Evolving energy efficiency programs to focus on CO2 reduction: Implications for program evaluation

Robert Bordner, EMI Consulting, Seattle, USA
Brett Close, EMI Consulting, Seattle, USA

ABSTRACT

Climate change poses an existential threat to our entire world ecosystem. In order to effectively combat climate change, energy programs will likely need to jettison historically-preferred program metrics and focus directly on reducing greenhouse gas (GHG) emissions. Evaluators play a key role in this transition by conducting effective evaluation, by minimizing the burden evaluation places on program participants, and by providing concrete, timely, and actionable results and recommendations. This paper lays out the need to make this transition to programs that are focused on GHG emissions reductions, details the benefits of that change in focus, and describes an evaluation approach that the authors developed to evaluate and better support the success of GHG emissions reductions programs.

INTRODUCTION

Program evaluation is intended, fundamentally, to support the success of public policy initiatives intended to better our society. The field of energy program evaluation serves, fundamentally, to support the evolution of a sustainable energy economy, and the field has readily evolved and adapted over the years to address the specific policy drivers of the programs relevant in time. These policy drivers have shifted over time and have included: saving costs associated with burning oil; avoiding the construction of the next nuclear power plant; minimizing SO_2 emissions; ensuring reliability of supply on the grid; meeting the demand of all new load growth; ensuring equity in access to affordable energy; and ensuring reliability of operation of the grid. We are now in the midst of a major policy shift for energy programs as our society strives to reduce greenhouse gas (GHG) emissions.

The authors believe that, just as energy programs are evolving to address climate change, so must the field of energy program evaluation evolve to support these changing needs. We first review the current policy drivers and need for evaluation, review relevant metrics for program evaluation, and provide a case study example based upon an innovative program that is underway in the State of California.

The Stakes for Effective Energy Program Design and Evaluation

With 38 percent of global GHG emissions sourced from electricity production (EIA 2019), efforts to reduce these emissions are a top policy priority. Much of this reduction will occur through increased use of renewable resources (e.g., solar photovoltaic (PV), wind, and hydro). However, energy efficiency programs, coupled with storage, provide multiple benefits – including minimizing and shaping system loads in a way that supports the integration of these renewable resources in to the broader energy system. Within this context, the goals of programs are not necessarily reductions of energy consumption (kWh) per se, but also very time- and location-specific reductions in demand (kW). These reductions and shifts in demand allow the better integration of renewables and, in turn, significant reductions in GHG emissions. The ultimate value of these programs is not energy savings, as measured through traditional avoided energy and capacity costs, but direct reductions in GHG emissions.

We Focus on What We Measure: Traditional Program Metrics and the Need to Change

Energy efficiency programs have long relied on energy consumption as a useful proxy to measure or estimate the true desired policy outcome. Beginning with the energy crises of the 1970s, and continuing with the increased awareness of environmental impacts from energy consumption in the 1980s and 1990s, metered end-user consumption was a handy metric to encapsulate the goals of programs: if customers could use fewer kWh or therms / joules, then they were better off and the world was better off. With the California energy crisis in the early 2000s, the high costs (both in terms of marginal generation and the risk of blackouts) among the peak hours added peak kW reduction to the list of metrics. This framework made sense because, although there were differences in marginal impacts between reductions to baseload (when a marginal generation source was likely to be a coal plant), shoulder load (when a marginal generation source was likely to be a combined-cycle gas plant), or peak load (when a marginal generation source was likely to be a gas peaking plant), all reductions in kWh were valuable reductions. The averages of impacts were imperfect, but close enough for the purposes at hand.

This focus on reductions in energy consumption then led to impact evaluation approaches that focused on relatively narrow interests: primarily, estimating the difference in power / wattage draw between baseline equipment and more efficient equipment. To the extent that the timing of the reductions was relevant, it often focused on a very limited number of high value hours around the system peak. Evaluations often took this peak coincidence factor as given and did not seek, due to the added expense, to update established values. Even building simulation modeling that produced 8760 load profiles were generally aggregated to yield one or two metrics: energy savings and demand reduction. Along with the narrow focus on kWh, as many US jurisdictions started to provide opportunities for utilities to earn profit based on the performance of their energy efficiency program, the goals of many evaluations became to ensure utilities were not getting credit for savings they did not deserve, rather than providing insight that would enable achieving improved savings. Measured precision was high (or assumed to be so through the use of building simulation models that reported results to many digits but had unknown statistical properties), and conservative assumptions to avoid over-estimating savings were considered the most reasonable assumptions given the high level of uncertainty about measurement of savings.

The framework of all reductions in kW or kWh being equal no longer meets the same needs as it once did. As time-varying renewables make up a greater and greater part of the generation mix, a kWh avoided during one hour can be vastly different from a kWh avoided during another hour. The State of California has even reached the point where during some hours the solar power generated is greater than the load, resulting in a situation where reducing energy consumption even further actually incurs a net cost to the utility (and its customers) and the long-term economic value of these resources may be distorted as a result. In a world where marginal emissions (and cost) vary so widely, it is critically important to consider the timing of the energy savings. Fortunately, the increasing prevalence of interval metered consumption data through the increasing deployment of advanced metering infrastructure (AMI) enables evaluators to estimate these time differentiated savings based on actual performance, and not rely on building simulation models constructed from assumed conditions.

Program approaches enabled by this formulation

In addition to focusing program goals on the true desired social outcomes, focusing on GHG emissions allows more flexible program approaches to meet customer needs. Traditional programs have focused on a single element of demand-side management (DSM). For example, a program would offer energy efficiency rebates and possibly engineering support, but if a customer also wanted to enroll in demand response (DR) and install solar photovoltaics (PV) and battery storage on-site, that customer would need to fill forms and take other steps to participate in separate programs that supported DR, PV, and storage. This design makes sense in the traditional regulatory framework: utilities had separate goals for their EE, DR, and on-site renewables programs that could be difficult to compare, let alone share between programs. This reflects the general problem of “program silos” and their lack of integration, especially with evaluation (see Vine 2008).

With a single unified metric, reduced GHG emissions, these program types can all be brought together into one delivery mechanism. While this change will likely require different types of regulatory oversight, and utilities will need to develop new program designs and re-organize internal structures and teams, it has the potential to greatly simplify participation for the customer. Moreover, addressing this coordination on the utility program side is likely to be much more efficient and effective than having each customer individually coordinate between the utilities’ programs.

Breaking down these silos can also support better project design by directly integrating these elements (Vine 2008). When energy efficiency, DR, and renewables are combined without coordinating them, there is a danger that unintended consequences of the disjointed program design will make projects less desirable to the customer. For example, installing renewables before enrolling in DR can make the customers' peak consumption look lower and reduce their eligibility under the DR program. Unifying the approaches can avoid those problems and allow utility program support to include helping customer optimize the design of integrated projects. The primary goal of these coordination elements, and of the overall program design, should be to enable the customer to make changes that make sense for them and, at the same time, benefit society. While it is important to have safeguards to ensure that ratepayer or other public funds are being well-spent, the stakes from global climate change are too great not to actively break down barriers to participation. Programs should be designed around meeting the customers' needs, rather than making things easiest for the utility and/or regulator/policymaker.

CASE STUDY: CLEAN ENERGY OPTIMIZATION PILOT (CEOP)

A prime example of a program that is shifting away from the traditional focus upon kWh and toward a focus on reductions in GHG emissions is the Clean Energy Optimization Pilot (CEOP) that is being designed and implemented by Southern California Edison (SCE). Overall, the CEOP program is designed to achieve three major goals. First, the program is designed to achieve significant and quantifiable GHG reductions. To do this, the CEOP program will develop a one-time, up-front customer agreement to provide incentives to participating customers that demonstrate annual GHG reductions at the campus level (pilot participants are all universities in SCE's service territory)). Incentive amounts will be determined by meter-based performance and will be technology-agnostic: each customer will define their own mix of GHG reduction approaches based on their needs and constraints. These approaches may include energy efficiency, DR, cogeneration optimization, on-site renewables, smart load growth, clean transportation, and energy storage.

Second, the CEOP program design seeks to improve the customer experience by developing a more flexible pay-for-performance framework to simplify requirements and allow each participant to determine what approaches work best for their organization. The CEOP program is designed to maximize customer engagement by minimizing the burdens of participation, while also aligning separate GHG reductions strategies (e.g., energy efficiency, renewables, etc.) under a single program design to enable participants to achieve the greatest reductions possible. The CEOP program participants, the University of California (UC) and California State University (CSU) campuses in SCE's territory, have already achieved significant improvements in their energy consumption with SCE's support over many years. While traditional prescriptive or custom-calculated program designs can be constraining, either due to uncertainty of energy savings or the difficulty of conducting complex engineering models, the CEOP program will allow the campuses to take additional steps to achieve deeper reductions in the emissions that do not fit neatly into either of the traditional program designs.

Finally, the CEOP program is designed to act as a learning opportunity for SCE and others. This approach utilized by this pilot differs from previous program designs, and SCE wants to be able to assess the effectiveness of the design while the program is being implemented and discover how it should be changed in order to scale from a pilot to a program for a broader set of customers. To support this assessment and potential scaling, the authors developed the evaluation framework that SCE filed in its application with the California Public Utilities Commission (CPUC) to run the pilot. The evaluation approach specifically aligns with the goals of the CEOP as discussed above. It is meant to (1) allow for flexible participant approaches, (2) impose a minimum burden on participants while providing the necessary rigor to ensure that funds are well spent, and (3) gain early and ongoing insight about how the program is operating to determine what improvements will best achieve the program's objectives.

AN EVALUATION FRAMEWORK FOR GHG REDUCTION PROGRAMS

The evaluation framework developed for the CEOP, and that will be applicable for other programs focused on GHG emissions reductions, involves quantitative meter-based estimates of GHG reductions supported by process evaluation. The impact evaluation framework uses statistical techniques for analyzing energy meter-based data to develop estimates of on-site GHG emission reductions. This approach extends the use of customer energy meter analysis in order to measure changes in GHG emissions. The majority of GHG emissions are from fossil fuel combustion, whether it be from vehicles, on-site power generation, or from purchased electricity or natural gas. As GHG emissions are not directly metered, this Pilot would use metered energy data as an indicator of the source of

GHG emissions and convert energy consumption to GHG emissions using emission-intensity metrics. This approach is aligned with principles outlined in the International Performance Measurement and Verification Protocol (IPMVP) Option C and ASHRAE Guideline 14, which use quantitative and statistical analysis to identify any changes in energy use using data from energy meters. It also aligns with similar guidance, such as the United Nations Framework Convention on Climate Change CDM methodology AMS-II, the Uniform Methods Project sponsored by the US Department of Energy (US DOE), and guidance provided to US Environmental Protection Agency (see, for example, Vine and Sathaye (2000).

This evaluation makes use of statistical techniques to guide data collection, the establishment of baselines, and normalization of these data. GHG reductions are estimated by comparing modeled emissions during the period before the program begins and modeled emissions during the program implementation, using hourly or daily meter data to achieve a high degree of rigor and robustness without representing an excessive burden to program participants. Energy consumption units are converted to GHG units using best-available hourly grid intensity factors for electricity and gas emissions factors for natural gas. The quantitative data are geared at estimating the impact of the program in kWh, kW, therms / joules, and GHG emissions. These would include site-level electric and gas meter data, as well as any sub-meter data available during the program. While the evaluation approach does not require customers to install any new metering equipment, the evaluation design does allow evaluators to use any and all available meter data as relevant. For programs or projects that target a well-defined set of customer segments (such as the campuses in the pilot described below), the approach also leverages the similarities between different buildings or sites to improve the estimation by aggregating the meter data into consumption segments that have similar properties.

In addition to quantitative data, the evaluation design uses qualitative data to inform the overall analysis. One goal of collecting qualitative data is to capture changes in energy use that are due to non-routine events, such as the addition or removal of large loads, tenant move-in/out, or square footage additions. These events would likely result in major changes to energy use, but are not due to program activity. If non-routine events were not accounted for in the analysis, there would be risk that results would be over- or under-stating actual impacts. Examples of qualitative data collection include changes to building operating schedules including shutdowns or unusual events, retrofits or major equipment changes, and occupancy changes. Qualitative information about what behavioral activities actually took place at the more granular level may also be useful for program planning purposes and to help participants manage or prioritize their behavioral activities. Additional qualitative information could also support the impact analysis by correlating achievements to behavioral activities. For example, evaluations may be able to show some resolution around what behavioral activities or actions happened within a specific timeframe that correlates with a measurable drop in emissions.

CEOP IMPACT EVALUATION APPROACH: DETAILED DESCRIPTION

The following section describes this overall process in greater depth. As noted previously, this work builds on the work of other evaluation practitioners and the description provided is not intended to be comprehensive. Rather, we walk through the process step-by-step to describe the key elements, discuss possible variations in cases where programs include transportation electrification or where AMI electric data are not available, and

Identify Available Data and Refine Analysis Plan

The evaluation begins with collection and cleaning of data. Note that, ideally, the evaluation team will be present at project inception to support formative research and to ensure that data for the impact evaluation are collected and tracked in the most efficient and minimally disruptive means possible. Data requirements for the evaluation include:

- Building use-type information (if the evaluation includes organizing loads by consumption segments)
- Hourly electric usage data
- Daily gas usage data
- Building conditioned square footage (if there is a possibility of new construction during the program period)
- Hourly electric vehicle charging data (if vehicle electrification is part of the program)

Once the evaluator has gathered the data, the evaluator should assess:

- Data granularity, including building sub-metering availability and energy meter data frequency. This impact approach is built on the assumption of having building-level meters and sub-meters reading at an hourly (for electric meters) or daily (for gas meters) level.
- Building characteristics data. This impact approach assumes that if building-level metered data is available, the use-type and conditioned square footage are also known.
- Availability of other data. The impact approach described below outlines how other data (e.g., operating hours, information about non-routine events) could be incorporated into the analysis, when it is available.
- Specificity of electric vehicle charging data and other fleet usage information. The approach to GHG reduction quantification assumes vehicle charging is metered separately or sub-metered. Unlike other sources of emissions, vehicle travel that changes from gasoline to electric will not be directly reflected in metered data for the pre-period and the post-period. The extent of the available information about fleet usage in the pre-period will influence what assumptions are necessary to estimate reductions in vehicle emissions based on differences in fleet efficiency and vehicle miles traveled over time.

If the data review identified any discrepancies between the assumptions documented in this Pilot Program Evaluation Plan and actual data available, the GHG reduction quantification approach discussed in the sections below would be refined to accommodate actual data availability.

Steps 1 and 2. Clean and Process Data

Data cleaning steps could include:

- Identifying and filling in missing data by interpolating for small amounts of missing data (a few hours) or excluded timeframes for larger amounts of missing data.
- Removing duplicates.
- Analyzing and recording any outliers to confirm they are data glitches and not due to something in the building, charger, or solar array.
- Aligning time and date stamps from the various data sources.
- Aggregating data as necessary for modeling (e.g., rolling 15-minute data to hourly, or hourly data to daily).

The outcome of these steps would be a data set that is ready for the remaining data analysis steps.

Step 3. Define Consumption Segments

If the program lends itself to categorizing loads into consumption segments, building meter data should be grouped into consumption segments that have similar energy usage patterns. For example, residential buildings will have a different energy use profile than office buildings or grocery stores. As such, the variables that drive energy use, such as temperature and occupancy hours, will be different for each segment of buildings. As such, the types of models (change point, spline, etc.) that result in the best model fit might also differ by use type. These groupings would also make square-footage standardization more meaningful, as energy use intensity is highly variable based on use-type.

While modeling each building separately may possibly result in a highly precise GHG reduction estimate per building, such analysis would necessarily increase the overall administrative burden and would be contrary to the goal of reducing administrative burden. An analysis that groups loads in meaningful ways would improve the modeling in cases where new construction occurs. This approach would also reduce uncertainty in the baseline models, thereby offering the ability to potentially detect smaller GHG reductions as compared to building-level analysis. Therefore, this approach would aggregate the energy use data for each group of sites into one dataset each for kWh/kW, and therms / joules.

Step 4. Create Normalized Energy Models

The analysis should use a normalized approach where a baseline energy model and a post-implementation energy model will be created using monitored data from the respective time periods. Both models would then be driven by a common data set, such as recent Typical Meteorological Year (TMY) temperature estimates. Therefore, four models for each segment would be created: baseline models for kWh/kW, post-implementation models for kWh/kW, baseline models for therms / joules, and post-implementation models for therms / joules.[1] For each model, the analysis would:

- Investigate possible independent driving variables such as temperature, time of day, day of week, and occupancy based on available data.
- Examine different model specifications including change point models, spline models, interactive variables, and trends to improve model fit and validity. These may include both ordinary least squares (OLS) regression models as well as machine-learning-based models.
- Test and correct (where possible) for autocorrelation.
- If necessary and possible, incorporate non-routine adjustments to account for changes in building use, occupancy, or other changes in energy use.

The inclusion of the trend term in the baseline model would reflect the pre-existing rate at which overall GHG emissions were being reduced at the site. The analysis would then be able to estimate the extent to which GHG emissions are being reduced beyond the reductions that were already occurring at the site. Because participants may have begun changing the trajectory of their GHG reductions prior to the start of the program, it would be important to allow that trajectory to be flexible in the pre-period models. That is, while a longer series of pre-period data would help in controlling for weather and other observable characteristics, estimating a fixed trend over a longer pre-period would attenuate estimates of the trend. Allowing the trend to be flexible over the pre-period would address this concern.[2] The model that provides the best fit to the data should be chosen, as determined by exhibiting the lowest NMBE and CVRMSE or other applicable measures.

Each model would then be combined with a common dataset, such as TMY temperature data or observed weather. The outcomes would be two 8760 profiles (for kWh and kW) and two 365 profiles (for therms / joules) for each segment that control and normalize for weather. In successive program years, the post-period would extend to ensure that reductions achieved are beyond what had been achieved in prior years.

Step 5. Calculate Time- and Location-Specific GHG Emissions by Segment

To calculate time- and location-specific GHG reductions, the models developed in the previous step should be used to derive an annual baseline and post-implementation load profile for each segment. Each time point (hour for kWh or day for therms / joules) would then be multiplied by a Grid Emissions Intensity (GEI) metric or a Natural Gas CO_2 Intensity (NGI) metric, as appropriate. The overall baseline would be the sum of the 8760 hourly electrical emissions values and the 365 daily gas emissions values. The calculation of post-period GHG emissions is equivalent, with post-period energy consumption replacing baseline consumptions. This approach would allow the CO_2 intensity per unit of electrical energy to change over time according to the SCE generation mix. The outcome of this step would be the total baseline GHG emissions and total post-period GHG emissions by segment in tons of CO_2.

Step 6. Program-Level Annual GHG Reduction

The final step in the quantitative analysis will be to calculate the annual GHG reduction at the program level. Annual GHG reduction would be calculated as the total baseline GHG emissions, minus the total post-period GHG emissions. The result would be an estimate of GHG emissions reduction in terms of tons CO_2 equivalent (tCO2e).

Incorporating Vehicle Electrification

If the program includes vehicle electrification, this element must be treated separately. Because electrification takes a non-metered source of emissions (gasoline tailpipe emissions), and makes it a metered source through an electrical meter, increased metered consumption is actually a reduction in emissions. To deal with the unobserved avoided gasoline consumption, pre-period emissions can be estimated using a conversion factor between gasoline and electric vehicle emissions based on a vehicle miles traveled metric.

---

[1] Note that in hourly models, kWh and average kW are equivalent and so would not require a separate model.

[2] A potential method for achieving this would be to estimate piece-wise continuous trends for each year of the pre-program model and then constrain the trend to be equal over successive previous years until model fit declines and residuals become unbalanced across time within the year immediately preceding the Pilot.

Dealing with Monthly Consumption Data

This evaluation methodology is based on modeling metered energy consumption at the 15-minute or hourly level for electricity and the daily level for gas. If only monthly data are available, the overall logic of the approach could be used, but it would lack the empirical basis and the statistical rigor of this approach. Instead, load profiles for the individual changes or building simulation models could be used to produce the emissions profiles that are developed in Step 4. This would obviate many of the benefits of the GHG approach discussed above in terms of the beneficial program design.

Process Evaluation

Process evaluation is an integral part of the overall evaluation of GHG reduction programs. We believe that a developmental evaluation approach is the most appropriate approach for conducting process evaluation of a program focused on GHG reductions. Pioneered by Michael Patton (2011), developmental evaluation “supports innovation development to guide adaptation to emergent and dynamic realities in complex environments.” Instead of considering themselves outsiders to the program design and implementation process, evaluators should work with program administrators to act as an internal resource. Instead of collecting and reporting results at one time in a summative fashion, our approach is designed to provide feedback throughout the program. This approach creates a decision-making framework for improving the value of evaluation results when engaging innovative program designs operating in complex environments. This approach is defined by several components, but those most relevant to the a GHG reduction program evaluation include:

- Engaging program stakeholders early and frequently,
- Embedding the evaluation as a core component of program design and delivery, and
- Providing ongoing feedback on the program’s performance.

By identifying key stakeholders at the start of the evaluation and by engaging with them early and frequently, evaluators will ensure that their results are truly useful in driving needed changes in the program. Using this framework, the process evaluation should have several key research objectives including (1) understanding participants’ experiences with the program, (2) tracking the influence of program incentives on customers’ decision-making, (3) identifying opportunities for pilot implementation improvement, and (4) identifying ways to improve future program designs.

CONCLUSIONS

While energy program evaluation is intended to support policy objectives, it is also possible for program evaluation to actually impede these objectives if not fully aligned. With a clear focus on GHG emission reductions, the program design and evaluation framework can be designed to explicitly support each. Specifically, scarce time and money allocated for evaluation can be used to collect and support metrics that are linked directly to the policy objectives at hand (i.e., GHG emission reductions, rather than kW or KWh savings) and – most importantly for our society – guiding these programs in ways that ensure overall success in reducing GHG emissions that contribute to global climate change. The CEOP program in California – and the evaluation framework developed for this program – supports these intentions and will provide important learnings for the entire energy program evaluation community.

REFERENCES

Energy Information Agency (EIA), Global Energy &CO2 Status Report. https://www.iea.org/geco/electricity/. Accessed 25 July, 2019.

Patton, M. 2011. Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use. The Guilford Press. N.Y., N.Y.

US Department of Energy (DOE) – Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures. https://www.nrel.gov/docs/fy18osti/70472.pdf Accessed 15 September, 2019.

Vine, E. "Breaking Down the Silos: The Integration of Energy Efficiency, Renewable Energy, Demand Response and Climate Change." *Energy Efficiency* 1: 49-63 (2008).

Vine, E. and J. Sathaye. "The Monitoring, Evaluation, Reporting, Verification, and Certification of Energy-Efficiency Projects," *Mitigation and Adaptation Strategies for Global Change*, 5(2): 189-216 (2000).