COMMERCIAL WHOLE BUILDING DEMONSTRATION

Joint Study Report

Pacific Gas and Electric Company
California Public Utilities Commission

CALMAC Study ID PGE0431.0

May 1, 2019

Acknowledgements

Pacific Gas and Electric Company is responsible for this demonstration project that was conducted jointly by Pacific Gas and Electric Company and the Energy Division of the California Public Utilities Commission. kW Engineering managed this project and David Jump guided and managed it from onset to completion. The study was developed as part of Pacific Gas and Electric Company's Energy Efficiency Programs. The project team would like to acknowledge the significant contributions of the following individuals without whom this project would not have been possible: Leo Carrillo, formerly of Pacific Gas and Electric Company; Paula Gruendling and Jeorge Tagnipes of the California Public Utilities Commission; Peter C. Jacobs of BuildingMetrics Inc.; Amit Kanungo of DNV-GL; Nikhil Gandhi of Strategic Energy Technologies, Inc., and Susan Norris, Phillip Broaddus, Alex Wortman, Billy Roderick, and many others from PG&E. We are grateful particularly to the customers of PG&E, as well as the evaluators, implementers, modelers, regulators, reviewers, and the many other individuals who contributed to the realization of this project. For more information, please contact Brian Arthur Smith of Pacific Gas and Electric Company at b2sg@pge.com.

Executive Summary

1. Introduction

In 2012, the California Public Utilities Commission (CPUC) requested its regulated investor-owned utilities (IOUs) to develop energy efficiency (EE) programs that focus on interventions that leverage comprehensive approaches to achieving deep savings in commercial buildings (Decision 12-05-015, 2012). In response to this request, Pacific Gas and Electric Company (PG&E) developed and launched the Commercial Whole Building Demonstration (CWBD) in 2014.

Several factors were considered in the development and launch of the CWBD. These included:

- California's IOUs including PG&E had substantially completed installation of advanced metering infrastructure (AMI) to customer buildings, which provided electric and natural gas energy use data in short time intervals. These data enable relatively fast feedback on a building's energy use patterns and day-to-day performance.
- To date commercial energy efficiency (EE) programs have not widely leveraged the availability of AMI data, nor the new analysis tools and techniques. Recent years have produced research on the accuracy and technical merits of M&V based on AMI data; however, evaluation protocols have not yet been updated to address the use of AMI data. This is due in part to the established measure-based regulatory framework and partly due to unfamiliarity with the more advanced analytical methods used to model energy use patterns in existing buildings based on AMI data among industry practitioners.
- Development and widespread use of energy management and information systems (EMIS) have helped building operators identify and address building performance issues and quantify avoided costs of installed energy efficiency measures.
- In 2015, the California Governor signed into law Assembly Bill 802 (AB 802, Williams, 2015) that directed the CPUC to allow IOU savings claims based on reductions in energy usage observed at the meter, normalized for factors that typically influence energy consumption. This methodology is known as normalized meter-based energy consumption ("NMEC"). This has provided the EE industry with new methods to estimate savings and track performance for multi-measure, deep savings commercial buildings projects and elevated the need to better understand 'meter-based' methods and their role in EE programs. Throughout this document, we refer to meter-based programs as those that provide this site-level NMEC measurement approach to quantifying savings.

While whole building M&V methods have existed for many years, California IOUs have not implemented commercial programs to achieve deep savings in existing buildings at scale. Consequently, there is a need to create a formal process (the Joint Study Process) involving multiple stakeholders to subject the methods to rigorous testing and to memorialize the results in a thorough report. The CWBD is a 12-building demonstration program ("Demo") designed to implement and evaluate multi-measure, deep savings projects and to pay incentives based on savings achieved over the year after implementation. Two whole-building M&V methods, IPMVP's Option C and Option D were used to quantify savings. Two EMIS vendors were engaged to provide Option C M&V and one technical consultant provided Option C analysis using two public domain models. CWBD Implementers provided the Option D analysis. Participation in the CWBD was invitation-only and was based on the customer's business sector, premise characteristics, and past energy usage.

PG&E and the CPUC initiated a Joint Study Process of five participants early in the program as customers were recruited and continued well past payment of the final incentives. Examination of the CWBD's program design, implementation processes and requirements, savings analysis, and results provided an opportunity to inform a framework for program design and evaluation moving forward. A limited sample of five of the twelve program projects were selected for the Joint Study; the remaining projects were part of a separate study undertaken by SBW Consulting. The Joint Study team is comprised of PG&E CWBD program administration staff, CPUC staff, technical consultants, and evaluators.

PG&E offered participants incentives based on savings determined by Option C methods. In all cases PG&E used savings estimated using the Option D method as the basis for the program's regulatory savings claims. Regulatory savings were claimed based on the measure-by-measure savings estimates from Option D since the CWBD was developed and implemented prior to passage of AB 802. Option D was added to the Joint Study after its launch after CPUC clarified that a savings claim resulting from the CWBD must be made according to regulatory rules. A back-casting approach was used in Option D in which simulations were calibrated to post-installation data.

This Demo provided a unique opportunity to compare results from each of these unique M&V methods, one essentially a whole-building "top-down" approach using statistical methods (Option C) and the other a measure-focused "bottom-up" approach using engineering simulations (Option D).

2. Study Objectives

The objectives for the Joint Study were narrowed to the four described below.

1. **Comparison of whole building program savings estimates between Option C and D approaches.**
   How do the savings estimates compare between the two methods and what are the major risks to whole building programs that each method addresses? What trade-offs are made when using one methodology over the other?

2. **Improvements to key program technical requirements and documentation.**
   What worked well and what needed improvement regarding the CWBD's technical and documentation requirements that are useful for future program designs and evaluation? What specific improvements are needed for successful implementation of Option C and Option D methods moving forward?

3. **Factors informing the design of future meter-based whole building performance programs.**
   What policies and procedures are required from both a regulatory and a program administrative perspective? What data and information are needed to support evaluation of a program using IPMVP Option C to determine savings in the California regulatory context?

4. **Addressing the emerging technical and regulatory questions necessary to scale such programs and support evaluation.**
   What changes to the existing California evaluation frameworks are appropriate for whole building programs? How can NMEC-based programs be designed to facilitate evaluation? What data are critical for IOUs to collect that will facilitate timely evaluation?

3. Results

3.1. Comparison of Savings Resulting from Option C and Option D

Figure 1 shows the comparison of the electric savings for each of the five study sites as estimated from Option D and Option C analyses. Option C results are based on the open source time of week and temperature (TOWT) modeling algorithm developed by Lawrence Berkeley National Labs. The figure compares both the weather-normalized and the actual savings estimates for the two methodologies. Similarly, Figure 2 shows the comparison of the natural gas savings estimates for the study sites as estimated from Option D and Option C TOWT analyses.

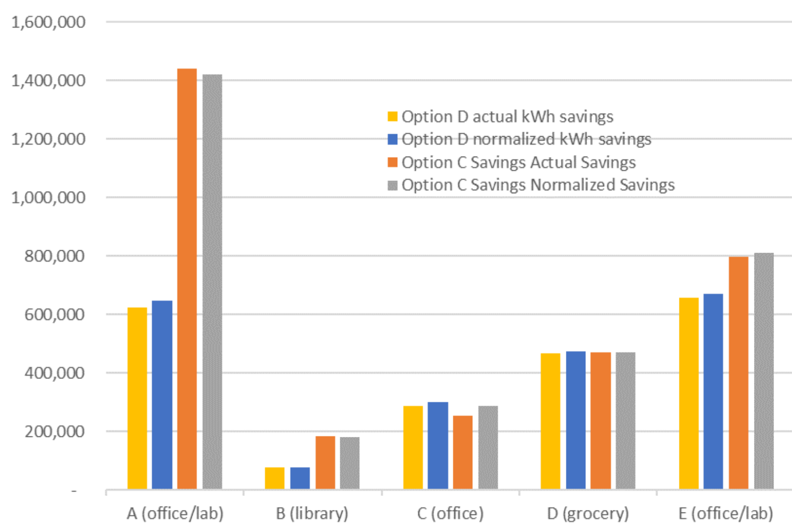

Figure 1: Comparison of Option C and Option D Electric Savings Estimate of 5 Study Sites

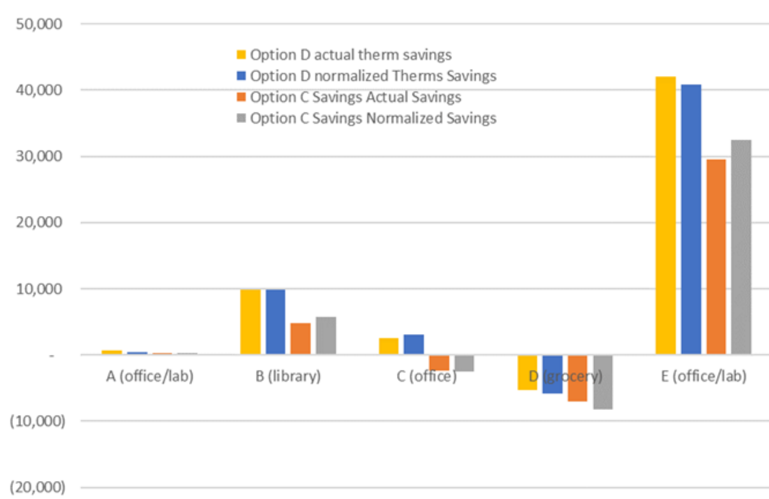

Figure 2: Comparison of Option C and Option D Natural Gas Savings Estimate of 5 Study Sites

From the two figures above, it is evident that overall the savings estimates for the five study sites didn't change significantly with the change of weather data from actual to normalized conditions. However, significant differences were observed between the Option C and Option D savings results. In three cases, the Option C electric savings estimates were higher than the Option D estimates. Overall, the Option D savings estimates were higher than the Option C estimates for gas savings.

The Option D simulations and Option C models provided kWh savings estimates that were comparable (within 5%) in two of the five buildings (Sites C and D) whereas there were large differences between the kWh savings estimates for the other three buildings. The key factors driving these differences in savings estimates between the Option C and D estimates were identified as non-routine events (NREs) confirmed for sites A and E, and possibly occurring at site B. The known NREs at sites A and E were simulated in the Option D method.

The Option C analysts (PG&E technical consultant and EMIS vendors) did not conduct site visits. They were provided the energy use data by PG&E to run their M&V analyses, but they did not collect data on building operation schedules, special events, or NREs. Thus, the analysts had no ability to adjust their models or check the accuracy of their estimates after having adjusted for these events.

The Option C and D methods did not produce similar estimates of natural gas savings for any of the five study sites. In three of the five sites, the Option C modeling algorithms could not reproduce the baseline gas usage patterns, resulting in poor goodness-of-fit (GOF) metrics. NREs were suspected as the cause of the differences with Option D in the two of the sites. Some of the Option D simulations did not calibrate well to gas usage. For these sites, the equipment was not defined appropriately in the simulations due to lack of equipment specifications and operational data having been incorporated into the models.

4. Discussion and Recommendations

4.1. Comparison of Option C and D Approaches to Whole Building Program Design

Option C and Option D represent fundamentally different approaches to estimating savings resulting from energy efficiency interventions. Both estimation approaches have advantages and disadvantages depending on the application. Some of these are discussed below.

- The Option C approach is generally lower cost because it bases savings estimates on the totality of energy savings by observing the changes in energy use before and after EE measures are installed. This meter-based approach works best for projects where savings are expected to be large as compared to the total energy consumed at the premise. A key challenge for Option C is when significant NREs occur at a facility during the reporting period, adding cost and uncertainty to the savings estimate.
- The Option D approach is more labor intensive—and therefore costlier—given its use of simulation software that requires considerable training to use, the level of effort and time to collect input data and information, and the iterative runs required to reach calibration and estimate savings for each EE measure. A key advantage is its ability to isolate savings estimates for individual measures and estimate interactive effects between measures.

Shortcomings of each approach, along with tradeoffs to consider when selecting an approach are summarized in Table E-1.

Table E-1. IPMVP Whole Building Option Selection

Shortcomings of Option C	Shortcomings of Option D	Tradeoffs Between Option C and D Methods
• Option C models have a risk that results do not reflect the actual savings achieved by the project – mainly due to NREs. • Models that don't meet goodness of fit criteria. • Projects that don't generate savings "above the noise."	• Ignoring or mischaracterizing behavioral (and operational) issues in whole building simulations • Simulation errors due to inaccurate or insufficient inputs • Simulation errors due to simulation engine limitations • Simulator's experience • Simulation calibration cost	• Data requirements • Non-program-related changes to buildings that confound the analysis • NRE adjustments

4.2. Recommendations to Improve Program Technical Requirements and Documentation

The CWBD was ambitious. It was developed to test many aspects of achieving comprehensive deep savings in participating buildings. The Demo employed unique entities to perform different activities. The administrator developed infrastructure to provide data to software vendors, technical consultants, and implementers. The CWBD began in 2013 and its final participant's incentive was paid in 2018. The Joint Study identified several elements that worked well and several areas that could be improved.

- Option C analysis by remote parties. While the electric energy models performed well based on goodness-of-fit metrics, the models did not account for building operation schedules that could have improved the models, nor for NREs that created biased estimates.
- Building pre-screening. Poor model goodness-of-fit metrics were obtained for natural gas models for four of the five study sites. Building pre-screening, whereby one year of historical meter and weather data would inform which buildings and energy sources are good candidates for Option C analysis, should be required for improved modeling outcomes.
- Use of performance period monitoring for NRE detection. Option C analysts should periodically download and analyze whole building meter data throughout the performance period to track whether the expected savings are materializing and assess whether NREs have occurred during the performance period.
- Level of detail "sweet spot" for Option D simulations. While it is generally time consuming to create fully geometric representation of buildings with all HVAC zones included, a "sweet spot" must be found that allows for an accurate representation of building characteristics, operations and loads presented to the HVAC systems affected by the measures but does not take an inordinate amount of time to create.
- Improve Option D calibration requirements. The CWBD procedures manual required verification report simulations to be calibrated with monthly data only. Building simulations can be improved

through calibration to hourly data to true up daily building operating schedules, identify daytypes, and unanticipated building shutdowns.

Specific technical improvements were identified that may improve the success of future whole building Programs:

- Development and standardization of NRE identification and impact quantification methods. NREs occurred in a significant number of the joint study sites. Given their uncertain characteristics, multiple methods should be used to identify and quantify NREs, including through analysis of AMI data, ongoing contact with building operators, through adjustments in whole building simulations, and through modifications to engineering calculations based on building data collected separately. More development work is needed in this area.
- Improve methods for estimating savings uncertainty for Option C methods. Further work is needed to develop an industry-accessible methodology to estimate savings uncertainty using advanced data models and short time interval data.
- Improve methods for estimating savings uncertainty for Option D methods. The estimation of saving uncertainty in Option D models has received less attention than Option C methods. Research has brought some statistical rigor to the estimation of savings uncertainty, but more work is needed.
- Industry standard calibrated simulation procedures for existing buildings. The CWBD's procedures for calibrating simulations fell short of simulating the study site buildings accurately. A search of building energy simulation industry organizations did not reveal an industry-consensus procedure for calibrating simulations of existing buildings. An industry-standard procedures document for Option D methods would greatly benefit the industry.

4.3. Design of Future Whole Building Performance Programs

Further development work is needed so that California will have the tools in place to administer and evaluate these types of programs. Several recommendations are offered to improve program management and improving program designs. Some of these elements that were identified are discussed below.

Program Management

- The greenfield subject matter of whole-building meter-based methods made it difficult for those without a statistics background to navigate. Recommendations for improved program management include:
    - Development of a matrixed approach to project management should be considered whereby representatives from multiple organizations having roles managing specific program functions are involved in program planning, including: customer targeting and outreach, field engineering, project implementation, data science, project technical review, evaluation, policy, regulatory reporting, and regulatory oversight.
    - Formal project planning tools and ongoing revisions to their outputs are essential for communication, recordkeeping, management and evaluability.
    - A "living" policy and procedures manual that details program requirements and processes that accommodates frequent updates.
    - A "data and file specifications and workflow requirements" document to facilitate program processes as customers are recruited, their energy use patterns are analyzed, baselines are

established, measures are identified, savings are quantified, and the program is evaluated. Key files include interval-level energy usage data, premise characteristics, key weather data, project data, and tracking and adjustments for NREs.

- Embedded M&V. In addition to an in-house evaluator on the team to represent the program administrator, third-party evaluators should be included at the planning stage to inform data collection requirements and procedures to ensure evaluability and conformance to regulatory requirements.
- Single repository of record for project files and a manageable directory structure.
- Weather data. The utility should be the sole source for weather data used in the analyses, including the mapping of treated premises to the appropriate weather stations.

Program Design Elements and Requirements for Success

Some of the most important design elements are:

- Program implementation plan incorporating embedded M&V. Having a program implementation plan (PIP) that explicitly identifies the points in the program where key information and data will be provided to regulatory staff and evaluation consultants is critical.
- Customer targeting, including use of an invitation only approach, pre-screening customers, and recommending program alternatives for unsuitable projects.
- Documenting program influence. Guiding principles for program design should be focused on effectively influencing potential participants to develop and implement multi-measure deep savings efficiency plans in their buildings.
- Documenting existing conditions prior to project acceptance. Proper documentation of existing conditions ensures that the baseline established routine operating conditions and is not artificially depressed from non-working equipment or artificially inflated from poorly maintained equipment.
- Need for a "Plan B" analysis plan. An alternate analysis strategy will need to be in place to meet customer needs when the Option C analysis is inconclusive or cannot be completed.
- Documentation of energy efficiency measures. Because the demo evolved to be a comparison between Options C and D, detailed information about measure-level savings estimates and expected useful life was captured. Moving forward with a meter-based program, capturing measure-level detail will be just as critical.
- Post-installation inspections. Given their reliance on post-installation data to populate their models, it is critical that on-site, post-installation inspections are built into the program design. These inspections should go beyond a simple physical inspection since proper operation of the measures can be difficult to verify without performing functional performance testing.
- Planning to address non-routine events (NREs). Two primary means to address NREs include: Ongoing structured interviews with building operators and software detection of NREs.
- Incorporate performance monitoring of energy use throughout the implementation and performance periods. Performance monitoring may be implemented in different ways, but it is essentially a comparison of actual energy performance to expected energy performance.

4.4. Addressing Emerging Technical and Regulatory Questions to Scale Programs and Support Evaluation

The site-specific Option C methodology of the CWBD was a direct response to the CPUC's call to IOUs in 2012 to develop a comprehensive approach to accomplish deep savings in commercial buildings. At that time—and until the passage of AB 802 in 2015—there was no methodology approved by the CPUC to use utility meters as a basis to determine energy and demand savings in commercial buildings.

There are two documents governing evaluation methods in the State of California: the California Evaluation Framework (2004) and the California Evaluator's Protocols (2006). Neither document provides a methodology for using data from utility meters to determine energy and demand savings for a specific premise whereby adjustments must be made for non-routine events. There is a recognition among evaluation professionals that the California Evaluation Framework needs a major update. A needs assessment in 2007 outlines a set of recommendations that include the addition of methodologies for NMEC analyses. Until the time that such an update is available, Southern California Edison has completed a first-generation NMEC Savings Procedures Manual for implementers that addresses much of the site-level analysis requirements. The Efficiency Valuation Organization, publishers of the IPMVP Protocols, has elected to develop its Advanced M&V guide based on this manual.

Several recommendations were made to improve program design to facilitate evaluation and increase the scale of efficiency through programs.

- Develop a standard set of procedures and requirements for the development and application of advanced M&V, including validating meter data, preparing data sets, matching weather data files to premises, selecting appropriate modeling algorithms, reporting model goodness-of-fit metrics and uncertainties, and reporting savings.
- Increase transparency and replicability to facilitate evaluation by use of well documented open source modeling algorithms. Use validation methods and standards such as those developed at LBNL when proprietary modeling algorithms are used.
- Resolve uncertainty estimation methods for hourly models and use savings uncertainty as a criterion in place of rules-of-thumb at both a project-level and a portfolio level, enabling projects with poor model fit but high savings as well as projects with good model fit but low savings to participate in meter-based programs.
- Always strive for deep savings, greater than the 15% targeted in CWBD, and allow inclusion of capital projects and BRO measures.
- Programs delivered using ratepayer funds under the jurisdiction of the CPUC must meet regulatory cost effectiveness criteria. Cost effectiveness calculations require a consistent set of metrics, including gross savings, project costs and net-to-gross ratios consistent with gross savings baseline assumption. The effective useful life of a multi-measure project is based on estimated savings-weighted EUL of the measures installed. Measure-level engineering estimates will be necessary to estimate project EUL and savings calculations done during project development will likely suffice for this purpose.
- Embedding M&V into the design and operation of a meter-based program facilitates the evaluation process by ensuring that programs are evaluable. Meter-based programs are data driven, and by their nature require data collection on the part of the program implementers to a level of quality required by evaluators. Given the time required for projects to complete the performance period

and enter the savings claim process, the ex-post evaluators will need to maintain contact with implementers throughout the project lifecycle. Data gathering, documentation of processes, and data analyses should be responsibilities that are shared between program implementers and evaluators.

5. Conclusion

5.1. Comparison of whole building program savings estimates between Option C and D approaches.

The predominant risk in whole building programs is the occurrence of NREs and their impacts on the energy savings, whether estimated through Option C or Option D. In this study, the team could estimate NRE impacts after it collected information from the site that enabled it to be simulated in the Option D model. The Option C analysts did not go on site and therefore were blind to any occurrences of NREs. A proactive approach to identify, document, and quantify non-routine events is critical for estimating site by site savings. All models, regardless of the GOF statistics, will converge on the wrong savings if NRE adjustments are not included.

The CWBD encountered issues with both Option C and Option D analysis. Programs should avoid a focus on a single analysis method and recognize the need for 'backup' analysis methodologies to meets the needs of individual projects and regulatory requirements.

5.2. Improvements to key program technical requirements and documentation.

Although there was significant development of the program policies and procedures manual, data specification documents, reporting templates, and other resources, the study team found weaknesses in their content, and lack of enforcement of their use.

It was also noted that there are several areas where consensus industry standards would be helpful: establishing a framework and methodology for identifying and quantifying NREs, improving uncertainty estimation methodologies for both Option C and D M&V methods, and developing an industry standard methodology for calibrating Option D simulations and estimating individual measure and well as whole building savings.

5.3. Factors informing the design of future meter-based whole building performance programs.

Deep savings whole building savings programs are long engagements with their participants, the CWBD engaged with its customers for up to five years. Consistent project management, reporting, and data archiving is critically important, as is maintaining consistency of procedures and technical review. Embedded EM&V concepts should be explored and included in program design to assure evaluators will have the information needed for their tasks.

The CWBD illuminated several best practices: identifying potential participants through analysis of usage data, pre-screening program participant to assure baseline model goodness of fit and accuracy, assuring participants willingness to invest in deep savings and await incentive payments based on performance, providing consistent energy data and weather files, providing good documentation of baseline conditions, documenting program influence, and establishing specific procedures for NREs.

5.4. Addressing the emerging technical and regulatory questions necessary to scale such programs and support evaluation.

There is a lack of an accepted evaluation methodology to use utility meters as a basis to determine energy and demand savings in commercial buildings. Neither of the two main evaluation guidance documents, the Evaluator's Framework and the Evaluator's Protocols, provides a methodology for using data from utility meters to determine energy and demand savings for a specific premise whereby adjustments must be made for non-routine events. In addition, reporting structures for meter-based programs don't fit well in deemed or custom reporting structures.

Several other considerations should be addressed to facilitate evaluation of meter-based projects and programs:

- Education and training on advanced M&V methods to reduce confusion and leverage its benefits appropriately.
- Open source advanced M&V models will be preferred, but methods to validate proprietary models should be adopted to increase the number of market actors.
- Research should be pursued in needed areas, such as developing savings uncertainty algorithms for M&V methods employing advanced modeling algorithms.
- Cost-effectiveness and useful life metrics should be broadened to evaluate the combined packages of deep saving measures installed in meter-based projects.

Finally, as meter-based projects have a long duration, successful and informed evaluation will only result when evaluation data and information requirements are embedded in program design. Embedded EM&V will be critically important to gain rapid feedback on program operation and effectiveness.

Table of Contents

1. Introduction and Background

In 2012, the California Public Utilities Commission (CPUC) requested its regulated investor-owned utilities (IOUs) to develop energy efficiency (EE) programs that focus on interventions that leverage comprehensive approaches to achieving deep savings in commercial buildings (Decision 12-05-015, 2012). Starting prior to 2012, IOUs and building owners had begun installation of the advanced energy metering infrastructure which provides the means to transmit and store electric and gas usage data measured in short time intervals. This high-frequency usage data can be analyzed and presented to provide not only insight on a building's energy use patterns, but also can provide fast feedback on day-to-day energy performance for building owners and operators. Advances in metering, data communications, and analysis methods have facilitated the widespread adoption of energy management and information systems (EMIS) (PG&E, 2013b), which have helped owners and operators use the high-frequency data to identify and address building performance issues, while quantifying energy savings and avoided costs associated with the installation of energy efficiency measures. These developments have potential to help the energy efficiency industry identify and capture more savings opportunities in California's existing building stock.

By 2013, California's electric smart grid infrastructure was substantially completed (CA Smart Grid Annual Report, 2014) but savings claims from IOUs from its EE programs (typically referred to as ex-ante savings values) continued to be based on measure-based savings values contained in workpapers workpapers or DEER[1] ("deemed" values) or based on unique calculations for each project ("custom" projects). Both deemed and custom approaches to estimating savings involve several assumptions about baseline conditions including codes, industry standard practices, equipment retirement and remaining useful life analyses, and other factors that have proven to be contentious. Verification of IOU savings claims (referred to as ex-post evaluation) by the CPUC has for many years included verification of the engineering calculations, short term metering, or use of monthly billing analysis methods. Applying these methods at scale is considered problematic. Engineering review and short-term metering methods are generally conducted on the sample of participants due to the high per-site costs and issues with customer intrusion. Billing analysis may be unreliable for program-level evaluation in some sectors.

In 2015, the California Governor signed into law Assembly Bill 802 (AB 802, Williams, 2015) that directed the CPUC to allow IOU savings claims that are based on reductions in energy usage observed at the meter, normalized for factors that typically influence energy consumption. The term for this measurement approach is normalized metered energy consumption (NMEC). The availability of granular usage data from the completed advanced metering infrastructure (AMI) network, combined with new measurement and verification (M&V) analysis tools and the directive of AB 802, have provided the EE industry with new methods to estimate savings for ex-ante reporting, track performance for multi-measure deep savings projects in commercial buildings on an ongoing basis, and verify savings estimates in ex-post impact evaluations. Throughout this document, we refer to meter-based programs as those that provide this site-level NMEC measurement approach to quantifying savings.

While whole building M&V methods have existed for many years, California IOUs have not implemented commercial programs to achieve deep savings in existing buildings at scale. "Whole-house" programs have existed in the residential sector but impact evaluations using billing analyses have resulted in

[1] Database for Energy Efficiency Resources (DEER). See www.deeresources.com

dismal realization rates (CPUC, 2017; CPUC 2016) for reasons too complex to mention here. The billing analysis approach used for ex-post verification in the residential sector has relied on the pre/post comparisons of large populations of treated and comparable, untreated residences (typically future participants, to control for self-selection bias). In the commercial sector, this comparison group method has not been well explored because of the heterogeneity of non-residential energy use, the frequency of non-routine events, and the challenges involved in identifying comparable premises, so site-specific methodologies are used.

To date commercial EE programs have not widely leveraged the availability of AMI data, nor the new analysis tools and techniques. Recent years have produced research on the accuracy and technical merits of M&V based on AMI data (PG&E, 2013a); however, evaluation protocols have not yet been updated to reflect the more widespread availability of time-granular energy use data. This is due in part to the established regulatory framework that was set up for measure-based deemed and calculated methodologies, with their incumbent requirements to incentivize only the additional savings beyond energy code and industry standard practices. It is also partly due to unfamiliarity among most industry practitioners with the more advanced analytical methods used to model energy use patterns in existing buildings, which include methods well beyond the commonly-known ordinary least squares regression approaches. Finally, the relative newness of the availability of AMI data has not allowed sufficient time to pass for academicians and other thought leaders to reach consensus on industry standards. AB 802's directive that programs shall provide incentives for savings from existing conditions baselines is expected to foster the growth of meter-based methods and make greater use of AMI data.

M&V based on time-granular data and advanced analytics is the next generation of billing analysis in the industry, and often referred to as "M&V 2.0" or "Advanced M&V." Potential benefits of Advanced M&V include the ability to estimate savings at all program participants with minimal intrusion, improved accuracy of saving estimations, rapid feedback on building performance, and automation of some elements of analysis. Requirements for implementing Advanced M&V are not fully specified beyond those used in billing analysis. As AB 802 elevated the need to understand 'meter-based' methods and their potential role in energy efficiency programs, regulators are working to determine their proper role in ratepayer funded efficiency programs.

In 2014, Pacific Gas & Electric Company (PG&E) developed and launched the Commercial Whole Building Demonstration (CWBD). The CWBD used advanced M&V methodologies as well as measure-based methods to quantify savings in its participant buildings. The CWBD provided the opportunity needed to study the merits and costs of each of these methods and help inform the role of meter-based methods in EE programs.

PG&E and CPUC initiated a Joint Study Process, described in Appendix B, early in the program as customers were recruited and continued well past making the final incentive payments to implementers after completion of the post-installation year. Examination of the CWBD's program design, implementation processes and requirements, savings analysis, and results provided an opportunity to understand its intervention strategies, and to inform a framework for program design and evaluation moving forward. A limited sample of five of the twelve program projects were selected for the Joint Study; the remaining projects are part of a separate study undertaken by SBW Consulting, a summary of which is contained in Appendix D of this report. The Joint Study team is comprised of PG&E CWBD program administration staff, CPUC staff, technical consultants, and evaluators.

Now that the CWBD is complete, this report describes the lessons learned and recommends best practices for future program designs targeting multi-measure deep savings projects using meter-based methods to account for savings. The report presents and discusses results to date from the in-depth review of the five study sites, identifies key risks and issues in deep savings meter-based programs, and describes necessary elements of a regulatory framework to help advance scaled implementation of such programs to meet California's aggressive energy efficiency goals.

2. CWB Demonstration Program Description

In response to the CPUC's directive for deep savings programs, PG&E developed the CWBD. Launched in 2014, the CWBD is a 12-building demonstration program designed to implement and evaluate multi-measure, deep savings projects and for demonstration participants to receive incentives based on savings achieved over the year after implementation. Energy savings were determined from two distinct whole-building M&V methods, one based on analysis of the meter data, and one based on calibrated computer simulations of the buildings. Both M&V methods used for estimating savings resulting from the CWBD are based on protocols from the International Performance Measurement and Verification Protocol (IPMVP) (EVO, 2016). Led by the U.S. Department of Energy, the IPMVP was initiated by a coalition of international organizations in the mid-1990s. Since its inception, the IPMVP has become the leading set of international M&V standards and provides four options for determining savings (A, B, C, and D). The meter-based option is Option C; the calibrated simulation option is Option D.

2.1. Participation Requirements

PG&E developed technical and economic requirements to select the CWBD participants that were likely to meet the objectives of an expanded program's goals. Participation in the CWBD was only on an invitation-only basis; candidate customers were selected or pre-screened among those already engaged by PG&E's internal sales team. Potential participants for the multi-measure deep savings CWBD projects were targeted using an approach built upon business sector, premise characteristics and past energy usage. The list of basic requirements for pre-screening for participation in the CWBD are listed below. These requirements were flexible, so that exceptions could be made if a candidate customer did not meet all requirements but was otherwise considered acceptable (for example, if energy use was high but square footage was lower than the threshold, or the presence of a data center exceeded the minimum load requirement).

- Own or lease a facility with at least 50 kW in electric demand or 100,000 kWh in electric usage, and have tenant improvement rights and at least five years of remaining term on the lease, if the facility is subject to a triple net lease
- Have at least 30,000 square feet of occupied space that were occupied by a single business entity
- Receive gas and electric service at the facility from PG&E (or, for electric only facilities, receive electric service from PG&E)
- Not have net metered distributed generation at the site or have plans to install such generation
- Not have a data center or other material industrial or agricultural load that may account for more than 10% of energy consumption
- Pay a California public purpose surcharge on their utility bills

- Provide PG&E with a minimum of 9 months of historical interval meter data for at least electric utility service at the time of application and continue doing so for no less than five years from the date of application
- Building operations must have been steady over time with energy use patterns that were largely dependent on weather conditions and building operation schedules, since data for these influences were readily collectable.
- Certify that no material changes to building conditions or operations are expected over the next 24 months, such as tenant additions or departures (only for multi-tenant buildings) or staffing additions or reductions
- Be willing and able to: (1) fund an efficiency project or any measures with estimated simple paybacks of two years or less (after incentives); and (2) to make a contractual commitment to implement by December 20, 2014 an energy efficiency project that may be conservatively expected to achieve at least 15% electric or gas energy savings over a subsequent 12-month period
- Participants must have been willing to receive a portion of their program incentives a year after installation of the measures.
- Participants were required to provide access to their buildings and provide data to the program administrator (PG&E), regulator (the CPUC) and their technical consultants for the duration of the engagement.

CWBD implementers were engaged by PG&E to conduct initial building assessments to identify potential measures and develop cost estimates. Proposals that were acceptable to PG&E were presented to customers by the CWBD's implementers, and customers who agreed to proceed with the projects contracted with PG&E to establish all terms and conditions of the engagement. Once the measures were in place, the implementers verified the installations.

2.2. CWBD Savings Methodologies

The two IPMVP methods used for the CWBD were Option C Whole Facility and Option D Calibrated Simulation. Appendix A includes a more detailed description of these methods based on context from IPMVP and requirements from the CWBD Policy and Procedures Manual. In CWBD, the Option C analyses used AMI data and were conducted by software vendors using proprietary methods and a technical consultant selected by PG&E using open source methods. The Option D analyses used monthly billing data and were conducted by CWBD implementers.

PG&E offered participants incentives based on cumulative whole-building gross savings determined either by Option C (when feasible) or Option D methods. In all cases PG&E used the Option D method to make the program's regulatory savings claims. Regulatory savings were claimed based on the measure-by-measure savings estimates from Option D since the CWBD was developed and implemented prior to passage of AB 802 and the CPUC required IOUs to base savings claims on above-code savings estimates. The Option D requirement was added after program launch after clarification was received that the CWBD must claim savings according to regulatory rules. These timely circumstances—the CPUC's call for the design of deep commercial retrofits in 2012, the completion of the SmartMeter build-out in 2013, the launch of the CWBD in 2014, and the passage of AB 802 in 2015—converged to provide a unique opportunity to compare results from each of these unique M&V methods, one essentially a whole-building "top-down" approach using statistical methods (Option C) and the other a measure-focused "bottom-up" approach using engineering simulations (Option D).

The CWBD implementers were responsible for developing and calibrating building simulation models that estimated measure-by-measure savings, referencing procedures in the CWBD policy and procedures manual developed for PG&E. Due to the late addition of the Option D savings estimation requirement, several projects were already installing their efficiency measures. These measures were identified based on initial building assessments by the implementers, and savings were estimated using non-rigorous means, such as similar project experience, non-calibrated simulations, or simple engineering calculations. To provide participants with partial incentives soon after the projects were installed, implementers were directed to develop a calibrated simulation within the first three months after installation. These initial simulations were calibrated to monthly baseline period energy use, and the calibration criteria was relaxed from ASHRAE Guideline 14 (ASHRAE, 2014) requirements. After the year-long performance period, the simulations were updated to include the new equipment and operation strategies of the efficiency projects and calibrated to post-installation period monthly energy use using the more rigorous calibration criteria. Table 1 below specifies the CWBD's calibration criteria for the coefficient of variation of the root mean squared error (CV(RMSE)) for Option D whole building simulations. This criteria is close to that found in ASHRAE, 2014.

Table 1: CWBD Option D Model Calibration Criteria[2]

Data Interval	NMBE	CV(RMSE)
Monthly baseline (VR1)	± 10%	± 30%
Monthly post (VR2)	± 5%	± 15%

Where:

$$NMBE\ (\%) = \frac{\sum_1^n (S_i - M_i)}{\sum_1^n M_i} \times 100 \tag{1}$$

$$RMSE_n = \sqrt{\frac{\sum_1^n (S_i - M_i)^2}{n}} \tag{2}$$

$$A_n = \frac{\sum_1^n M_i}{n} \tag{3}$$

$$CV(RMSE_n) = \frac{RMSE_n}{A_n} \times 100 \tag{4}$$

where:

M = Measured data
S = Simulated data during the same time period
i = interval number
n = number of data points in time period (12 for monthly comparisons)

---

[2] The Policy and Procedures Manual specified monthly CV(RMSE) calibration criteria of +/- 30% for VR1 models and +/- 15% for VR2 models, recognizing that the implementers would have better information about the building once the VR2 report was completed.

Measure-by-measure savings were determined by successively altering the simulation inputs of each affected system, first to the efficiency code level, then to describe the existing equipment as it was found in the baseline period, with each simulation run. The Option D calibrated simulation savings claims were based on this back-cast methodology.

The CWBD implementers produced two post-installation reports:

1. Post-Installation Verification Report (VR1 Report). This report documents the installed measures, and their individual savings based on analysis of measures from an Option D simulation calibrated to baseline data.
2. Post-Monitoring Verification Report (VR2 Report). This report provides the savings of each installed measure based on analysis of measures from an Option D simulation calibrated to post-installation energy use data. Simulation calibration criteria were more stringent than for the VR1 Report. Savings were back-cast to baseline conditions in a specified order (load, systems, plant) prescribed in the program procedures manual. Results of the Post-Monitoring Period Verification Report were used by PG&E to claim savings according to existing CPUC rules in effect for custom measures.

To run Option C analysis, two pre-qualified third-party energy management and information system (EMIS) software vendors were engaged. The vendors were selected based in part on a modeling software testing method developed under a project funded by PG&E's Emerging Technologies Program (PG&E, 2013a). The test examined the accuracy of EMIS modeling software using a data set of electric interval-level usage data from approximately 70 commercial buildings.

The CWBD project team engaged with a consultant to run two public domain modeling algorithms (temperature and time of week (TTOW) and a mean week (MW) algorithm[3] along with the two separate EMIS software vendors to run two proprietary modeling algorithms. Each modeling algorithm was also tested for its accuracy in predicting energy use of the CWBD participant buildings. This enabled CWBD to test how accurately four different models could predict energy usage in diverse types of commercial sites and provide some experience with working with EMIS software vendors.

The public domain modeling algorithms were considered accurate for predicting energy use in the CWBD buildings if the CV(RMSE) and the NMBE met the criteria shown in Table 2. These criteria are taken from ASHRAE, 2014. CV(RMSE) and NMBE are determined as described above, the net determination bias error is determined from equation (5).

$$NDBE = \sum_{i=1}^{n}(M_i - S_i)/M \qquad (5)$$

Where: $M = \sum_{i=1}^{n} M_i$

[3] Both modeling algorithms are described in PG&E, 2013a.

Table 2: CWBD Option C Model Goodness of Fit Criteria

Data Interval	NMBE	CV(RMSE)	NDBE
Hourly or Daily	± 10%	± 30%	0.005%

Finally, an "ensemble" model, developed by PG&E that was based on the results of each of the four modeling algorithms presented above, was used to make the best Option C savings estimate.[4] The ensemble model result was used for program incentives to the customer, when incentives were based on Option C.

2.3. Data Requirements and File Specifications

A detailed "Data and File Specifications, and Workflow Requirements" document was produced for PG&E in 2015 to facilitate the CWBD's processes from initial customer recruitment through savings quantification. It described participant, energy usage, weather, and other file specifications, contents, and formats. It described key data and information file management responsibilities for PG&E, its vendors, and the implementers.

A secure file transfer server was used as the central repository for all data provided by PG&E as well as data and analysis files supplied by the software vendors, technical consultants, and implementers. The file specifications document described the key responsibilities of each party, what files to produce and store on the secure server, and directory structures and locations for file storage. The file specifications document was intended to facilitate automated downloading and uploading of data files throughout the duration of the CWBD.

3. Study Objectives

Given the developments in data technology, analysis methodologies, and legislative directives that have taken place since 2013, and considering the broad number of potential objectives this study could address, the objectives for the Joint Study were narrowed to the four listed below.

1. **Comparison of whole building program savings estimates between Option C and D approaches.**
   How do the savings estimates compare between the two methods and what are the major risks to whole building programs that each method addresses? What trade-offs are made when using one methodology over the other?

2. **Improvements to key program technical requirements and documentation.**
   What worked well and what needed improvement regarding the CWBD's technical and documentation requirements that are useful for future program designs and evaluation? What specific improvements are needed for successful implementation of Option C and Option D methods moving forward?

3. **Factors informing the design of future meter-based whole building performance programs.**

[4] Ensemble modeling "blends" the results from the four unique Option C models. The process often reduces errors and leads to a more accurate result. It is often used in machine learning approaches.

What policies and procedures are required from both a regulatory and a program administrative perspective? What data and information are needed to support evaluation of a program using IPMVP Option C to determine savings in the California regulatory context?

4. **Addressing the emerging technical and regulatory questions necessary to scale such programs and support evaluation.**
   What changes to the existing California evaluation frameworks are appropriate for whole building programs? How can NMEC-based programs be designed to facilitate evaluation? What data are critical for IOUs to collect that will facilitate timely evaluation?

4. Study Process Methodology

The Joint Study buildings include two medium-size office buildings, an office/lab building, a grocery store, and a public library. The buildings are located throughout the PG&E service territory in different climate zones. Each of the five buildings operate on regular schedules with typical space conditioning equipment, lighting, and plug loads, and had no production equipment during the periods of the study. Table 3 provides a summary of the study sites.

Table 3. CWBD Study Sites.

STUDY SITE NO.	BUILDING TYPE	GEOGRAPHICAL LOCATION	SQUARE FOOTAGE	ANNUAL KWH	ANNUAL THERM
A	Office/Lab	South Bay (Santa Clara County)	50,000	4,800,000	4,200
B	Library	SF Peninsula (San Mateo County)	20,200	348,000	11,00
C	Office	West Sacramento (Yolo County)	35,300	811,000	3,100
D	Retail Grocery	Central Valley (Fresno County)	28,100	1,900,000	43,400
E	Office/Lab	South Bay (Santa Clara County)	106,000	2,280,000	58,500

The four main activities of the Study Process are described below.

Activity 1. Project Documentation, Data and Energy Model Reviews. The study team received the study site documentation from the implementers, including the simulation files and data collected during on-site visits. The implementers' reports and energy models were extensively reviewed by the study team.

Activity 2. Meetings with CWBD Program Implementer (optional). Usually conducted in combination with Activity 3, and to become familiar with the specifics of each project, the study team met with CWBD implementers to review the reports and models, understand the data collected, and discuss individual site operations and equipment.

Activity 3. Site Visits and Customer Interviews. The study team met with facility personnel, implementers, and PG&E program managers during on-site visits to inspect the building and its equipment. These site visits usually began with meetings with facility owners and operators to discuss their understanding and experiences with the CWBD, and whether they found it a beneficial approach to improving the efficiency of their building operations.

Activity 4. Study Process Assessment. The study team reviewed and analyzed all the information collected to address the study questions described earlier in this report, assess lessons learned, and complete this report.

The Study Process procedures, as described in Appendix B, provided PG&E and CPUC Staff with first-hand site-specific customer feedback and information, both qualitative and quantitative in nature, to help shape future program design and evaluation considerations. Since future whole building programs will need to include savings calculation approaches and M&V methodologies in the program design to be successful, a parallel ex-post M&V analysis was conducted. This report's recommendations attempt to balance program design inputs, technical requirements, and regulatory requirements with the size of potential energy savings. The report considers the degree of effort required to produce these savings, their data collection requirements, M&V planning, and the rigor of analysis necessary to produce a scalable program.

5. Results/Findings

The Option C and Option D savings methods used at five out of the twelve CWBD sites were studied to inform the merits and drawbacks of each to help inform future program design. This section of the report first presents the findings from each analytical approach and then discusses their limitations in estimating whole building savings.

Procedurally, the study team received the VR1 Reports for each study site and conducted an in-depth review of the associated calibrated simulations and savings estimates. The study team identified several issues with the simulations and savings estimates. The list of issues was provided to each implementer so that they would be addressed in the VR2 Reports and simulations. The study team reviewed the VR2 Reports and simulations and worked with implementers to improve them to determine the final savings estimates.

The Option C estimates were provided by the technical consultant and the two EMIS software vendors to PG&E after completion of the performance phase for the study sites. The open source public domain models run by the technical consultant were reviewed by the study team. The following discussion in based on the final Option C and D savings estimates.

5.1. Comparison of Savings Resulting from Option C and Option D

Figure 1 shows the comparison of the electric savings for each of the five study sites as estimated from Option D and Option C analyses. Option C results are based on the open source TOWT modeling algorithm. The figure compares both the weather-normalized and the actual savings estimates for the two methodologies. Similarly, Figure 2 shows the comparison of the natural gas savings estimates for the study sites as estimated from Option D and Option C analyses.

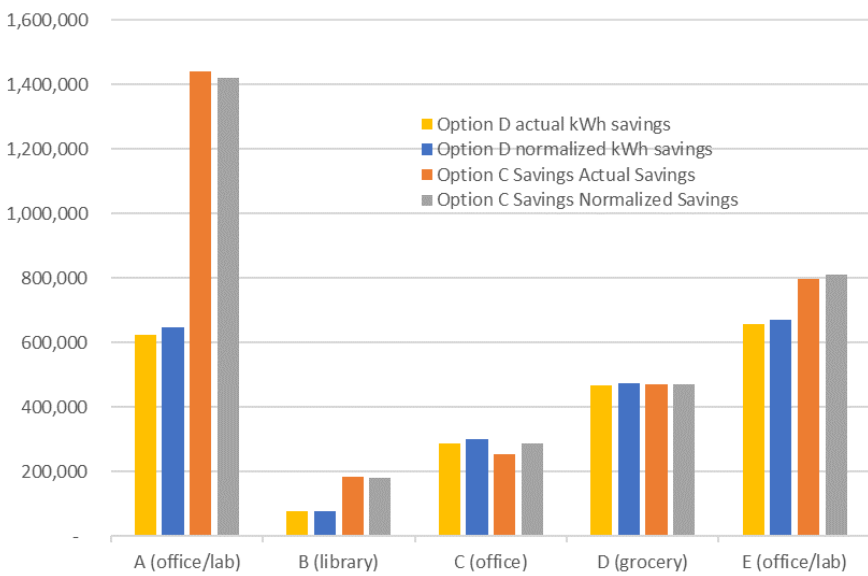

Figure 1: Comparison of Option C and Option D Electric Savings Estimate of 5 Study Sites

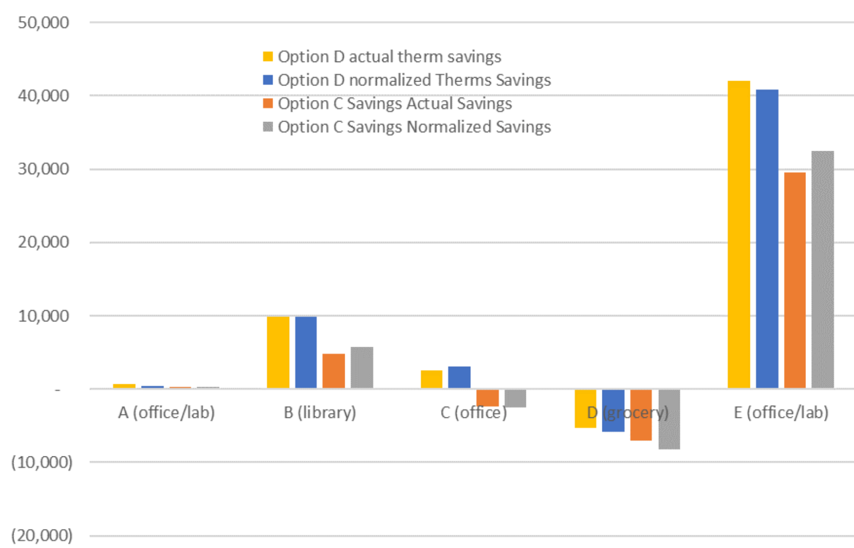

Figure 2: Comparison of Option C and Option D Natural Gas Savings Estimate of 5 Study Sites

From the two figures above, it is evident that overall the savings estimates for the five study sites didn't change significantly with the change of weather data from actual to normalized conditions. However, significant differences were observed between the Option C and Option D savings results. In three cases, the Option C electric savings estimates were higher than the Option D estimates. Overall, the Option D savings estimates were higher than the Option C estimates for gas savings. The details of the difference in results across the methods are enumerated below:

5.2. Electric Savings: Detailed Comparisons

The Option D simulations and Option C models provided kWh savings estimates that were comparable (within 5%) in two of the five buildings (Sites C and D) whereas there were large differences between the kWh savings estimates for the other three buildings. The key reasons driving these differences in savings estimates between the Option C and D estimates are listed below:

a. Two of the sites (A and E) experienced non-routine events (NREs) that were not taken into account in the Option C models whereas the Option D simulations were adjusted for them.

    - For site A, multiple NREs were identified that had occurred either during the performance period or in the construction period of the project. The first NRE was a replacement of an old rooftop package air conditioning unit with a higher capacity unit which wasn't part of the energy efficiency project. The second major NRE was the removal of 70 kW of lab equipment from the building during the construction period of the project. The final NRE for site A was an addition of a few electric car chargers to the building load.
    - For site E, the difference in savings as determined by Option C and Option D methods could be due to a temporary building shut down during the final two weeks of the year in the post installation period. The Option D simulation adjusted for this two-week shutdown whereas the Option C model didn't account for this in the analysis.

b. The reason(s) behind the discrepancy observed between the Option C and Option D savings estimate for site B is unknown. The Option D energy simulation calibrated well with the performance period billing data, however the baseline Option D simulation fell way short of meeting the calibration criteria. After reviewing both the models, the site data and the load profiles in detail, the study team suspects that the difference in savings estimates could be due to an NRE (i.e., load addition to the building during the performance period) that wasn't documented by the implementer during the site inspection. In this case, an Option D forecasting approach could have helped in identifying the unknown NRE in the building. In this approach, baseline model could have been created with baseline building characteristics and operating conditions and calibrated to the baseline period billing data. Then, measures should have been applied to the baseline model and finally calibrated the model to reporting period billing data. This could have detected the reporting period NRE which went undetected in the current back-casting approach.

c. The Option C analysts (PG&E technical consultant and EMIS vendors) did not conduct site visits. They were provided the energy use data by PG&E to run their M&V analysis but did not collect data on building operation schedules, special events, or NREs. Thus, they had no ability to adjust their models in response to these events.

d. Some NREs were identified during the study team site visits. At site A, the study team determined that the packaged HVAC unit was replaced, and the customer removed lab equipment and added electric vehicle chargers. With this information, the impact of the NREs could be modeled in the simulations.

For all five study sites, the Option C approach forecasts baseline conditions to the reporting period, while the Option D approach back-casts the reporting conditions (including the NREs) to the baseline period. When an NRE occurs during construction, the back-cast Option D approach will include its impact (because it is calibrated to post-install conditions) while the Option C approach will not and savings estimates will be different for the two approaches. Comparing the results helped identify that an

NRE occurred in cases where the team did not identify them previously. Figure 3 illustrates this comparison.

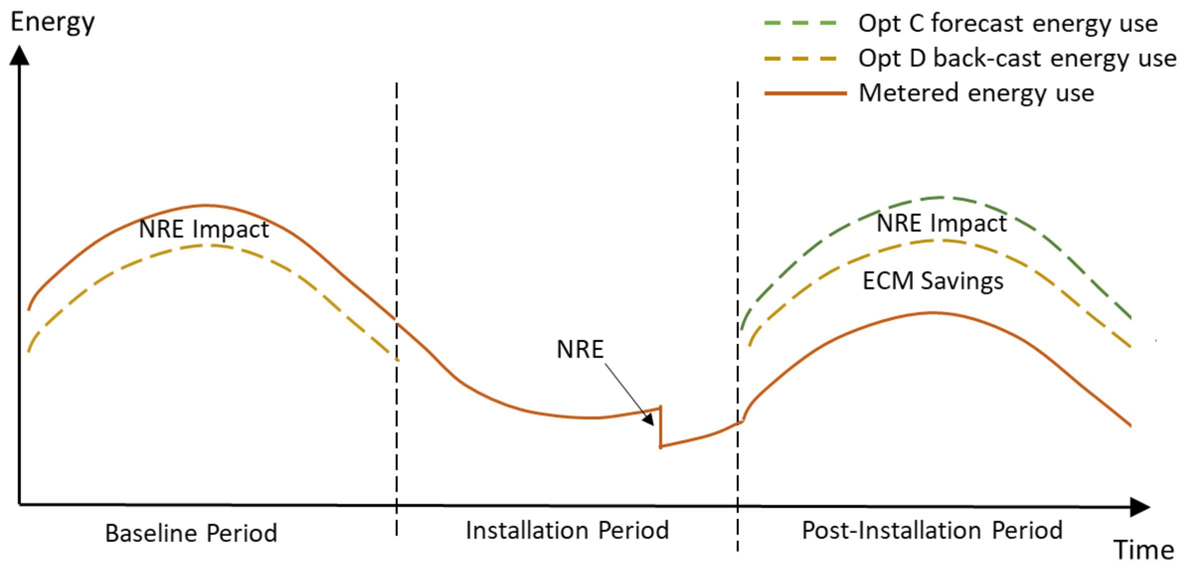

Figure 3. Illustration of NRE identification in CWBD Option C Forecast and Option D Back-cast Approaches.

5.3. Gas Savings: Detailed Comparisons

The Option C and D methods did not produce similar estimates of savings for any of the five study sites. In three of the five sites, the Option C modeling algorithms could not reproduce the baseline gas usage patterns, resulting in poor goodness-of-fit (GOF) metrics. NREs were suspected as the cause of the differences with Option D in the two of the sites. Some of the Option D simulations didn't calibrate well to gas usage. The equipment was not defined appropriately in the models in terms of specifications, schedules and operational performance due to a lack of site-specific data which made it difficult to finely calibrate the Option D models.

a. For site A, the Option C normalized gas savings estimates were lower than Option D normalized gas savings estimates. It appears that a boiler in the building was shut off during summer, and this event was not taken into account in the Option C analysis whereas the Option D simulation included the boiler shutdown. The Option D simulation calibrated with the performance period billing data with a MBE of 3% and CV-RMSE of 7% whereas the Option C model goodness of fit statistics were of 77% R^2 and 51% CV-RMSE. The Option D calibration metrics fell within the Joint Study's acceptable calibration criteria whereas the Option C metric didn't meet ASHRAE's goodness-of-fit criteria.

b. In site C, the Option D simulation estimated positive gas savings whereas the Option C model estimated negative gas savings although both the Option C and D methods used same meter data in the analyses. Figure 4 below shows the gas usage data for the baseline and post-installation periods. The baseline period (the data from which the Option C baseline model is developed) shows low usage throughout the year except for a few very high peaks in the spring and winter. The post-installation period (from which the VR2 Report Option D simulation was calibrated) shows a higher and more distributed gas usage pattern throughout the year. This increase in therm usage was expected from this site in the performance period because the personal electric room heaters were removed from the site in the performance period and space heating was provided by the furnace instead. However, neither the Option C model nor the Option D simulation were able to predict their respective usage patterns, and therefore had poor GOF or calibration metrics. The study team suspects that either an operational shift or random usage of the heating system or bad usage data prevented development of acceptable baseline models.

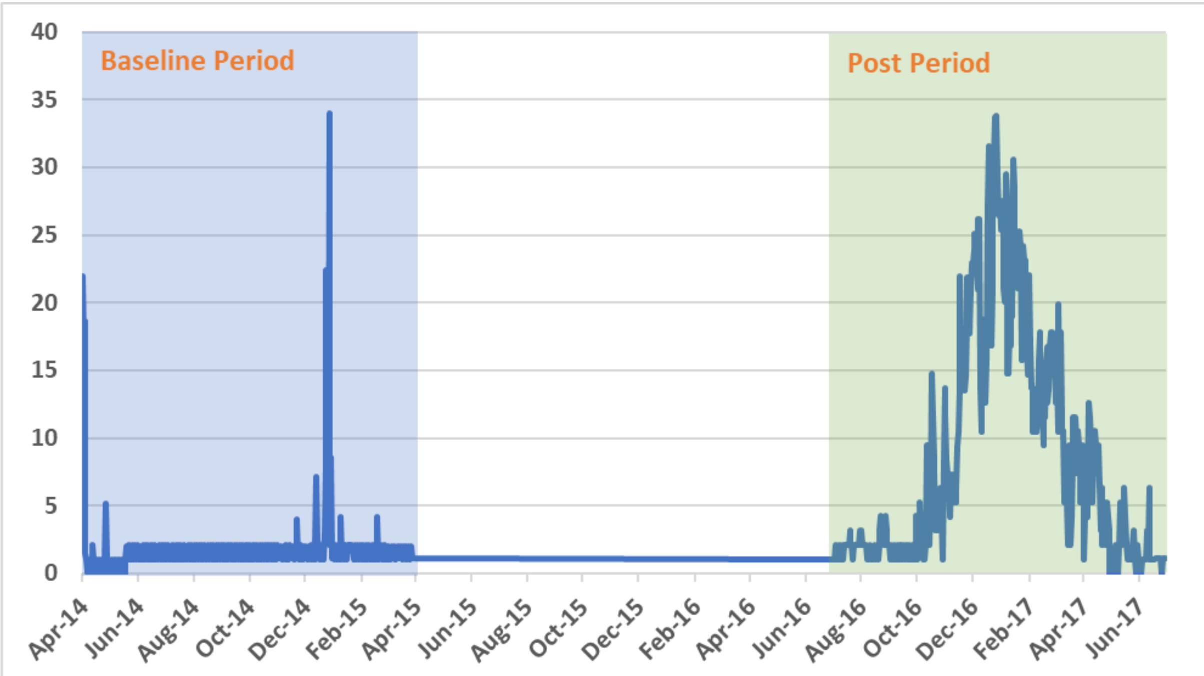

Figure 4. Site C Natural Gas Use.

c. For site D, both the Option C model and Option D simulation estimated negative gas savings because of an undocumented NRE: we postulated that the heating capacity of the building's air handling unit was increased in the post-installation phase to meet the building's space heating load. This was determined from the high unmet hours in the Option D simulation model in absence of site specific information. Even after adjusting the furnace capacity by using the auto sizing feature of the eQUEST modeling tool in absence of the actual size of the heating coil, the Option D model estimated higher negative savings because the Option D baseline model wasn't calibrated well to the baseline gas consumption even after adjusting the heating coil capacity. The study team suspects an additional undocumented NRE in the performance period (possibly gas load addition) could be the reason behind the increase in gas usage in the performance period for this site.

5.4. Calibrated Simulation (IPMVP Option D) Estimation Approach Findings

The team's technical consultants reviewed the eQUEST calibrated simulations for the five study sites and found significant problems. These included:

- Inadequate project data and documentation. In some cases, documentation such as as-built mechanical drawings, equipment specifications, cut sheets, and lighting plans were missing. Without a complete set of project documentation, it was difficult to understand the quality of the simulation and its resulting savings estimates.

- Simulations not built to the building end-use conditions. The team found that limited building characteristics data were used in the simulations. Primarily, the simulation software was built using many default settings, with only a few specific-site specific parameters used. Furthermore, HVAC systems and HVAC zoning in some of the simulations did not match the actual systems and zones in the buildings. Building end-use conditions have significant impact on the energy consumption of buildings which in turn affect savings estimates.
- Measures not well defined in the simulations. The measure parameters were not well defined in some cases and this impacted some of the measure savings estimates. For example, a parametric run simulated a chilled water pump variable-frequency drive (VFD) measure with three-way valves when in actuality a two-way valve was in use. Similarly, for a VFD fan measure, the zone minimum flow ratio was set too high. In some cases, the simulated lighting power density (LPD) was different than the actual LPD.
- Calibrated simulations not representative of building characteristics and operations. Although the simulations were calibrated to monthly usage data, the team's review revealed the simulations did not represent actual building characteristics, systems, and operations when compared to the hourly or daily consumption data. Therefore, calibrating incorrect simulations to the building usage data could have a substantial impact on the simulated savings results. In addition, the projects did not maintain a calibration log of adjustments performed for each simulation run. The lack of a calibration log made it difficult to track the input changes for calibration. This information would have proven useful for technical review.

Although the CWBD policy and procedure manual provided a step-by-step procedure for developing and calibrating simulations, the procedures did not mandate that building inputs for systems and operations be based on as-built drawings, equipment specifications, and monitored operational data. As a result, the simulations were not representative of the actual buildings or their operational characteristics. These issues illustrate the degree of difficulty in using Option D methods on existing buildings. Many best practices in calibrating simulations were identified for future program applications, including collection and use of building documentation, operational data on building end-uses, and procedures designed to assure that end-uses are properly represented in the calibrated simulations. Because the CWBD was a demonstration to test the validity of whole building approaches, the Joint Study team's technical consultants spent considerable effort in further developing each site's simulations with the goal that they fairly represent the actual buildings and their operations. As part of this task, the technical team reviewed the implementers' Option D simulations thoroughly, compared the simulation input parameters with the actual building parameters collected on-site, made adjustments to the simulations as required and finally attempted to calibrate the simulations to the post installation billing data. However, for some sites it was difficult to calibrate the simulations due to lack of availability of adequate on-site data. Therefore, it is critical for the implementers to diligently collect sufficient data during both the baseline and the performance period to support appropriate model development and calibration. It is extremely difficult for evaluators to go back to the sites and gather additional information regarding the building and the measure performance, especially as it relates to the baseline condition.

Once calibrated, the technical team tried to recreate the baseline building by first removing the measures from the as-built simulation wherever known NREs occurred to reflect the actual baseline conditions. Second, the baseline buildings were simulated with the base period weather data to determine the baseline energy usage of the sites. Finally, the baseline simulated energy consumption of the sites were compared with the baseline period billing usage to check how well the baseline simulation calibrated with the billing usage. At that point, if Option D simulations were found to be far off the calibration criteria, the technical team would adjust the baseline simulations further to bring them closer to the calibration range. These adjustments were based on the information collected from the implementers regarding the baseline condition of the buildings, and/or based on reviews of the end-use data and/or billing data from the baseline period. However, for some sites, calibrations were precluded due to unavailability of adequate data from the sites. Table 4 shows the final calibrations achieved for the five Option D simulations for both base and post periods.

Table 4: Base and Post Option D Calibration Results

	Option D				
		kWh		Therms	
PG&E Site ID	Site ID	NMBE	CV-RMSE	MBE	CV-RMSE
100024	A -Base	13%	14%	-7%	14%
	A -Post	5%	6%	3%	7%
100052	B -Base	30%	31%	-32%	54%
	B -Post	-2%	7%	45%	52%
100054	C -Base	1%	7%	-293%	491%
	C -Post	-1%	19%	68%	99%
100014	D -Base	13%	3%	-7%	14%
	D -Post	5%	6%	3%	7%
100044	E -Base	5%	7%	-4%	37%
	E -Post	-4%	10%	0%	32%

Comparing the joint study team's calibration targets with the calibration results of the study shows that the technical team was able to calibrate the simulations close to ASHRAE calibration criteria for most of the sites. The few sites that we were not able to calibrate lacked sufficient data to make informed calibration adjustments.

5.5. Whole Facility (IPMVP Option C) Estimation Approach Findings

The study team reviewed the two public domain algorithms used for the Option C analyses of the study sites. The software vendor's proprietary modeling algorithms were unavailable for review. One vendor worked with the study team to provide results as requested, while the other vendor stopped working with the team and removed the data after its contract expired.

Collectively, the technical consultant that ran the open source models and software vendors provided the required data and results of their Option C analysis as specified in the Data Specifications document, but not consistently in the specified file formats. The Data Specifications document did not require

Option C providers to produce savings estimates under long terms average weather[5] conditions, it required savings only under post-installation period conditions. The PG&E-developed ensemble model used each provider's adjusted baseline use predictions (baseline use under post-period conditions) combined with their GOF metrics and each participant's post-installation energy use, to determine the Option C savings for each participant.

The study team elected to focus the study on the Option C analysis using the two public domain models, and results from the one vendor who continued to support the study process. As the ensemble model was not documented and was based only on savings under post-installation conditions and not normalized conditions, it was dropped from the study.

As described in section 2.2, the two public domain algorithms were Lawrence Berkeley National Laboratory's temperature and time of week model (TTOW) and the mean week (MW) modeling algorithms. The TTOW model is accurate for regularly-scheduled, temperature-sensitive buildings, as it includes the effect of the time-of-week and the ambient temperature in its modeling algorithm. The MW algorithm is much simpler as it employs a simple average of energy use for each time of the week. Note that the modeler can choose the hour of week or day of week as the time of week variable in each modeling algorithm.

Each modeling algorithm may be used with the entire baseline data set to develop a baseline model, or the data sets may be grouped into distinct modes of operation, with individual models for each mode developed from the grouped data. Alternatively, an indicator variable may be used to distinguish the modes of operation, taking on a value of 1 or zero to identify each operation mode. This is a technique used to improve model goodness of fit and accuracy when there are distinct and identifiable modes of operation. Building operation modes can be occupied and unoccupied periods, equipment outage periods, or building closure periods, and so on. The open source and proprietary models were run using a black box approach, with no accounting of different operation periods. This is consistent with running the analysis remotely, with no information from the building to define different operation periods.[6]

The significant independent variable for the TTOW and proprietary algorithms was the ambient dry-bulb temperature. One year of prior energy use and temperature data were used to create baseline models. Building 'predictability' was indicated by how well the models reproduced their baseline period usage profiles (hourly use for electric and daily use for natural gas). Model accuracy was determined by minimizing their random and bias errors of prediction, as quantified with CV(RMSE) and NDBE[7] respectively. For the Joint Study sites, the buildings were considered sufficiently predictable when the CV(RMSE) and NDBE were below the threshold criteria of 25% and 0.005% respectively (ASHRAE, 2014).

PG&E had engaged another evaluation consultant to review the Option C models and Option D simulations, in an 'Early EM&V' process. Their scope included review of all CWBD participant's Option C analysis, and review of eight Option D savings analysis (one overlapping project – Site E – and the other

[5] CPUC reporting requirements stipulate long term average savings based on CTZ2010 data from the California Energy Commission for 16 standard Title 24 climate zones

[6] This is not to say that different modeling approaches cannot account for different operation periods, however open source TTOW and MW models were known not to account for them, while it remained unknown whether the proprietary model accounted for them in its modeling approach.

[7] CV(RMSE) is the coefficient of variation of the root mean squared error, NDBE is the net determination bias error. Both are defined in BPA (2012).

seven participants not in the joint study). The joint study and early EM&V study ran in parallel, and preliminary findings were shared. One preliminary finding was that there was no consistency in the prepared meter data for each site. As originally planned, PG&E's data teams uploaded electric and natural gas usage data each day on the secure file server (electric data in 15-minute intervals, gas usage in daily intervals). The uploaded files were compressed flat files of each participant's data by account number. The plan was for vendors and technical consultants to download data from the server, unpack and sort it for each participant. This was followed by the vendors only. The open source model technical consultant made separate data requests to PG&E to receive participant data. Each party collected their own weather data. Implementers obtained data from PG&E or through their participant's own access to data. In addition, each party conducting Option C analysis used their own techniques for inspecting the data, identifying gaps, outliers, and repeated values, as well as their own methods for treating these issues in preparation of the data for analysis. As a result, not all the baseline and post-installation energy use and weather data sets were consistent. The Early EM&V consultant collected the meter and weather data for each site from PG&E, inspected and treated the data issues described above, and prepared the data sets for analysis. The prepared data sets for the study sites were provided to the joint study team.

The study team reviewed the open source and the single vendor's proprietary baseline models goodness-of-fit metrics (NDBE, CV(RMSE), R^2), savings, and savings uncertainties (savings were estimated for post-period only, not under normalized weather conditions). No source code for the open source or proprietary algorithms were obtained for review, which prevented direct verification of their results (e.g. use the same data and algorithm to see if the same results are obtained).

The study team used the raw data prepared by the early EM&V team to re-run the TTOW and MW models for each study site. No accounting for building operation schedules or NREs were made. The electric models were developed from electric data converted to hourly intervals while the natural gas models were developed from the daily gas data.

Table 5 below shows the results of the analysis. The table shows the baseline goodness-of-fit metrics for the electric and natural gas baseline models, highlighting in red where the CV(RMSE) and NDBE metrics failed to meet the model acceptance criteria (CV(RMSE) < 25% and NDBE < 0.005%). For each model, the avoided energy use (savings under post period weather) and normalized savings (savings under TMY weather) are shown. The participating vendor's analysis results were also included, which did not include estimates of normalized savings.

Table 5. Comparison of Option C Model Estimates of Actual and Normalized Savings

		Electric						Natural Gas					
		Baseline Model						Baseline Model					
Site	Model	CV(RMSE)	NDBE	R^2	Savings (kWh)	Savings (% of Baseline)	Normalized Savings (kWh)	CV(RMSE)	NDBE	R^2	Savings (therms)	Savings (% of Baseline)	Normalized Savings (therms)
A	TTOW	4%	-	78%	1,440,717	30%	1,419,842	51.4%	-	77%	285	7%	221
	MW	7%	-	44%	1,429,574	30%	1,425,659	107.4%	-	1%	815	19%	834
	V1	5%	0.001	65%	1,436,084	30%		40.6%	-	72%	325	8%	
B	TTOW	19%	-	84%	183,233	53%	180,970	36.1%	-	74%	4,798	43%	5,737
	MW	23%	-	77%	187,063	54%	186,573	71.2%	-	0%	5,040	46%	5,029
	V1	22%	0.001	31%	177,038	51%		21.3%	-	83%	4,487	41%	
C	TTOW	15%	-	43%	252,723	31%	288,644	126.4%	-	11%	-2,385	-77%	-2,512
	MW	17%	-	21%	244,136	30%	280,703	133.6%	-	1%	-2,450	-79%	-2,818
	V1	11%	0.001	3%	292,811	36%		87.1%	0.003	0%	-2,206	-71%	
D	TTOW	6%	-	93%	472,412	25%	469,370	10.3%	-	83%	-7,024	-16%	-8,245
	MW	12%	-	71%	507,457	27%	507,075	25.4%	-	0%	-8,695	-20%	-8,695
	V1	6%	(0.001)	51%	485,973	26%		8.6%	-	73%	-7,399	-17%	
E	TTOW	12%	-	89%	792,751	35%	810,356	15.7%	-	82%	29,501	50%	32,441
	MW	17%	-	75%	808,480	35%	808,690	27.5%	-	44%	29,860	51%	32,351
	V1	-	-	-	-	-	-	14.4%	-	23%	35,795	61%	

Results of the analysis are summarized separately for electric and natural gas below.

Option C Electric Results

- The open source electric baseline models each met the goodness-of-fit criteria for CV(RMSE) and NDBE, while the vendor's model did not meet the bias error metric.
- The electric savings are consistent for each study site, with similar savings results from each model. However, since the Option C analysis did not take into account the impact of NREs, the savings for sites with electric NREs (A and E) are biased.
- Estimations of normalized electric savings were not significantly different and were not consistently higher or lower than savings estimated for the post-installation period.

Option C Natural Gas Results

- The vendor's model met the CV(RMSE) criteria in three of the five cases, the TTOW model in two of the five cases, and the MW model in none of the cases.
- Two sites (A and C) had low annual baseline usage (under 5,000 therms).
- Two sites (C and D) showed negative savings.
    - Figure 4 shows low baseline usage for Site C which may have been a problem with the data or an unusual building operation.
    - Figure 5 shows that gas use for Site D clearly increased in the post-installation period.
- The gas savings were less consistent for each study site than for the electric savings, with estimates based on the MW model more different than estimates from the TTOW and vendor model. Savings estimates did not account for NREs in the gas models.
- Estimates of normalized gas savings were less significantly different and were more consistently higher than gas savings estimated for the post-installation period.

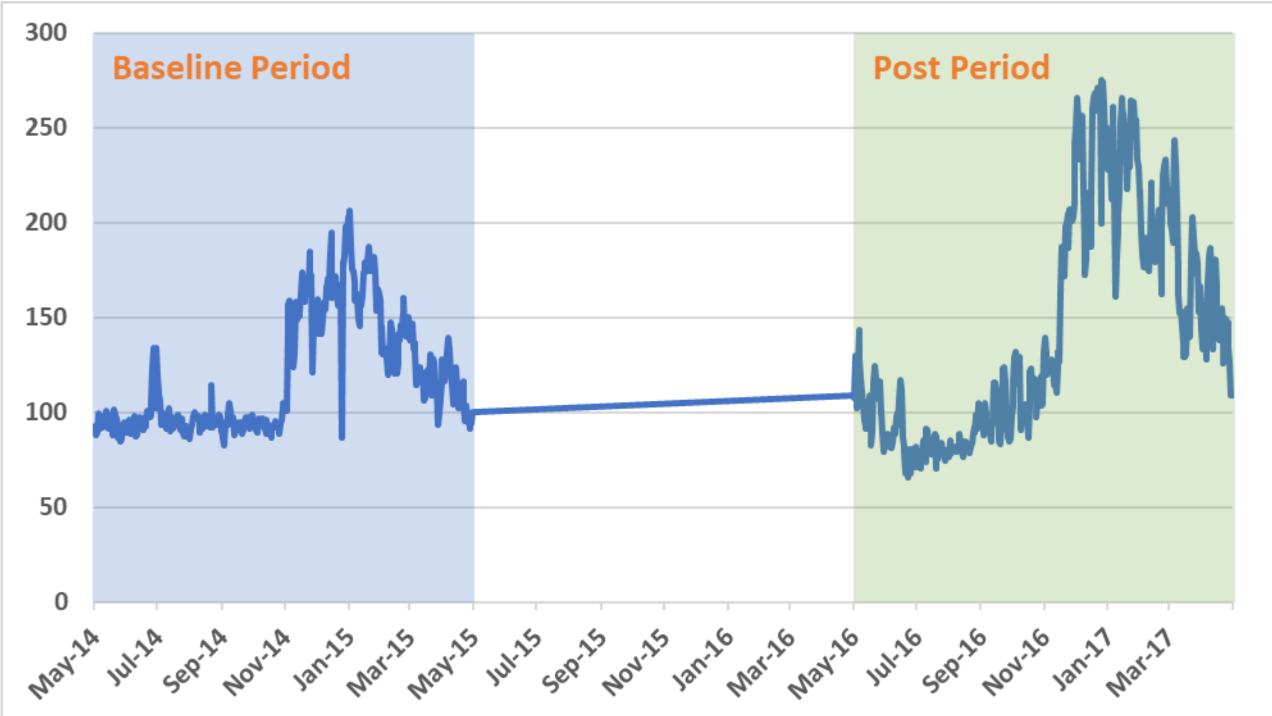

Figure 5. Site D Baseline and Post-Installation Gas Usage.

Overall, the Option C natural gas savings analysis was disappointing. Factors that led to poor results included difficulty in modeling natural gas use in most cases, and the presence of NREs and low savings in some cases. Except for accounting for the impacts of NREs in the electric savings for some sites, overall the Option C analysis for electric savings worked well. The study team noted that some of the impacts of the issues encountered may be lessened with

- improved data management practices,
- pre-screening to assure the energy use is predictable within goodness of fit criteria, and
- assuring Option C analysts are more informed about the buildings to understand operations and to identify and quantify impacts of NREs

5.6. Uncertainty in Option C and Option D Models

A rigorous comparison between Option C and Option D models should include an assessment of the uncertainty of the model predictions. Differences in savings estimates that fall outside the range of the confidence intervals of both models are statistically significant, while differences that are within confidence intervals are not. The Joint Study team however did not attempt to estimate uncertainty due to several technical issues with the savings uncertainty process, as described below.

ASHRAE Guideline 14-2014 contains a method for calculating the uncertainty of Option C models. The analysis is based on the number of data points in the pre and post period, the model goodness of fit metrics, and empirical constants. The document covers weather independent models, weather-based regression models with uncorrelated residuals (such as monthly models), and weather-based regression models with serially correlated residuals (such as daily and hourly models). Since the Option C models used in this study were all based on daily or hourly data, the issue of serial correlation is an important consideration in the calculation of model uncertainty.

Serial correlation is accounted for as a reduction in the number of independent data points in the uncertainty calculation. The autocorrelation equation assumes the current value is influenced only by the previous value (lag -1 autocorreleation), where the hourly energy consumption most commercial buildings is influenced by the temperature history over several prior periods. The ASHRAE formula also includes an empirical constant with embedded assumptions that are not followed in practice by most models. These issues tend to underestimate the model uncertainty (BPA, 2017). Given the known issues with the ASHRAE Guideline 14 uncertainty method, the Joint Study team decided not to calculate the fractional savings uncertainty in the Option C models[8].

The uncertainty estimation process in Option D models is much more complex. ASHRAE Guideline 14-2014 provides model calibration criteria for monthly, daily and hourly models, but unlike Option C models, the goodness of fit of the Option D model to the actual consumption data is not directly related to the uncertainty in the estimated savings. Note for most Option D models, it is possible to calibrate the model to observed data using a variety of model input combinations, making it difficult to identify which set of model inputs correctly represents the building.

Option D model uncertainty can be calculated by conducting a sensitivity analysis on model inputs to identify inputs most influential on the savings calculations. The uncertainty in the most influential model inputs is established from the measurement errors of directly measured inputs or estimated based on a subjective assessment of the reliability of secondary data sources. These uncertainties can be propagated through the model using a numerical method to vary the inputs according to their respective uncertainties and observe the variations in the savings results. The lack of a unique solution to model calibration further complicates this process, requiring running the uncertainty analysis across several possible sets of inputs yielding a calibrated model, and aggregating the results across multiple calibration solutions. Due to the complexities involved, a fractional savings uncertainty analysis of the Option D models was not attempted.[9]

6. Discussion and Recommendations

In this section, we address the research questions based on our findings and experience with the CWBD through the Joint Study process.

6.1. Comparison of Option C and D Approaches to Whole Building Program Design

Option C and Option D represent fundamentally different approaches to estimating savings resulting from energy efficiency interventions. An extensive description of each approach is provided in Appendix A, however a simple way to understand the difference between the two approaches is presented below.

---

[8] ASHRAE approved a committee in 2017 to revise Guideline 14-2014. Uncertainty in hourly billing models is currently under discussion.

[9] See Reddy, et al. (2007 Part 1) for a more complete description of FSU calculations for Option D models. Reddy, et al. (2007 Part 2) provides an example of an Option D FSU calculation of +/- 30% for a calibrated office building model.

Option C is a statistical approach that forecasts future energy use based on historical energy use without taking into consideration the equipment installed at the premise. At the end of a monitoring period, adjustments are made to control for observable differences in variables known to influence energy use (such as weather) then the actual energy use is compared to the forecast, and the difference between the forecasted energy use (without the measures installed) and the observed energy use (with the measures installed) is the energy saved by the intervention.

Option D is an engineering approach that uses software to build a simulation of energy use for the premise using a number of inputs that describe installed equipment, weather, and occupancy patterns. Once the simulation has been calibrated—validated to ensure that it represents the energy use and load shapes[10] of the premise, the next step is to build upon the model to reflect the savings estimated for each of the efficient measures installed. In CWBD, the final savings values were determined using a back-cast procedure. The resulting building simulation with all measures included should estimate the baseline use, if all differences (including adjustments for observable differences known to influence energy use such as weather) have been properly accounted. Figure 3 illustrated the differences in the Option C forecast and Option D back-cast method.

Both of these estimation options have their own advantages and disadvantages depending on the particular application. Some of these are discussed below.

- The Option C approach is generally lower cost because it bases its estimate of savings on the totality of energy savings by observing the changes in energy use before and after the measures are installed. This meter-based approach works best for projects where the savings are expected to be large as compared to the total energy consumed at the premise. A key challenge for Option C is when significant NREs occur at a facility during the reporting period (IPMVP Core Concepts, 2016, pg. 29), adding cost and uncertainty to the savings estimate.
- The Option D approach is more labor intensive—and therefore costlier—given its use of simulation software that requires considerable training to use, level of effort and time to collect input data and information, and iterative runs to reach calibration and estimate the savings for each measure. A key advantage is its ability to isolate the savings estimates for individual measures and estimate the interactive effects between measures.

Shortcomings of the Option C Approach: A Deeper Discussion

<u>Option C models have a risk that results do not reflect the actual savings achieved by the project</u>. The Option C models in the demo all provided relatively close results to each other for kWh savings, with baseline models that met the program goodness of fit criteria. However, due to the omission of needed baseline adjustments due to NREs, the Option C models converged on the wrong answers for buildings where NREs occurred. Omission of baseline adjustments for NREs is a major risk to Option C models generally, and the CWBD was not an exception. Program implementers must be diligent in their identification of NREs not only during the baseline period but also in the performance period and in the intervention (measure installation) period of the project and modelers must take care to estimate their impact with sufficient rigor.

---

[10] Load shape verification requires calibration to hourly data, which was not a requirement of the CWB pilot program.

Models that don't meet GOF criteria. Although the buildings in the pilot study were selected based in part on the expectation that their operations would be stable and their energy use would be predictable, only two of the five sites had gas models that meet the baseline model goodness of fit criteria. Assessing the GOF criteria of the baseline models using historical data prior to accepting a building into an Option C program could potentially mitigate this risk.

Projects that don't generate savings "above the noise." No matter how advanced a model may be, sometimes model fit will be poor and unable to estimate savings within the requisite uncertainty. Program designers should have a backup plan that is explicitly outlined in contract terms for calculating incentives—and when it will be used—when Option C methods fail.

Shortcomings of the Option D Approach: A Deeper Discussion

Ignoring or mischaracterizing behavioral (and operational) issues in whole building simulations. Building simulations attempt to characterize occupant behavior through a series of fixed and scheduled inputs. It is generally not feasible to capture the full range of occupant behavior in the simulation inputs; the simulations attempt to characterize average behavior. Initial inputs into a simulation are generally developed from on-site surveys and occupant interviews, and occupant recall of building operations and events affecting energy consumption may be incomplete. If simulation calibration is done using interval rather than monthly consumption data, it may be possible to make some schedule adjustments, but assigning the adjustment to the correct equipment or end use may not be possible in the absence of additional end-use metering which can add significant time and cost to a project.

Simulation errors due to inputs. Accurate estimates of energy savings depend on accurate development of simulation inputs, especially for inputs related to the installed measures. Data gathering generally occurs during an on-site survey, where building plans are examined, and the physical characteristics of the building are observed. This "snap shot" of the building may not represent the relevant state of the building, especially if the on-site survey is done during project development, but the savings are back-cast from a post-construction calibrated simulation, as was done in the CWBD. It may not be possible to locate plans for existing buildings, plans may not reflect the current state of buildings, and it may not be possible to physically inspect aspects of the building design due to lack of time or access. In some cases, it may be difficult to capture the baseline conditions of building in the post construction period as the building may have changed substantially from the baseline period. Therefore, some of the simulation inputs may be estimated based on the simulator's experience or based on software default values that may not represent the actual building. There may be errors in the development of the model due to inaccurate takeoffs of building characteristics from the plans and site survey, lack of rigor on the part of the modeler, or inaccurate or inadequate documentation of measure performance characteristics.

One of the main difficulties in CWBD was establishing the performance of existing equipment that was replaced by the program. Obtaining performance specifications for legacy equipment and estimating performance degradation due to equipment aging were problematic. Depending on the timing of the onsite survey, it was not always possible to physically observe the existing equipment or obtain performance specifications.

Simulation errors due to simulation engine limitations. Building energy simulations have limitations and are not designed to estimate the savings of all possible energy efficiency measures. In some cases, the simulator may choose to employ a workaround or "trick" when estimating savings for a measure not explicitly addressed by the simulation engine. The results from a workaround may be suspect and highly dependent on the simulator's ability to represent adequately the measure performance within the limitations of the engine. The hybrid evaporative condenser measure in site D is an example of a measure not explicitly simulated in eQUEST.

Simulator's Experience. Another issue in Option D simulations that should be taken into consideration is simulator's experience in creating and calibrating the energy simulations. In some cases, the building simulations were not built appropriately due to the limited experience of simulators. Since a well calibrated simulation would depend on the simulator's experience with the building simulation engine, a requirement that simulators have more than a basic level of building simulation experience and understanding of building mechanical systems could minimize some these issues.

Simulation Calibration Cost. The process of creating a building simulation is not an easy task, even for a building that doesn't require calibration. Calibrated simulations are far more complicated, time consuming and resource intensive to build than uncalibrated ones.

Tradeoffs Between Option C and Option D Methodologies

Data requirements. Data requirements for Option D models are much more substantial than for Option C models. Under Option D, the building physical characteristics and operations must be adequately characterized into large set of simulation inputs, and these data need to be verified during an onsite survey. Note that a certain portion of the building characteristics data may be gathered as a necessary component of project development, especially when simulation is used to develop an upfront estimate of potential project savings. Using the same software package for both project development and performance estimation will reduce simulation development and data gathering effort.

Under Option C, data requirements are limited to the cleaned billing or AMI data, and independent variables such as weather and building schedules. Onsite inspections of the facility are generally not required to obtain these data. Onsite inspections of building measure installation may occur to monitor contractor installation quality, and these inspections can be used to verify meter numbers and other independent data as necessary. Additional data gathering may be needed to make baseline adjustments or characterize non-routine events in the performance period, although the data requirements generally are less intensive than those required for a full Option D analysis.

Non-program-related changes to buildings that confound the analysis. Adding or removing loads, and changes to the building occupancy, business operating hours or HVAC system operating strategies that are not related to the program, will be incorporated into the savings estimate unless they are explicitly removed through model specification or baseline adjustment(s). Continuing building upgrades, either self-funded or through participation in other efficiency programs, also will be incorporated in the performance period consumption, unless they are removed through model specification or adjustment(s). By design, Option C analysis does not attempt to identify and associate energy savings with specific building interventions, while Option D analysis should tie these adjustments explicitly to them.

Uncertainty. ASHRAE Guideline 14 provides a method for estimating the savings uncertainty (known as fractional savings uncertainty) in Option C analysis. It is applicable to linear ordinary least squares regression models and incorporates a correction factor for autocorrelated data. Generally, the method works well for models based on monthly and daily time interval data, but recent work (PG&E, 2018) has shown that it is unreliable for models based on hourly data.

The fractional savings uncertainty for Option C models not requiring engineering adjustments can be estimated from model statistics, while the savings uncertainty from Option D simulations is generally unknown. Properly calibrated Option D simulations are constrained by physics and generally are accepted as reasonable engineering estimates of savings, but the true uncertainty generally cannot be assessed without extensive analysis. For models with poor goodness of fit statistics and/or small fractional savings, the uncertainty from Option C models may be too high to provide a reliable estimate, requiring substitution of an alternative estimate that may involve Option D or other analysis. This pilot had the luxury of both sets of estimates and was able to substitute the Option D savings when the Option C analysis did not work out.

<u>NRE adjustments</u>. Both Option C and Option D analysis require active engagement with the customer to identify and quantify NREs and estimate the necessary adjustments. Both methods require that NREs be identified and their impacts quantified. Engineering calculations based on site-collected data, as well as the use of indicator variables within the regression equation, are methods that may be used when whole building savings are quantified with Option C. When savings are quantified with calibrated simulations, the simulation may be used to determine their impacts. In some cases, NREs may be simulated as independent parametric runs, similarly to how individual measure savings are simulated.

Back-casting the Option D results from a well-specified and post-calibrated model will eliminate the need for baseline adjustments when NREs occur in the installation period since the model is based on performance period operations and the savings are due only to changes in equipment efficiency from baseline to as-installed. Calibrating the back-cast baseline model to actual baseline consumption may reveal modeling errors or identify potential NREs. This approach was taken during the Option D model calibrations in the demo and revealed a combination of modeling errors and differences between the baseline and as-constructed projects that were previously undetected.

6.2. Recommendations to Improve Program Technical Requirements and Documentation

As a first-of-its-kind efficiency program, the CWBD was ambitious. It was developed to test many aspects of achieving comprehensive deep savings in participating buildings. The demo employed unique entities to perform different program-related activities. It included both Option C and Option D methodologies to estimate whole building savings and used the Option D estimates to claim savings as required under then-current regulatory rules. The demo developed infrastructure to provide data to software vendors, technical consultants, and implementers. The CWBD began in 2013 and its final participant's incentive was paid in 2018.

From inception to final payment, the joint study identified several elements of the CWBD that worked well, and several that could be improved for better program efficiency, greater rigor and empirical support in its savings estimates, and to mitigate risks encountered during implementation. The following are key elements that worked well, and areas that could have been improved.

CWBD Elements – What Worked Well and What Could Have Been Improved?

Option C analysis by remote parties. Option C analyses were performed by parties that were not familiar with ongoing operations of the treated buildings. While the electric energy models performed well based on goodness-of-fit metrics, the models did not account for building operation schedules that could have improved the models, or for NREs that created biased estimates. In the case of NREs, the proprietary software vendors had anomaly detection algorithms that implementers (or PG&E) could have used to flag NREs in a timely manner, thereby enabling an investigation of and proper adjustment for the NRE. As it turned out, none of the Option C analyses accounted for NREs although three of the five study buildings had them. Moving forward, whole building programs will benefit from designs that ensure coordination between administrators, implementers, and facility managers to identify, document, and account for NREs in the models. Leveraging anomaly detection algorithms to flag potential NREs and then investigating them would have resulted in better models and better estimates of savings in the demo.

Comprehensive whole building programs require an onsite survey or energy audit during project development to identify energy savings opportunities. To leverage them fully, analysis of baseline consumption data and identification of potential NREs should be done prior to conducting onsite surveys. Then onsite surveys could be used both to identify potential measures and to understand the drivers of unexpected variations in facility energy use so that implementers can determine whether they are caused by NREs, and if so, to make non-routine adjustments in the models.

Building pre-screening. Poor model goodness-of-fit metrics were obtained for natural gas models for four of the five study sites. In one site, the problem may have been a problem with the consumption data, but in the other three sites, the gas usage was not 'predictable' within the CWBD's criteria by any of the modeling algorithms employed. For these sites, Option D analysis was used to determine participant incentives. Building pre-screening, whereby one year of historical meter and weather data would inform which buildings and energy sources are good candidates for Option C analysis, should be required for improved modeling outcomes. In the cases where building pre-screening suggests poor model fit—or when poor model goodness-of-fit metrics are obtained after the fact—it is critical that alternate strategies for verifying savings have been identified (such as Option D analysis).

Use of performance period monitoring for NRE detection. Option C analysts should periodically download and analyze whole building meter data throughout the performance period to track whether the expected savings are materializing and assess whether NREs have occurred during the performance period. Regular performance period monitoring will help with the early identification and resolution of NREs and increase the likelihood of favorable project outcomes. Although the CWBD implementer contracts included tasks for ongoing performance period monitoring, these tasks were not fully implemented, so this benefit was not provided or explored.

Savings claims need to be normalized to long term average weather data. Before savings can be normalized, performance period models first must be constructed, and then weather-related adjustments can be applied. Performance period models should be used to detect performance period NREs and identify erosion in energy savings due to problems with measure performance, other building equipment or operations.

Level of detail “sweet spot” for Option D simulations. Several of the Option D simulations created for this pilot utilized the eQUEST “Wizard” input mode, which relied heavily on eQUEST default values which created simplified simulations that did not follow the HVAC zoning of the actual buildings. Simulation inputs relating to specific HVAC measures installed in the buildings were hard to identify. While it is generally time consuming to create fully geometric representation of buildings with all HVAC zones included, a “sweet spot” must be found that allows for an accurate representation of building characteristics, operations and loads presented to the HVAC systems affected by the measures but does not take an inordinate amount of time to create. Simulation model guidelines for existing buildings that emphasize a well specified model for the systems affected by the measures, while providing a simplified representation of the rest of the building should be developed. These guidelines should make it easier to generate the initial Option D simulations and will simplify review.

Improve Option D calibration requirements. The CWBD procedures manual required VR1 (post-installation, calibrated to baseline data) and VR2 (post-monitoring, calibrated to post-installation energy use) verification report simulations to be calibrated with monthly data only. Building simulations can be improved through calibration to hourly data to true up daily building operating schedules, identify daytypes, and unanticipated building shutdowns. Simulations can be further improved through calibration at the subsystem level using control system trend data, so that energy use in building subsystems affected by the measures are accurately represented. The CWBD procedures manual included procedures for adjusting simulation inputs to determine individual measure savings, however it did not fully describe the requirement to check the back-casted baseline energy use with metered baseline data. This comparison can also be used to potentially identify non-routine events during the baseline period.

Include all required analytical approaches at program launch. The regulatory requirement for implementers to conduct Option D analysis came after program launch when many participants had begun installation of measures. The program administrator reviewed the different options for meeting regulatory requirements and determined to use a back-cast Option D approach, as the opportunity for documenting existing conditions baseline equipment and operations was lost. Future whole building deep savings programs should be designed with all the regulatory requirements identified and select the best methods accordingly. In addition to considering an Option D forecasting approach, retrofit isolation Options A and B should be considered.

Data requirements and file specifications. The “Data and File Specifications, and Workflow Requirements” document contained critical information that laid the foundation for the CWBD. Although well designed, there was no consistent stewardship of the document throughout the study process; this represents a major shortcoming of the management of the CWBD. Responsibility for verifying that the data requirements and file specifications were being followed, and for updating the document, were not clearly communicated to the succession of project managers taking responsibility for the CWBD. Updates to reflect key additions, changes and clarifications to data requirements and file specifications were not distributed to the parties participating in the CWBD consistently. In part this critical problem could have been solved by the addition of file sharing: establishing a central document server to share and update information between the organizations on the team may have addressed the lack of stewardship of the document—the essence of the project given its data intensive nature. As a result, the document could not be relied on as a source of information for the demo. Among the critical missing information was the source of weather data, mapping of premises to weather data sources, and processes for addressing errors in raw data files (including missing data and cleaning anomalous energy data such as zeros and spikes). Tracking down the missing information (and verifying that different parties were using the same information sources) prolonged the data analysis phase significantly.

Substantiating Influence.[11] Participation in the CWBD was by invitation only. Among other qualification criteria (e.g., building size and end use, model fit of past energy use, no participation in custom or deemed programs in prior two years), customers had to complete a "Customer Attribution Questionnaire." The five questions on the form were open-ended and asked about the project scope and the problems the customer sought to address by participating. The form requested organization contacts as well. The completed questionnaires were scanned and stored in project files on PG&E servers. Contrary to best practice, standard net-to-gross instruments were not administered. The questions were not carefully designed: they did not make use of pre-tested, multiple-item scales that would have captured the complexity of organizational decision-making. Moving forward, we recommend that a comprehensive questionnaire be used to substantiate the influence of the program on the installation of equipment and changes to behavior and retro-commissioning and operation of existing equipment that comprise a whole building intervention.

Specific Improvements for Successful Implementation of Option C and D Approaches

The CWBD was a first of its kind program to achieve comprehensive and deep savings in whole building projects using both Option C and Option D methods in the commercial sector. Through review of its program procedures and results from selected participant sites, the joint study team identified specific technical issues that should be addressed to improve the likelihood of success of future whole building programs. These issues are described below.

Development and standardization of NRE identification and impact quantification. The study team identified NREs in electric energy use in three study sites and NREs in gas energy use in two study sites. While five is a small sample and inconclusive of the general population, it illuminates the risk that NREs pose to whole building efficiency programs: NREs may occur any time (from the baseline period through the performance period), may vary in magnitude from insignificant to large, and the energy use characteristics may be simple or complicated. Given their uncertain characteristics, multiple methods may be used to identify and quantify NREs, including through analysis of AMI data, through adjustments in whole building simulations, and through modifications to engineering calculations based on building data collected separately. Some work has been done on detecting and adjusting for NRE impacts (LBNL, 2018) and on providing guidance for identifying and quantifying their impacts (SCE, 2018) when using Option C approaches. More development work is needed in this area.

Note, when estimating impacts of NREs, an engineering adjustment to a project baseline of unknown uncertainty will cast doubt on the overall precision of an Option C savings estimate. Non-routine baseline adjustments should be made according to engineering best practices to minimize bias and uncertainty in the adjusted savings. If possible, an uncertainty analysis of the engineering estimate should be incorporated into the overall model uncertainty. Methods should be developed for estimating uncertainty of NREs that use engineering calculations.

[11] Although customer survey data about program influence were collected, a net savings analysis was not in scope of this program and was not conducted.

Improve methods for estimating savings uncertainty for Option C methods. While some valuable work has been published in this area (ASHRAE, 2014), its limitations have become apparent as the use of short-time interval data for energy savings estimates has grown. The principal limitations include use of non-linear empirical modeling in Option C analysis where the established methodologies do not apply (as they are based on linear regression techniques), as well as the effect of autocorrelated data, which has been shown to render uncertainty formulations meaningless (PG&E – LBNL, 2018 and BPA, 2017). Further work is needed to develop an industry-accessible methodology to estimate savings uncertainty using advanced data models and hourly interval data.

Improve methods for estimating savings uncertainty for Option D methods. The estimation of saving uncertainty in Option D models has received less attention than Option C methods. As many engineering methods, the uncertainty in the results is often based on rules-of-thumb generally applied to engineering analysis. Research (Reddy et al. 2007) has brought some statistical rigor to the estimation of savings uncertainty, but more work is needed. A recent addition to the OpenStudio software interface to EnergyPlus includes a savings uncertainty module, but the process is computationally intensive and relies on greater acceptance of OpenStudio and EnergyPlus in the simulation community.

Industry standard calibrated simulation procedures for existing buildings. The CWBD's procedures for calibrating simulations fell short of simulating the study site buildings accurately due to the reasons discussed previously: use of eQUEST "Wizard" input mode, inputs not being based on actual building data and information, calibration with whole building monthly data only, and a lack of data for calibration at the building subsystem level. A search of building energy simulation industry organizations did not reveal an industry-consensus procedure for calibrating simulations of existing buildings. Prior work (ASHRAE RP 1051-2002) was referenced in the CWBD procedures for the order of simulating energy efficiency measures. An industry-standard procedures document for Option D methods would greatly benefit the industry.

Accounting for EVs and onsite generation. Electric vehicles (EV) and onsite generation of electricity by photovoltaic (PV) systems are increasingly popular and will need to be considered as these programs are scaled. Although the demo sites were hand-selected for ease of modeling and predictable energy consumption, EVs and PVs were present in two of the five buildings. EV charging stations were added to Site A after the project was completed, requiring a non-routine baseline adjustment. A grid-tied PV system included in Site D complicated the Option C model development and calibration of the Option D model. A method for obtaining PV system production data or incident solar radiation data at the frequency required for the model will need to be developed as part of the measurement and verification (M&V) plan. Similarly, a method for obtaining EV charging consumption data may be needed to make an adjustment to an Option C model or to calibrate an Option D model.

6.3. Design of Future Whole Building Performance Programs

From the joint study team's perspective, there are several elements needed to run a meter-based, comprehensive whole building energy efficiency performance program successfully. The study team focused on the technical elements of the CWBD program, but recognized that regulatory and program requirements, as well as other elements, must be considered. As described above further development work is needed so that California will have the tools in place to administer and evaluate these types of programs. Some of these elements that were identified over the course of running the demo are discussed below.

Program Management

The demo began at a time when most deemed and custom projects used savings values pulled either from the Database for Energy Efficiency Resources ("DEER") or from engineering calculations. Relatively few working in the California IOU-sponsored commercial program market had experience building Option C-based statistical models of energy use, and rulesets governing their use were nonexistent.[12] The greenfield subject matter made it difficult for program managers without a statistics background to navigate. Steep learning curves, limited knowledge management tools, and other organizational factors led to multiple changes in the demo's program managers with poor handoffs between them. The result was inadequate contractor management, contracts with software vendors lapsing without verifying receipt of completed deliverables, timelines slipping, and, in certain cases, the inability to complete some of the tasks initially scoped.

Experience with the CWBD underscored the critical role of program management in the potential success of an Option C-based program. Several recommendations resulting from the CWBD are discussed below.

- Program planning phase. Since there can be no guarantee that any one individual will remain in a role over the course of long engagements required by Option C-based projects, thoughtful planning of process flows and planning for formal project management handoffs should be part of the planning and initiation phases. A matrixed approach to project management should be considered whereby representatives from multiple organizations having roles managing specific program functions are involved in program planning. For the demo, these key functions were: customer targeting and outreach, field engineering, project implementation, data science, project technical review, evaluation, policy, regulatory reporting, and regulatory oversight.
- Planning outputs. Formal project planning tools and ongoing revisions to their outputs are essential for communication, recordkeeping, management and evaluability. The planning documents should be accessible to team members on a shared drive. Planning documents should be updated by the program manager or a delegate as processes change throughout the administration of the program. They include:
    - A cross-functional program process flow chart indicating the actions, decision, and responsibilities for each organizational team involved in the program.
    - A corresponding RACI chart[13] that identifies entities, roles and individuals that will be responsible, accountable, consulted, and informed of key actions, decisions, and deliverables related to the program. This chart should serve as a key "quick reference" document given its accessible format.

[12] Option C analysis based on monthly billing data has been a staple of the Energy Services Company (ESCO) Performance Contracting market but has not been used to estimate ex-ante savings claims for ratepayer funded custom energy efficiency projects.
[13] Responsible, Accountable, Consulted and Informed (RACI), which are used to inform individuals of their level of responsibility for each step on a project or program.

- A policy and procedures manual that details program requirements and processes. The manual should contain appendices containing project forms and report templates. Given the greenfield nature of Option C analysis and regulatory environment, this manual is likely to be a “living” document with frequent updates. The project manager should take responsibility for its stewardship and ongoing revisions with the assistance of subject matter experts. (As detailed in a prior section of this document, a key learning from the CWBD is that a policy and procedures manual that is not followed is of little use. The CWBD's Policy and Procedures Manual did a reasonable job of laying out the demo's processes, but it wasn't systematically followed by the implementers, particularly the pre- and post-implementation modeling report templates.) The program design should include a requirement for a policy and procedures manual, and the program budget needs to provide adequate resources to meet the requirements that the manual establishes.
- A “data and file specifications and workflow requirements” document to facilitate program processes as customers are recruited, their energy use patterns are analyzed, baselines are established, measures are identified, savings are quantified, and the program is evaluated. Key files include interval-level energy usage data, premise characteristics, key weather data (such as mapping to NOAA weather station and time-stamped dry bulb temperature, wind speed and solar radiation), project data, and tracking and adjustments for NREs.

- <u>Embedded M&V.</u> In addition to an in-house evaluator on the team to represent the program administrator, third-party evaluators should be included at the planning stage to inform data collection requirements and procedures to ensure evaluability and conformance to regulatory requirements. Evaluators should be included in all phases of a project’s life cycle including initial qualification, documentation of existing conditions, policy and influence review, identification and treatment of non-routine events, site visits and post-installation monitoring, incentive payments, and savings reporting.
- <u>Data management.</u> The multiple parties involved with the CWBD—PG&E, its implementers, the CPUC and their evaluators—may have started with some common files at the onset of projects, but as files were modified by one party the updated files were not distributed to others on the team. The lack of a consistent set of project files—due in large part to the lack of a shared server containing all the files necessary for analysis—created considerable confusion and resulted in organizations using different files (e.g., inconsistent sources of weather data) and making unique edits to files (e.g., independently cleaning interval data) without documenting the changes or providing the edited files to others on the team. The result was multiple analyses which could not be compared, confusion, duplicated work, missed deadlines, and the need for repeated transfers of large files so that analyses of energy savings could be repeated.
    - <u>Single repository of record for project files</u>. Moving forward, all project files should be accessible within shared, password-protected partitions on a shared server maintained by the program administrator with a clear protocol governing safeguarding of customer information, access validation, data security, and change privileges. If IT budgets, data security or other constraints make it impossible for a shared server to be used, then the program administrator must be the source of record for all data files to be used in project-level analyses with the files transferred via an enterprise secure file transfer (ESFT) system.

- Manageable directory structure. When designing the structure of a file storage system (whether shared or not), the program administrator should avoid using an excessively-complicated directory structure: the one designed for the CWBD was hard to navigate and resulted in difficulty in locating files and version control challenges. Detailed instructions and training for accessing, safeguarding customer data and safeguarding project files should be included in the data and file specifications and workflow requirements document. This document should be reviewed and agreed to by all parties at the onset of a program.
- Protocol for cleaning of interval-level energy usage data. Interval data should cover the complete period being investigated, be complete, be without empty intervals, and not contain duplicate entries. In practice, interval data files can contain empty and duplicate entries on occasion. Other problems include occasional anomalous spikes, both negative (zeros) and non-zeros, which should be investigated and tagged as anomalous prior to the data being released to third parties. Given that the utility is the original source and repository of record for usage data, the utility should be the provider for cleaned interval-level data to be used by all parties involved in Option C-based projects. So that interval data used in analyses can be traced back to data of record that will be the subject of evaluator and regulator data requests at any time, the utility should document and publish its data format and its cleaning protocols. The data cleaning protocol should contain specific rules for the treatment of empty intervals (such as unexpected missing reads, as well as expected missing reads such as "spring forward" hours at the start of daylight savings time), anomalous zeros and spikes, non-zero spikes that fall below the trend, duplicate rows and missing dates. Moreover, cleaned data files should have transformation rules appended so that the "raw" data file can be reproduced from the clean data file so that it can be traced back to the system of record.
- Weather data. The utility should be the sole source for weather data used in the analyses, including the mapping of treated premises to the appropriate weather stations. Protocols for determining premise-to-weather station mappings should be agreed to in advance by all parties. Data specifications and formatting should be standardized and compatible with participating software packages. Weather data formatted for use with eQUEST (or DOE-2) and EnergyPlus simulation packages should also be provided.

Program Design Elements and Requirements for Success

As PG&E plans to move forward with a full-fledged Commercial Whole Building Program, several findings from the demo are informing the design of its in-house program and may be useful to other program administrators. The key recommendations are listed below, listed in the order of program implementation:

1. Program implementation plan incorporating embedded M&V. The promise of "embedded M&V" relies on data of sufficient quality—delivered at the appropriate time—to meet the needs of a timely evaluation to permit changes to be made to the program implementation. Having a program implementation plan (PIP) that explicitly identifies the points in the program where key information and data will be provided to regulatory staff and evaluation consultants is critical. By building embedded M&V into the PIP, the points in program implementation when program design assumptions established in the program theory and logic model (PTLM) can be tested—and when modifications to the PTLM can be made—are made explicit. A "living PIP" will make transparent

when data elements will be shared, when program design assumptions will be tested throughout the program lifecycle, and when early evaluation feedback can be incorporated to improve program design.

2. Customer targeting. There are several possible approaches that may be taken. Two of these are discussed below.

   a. When possible, use an “invitation” approach. The CWBD used an invitation-based approach whereby PG&E identified and targeted customers most likely to qualify and benefit from a whole building approach. This approach requires identifying promising sectors, assessing historic energy use to assess goodness-of-fit statistics of Option C models and verifying utility records to ensure that customers have not participated in other programs that would disqualify them. One benefit of this approach is that it safeguards against high levels of free ridership because the impetus to participate originated with the utility rather than with the customer. While an invitation-based approach can be challenging to implement, proactively inviting a pre-identified set of customers with premises having high levels of energy use, certain building characteristics, and within market segments and climate zones has an additional benefit of focusing sales efforts.

   b. Assess feasibility of “scoring” interval data from the universe of commercial customers. This approach entails mapping commercial meters to premises and generating goodness-of-fit statistics for Option C modeling. The difficulty of this approach depends in large part on how well a utility maps meters to specific premises. A key advantage of this approach is that a sales rep can simply “look up” a premise to know whether it is suitable for Option C modeling.

   c. Create "customer journeys" for interested yet unqualified customers. To avoid customer disappointment, specific program alternatives should be prepared to be presented to customers expressing an interest in participating in a whole building program, but who are deemed unsuitable for a commercial whole building approach. Among the reasons why a customer may not be suited for a whole building intervention include: the "basket" of measures they are willing or able to adopt are unlikely to realize sufficient savings to be reliably detected at the meter, the measures desired would not qualify for inclusion, unacceptable submetering requirements, and/or prior program participation.

3. Need for a self-generation policy. The demo unsuccessfully attempted to exclude buildings with self-generation, and this prohibition becomes increasingly untenable as rooftop solar and other types of self-generation become more commonplace. The restriction should be limited to customers that are unable to provide access to self-general production data at the time of applying for program application at the level of granularity required to produce a model with adequate goodness-of-fit metrics.

4. Documenting program influence. The joint study uncovered several issues to be addressed so that meter-based programs can be successful. Guiding principles for program design should be focused on effectively influencing potential participants to develop and implement multi-measure deep savings efficiency plans in their buildings. Whether or not the invitation-only approach outlined above is feasible, the emphasis should be on engaging customers through the design of a unique and comprehensive suite of whole building measures—including behavioral, operational, and retrocommissioning measures where applicable—and assuring that the appropriate package of

measures is selected, implemented, and maintained over their expected lifetimes. By offering a suite of measures that would otherwise be difficult to obtain outside of the program, the burden of demonstrating influence will be lessened. That said, program services should apply resources where they are most effective for achieving savings and be as streamlined and as uncomplicated as possible, to promote applications at scale. Granted, methods are needed to document a program's influence on participant behavior, whether through a modification of existing net-to-gross surveys, or development of an entirely new methodology. These efforts should be done in cooperation with ex-post evaluators to ensure the data needs of impact evaluations are met.

5. <u>Documenting existing conditions prior to project acceptance.</u> The demo did attempt to document baseline conditions to ensure that baseline energy use reflected what was typical for the premise given the equipment that was in place prior to the intervention. Proper documentation of existing conditions ensures that the baseline established routine operating conditions and is not artificially depressed from non-working equipment or artificially inflated from poorly maintained equipment. Moving forward there are opportunities to improve these practices. For example, physical inspections should be required to ensure that all equipment found on site is inventoried, verified to be in proper working order, and is being operated on a regular basis. Although the documentation requirements for the demo were largely driven by the Option D simulations, a minimum documentation requirement should be maintained for projects using the Option C approach to ensure that the Option C approach does not undercount-or overcount-savings due to non-working, non-operating, and/or poorly maintained equipment.

6. <u>Changes to requirements for baseline data.</u> It is possible to model premise energy use with less than a full year of energy usage data, particularly for premises with highly seasonal energy usage patterns and for those in temperate climates. When statistical evidence can be presented to substantiate when less than a full year of data are required to represent the typical temperature conditions at a specific premise ("coverage factor"), the requirement for one year of baseline data could be waived.

7. <u>Establishing premise boundaries, meter mapping, and metering requirements prior to project acceptance.</u> The buildings in the demo were selected in part because they were deemed "well behaved": an examination of historic interval data confirmed that a statistical model of energy use with decent goodness-of-fit statistics could be created without requiring multiple independent variables (other than weather) to explain building energy use. Modeling energy use for the buildings in the demo was straightforward because each premise represented a simple use case for mapping meters to buildings: there was one gas meter and one electric meter per premise. There are many premises with more complex metering including multi-tenant buildings that typically have multiple sets of meters. Conversely, there are campuses where multiple buildings are served by a single set of meters; in such cases, or when energy efficiency measures focus on certain building subsystems, sub-metering may be necessary to establish a baseline prior to initiating a project using Option C modeling. It is critical to identify all the meters serving a premise prior to greenlighting a project, and to ensure that the meters serve only the premise being treated. Moreover, it is critical to ensure that the energy efficiency measures being proposed are likely to sufficiently reduce energy use so that it can be measured by the meters in place-or to specify that submeters will be required prior to project acceptance.

8. Need for a "Plan B" analysis plan. Implementers have the expectation that projects that are greenlighted to proceed for an Option C-based whole building program will meet minimum savings threshold. If the projects do not meet savings thresholds, the uncertainty in the estimates may become excessive. In such cases it may be difficult to detect if any savings occurred, and whether a performance payment for small but non-zero savings can be made. It may also be impossible to correct for non-program effects in an Option C model due to a lack of data required to calculate the required adjustment. An alternate analysis strategy will need to be in place to meet customer needs when the Option C analysis is inconclusive or cannot be completed. This implies a program requirement be in place for a minimum set of data collection to document the baseline building characteristics regardless of the analysis approach in the event an alternate method is needed, such as IPMVP Options A, B or D.

9. Project development. After customer participation agreements are signed, conduct an energy audit (ASHRAE level 2) to identify deep savings from multiple measures that may be installed within the project's timeframe. Savings analysis for identified measures may use Option D simulations, or engineering calculations. Audits should include effective useful life of measures and overall cost-effectiveness of the package of measures. Measure savings are gross savings estimates only with an existing equipment baseline, in accordance with AB 802. During the audit verify meter numbers and inquire about potential NREs. Upfront estimates of energy savings must be credible and of sufficient rigor to obtain customer commitment to fund the project. A site-specific M&V plan should also be created as a component of the overall project development

10. Documentation of energy efficiency measures. Because the demo evolved to be a comparison between Options C and D, detailed information about measure-level savings estimates and expected useful life was captured. Moving forward with a meter-based program, capturing measure-level detail will be just as critical. Even though savings may be measured at a whole building level, measure-level information will be required during the project development phase to estimate project savings and costs; and sell the project to the building owner. Measure-level savings estimates will provide information needed to estimate measure-weighted effective useful life (EUL) for savings reporting and project cost-effectiveness purposes. There was an extensive reporting template that implementers participating in the CWBD were required to follow. Adapting this for use moving forward would be a best practice for this purpose.

11. Post-installation inspections. Given their reliance on post-installation data to populate their models, it is critical that on-site, post-installation inspections are built into the program design. These inspections should go beyond a simple physical inspection since proper operation of the measures can be difficult to verify without performing functional performance testing. Functional performance testing can be conducted by monitoring the affected equipment through Energy Management System (EMS) trend logs or installing standalone instrumentation for sufficient time to verify proper operation. Acceptance testing conducted under Title 24 for projects requiring a permit may be helpful for verifying correct operation of measures.

12. Project quality assurance. After measure installation, implementers should employ functional performance testing[14] of measures to verify correct installation and operation. To ensure customer value, verification of all installed measures will provide assurance that the anticipated savings will materialize (or trigger further work to rectify problems with installation and operations). Implementers should encourage customers to integrate functional performance testing into their established operations and maintenance practices to improve measure savings persistence.

13. Planning to address non-routine events (NREs). The study team found that NREs occurred in three of the five study projects; NREs may occur in a similar proportion of programs using an Option C approach. The CWBD did not adequately plan for identifying and dealing with NREs: there was no systematic study of baseline data for detection of potential NREs, no use of software to monitor energy usage, and no systematic plan for dealing with NREs that were identified. Several suggestions for addressing these shortcomings are discussed below.

    a. Ongoing structured interviews with building operators. Interviews with building operators are critical at every project phase. Prior to project acceptance, initial structured conversations with building operators will help develop rapport, confirm number and location of utility meters and sub-meters, and potential data streams to help explain energy use (e.g., from energy management systems, building operations schedules, recent or planned changes in occupancy and commissioning or decommissioning of equipment). After project acceptance, baseline period consumption data should be analyzed, and any unexpected usage patterns should be explored with site personnel during the energy audit/onsite survey phase of the project to discuss the sources of potential NREs. Throughout the project implementation and monitoring periods, periodic scheduled "check-in" interviews will maintain rapport, identify potential changes in building usage that may affect building use, and encourage building operators to reach out to implementers to report significant changes in building use that may impact energy use. In addition to these scheduled interviews, additional interviews should be scheduled when automated tools identify anomalies in energy use suggest likely NREs.

    b. Software detection of NREs. Although there are no standardized statistical methods for detecting NREs in an automated fashion, we note that public domain software tools such as those that have been developed by Lawrence Berkeley Labs (LBNL GitHub, 2018) are promising for flagging likely NREs observable in patterns of interval data for further examination. An assessment of these tools is beyond the scope of this project, but we anticipate that statistical identification of potential NREs will be used together with structured interviews with building operators as discussed above.

14. Ongoing project performance tracking. Incorporate performance monitoring of energy use throughout the implementation and performance periods. Performance monitoring may be implemented in different ways, but it is essentially a comparison of actual energy performance to expected energy performance. One method of implementation is to download energy use data

[14] Functional performance testing (FPT) here means testing the systems and equipment affected by the efficiency measures to assure they are installed correctly and perform according to their requirements. FPTs are part of California Title 24 acceptance testing, as defined in section 120.7 of the Building Commissioning Guide (https://www.energy.ca.gov/title24/2013standards/nonresidential_manual.html): Functional performance tests shall demonstrate the correct installation and operation of each component, system, and system-to system interface in accordance with the acceptance test requirements.

frequently (e.g., every 4 to 8 weeks) and perform savings analysis to verify whether savings are accruing as expected. Study energy usage data to identify potential NREs, interview building operators to verify their occurrence, and quantify their impacts as applicable. When savings are not accruing as expected, identify and address causes related to deficient performance and maintain focus on energy savings performance by building on-going energy consumption tracking into standard operating procedures. Continue to proactively engage customer on issues affecting energy consumption.

6.4. Addressing Emerging Technical and Regulatory Questions to Scale Programs and Support Evaluation

Considerations for Evaluation of Commercial Whole Building Programs

The site-specific Option C methodology of the CWBD was a direct response to the CPUC's call to IOUs in 2012 to develop a comprehensive approach to accomplish deep savings in commercial buildings. At that time—and until the passage of AB 802 in 2015—there was no methodology approved by the CPUC to use utility meters as a basis to determine energy and demand savings in commercial buildings. Savings for non-residential projects were determined by either deemed values in approved work papers or developed through engineering calculations including building energy simulation software (that is, IPMVP Option D). The Option D approach was added to the CWBD after its launch to enable PG&E to claim savings for the demo and to provide a basis to compare savings estimates from the two approaches.

There are two documents governing how evaluation is conducted in the State of California:

- The California Evaluation Framework (TecMarket Works, et.al., 2004) describes the context and purpose of evaluation in California. The Framework provides detailed descriptions of evaluation methodologies and it still relevant 15 years after its initial publication.
- The Evaluation Framework informed the California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals (TecMarket Works, et.al., 2006) which is the governing document for evaluation methodologies in California. The Protocols was adopted by ruling in 2006.

Neither document provides a methodology for using data from utility meters to determine energy and demand savings for a specific premise whereby adjustments must be made for non-routine events. Instead, the Evaluation Framework describes *billing analysis* whereby savings is determined for a *population* of premises. Although consistent with the general principles of IPMVP Option C at an individual premise level, billing analysis aggregates consumption across many premises to control for changes unrelated to the energy efficiency measures.

There is a recognition among evaluation professionals that the California Evaluation Framework needs a major update. Using a series of interviews completed in 2017, a needs assessment outlines a set of recommendations that include the addition of methodologies for NMEC analyses (Research into Action, 2017). Until the time that such an update is available, Southern California Edison has completed a first-generation NMEC Savings Procedures Manual (SCE, 2018) for implementers that addresses much of the site-level analysis requirements. The Efficiency Valuation Organization, publishers of the IPMVP

Protocols, has recently elected to develop its Advanced M&V guide from this manual, and so it is as good a starting point as any for site-level NMEC analysis methods and requirements.

Establishing acceptable "Advanced M&V" methods and requirements for site-level NMEC

The use of interval-level energy usage data may soon displace the use of monthly billing data as more jurisdictions adopt the metering infrastructure to collect, transmit, and store the significantly larger quantity of data. The industry must agree upon a standard set of procedures to correct and prepare high-frequency energy usage data. Efforts to standardize these procedures have been made in the CalTRACK process (CalTRACK 2.0, 2018). The more detailed the industry standard, the fewer issues that technical reviewers and evaluators will need to address. Methods and requirements for so-called "advanced M&V" include: validating meter data, preparing data sets, matching weather data files to premises, selecting appropriate modeling algorithms, reporting model goodness-of-fit metrics and uncertainties, and reporting savings. Several key requirements for site-level meter-based program evaluation are discussed below.

Open Source Model Transparency

Given that NMEC is a relatively new approach to determining savings, transparency and replicability of energy modeling is critical. Sufficient transparency of methods will enable regulators and evaluators to replicate results. We recommend the use of well documented open source, public domain models for transparency and replicability. So that savings estimates can be reviewed and verified, savings reports should document data cleaning processes, contain a detailed analysis approach narrative and provide access to well documented source code.

Validation Protocols for Proprietary Models

We recognize that proprietary models provide features and ease-of-use sought by some commercial customers. However, proprietary models tend to be difficult to evaluate because their source code is not openly available. Further complicating the ability of regulators and evaluators to replicate results is the lack of a well-accepted validation protocol. Research underway by Lawrence Berkeley Laboratory and other independent organizations should strive to produce a validation standard that can be adopted by the CPUC. Standardized non-disclosure policies adopted by the industry would facilitate the review of models produced by proprietary methods while safeguarding intellectual property rights.

Uncertainty Criteria

Once the key issues associated with fractional savings uncertainty (FSU)[15] calculations for hourly models are settled, FSU criteria should be established at the site (project) and/or aggregate (program) level. Current site-level rules-of-thumb regarding a minimum savings fraction of 10% and the 20% CV(RMSE) specifications included in ASHRAE Guideline 14 imply an FSU of approximately +/- 30%[16] at 90% confidence. Substituting an FSU criterion would allow projects with poor CV(RMSE) statistics, but with high fractional savings, or projects with low fractional savings but with good CV(RMSE) statistics to

[15] FSU is also known as relative precision.

[16] Based on a daily model with no auto-correlation using ASHRAE Guideline 14 calculations. Note, the analytical techniques to estimate FSU presented in ASHRAE Guideline 14 are under review.

qualify. Uncertainties at the program level will likely be lower than the individual project uncertainties as the errors in the individual estimates “compensate” at the program level. Establishing FSU at the program level would allow program implementers to establish their own project-level criteria, provided that the aggregated uncertainty of all projects in the program met the program-level criterion.

Deep Savings

The CWBD was developed to facilitate deep savings. The projects were designed to meet program targets to save 15% or more of total premise consumption. The following recommendations follow from analyses of the CWBD projects:

- Models developed from interval data may meet FSU criteria with lower total savings fractions, but whole building programs with multiple projects with lower savings fractions run counter to the objective of obtaining deep savings. Deep retrofit projects should strive to obtain greater savings than the nominal 15% threshold.
- The projects in the pilot contained a mix of capital and BRO measures. While BRO measures can be an important aspect of whole building projects, projects with most savings resulting from capital improvements are essential for achieving savings fractions exceeding the nominal 15% threshold and for assuring project-level savings with sufficiently long EULs to meet cost effectiveness thresholds.
- Focus on setting savings thresholds based on lifecycle savings rather than on first year savings. A lifecycle focus will encourage a focus on more capital measures and capital-focused projects.

Ex-Ante and Ex-Post Net and Gross Savings Estimates

The savings estimates resulting from a site-specific approach such as that used for the CWBD is gross savings. Gross savings are calculated for program participants relative to their predicted future energy usage absent participation in the program. On October 11, 2018 the CPUC issued resolution E-4952 that provides updates to the DEER database which provides an ex-ante net-to-gross value of 0.95 for non-residential meter-based projects which install multiple measures. Consequently, the ex-ante net savings estimate for site-specific meter-based programs is the gross savings estimate multiplied by 0.95. The ex-post gross and net savings estimates are determined by impact evaluations that are directed by the CPUC.

Cost Effectiveness Criteria

Programs delivered using ratepayer funds under the jurisdiction of the CPUC must meet regulatory cost effectiveness criteria. Calculating cost effectiveness presents several challenges unique to programs using an NMEC savings estimation approach.

Cost effectiveness calculations require a consistent set of metrics, including gross savings, project costs and net-to-gross ratios consistent with gross savings baseline assumption. Meter-based programs by their nature use existing conditions as the gross savings baseline. Measure costs need to be estimated consistently, requiring full costs rather than incremental costs. Net savings methodologies must also reference the same baseline as gross savings.

The effective useful life of a multi-measure project is based on estimated savings-weighted EUL of the measures installed. The measures and their relative contribution to project savings will need to be estimated prior to project implementation to obtain an estimated project-level EUL. NMEC procedures do not estimate savings by measure, so measure-level engineering estimates will be necessary to estimate project EUL. Savings calculations done during project development will likely suffice for this purpose without adding significant additional project expense.

Measure Persistence

Projects undertaken in the CWBD contained a combination of capital measures and BRO measures. The persistence of the BRO measures is not well known, and consequently the CPUC has deemed the effective useful life of BRO measures at two years. Ongoing monitoring of projects using functional testing and NMEC analysis beyond the contractual performance period can demonstrate measure-level persistence at a site. Credible data from functional testing will be a necessary precursor to any policy changes to demonstrate to regulators whether savings endure for BRO measures beyond this deemed EUL of two years.

Embedded M&V

Embedding M&V into the design and operation of a meter-based program facilitates the evaluation process by ensuring that programs are evaluable. Meter-based programs are data driven, and by their nature require data collection on the part of the program implementers to a level of quality required by evaluators. The data elements, collection protocols, data cleaning procedures, and documentation should be established in coordination with the evaluation team—then verified by the evaluation team early in the data collection process. The promise of "embedded M&V" relies on data of sufficient quality to meet the needs of the evaluation. The specific *data needs* for evaluation that need to be established at the onset of the project include:

- Metered data (raw and cleaned)
- Independent variable (weather) data (raw and cleaned)
- Software version used to conduct the analysis
- Site characteristics data
- Validation of meter numbers appropriate for the analysis
- Measure savings analysis done during project development to inform EUL calculations
- Data and analysis used to identify and calculate non-routine adjustments (as needed)
- Documentation of program engagement to facilitate net savings analysis

The site characteristics data requirement for Option D analysis exceeds what is required for Option C and may not be routinely collected. For example, existing equipment baseline efficiency and condition information may be required to inform the Option D analysis but would not be required for an Option C analysis. One challenge of the CWBD was that the Option D analysis was "layered on" after the start of the project and much of the data required for Option D analysis was not initially scoped into the project. In the event an Option D analysis needs to be conducted, it should be planned at the onset of the project since key data may be required that cannot be obtained during an ex-post site visit.

If projects will rely on independent data other than weather (such as production quantities for process-driven projects), implementers and evaluators will need to plan to collect it at a frequency consistent with the analysis. Note that the frequency of production data may be a limiting factor in the time scale of the overall analysis; this limitation may require aggregation of interval usage data to match the time scale of the production data. Careful documentation of data sources and all data cleaning activities is critical.

Given the time required for projects to complete the performance period and enter the savings claim process, the ex-post evaluators will need to maintain contact with implementers throughout the project lifecycle. This is a departure from the typical evaluation process and is inherent in the need for embedding M&V into program implementation processes. For NMEC to be successful, data gathering, documentation of processes, and data analyses should be responsibilities that are shared between program implementers and evaluators.

The specific *operations processes* necessary for evaluation of a whole building programs include:

- *Prior to premise consideration.* Ensure the suitability of premises for participation by: verification of sufficient historical interval data and determining whether it is well-suited for modeling without the need for multiple independent variables; providing a clear-cut definition of premise boundaries, generation sources, mapping of utility meters and identifying whether need exists for submetering. Analytical techniques to identify applicable meters and match them to a premise will not be 100% accurate. Inconsistencies in building address and account address can make meter matching difficult without a conducting a physical inventory of the meters.
- *Prior to project acceptance.* Develop a high-level M&V plan; verify data from submeters as applicable; assess baseline model and regression model from implementer; review attribution documentation, and complete attribution and passing project for technical review. The model must be able to predict baseline consumption within reasonable goodness-of-fit parameters. Buildings where the baseline models cannot meet the GOF statistics will need to be eliminated from consideration.
- *Prior to measure installation.* Plan for controlling for non-routine events by initiating relationship with building operator and ongoing monitoring of interval data. Meter-based programs require ongoing customer engagement to identify and characterize potential NREs. Make plans for functional testing of existing equipment as applicable.
- *After measure installation and throughout the performance period.* Monitor interval data to confirm drop in premise energy use; execute plans for functional testing as applicable; continue execution of plan for controlling for non-routine events by maintaining contact with building operators, monitoring interval data for non-routine events, and adjusting statistical models as necessary; estimate project-level energy savings and performance payments, and report program-level savings to regulators. Buildings with sufficiently accurate baseline models may be subject to non-routine events that are difficult or impossible to characterize. These buildings will need to use an alternative method for estimating savings if NRE adjustments are not possible.
- *Throughout the project lifecycle.* Maintain contact with key stakeholders including regulators, implementers, technical advisors, and reviewers.

How can meter-based programs be designed to facilitate evaluation?

Figure 6 is a conceptual diagram that illustrates a process flow showing responsibilities and handoffs between customers, program administrators/implementers, and evaluators/regulators for a typical site-specific, whole building meter-based program. While not all activities are shown, the diagram attempts to capture the critical ones.

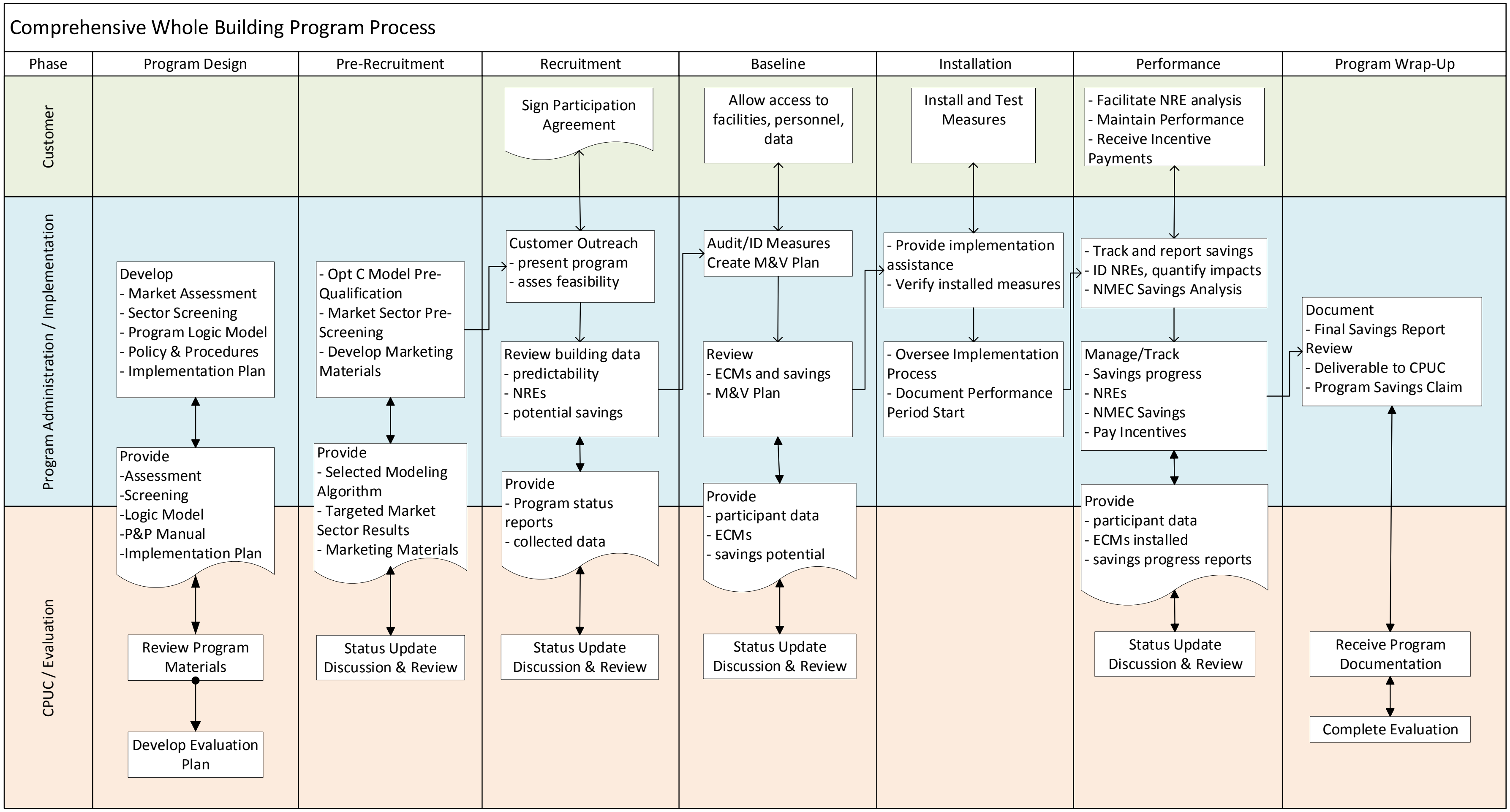

Figure 6. Illustrative Process Flow for Facilitating Embedded EM&V

Independent ex-post impact evaluations of whole building programs

Independent ex-post impact evaluations of whole building programs can use a variety of methods based on the objectives of the evaluation. Whole building programs have been a part of efficiency portfolios for some time, and some combination of Option C or Option D analysis is generally applied to the evaluation. Evaluation studies are generally conducted on a sample basis, where the uncertainty in the evaluated savings are based on the sample statistics. Meter-based programs offer some intriguing options for ex-post evaluation studies. The overall evaluation framework will likely include the following elements:

- Review of implementer Option C gross savings analysis
- Independent Option C or Option D gross savings analysis
- Independent assessment of net savings

These activities will likely be done on a sample basis. Based on the results of the pilot, the review of the Option C models was non-trivial and would not likely be done on the full census of participants. As described above, the Option C reviews included validation of the interval data, examination of the data cleaning process, examining model source code, re-running the Option C analysis and comparing results to reported values.

An independent Option C analysis may also be conducted to validate the results of the implementer analysis. Both proprietary and open source methods may benefit from an independent analysis. Validating the open source methods by simply examining source code may not be effective.

Sites in the program not suitable for Option C analysis may be evaluated using other IPMVP options, such as Option A (partially measured retrofit isolation), Option B (fully measured retrofit isolation) or Option D analysis. Whole building multi-measure projects with significant measure interactions will likely use Option D. This analysis would be done on a sample basis drawn from projects not meeting the Option C FSU criteria established for the evaluation.

7. Conclusion

The focus of this study was limited to the four research questions identified in section 3. The study team identified several issues in conducting deep savings whole building programs utilizing the two IPMVP Options C and D. While the Joint Study was not conclusive because it included only five of the CWBD participant sites, it did identify numerous issues while addressing each research question.

1. Comparison of whole building program savings estimates between Option C and D approaches.

After much effort to improve Option D simulations, for sites with no NREs, electric savings estimates generated by Option C and Option D were acceptably close in three of five cases. No good comparisons were found for the gas savings estimates, due to data quality issues as well as an inability of analytical models to accurately model gas usage patterns. The NRE impacts in the two site's electric savings estimates were estimated by the Option D simulations when significant information about them was known.

The predominant risk in whole building programs is the occurrence of NREs and their impacts on the energy savings, whether estimated through Option C or Option D. In this study, the team could estimate

NRE impacts after it collected information from the site that enabled it to be simulated in the Option D model. The Option C analysts did not go on site and therefore were blind to any occurrences of NREs. A proactive approach to identify, document, and quantify non-routine events is critical for estimating site by site savings. All models, regardless of the GOF statistics, will converge on the wrong savings if NRE adjustments are not included.

The CWBD encountered issues with both Option C and Option D analysis. Programs should avoid a focus on a single analysis method and recognize the need for 'backup' analysis methodologies to meets the needs of individual projects and regulatory requirements.

2. Improvements to key program technical requirements and documentation.

Although there was significant development of the program policies and procedures manual, data specification documents, reporting templates, and other resources, the study team found weaknesses in their content, and lack of enforcement of their use. Particularly the Option D calibration requirements were too loose and allowed many unacceptable shortcuts to be made. When Option C gas models failed due to lack of good data, or poor goodness of fit, no back-up actions were taken. The use of a secure file transfer system to exchange energy use data and project information, was not consistently followed. These issues led to several hours of confusion and loss of critical information that impeded this effort.

It was also noted that there are several areas where consensus industry standards would be helpful: establishing a framework and methodology for identifying and quantifying NREs, improving uncertainty estimation methodologies for both Option C and D M&V methods, and developing an industry standard methodology for calibrating Option D simulations and estimating individual measure and well as whole building savings.

3. Factors informing the design of future meter-based whole building performance programs.

Deep savings whole building savings programs are long engagements with their participants, the CWBD engaged with its customers for up to five years. The CWBD evolved over this time as well, for example adding in the Option D simulation requirements. Consistent project management, reporting, and data archiving is critically important, as is maintaining consistency of procedures and technical review. It is problematic to expect evaluation teams will have sufficient data to perform their tasks if program data is provided at the end of the performance periods. Instead embedded EM&V concepts should be explored and included in program design to assure evaluators will have the information needed for their tasks.

In terms of meter-based programs, the CWBD illuminated several best practices: identifying potential participants through analysis of usage data, pre-screening program participant to assure baseline model goodness of fit and accuracy, assuring participants willingness to invest in deep savings and await incentive payments based on performance, providing consistent energy data and weather files, providing good documentation of baseline conditions, documenting program influence, and establishing specific procedures for NREs.

4. Addressing the emerging technical and regulatory questions necessary to scale such programs and support evaluation.

In pursuing the joint study, several considerations for evaluating meter-based whole building deep savings programs in California became apparent. A primary consideration was the lack of an accepted evaluation methodology to use utility meters as a basis to determine energy and demand savings in commercial buildings. Neither of the two main evaluation guidance documents, the Evaluator's

Framework and the Evaluator's Protocols, provides a methodology for using data from utility meters to determine energy and demand savings for a specific premise whereby adjustments must be made for non-routine events. In addition, reporting structures for meter-based programs don't fit well in deemed or custom reporting structures. We echo the recently completed needs assessment for updating California's Evaluation Framework to include methods for evaluating meter-based site-specific projects and programs.

Several other considerations should be addressed to facilitate evaluation of meter-based projects and programs. We recommend education and training on advanced M&V methods to reduce confusion and leverage its benefits appropriately. To maintain transparency, open source advanced M&V models will be preferred, but methods to validate proprietary models should be adopted to increase the number of market actors. Research should be pursued in needed areas, such as developing savings uncertainty algorithms for M&V methods employing advanced modeling algorithms.

Cost-effectiveness and useful life metrics should be broadened to evaluate the combined packages of deep saving measures installed in these projects. Finally, as meter-based projects have a long duration, successful and informed evaluation will only result when evaluation data and information requirements are embedded in program design. Embedded EM&V will be critically important to gain rapid feedback on program operation and effectiveness.

References

ASHRAE, 2014, Guideline 14-2014, Measurement of Energy and Demand Savings, American Society of Refrigeration, Air-conditioning, and Heating Engineers, 2014, available at www.ashrae.org.

Assembly Bill 802, Williams. Energy Efficiency, October 8, 2015. Bill information available at: https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201520160AB802.

Bonneville Power Administration (BPA), August 2017, "Uncertainty Approaches and Analyses for Regression Models and ECAM," Prepared for Bonneville Power Administration by SBW Consulting, August 2017.

California Public Utilities Commission (CPUC) 2006, California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals (Evaluator's Protocols), April 2006, available at: http://www.cpuc.ca.gov/General.aspx?id=5399.

California Smart Grid Annual Report to the Governor and the Legislature, May 2014, California Public Utilities Commission (CPUC), available at http://www.cpuc.ca.gov/WorkArea/DownloadAsset.aspx?id=3292.

CalTRACK 2.0, 2018. "CalTRACK Methods, version 2.0", Available at: http://docs.caltrack.org/en/latest/methods.html

CPUC, 2016, Focused Impact Evaluation of the 2013-2014 Home Upgrade Program, CPUC CalMAC Study ID CPU0118.01, May 2, 2016, available at: http://www.calmac.org/publications/CPUC_HUP_Focused_Evaluation-FINAL_05-03-16atr.pdf.

CPUC, 2017, Final Report: 2015 Home Upgrade Program Impact Evaluation, CalMAC Study ID CPC0162.01, June 23, 2017, available at: http://www.calmac.org/publications/RES_5.1_HUP_FINAL_REPORT_ATR_06-30-17.pdf.

Decision 12-05-015, 2012, Decision Providing Guidance on 2013-2014 Energy Efficiency Portfolios and 2012 Marketing, Education, and Outreach, available at: http://www.calmac.org/events/Decision_12-05-15.pdf.

Efficiency Valuation Organization (EVO), 2016, International Performance Measurement and Verification Protocol (IPMVP), Core Concepts, October 2016, available at: www.evo-world.org.

Lawrence Berkeley National Laboratory (LBNL), 2018, GitHub Non-Routine Event Detection and Adjustment Tool, https://github.com/LBNL-ETA/nre.

Matthieu, J.L., P.N. Price, S. Kiliccote, and M.A. Piette, "Quantifying Changes in Building Electricity Use, With Application to Demand Response," IEEE Transactions on Smart Grid, 6 2:507-518, 2011.

PG&E, 2013a, Functional Testing Protocols for Commercial Building Efficiency Baseline Modeling Software, Emerging Technology Report ET12PGE5312, September, 2013, available at: https://www.etcc-ca.com/reports/commercial-building-energy-baseline-modeling-software-performance-metrics-and-method-testing.

PG&E, 2013b, Energy Management and Information System Software Technology Assessment: Considerations for Evaluating Baselining and Savings Estimation Functionality, Emerging Technology Report ET12PGE1311, September, 2013, available at: https://www.etcc-ca.com/reports/energy-management-and-information-system-software-technology-assessment-considerations.

PG&E, 2018 Energy Management and Information System Software: Testing Methodology and Protocol, Emerging Technology Project ET13PGE7161, January, 2018, available at: https://www.etcc-ca.com/reports/energy-management-and-information-system-software-testing-methodology-and-protocol.

Reddy, Maor and Panjapornpon. "Calibrating Detailed Building Energy Simulation Programs with Measured Data— Part I: General Methodology," HVAC&R Research,Vol. 13, No. 2, March 2007.

Reddy, Maor and Panjapornpon. "Calibrating Detailed Building Energy Simulation Programs with Measured Data— Part II: Application to Three Case Study Office Buildings," HVAC&R Research,Vol. 13, No. 2, March 2007

Southern California Edison (SCE), 2018. "Normalized Metered Energy Consumption Savings Procedures Manual," version 1.01. Emerging Technology Report no. ET15SCE1130. Available at: https://www.etcc-ca.com/reports/normalized-metered-energy-consumption-savings-procedures-manual

TechMarket Works, 2004, The California Evaluation Framework, prepared for California Public Utilities Commission and the Project Advisory Group, June 2004. Available at: http://www.calmac.org/publications/California_Evaluation_Framework_June_2004.pdf.

TechMarket Works, 2006, California Energy Efficiency Evaluation Protocols: Technical, Methodological and Reporting Requirements for Evaluation Professionals, April 2006. Available at: http://www.calmac.org/publications/EvaluatorsProtocols_Final_AdoptedviaRuling_06-19-2006.pdf

Research into Action and Navigant Consulting, 2017, "California EM&V Framework Refresh Needs Assessment, Draft Report," August 2017. Available at https://pda.energydataweb.com/api/downloads/1904/EMV%20Refresh%20Draft%20Report%20v9_clean%20(002).docx

Appendix A. Whole Building Savings Methodologies

Option C Whole Building. Under Option C, savings are determined from the predictions of what energy use would have been had the measures not been installed, as compared to the actual post-install period energy use as measured at the meter. Baseline period energy use and independent variable data are used to develop empirical models, which may be regressions or more advanced modeling algorithms. The model's independent variables are parameters that are expected to have the most influence on a building's energy use. Ambient temperature and building operating schedules are commonly used as independent variables. Once the baseline model has been developed, independent variables from the post-installation period are used in it to determine the predictions of what baseline use would have been as described above. Savings determined from the difference in the adjusted energy use (baseline model prediction under post-installation conditions) and the measured energy use from the meter are called avoided energy use by IPMVP. Normalized savings may also be estimated and require a second model be developed from post-installation period data, and both baseline and post-installation models used to predict their usages under a selected set of normal conditions, which usually include long-term average temperature data.[17]

The advantages of the Option C approach are that it requires relatively few streams of energy use and independent variable data, which are generally easy to obtain. Given the numerous M&V 2.0 modeling algorithms available, it is a straightforward process to develop and assess models prior to implementation, then carry out the savings analysis after the post-installation period has been completed. The Option C process is well known, follows a standard methodology and savings calculations may be completely transparent. It quantifies gross savings for all measures affecting energy use observable at the utility meter, including any interactive effects between equipment and systems. Key limitations of Option C are that it does not yield estimates of savings impacts for individual measures, which impacts the calculation of their life cycle savings and cost-effectiveness, and that it does not quantify savings from a baseline defined by state codes or industry standards. A major risk of Option C is that it does not adjust for impacts from other events that affect building energy use, called non-routine events (NREs). These non-routine adjustments must be estimated in addition to the savings analysis (EVO 2016).

Option D Calibrated Simulation. The Option D approach involves development of building simulations that are calibrated to energy usage and building operations data. Building simulation software packages can be thought of as a series of interrelated physical and engineering models that represent how energy flows into and throughout a building. Simulation inputs describe the physical building, its mechanical and electrical systems, and building operations. For M&V purposes these are physical models, analogous to the empirical models developed under Option C. The CWBD team used eQUEST,[18] a well-known public domain building simulation software package, as it had extensive capabilities to represent actual building geometry, construction materials, envelope characteristics, shading, equipment specifications, floor plans, and interior zoning. It accepted hourly data files describing ambient weather conditions and

[17] Normalized savings are generally stated under typical conditions (e.g. CA Climate Zone 2010 climate data), while avoided energy use are savings achieved in the post-installation period. Normalized savings are useful for long-term forecasting purposes. Under Option C, normalized savings are estimated with the use of a baseline and a post-installation period model, while avoided energy use requires only a baseline period model.
[18] Available at https://energydesignresources.com/

solar insolation. The software produced hourly and monthly estimates of energy use for the whole buildings as well as major end-uses. These outputs facilitated the calibration process, as the CWBD required simulations to be calibrated based on monthly electric and natural gas usage data.

The advantages of the Option D approach are that it serves multiple purposes: it can provide initial estimations of individual measure savings for cost-effectiveness considerations; it can produce whole building gross savings estimates as well as measure-level savings estimates, and it can produce to-code and above-code savings estimates for regulatory purposes. The disadvantages of Option D include extensive data requirements such as building plans, equipment specifications, and operational data; skilled building modelers, and a lot of time to develop the models. Due to software limitations, some measures cannot be simulated. Information and data for the entire building, not just the affected systems and equipment, are required. Simulators must be experienced and thoroughly familiar with the simulation software to use it correctly. Much time is required to develop the proper inputs and iterate simulation runs until calibration is achieved. Technical reviewers must possess these same skills to review the completed simulations effectively. Each of these issues drives up labor costs, making the Option D approach costlier than Option C alternatives. As with Option C models, non-routine events are also a major risk when using Option D simulations: there is no "automatic" way to account for them.

The simulations may be calibrated to equipment installed and operating conditions in the baseline period, and inputs that describe code requirements and the efficiency measures added to forecast savings. Conversely, the simulations may be calibrated to post-installation conditions with all the efficiency measure inputs included, then successively removed one at a time to obtain measure-by-measure savings estimates in a back-casting approach. A back-casting approach was used in the CWBD to calculate regulatory savings.

To meet regulatory requirements, the implementers estimated each measure's to-code and above-code savings, which required two simulation runs per measure. It also required that the measure simulation order be consistent (e.g., first measures affecting building loads, then systems, and finally central plant measures). For the back-casting method used, the two-step simulation runs for each measure were in reverse order (plant, systems, load).

The CWBD only required calibration to monthly post-installation reporting period electric and natural gas billing data, requiring that the CV(RMSE) and MBE be less than 15% and 5% respectively.[19] Rather than calibrating each simulation's estimations of building end-uses based on metered data, the CWBD required only that each simulation's end-use estimations be compared against California End-Use Survey (CEUS) data (CEC 2006) to assure each building's energy end-uses were reasonable.

---

[19] The random error metric CV(RMSE), which is the coefficient of variation of the root mean squared error, and the bias error metric MBE, which is the mean bias error, were used as calibration criteria. The monthly calibration requirement was obtained from ASHRAE Guideline 14-2002.

Appendix B. Study Process Methodology Description

To select sites for the "deep dive" review, PG&E assessed its participating CWB customers and selected five (5) sites for review. The customer contacts and the associated CWB Implementers associated with the five sites were notified of their inclusion in the group. Furthermore, the Customers were advised of the additional contacts with ED Staff and PG&E that would be required in order support the Study Process (e.g. customer interviews, site-visits, etc.). PG&E's selection criteria was driven by several considerations, including the magnitude of expected energy savings, diversity of represented building type, and the diversity of energy efficiency measures (EEM) by type. Priority was given to projects for which incentives were tied to energy savings determined through pre/post analytics (Option C).

TABLE B.1. Study Site Selections

STUDY SITE NO.	BUILDING TYPE	GEOGRAPHICAL LOCATION	SQUARE FOOTAGE	ANNUAL KWH	ANNUAL THERM
A	Office/Lab	South Bay (Santa Clara County)	50,000	4,800,000	4,200
B	Library	SF Peninsula (San Mateo County)	20,200	348,000	11,00
C	Office	West Sacramento (Yolo County)	35,300	811,000	3,100
D	Retail Grocery	Central Valley (Fresno County)	28,100	1,900,000	43,400
E	Office/Lab	South Bay (Santa Clara County)	106,000	2,280,000	58,500

As preparation for the Study Process, PG&E provided two types of materials: program materials that pertain to CWBD, and project-specific materials for the Study Process projects. Project materials for all Study Process sites were made available for review through a secure FTP site.

Program Materials

- CWB Policies and Procedures Manual
- Data Specifications Document
- Foundational research (e.g., 2013 ET Technology Assessment study (LBNL, 2013, PECI, 2013)

Project Materials

PG&E provided the ED Staff with project materials related to a sites enrollment in CWBD, as well as the energy saving analysis and supporting report documents, including:

- Enrollment materials
  - Program Application
  - Assessment / Scoping Report
  - Project energy savings analysis
  - Incentive Agreement
- Post-Installation Verification materials

    - VR1 Report (includes information collected through the Post-Installation Inspection)
    - Project energy savings analysis
    - Modeling Output Tables
    - eQUEST calibrated simulation input files
- Post-Monitoring Verification materials
    - VR2 Report
    - Modeling Output Tables
    - eQUEST calibrated simulation input files
- Other related project materials

ED Staff reviewed the project site materials and compiled a findings document summarizing the results of their review. This memorandum was sent to PG&E and was used by the utility to provide ED Staff with any additional information and to inform subsequent projects.

B.1 Study Process Activities

For each Study Site, PG&E arranged for review and discussion of project documentation data and energy models. PG&E also arranged additional "optional" activities that were subject to ED Staff interest and availability and in some cases, customer approval as follows:

Activity 1: Project Documentation, Data and Energy Model Review

This activity entailed the review of Study Process site project documentation, data and energy modes. A full complement of project documentation and materials were provided for review and covered each of the three project phases:

- Phase 1: Pre-Installation Commitment (Enrollment, Assessment and Scoping)
- Phase 2: Post-Installation Verification
- Phase 3: Post-Monitoring Verification

The documentation included simulation input and output files, simulation program (eQUEST) version number, and all data requested of its implementers that was used to calibrate the models and to determine savings for each measure.

Within four weeks of delivering Phase 2 documentation, data and energy models for each Study Process project, PG&E arranged a call or meeting between ED Staff, PG&E, the Technical Consultant and the assigned Program Implementer for that project to discuss the documentation, data and energy models. ED Staff questions and inquiries about the documentation and data were provided in writing.

Activity 2: Meet with CWBD Program Implementer (optional)

Upon request, PG&E arranged a conference call or meeting with the individual CWBD Program Implementer responsible for the investigation, engineering calculations and modeling, energy efficiency recommendations, scoping assistance, project commitment documentation preparation, technical assistance, site monitoring, etc.

Activity 3: Conduct Site Visit and Customer Interview

ED Staff and PG&E visited all four selected study sites to interview customers, perform facility walkthroughs, and inspect the implemented measures. The site visits also included efforts to identify additional facility information to further inform or clarify details included in verification reports, energy savings analysis, and data collection.

For ED, it offered an opportunity to visit with customers and learn their motivations for implementing the projects, how the experience differed from previous EE program participation, and discuss areas for improvement for a future program. For PG&E staff, the on-site review presented an opportunity to revisit initial project development and analysis steps, and most importantly, solicit customer feedback. Additionally, for project Implementers, on-site visits gave all participants the opportunity to understand the challenges and approaches taken to analyze and model the implemented measures.
Through this process, ED Staff and PG&E discussed and reviewed recommendations to CWBD project analysis, and used project findings to target and facilitate program-level discussions during meetings and check-in calls. The project site reviews informed both existing projects as well as future considerations for a scalable CWB Program.

Activity 4: Study Process Lessons Learned

PG&E and ED Staff scheduled meetings to review each Study Process project to discuss lessons learned, including opportunities to improve program design, program policies and procedures, program implementation quality assurance, data collection procedures or other aspects of the CWBD Demonstration to better meet regulatory needs and meet quality assurance requirements.