due to program approach of reporting negative savings results as zero.

The evaluation team was concerned that BPA's reporting convention treats negative and positive savings estimates inconsistently. Positive savings estimates were just as likely to exhibit error as negative savings estimates, and the sign of the savings estimate should not be the reason for accepting or rejecting it.

Program administrators and evaluators of other programs should have policies established for those facilities and years that have negative estimates of savings.

The Best Program Model may not be the Best Evaluation Model. As noted above, the evaluation found the program's MT&R modeling was reliable and the evaluation recommended that the program continue to use the forecast method on a site-by-site basis. The forecast model methodology, as described in the Program Background section above is the industry standard for program implementation because it is transparent and easy-to-understand and allows for implementers and facilities to track progress in real-time and make adjustments as needed. Yet, the evaluation industry has typically used pre-post models, as it is hypothesized that including all of the available facility data from before and after program implementation will produce a more accurate estimate of facility consumption. The pre-post-method is introduced and applied to industrial SEM evaluation by Luneski (2011). Although the BPA evaluation began by using pre-post models as the default methodology, the BPA program team had concerns with the evaluation using a different savings estimation method. Therefore, the evaluation team conducted a comparison of the two methodologies for ten facilities, which is described in Bernath (2017). This analysis produced very similar savings estimates for most facilities. Due to this, the evaluation made a mid-course correction and changed the default approach to the forecast method. Yet, the comparison also showed that there are circumstances when it may be preferable to use a pre-post model.The evaluation team created a decision tree, which is included in the evaluation report, to guide the evaluator in deciding when the pre-post method should be used.

Reporting Policies and Multi-Year Engagements Add Complexity. One of the innovative areas for this evaluation was to assess the year-by-year trends of savings from the program. Because facilities are entering, exiting, and re-enrolling in the program each year, the program-level results can shift suddenly due to a change in composition of the sample rather than a change in program performance. It is because of changes in composition of the sample that Figure 2 above separates those facilities with 3 year and 4 year performance periods. When all sites are combined, any differences between facilities participating for three years and facilities participating for four years would lead to a bias between the fourth-year result and the other three years. Additionally, the sample size would drop in the fourth year, which may not be evident in the graphic.

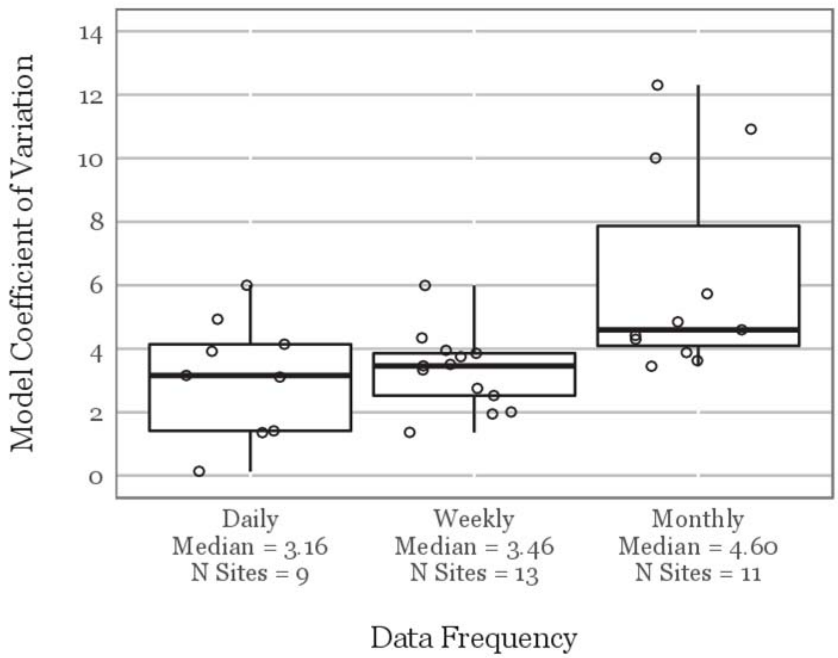
The evaluation also found that the calculation of realization rates is dependent on the view of the program administrator with respect to the definition of savings. For example, BPA's program used an incremental savings approach in its reporting of savings prior to 2014. The underlying premise of this approach, which subtracts the previous year savings from the current result, is that the most recent estimate of savings is the final result. If this is the policy framework, then the evaluation only needs to calculate a realization rate on the final year savings and not on each year. Yet, if the program is reporting savings on a one-year or average basis, then realization rates can be calculated annually and the overall realization rate is the average across all sites and years.

Finally, when the program or evaluation re-baselines[8] the facility energy consumption regression model, some consideration must be paid to reporting policies. For example, the BPA program after 2014 moved to a policy of re-baselining all facility energy models every two years. This allows the program team to reduce the significant effort of tracking capital projects, programmatic and non-programmatic activities, and data streams for five or more years, as well as periodically reassessing the SEM engagement and verification activities where appropriate. This policy makes program tracking easier, but needs to ensure that savings for the SEM program are not undercounted. For example, assume a facility saves a stable 100 MWh across four years of program participation. If the facility re-baselines after year 2, the program would report 100 MWh in years 1 and 2 and 0 MWh in years 3 and 4. This approach can be accounted for in reporting policies, but organizations need to be careful as some policies could discount the savings achieved by SEM (e.g., averaging savings across 4 years would lead to 25 MWh).

[8] Re-baselining occurs when the initial pre-participation baseline no longer applies due to changes at the facility and a new baseline must be created.

More Granular Data is Better, but Monthly isn't a Deal Breaker. The evaluation team found that the median facility energy model coefficient of variation[9] was lowest (best) for daily and weekly models and higher for monthly or bi-monthly models. Figure 4 shows boxplots of the model CV for evaluated facilities by the frequency of the facility energy consumption data (daily, weekly, and monthly). Boxplots show the quartiles, where the middle band represents the median. These results suggest that program managers and evaluators should attempt to collect high frequency energy consumption data whenever possible. Sometimes, however, production data will be the limiting factor, as they may only be available at lower frequencies.

Figure 4. Model Coefficient of Variation by Frequency of Data



Conclusions

The evaluation discussed in this paper led to new insights about the reliability of different SEM savings estimation methods, estimation of SEM savings uncertainty, and causes of negative savings estimates.

Additionally, the evaluation provides guidance for future evaluations and program designs, such as being prepared for site-specific variation and negative savings estimates, considering the tradeoffs for capital project attribution, choosing the best evaluation model, and lining up reporting policies with evaluation results.

Due to these complexities, we recommend early coordination between evaluators and program implementers. The purpose of this is to share updates and, when needed, to collaborate and seek solutions to thorny and complex topics, while maintaining independence of evaluation efforts.

Although this evaluation incorporated important innovations and provides useful insights, we believe there remains a large opportunity for the evaluation community to conduct research to improve understanding of SEM programs and SEM evaluation, such as: the feasibility of sampling, impact of uncertainty in capital projects, when should pre-post models be used, program design impacts on persistence, and how to ensure no bias in

[9] Ratio of the model root mean square error to mean response. A large CV indicates a model with high prediction uncertainty. A low regression CV indicates that the model can explain more of the variation in facility energy consumption. When a model explains most of the variation in a facility's energy consumption, there is greater likelihood of detecting savings statistically. The evaluation team computed all model CVs from regressions estimated with baseline period data.

savings reporting. Additional topics for research include: cost-effectiveness, persistence of SEM savings after a facility ends program engagement, and whether shorter re-baselining periods (two years versus three to five years, or more) change evaluation results.

References

Amundson, Todd. 2015. "Promoting Reliability of Industrial SEM Savings Through Multi Year Engagements". *ACEEE Industrial Summer Study 2015*

Bernath, Andrew. 2017. "Clash of Methodology: A Model Building and Comparison Simulation for Estimating Energy Savings Resulting from Strategic Energy Management Programs". *IEPEC 2017, forthcoming.*

Cadmus Group. 2013. "Energy Smart Industrial—Energy Management Pilot Impact Evaluation". Prepared for Bonneville Power Administration.

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

Koran, Bill. 2017. "A Comparison of Approaches to Estimating the Time-Aggregated Uncertainty of Savings Estimated from Meter Data". *IEPEC 2017, forthcoming.*

Ochsner, Heidi. 2015. "Improvements in SEM Program Impact Evaluation Methods: Lessons Learned from Several Recent Projects". *IEPEC.*

Ochsner, Heidi, and Tolga Tutar, and Erika Kociolek, and Steve Phoutrides. 2015. " Does SEM Achieve Verifiable Savings? A Summary of Evaluation Results". *ACEEE.*

SBW and Cadmus Group. 2017. "Industrial Strategic Energy Management (SEM) Impact Evaluation Report". Bonneville Power Administration.

U.S. Department of Energy (2016a). *Qualified Energy Savings Measurement and Verification Protocol for Industry*.

U.S. Department of Energy (2016b). *Superior Energy Performance Certification Protocol for Industry*.