IPMVP's Snapshot on Advanced Measurement & Verification

WHITE PAPER

***Prepared by*:**

Lia Webster
Facility Energy Solutions LLC

With Contributions from:

Jessica Granderson
Samuel Fernandes
Eliot Crowe
Shankar Earni
Lawrence Berkeley National Laboratory

JANUARY 2020

IPMVP's Snapshot on Advanced Measurement & Verification

JANUARY 2020

Cover page picture by Dan Formsma – Unsplash

V1.0

TABLE OF CONTENTS

Authors & Acknowledgements

Prepared by

Lia Webster, Facility Energy Solutions LLC

Contributors

Jessica Granderson, Samuel Fernandes, Eliot Crowe, and Shankar Earni, Lawrence Berkeley National Laboratory

Industry Input

The team thanks the following individuals for their technical review and input:

Todd Amundson, Emily Cross, Devan Johnson, David Jump, Bill Koran, Steve Kromer, Josh Rushton, and Hassan Shaban.

Acknowledgments

Efficiency Valuation Organization (EVO) is grateful to the following organizations for their support:

- Seattle City Light
- Bonneville Power Administration

Lawrence Berkeley National Laboratory's contribution to this work was supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, Building Technologies Office, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The authors thank Sarah Zaleski for sponsoring this research.

INTRODUCTION

Advanced measurement and verification (M&V), or 'M&V 2.0', is a strategy for verifying savings from energy projects which has received a great deal of attention over the last five years. Spurred by the abundance of high frequency metered energy use data from advanced electric meters for the 'Smart Grid' and associated analytic software tools, the past decade has seen much research, technology development, regulatory mandates, and pilot-level program deployment of advanced M&V methods. Despite this, there are still some questions and misconceptions among the different stakeholders of advanced M&V (e.g., utility program managers, energy service providers, M&V practitioners, and efficiency project investors) regarding the proper application, technical challenges, alignment of advanced M&V approaches with best-practice M&V principles.

This paper represents a snapshot of advanced M&V technical state of the art and current industry activities and is written to provide the groundwork for EVO's upcoming IPMVP Application Guide on advanced M&V strategies (scheduled for publication in Spring 2020).

This paper is organized as follows:

- Section 1 provides background on advanced M&V, including drivers and use cases.
- Section 2 highlights recent research findings, regulatory actions, and utility pilot programs.
- Section 3 covers the technical developments in advanced M&V tools, including the introduction of new open-source methods and the launch of EVO's tool testing portal.
- Section 4 outlines the key open issues.
- Section 5 discusses the implications of the changing industry on M&V and adhering to IPMVP principles when using advanced methods.
- Section 6 concludes with key findings and anticipated next steps.
- Additional details are provided in the Appendix.

1. BACKGROUND

Standard M&V methods[1] have been established for several decades as a means to quantify the impacts of energy efficiency projects. Over the past decade, there has been increased interest in "advanced M&V" (sometimes referred to as M&V 2.0)[2]. Advanced M&V (AM&V) applications are characterized by:

[1] Use of energy meter data in finer time scales with near real-time access; and

[2] Processing large volumes of data via advanced analytics, to give more accurate and timely feedback on energy performance and savings estimates.

These approaches are intended to be conducted more quickly, more accurately, and potentially at a lower cost than traditional methods.

Utilizing meter data to determine efficiency project savings is not new; the International Performance Measurement and Verification Protocol (IPMVP), defines the basic approach as shown in Equation 1.

Equation 1: IPMVP's General M&V Equation

Savings = (Baseline Period Energy – Reporting Period Energy) ± Adjustments[3]

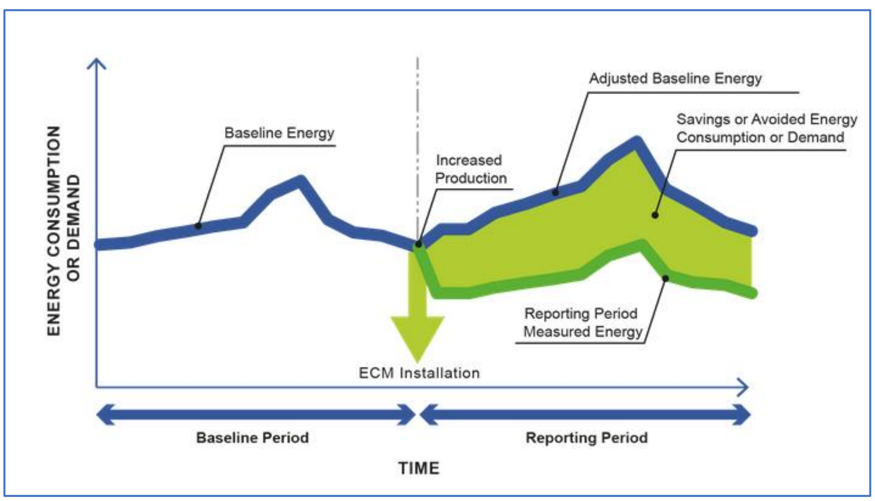

Figure 1: Whole Building energy savings, IPMVP Core Concepts 2016

As shown in and Figure 1, the baseline period may be the 12 months before the start of an efficiency project, the reporting period falls after the completion of the efficiency project, and the 'Adjustments' may

[1] These M&V Options are defined by IPMVP and include retrofit isolation (Options A & B), whole building metered energy use (Option C), and calibrated simulation (Option D).
[2]See The State of Advanced Measurement and Verification Technology and Industry Application, 2017
[3]Equation 1 from IPMVP's Core Concepts October 2016. See Appendix for related IPMVP equations for *Avoided Energy Consumption* and *Normalized Savings*.

be “routine” to account for expected changes in independent variables (e.g., weather and production) or “non-routine” to account for unplanned changes in other site conditions (e.g., shift in building use).

This basic approach is unchanged for advanced M&V, but the increased availability of interval meter data offers several benefits, such as:

- Verifying savings in a shorter timeframe (e.g., less than three months after efficiency project completion, depending on source and timing of savings).
- Visibility of savings at a lower threshold (e.g., ability to see 5% savings using hourly meter data, as opposed to needing >10% savings if using monthly data).
- Ability to quantitatively characterize energy savings seasonally by time-of-day and/or day-of-week.

Advanced M&V also offers ancillary benefits, such as improved monitoring of savings, providing feedback to building owners and energy managers about energy use at their facility, and allowing utilities and other program implementers to more closely monitor project performance and course-correct when issues occur. These enhancements are largely enabled by software providing powerful analytical and visualization capabilities of metered energy data supporting qualified M&V practitioners and building energy engineers.

The basics of conducting meter-based M&V are the same as published in the first IPMVP in 1997, as are the key technical limitations:

- The savings must be larger than the modeling error, and consistently larger than energy fluctuations at the facility throughout the year.
- The range of reporting period operating conditions should not stray outside of the operating conditions observed in the baseline period.
- Meter-based methods will include the effects of all changes occurring within the facility, with changes unrelated to the targeted project appearing as either increased or decreased savings.

Research and development efforts today remain focused on overcoming these constraints.

Advanced M&V’s Current Drivers and Use-cases

The evolution of advanced M&V over the past decade has been driven by several interrelated factors, including:

- Advanced grid management: Improved electric grid management for national security and resiliency requiring quicker feedback and improved monitoring and control of grid assets is a primary driver of smart meter infrastructure investments. Advanced Metering Infrastructure (AMI) meters are a prerequisite and provide the high-frequency revenue-grade energy data utilized in advanced M&V.
- Evolving Public Policy: Legislative actions, executive orders, and utility commission rulings have driven policy that calls for the use of utility meter data to account for demand-side management

program impacts. Utilities are using advanced analytics to manage the evolving generation mix and to measure performance against carbon reduction targets. Increasing the application of time-of-use rate schedules has raised awareness that the time energy savings occur is significant for carbon reduction efforts.

» Software Innovation: The availability of AMI data has been leveraged by private industry resulting in the rapid development of software and analytical tools, including various Energy Information Management Systems. Industry competition and a growing market have spurred rapid advances in software capabilities. Development is continuing with private, government and industry efforts.

Given the various drivers of advanced M&V – and the numerous industry stakeholders connected with those drivers – it is helpful to think of advanced M&V more as a set of applications rather than as a single approach. Now that the core modeling approaches and methods are more mainstream, the potential of advanced M&V can be realized through various use cases, each having its own emphasis, including:

» Performance tracking & cost reduction for building owners and energy managers: Advanced M&V is used to assess ongoing building energy performance, to reduce time-of-use charges, to flag operational anomalies, and to compare energy use to past performance.

» Pay-for-performance (P4P): Utilities and Energy Services Companies (ESCOs) use meter-based M&V as a primary means of quantifying savings and for establishing financial agreements for specific projects. Although this method can be more accurate than some traditional methods, complexities often arise in multi-year engagements due to changes in energy unrelated to the specific energy project(s).

» Aggregated approach: Advanced M&V is used across many sites, and the individual results (or the overall data) are aggregated. This approach targets a savings goal for the portfolio along with a maximum fractional savings uncertainty, and often has lower rigor for site-level accuracy, relies on a relatively homogenous population, and has no mechanism for site-level resolution of anomalies. They are used by utilities for residential programs and by energy aggregators for grid-level reporting.

» Utility embedded M&V: Utilities or their third party program administrators may engage traditional estimation methods to claim savings but use advanced M&V in parallel to see where the savings are not fully realized or apparently overachieved, thereby giving them a chance to react and investigate where needed.

» Third-party embedded EM&V: Similar to utility approach above, but in the context of program evaluation (EM&V) to satisfy regulatory compliance requirements. There may be more structure and stringency (e.g., with sampling approach and control groups), and advanced M&V may be used with other evaluation strategies.

While software can automate many of the data analysis steps of the M&V process, the advanced M&V process as a whole cannot be fully automated, and it is not a ‘one size fits all’ approach that makes existing M&V methods obsolete.

Requirements for stable building operations and levels of energy savings sufficient to consistently be seen in the model limits the use of advanced M&V. This method excels in certain projects and program situations (e.g., projects with a high level of savings and accurate baseline models) but, in most cases, requires a “human in the loop” to resolve limitations and contextualize results generated– as do other M&V approaches.

2. INDUSTRY DEVELOPMENTS IN ADVANCED M&V

The last decade has seen a drive for advanced M&V coming from three main directions: efficiency industry stakeholders conducting *research and development* efforts to define the new practice, *utilities and their regulators* striving for meter-based savings reporting and trying new program approaches, and *software developers* leveraging advanced M&V as a component of a broader set of analytical features. Each of these industry drivers is summarized below.

RESEARCH AND DEVELOPMENT

The emergence of advanced M&V raised many questions around technical rigor, accuracy, consistency, transparency, etc. Various organizations and industry groups have been working through these questions to develop industry-accepted tools and guidance that will build confidence around the adoption of advanced M&V techniques. Through these efforts, many technical and process-related issues have been resolved, and others are in progress of being addressed (see examples outlined in Table 1).

Table 1: Summary of Research Areas in Advanced M&V

Advanced M&V Research Area	Current State	Future Directions
Developing modeling algorithms and software tools	▪ Ongoing development of proprietary and public AM&V methods and tools	▪ Ongoing optimization of methods and tools. ▪ Filling gaps in the functionality of tools (e.g., calculate normalized savings[4] using TMY weather data). ▪ Development of user interfaces for open-source tools
Establish testing methods to validate the accuracy of public and private software tools	▪ Studies compared the accuracy of AM&V to other methods[5]. ▪ Launched EVO/LBNL's online tool testing portal[6]	▪ Expansion of model testing datasets, including targeted customer data for tool comparisons

[4] *Normalized savings* are determined by adjusting both the baseline energy and reporting period energy to a common set of conditions, often typical meteorological year (TMY) weather data used to represent average long-term conditions. Normalized savings are used for utility planning, and to compare savings performance consistently year-to-year. Otherwise, *avoided energy consumption* is determined based on the conditions during the reporting period.

[5] PG&E's Commercial Whole Building Demonstration Project compared AM&V side by side with traditional M&V approaches in selected utility projects.

[6] EVO's Advanced M&V Tool Testing Portal compares the accuracy of tools using a protocol developed by Lawrence Berkeley National Lab with a dataset from commercial buildings.

Advanced M&V Research Area	Current State	Future Directions
Calculating uncertainty in savings estimates	▪ Autocorrelation confounds uncertainty estimates using high-frequency energy data; Correction factors to Fractional Savings Uncertainty (FSU) for hourly/daily data were adopted, but uncertainty is underestimated. ▪ Published IPMVP application guide on uncertainty assessment[7]	▪ Explore alternate methods to calculate FSU. ▪ Identify other uncertainty criteria and build consensus within industry
Detecting non-routine events (NREs)	▪ Manual review of time series and CUSUM charts. ▪ Some initial research on data-driven statistical approaches	▪ Revise methods to reduce false positives. ▪ Develop categories and examples of NREs for *IPMVP Application Guide on AM&V.* ▪ Incorporate interval data strategies with site-level fault-detection tools
Conducting non-routine adjustments (NRAs)	▪ Initial research on data-driven approaches. ▪ Manual analysis strategies identified. ▪ Integration of calibrated modeling with AM&V[8]	▪ Continued research; Development of additional non-routine event indicators, including impact thresholds. ▪ Develop examples of common NRAs for *IPMVP Application Guide on AM&V*
Performing real-time M&V	▪ Day-behind energy data acquisition (by proprietary applications via API). ▪ Availability of data through the 'Green Button' initiative. ▪ Automated execution of routine adjustments using current weather to calculate Adjusted Baseline Energy; Tracking savings	▪ Improved access to AMI data. ▪ Increased automation in the calculation. ▪ Combine assessment of energy efficiency, distributed generation, and demand response measures
Performing accelerated program level EM&V	▪ Coordinate AM&V with EM&V. ▪ Embed EM&V monitoring	▪ Evaluation outcomes for programs aggregating lower-accuracy site results will inform future directions
Establishing best-practice technical guidance	▪ Various M&V Guidelines, including utility pay-for-performance programs[9]	▪ Publication of *IPMVP Application Guide on AM&V*

[7] See Uncertainty Assessment for IPMVP, 2018

[8] DeltaMeter is a proprietary AM&V tool based on a building simulation model and meter data analyses. This 'hybrid' tool provides a performance-period target without needing to wait and develop a performance period model; non-routine adjustments are made by adjusting the simulation model.

[9] Detailed guidance on meter-based M&V approaches is available from: California Commissioning Collaborative's Guide on Verifying Savings from Commissioning Existing Buildings; California Public Utilities Commission's Site -Level NMEC Guidance.

The work in these areas is supported by the United States Department of Energy (U.S. DOE), Berkeley Lab, various utilities, energy engineering firms, and industry organizations such as EVO[10], NW RTF[11], ASHRAE[12], and Linux Foundation[13].

Regulatory & Utility Activities

Meter-based energy efficiency programs serving large numbers of participants are still relatively new to regulators and utilities, but there are established and emerging programs in various regions. Strategic Energy Management (SEM) is a high-engagement program approach popular with large industrials and some commercial customers, particularly in the Pacific Northwest region, but expanding nationally. SEM typically monitors metered consumption over a multi-year engagement period and determines energy savings from a baseline energy model adjusted for independent variables, often production and weather variations.

Beyond SEM, California, and New York are leading the way: legislative actions in CA began in 2015 and established a focus on meter-based approaches to quantify utility program savings. The changes in California, detailed in Table 2, allow the use of *existing conditions* as the baseline and eliminates previous complications to determine 'above-code' savings for specific non-retrofit programs. CA has also issued programmatic and technical guidance in support of the 2015 legislation. New York State is another example where advanced M&V has been encouraged through regulatory language, and pilot projects are underway.

Not surprisingly, advanced M&V approaches are increasing in popularity in utility programs throughout the states, and across all customer sectors. Multiple pilot programs, feasibility studies, and comparative studies have been conducted. The Consortium for Energy Efficiency (CEE) published a summary[14] of utility pilot programs and case studies on meter-based approaches. Challenges with meter data aggregation and handling, expanding programs to small-to-medium businesses, and developing norms and best practices for the statistical validity of savings estimation were highlighted.

Beyond completed pilot programs, there are several active and pending programs across the United States and in Canada that are utilizing advanced M&V, some of which are listed in Table 2. These meter-based programs vary in program design details such as energy measures included, market sectors targeted, and customer engagement schemes. They also report a mix of avoided energy use and normalized savings.

---

[10] EVO or Efficiency Evaluation Organization, a non-profit organization which publishes the (IPMVP).
[11] NW RTF or Northwest Regional Technical Forum is a technical advisory committee to the utilities in the Northwest.
[12] ASHRAE Kaggle 'Great Energy Predictor Shootout III'
[13] Linux Foundation Energy (LFE) manages Energy Market Methods Consortium, which includes CalTRACK.
[14] See report Comparative Analysis of Meter Data-Driven Commercial Whole Building Energy Efficiency Programs, 2018.

Table 2: Utility Programs Utilizing Advanced M&V

State/Province	Utility or Sponsor	Program Name	Sector
BC (Canada)	BC Hydro	Strategic Energy Management	Commercial
CA	BayREN	Pay for Performance	Commercial
CA	SoCalREN	Metered Savings Program	Public Agency Program
CA	SCE	SCE Public Sector Performance-Based Retrofit High Opportunity Program	Commercial / Non-residential
CA	PG&E	Pay for Performance	Residential
CA	PG&E	NMEC meter-based savings platform	Commercial and Industrial
DC	DC Sustainable Energy Utility	Pay for Performance (P4P)	Commercial
IL	ComEd and Nicor Gas	Strategic Energy Management	Commercial
MA	National Grid	Pay for Performance for Monitoring-Based Commissioning and Retro-Commissioning	Commercial
MI	DTE Energy	Strategic Energy Management	Commercial
NJ	State of NJ's Clean Energy Program	Pay for Performance Existing Buildings[15]	Small Commercial & Multifamily
NY	NYSERSDA, Con Ed	Business Energy Pro - P4P Pilot	Small Commercial
OR	Energy Trust of Oregon	Pay for Performance Pilot	Residential
OR	Energy Trust of Oregon	Strategic Energy Management	Commercial & Industrial
RI	National Grid	Pay for Performance for Monitoring-Based Commissioning and Retro-Commissioning	Commercial
VT	Efficiency Vermont	Deep Retrofit	Commercial
VT	Efficiency Vermont	Continuous Energy Improvement (Strategic Energy Management and Commissioning Existing Buildings)	Commercial & Industrial
WA	Seattle City Light	Deep Retrofit Pay for Performance	Commercial
WA, OR, ID, MT	Bonneville Power Administration, Idaho Power, PacifiCorp, Puget Sound Energy	Strategic Energy Management	Commercial & Industrial

[15] 'Hybrid' program uses simulation models and meter-based analyses.

Table 3: Percent of Utility Meters using Advanced Metering Infrastructure (AMI) in Selected States (DOE 2018)

States with AM&V Legislation or Program	% AMI Meters
DC	99%
ME	92%
IL	90%
MI	89%
CA	83%
OR	81%
VT	81%
AZ	79%
ID	71%
MO	36%
AR	34%
WA	27%
VA	26%
MT	23%
CT	17%
NM	12%
MA	5%
NY	3%
RI	0%
All States	***56%***

PROGRAM APPROACHES

Behind the differences in program design, two general approaches are being taken:

1. **Focus on projects.** These programs strive to accurately determine savings for individual projects, typically at commercial and industrial sites. The rigorously determined site-level savings, including non-routine adjustments, roll-up to result in accurate program level savings. These programs typically follow established IPMVP M&V methods and support project-level monetary transactions (e.g., pay-for-performance).

2. **Aggregated methods.** These programs focus on a large number of similar buildings and prioritize estimating the total aggregated savings over verifying project-level results. Individual sites, typically single-family homes, may have poor baseline model statistical fit criteria (e.g., Cv(RMSE) ~50% to 100%), but focus on achieving an overall accuracy for the cohort based on the calculated precision or 'fractional savings uncertainty.'[16] Additional observations on these methods include:

   » Non-routine events (NREs) at project sites are usually ignored, and control groups are used to account for impacts from societal trends and non-routine events.

[16]Overall determination of uncertainty that relies upon Fractional Savings Uncertainty calculations using frequent energy data, e.g., hourly or daily, has found to be unreliable and to underestimate savings uncertainty.

- Non-routine adjustments (NRAs) to savings are not made at the site level, and their impacts are potentially left to the impact evaluation stage. It has been suggested that NREs cancel out over a large population of project sites, but this assumption is unfounded.
- Aggregated approaches might be effective for uniform residential or small commercial portfolios, although efficacy concerns arise when lower-accuracy site-level savings are combined.
- Large commercial and industrial sector applications, however, require savings adjustments be made for non-routine events affecting individual projects.
- Aggregated approaches may lower costs and potentially improve program-level savings estimates over deemed savings values typically used in residential programs.

Regardless of the program approach, the need for 'verification' elements such as operational verification and tracking site changes is not deferred by analyses of whole-building or industrial plant energy data. IPMVP methods require confirmation that the energy measures are installed and have the 'potential to perform' before savings can be claimed. Even when IPMVP adherent M&V is not required, program evaluators will need to validate these top-down energy analyses using project-specific details (this is typically performed on a statistically representative sample from the population of program participants).

SOFTWARE DEVELOPMENT

Advanced M&V functionality can be delivered via custom code, open-source/free code, free software tools, or as a component of proprietary energy management and information systems (EMIS) software. While no approach can be considered "fully automated," these various offerings can automate many steps of the process, enabling streamlined and labor-efficient application of advanced M&V.

The development of open-source tools and methods has been slower than proprietary EMIS software but has made substantial progress in the last two years. The work on these public-sector tools is ongoing and rapid and has resulted in the availability of a suite of 'free' tools and software methods. The five newest such 'tools', including 'NMECR', released in November 2019, are described in the following section and characterized in the Appendix. A 2017 Berkeley Lab paper[17] categorized the M&V features of 16 tools, 14 of which are proprietary software. The 2017 study and this review characterize these tools by market sector, the model type used, frequency of energy data used, level of user adjustments, level of statistical reporting, display of model equations, among other features.

These industry advances in R&D, regulatory efforts, program approaches, and software tool development are driving the need to return to IPMVP guidance, review principles and clarify definitions, clearly articulate best practice, and anticipate future applications.

[17] See Granderson & Fernandes The State of Advanced Measurement and Verification Technology and Industry Application, 2017.

3. Developments in Advanced M&V Best Practices

The technical state of the art for advanced M&V includes consideration of the variety of model types in use, issues with savings uncertainty, market options for advanced M&V tool testing, and considerations for non-routine events and related adjustments. Each is summarized below.

Variety of Models

The savings calculated by advanced M&V tools are based on the type of empirical model used, the interval of energy data used, independent variables included, and the specific technical adjustments made in applying the tool. The savings calculated for a given site, and the uncertainty in the estimates, will vary depending on the tool used and the approach taken by the practitioner. The extent of the difference in savings results from different methods is not well known, but variances are thought to be low[18].

Although tools[19] vary substantially, they are generally based on two model types – change-point and time-of-week and temperature (TOWT). These models are based on linear regressions of energy use to outdoor air temperature and are popular with practitioners as they have proven effective, are intuitive, and limit overall predictive bias[20].

1. **Change-point models**. Originated by ASHRAE Research Project 1050[21] in 2002, change-point models are piece-wise linear models of energy use for segments of outdoor temperatures (1-parameter or average, 2-parameter, or linear, up to 6-parameter), shown in Figure 2, and provide clear system-level performance indicators.

   The number of 'parameters' needed varies by a building's or industrial plant's individual load-shape(s). The simplest model (with the least number of 'parameters') that fits the data should be used. If a model with too many 'parameters' is used with too few data points, 'overfitting' may occur and result in a biased model.

2. **Time of Week and Temperature (TOWT) models.** Developed by Berkeley Lab, TOWT models use hourly data to create models for high and low use hours and use an indicator variable for each hour of the week. It does not use change-point models but creates a series of piece-wise linear and continuous temperature relationships using temperature bins. The method does not inherently account for holidays or other operational periods (e.g., Holidays). Software developers have modified TOWT, so there are several distinct versions in current use. In this paper, 'TOWT' always refers to the version published by Berkeley Lab, whereas 'TOWT_OpenEE', 'TOWT_UT3' and 'TOWT_NMECR' refer to various modifications.

---

[18] Authors compared savings estimates using ECAM and RMV2.0 for 18 buildings from one pilot program. The differences for site-level savings estimates were less than 2.5% in all cases.
[19] Proprietary software tools are not included since detailed information is not available on their methods and algorithms.
[20] Linear regressions using OLS methods will limit bias except when collinearity exists.
[21] The Inverse Modeling Toolkit was developed using monthly data *ASHRAE 1050-RP,* Development of a Toolkit for Calculating Linear, Change-point Linear and Multiple-Linear Inverse Building Energy Analysis Models, 2002.

Machine learning methods are also being developed, with artificial neural network being the most widely used method for building energy modeling. Another well-known machine learning method is the support vector machine (e.g., Gradient Boosting Machine or GBM[22] in Berkeley Lab's RMV2.0 tool, described below). With the increased interest in machine learning models, more such models will likely be released in the future.

Modeling approaches vary by project, but selecting the most accurate model typically requires evaluating multiple model forms to determine the best option. The selection should be based on both statistical criteria and confirmation of expected data relationships. The relationships between outdoor air temperature and heating and cooling loads in buildings are fundamentally linear, although temperature responses vary by building and operating mode. This tie with the known physics of buildings contributes to the industry's proclivity towards using these linear models; other mathematical relationships can exist if energy loads are driven by other factors (e.g., production processes with large variable-speed motors). Although these popular model forms have proven effective for most buildings, a "one-model-fits-all" approach is not best-practice nor adherent to IPMVP principles.

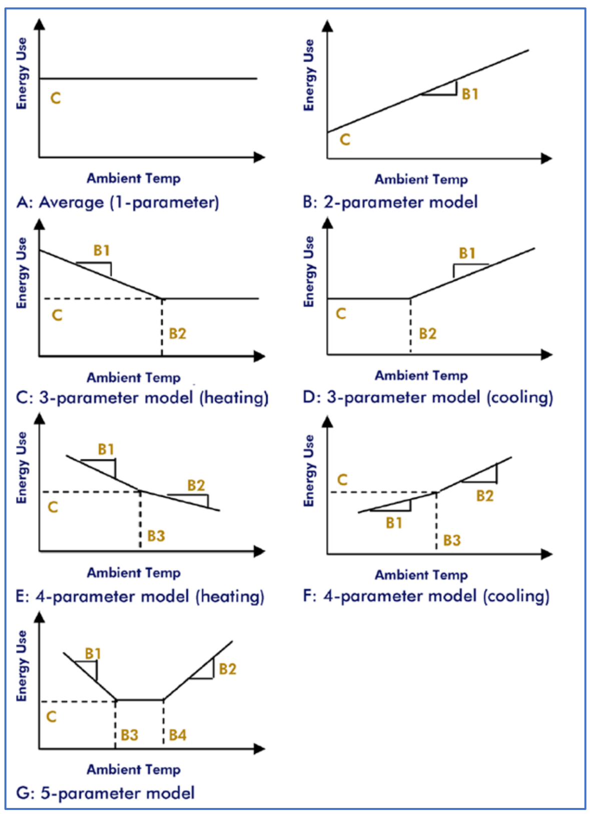

Figure 2: Change-point Models (CP-1 to CP-5) from CCC M&V Guide (CP-6 not shown)[23]

[22] More details are available at: https://github.com/LBNL-ETA/RMV2.0
[23] From the California Commissioning Collaborative

Free AM&V Tools

A detailed look at the ins-and-outs of five freely available tools shows how advanced M&V tools have been evolving over time, adding features to cover gaps and improve the application of the models. These tools have varying degrees of learning curves, depending on the practitioner and the tool. Although some include a user-interface and are relatively straightforward, others require skills in executing software code. Regardless of the tool used, all require an understanding of the underlying statistics and options for evaluating and improving results.

Although some of the model development features are automated[24], most require several judgment-points when developing a model. Decisions may include the type of model, data-increment (e.g., hourly, daily, monthly), and the number of day-types needed to represent the building's load profiles.

Each of the five tools examined (ECAM, RMV2.0, OpenEE Meter, UT3 M&V Module, and NMECR) includes nuances and modifications which are fundamental to their efficacy. Described below and detailed in Table 7 (see Appendix), all of these tools are free, and most are open-source.

ECAM. Currently available through SBW Consulting[25], this open-source tool is appropriate for M&V of commercial projects. It is accessed via an Excel add-in, which includes a user-interface. The tool accepts 15-minute utility data to create change-point models based on hourly, daily or monthly data. ECAM calculates both avoided energy consumption and normalized savings.

ECAM recommends day-types and develops load shapes to confirm them, accepts annual holiday schedules, allows custom day-types, defined occupancy periods and start-up and shut-down phases. Individual change-point models are developed for each day-type and occupancy mode (e.g., Weekdays-Occ, Weekdays-Unocc, etc.), and then combined into a single model.

RMV2.0. Developed by Berkeley Lab, this open-source tool is appropriate for M&V of residential and commercial projects. It is accessed via R-Studio and includes a user-interface accessible via a web browser. The tool requires pre-processed utility data to create TOWT or GBM models based on hourly data. RMV2.0 calculates avoided energy consumption.

RMV2.0 implements the original TOWT model, which includes weighting adjustments intended for demand response models. The Gradient Boost Machine (GBM)[26] modeling option is also included in the tool.

OpenEE Meter. Developed by OpenEE/Recurve, this open-source tool is appropriate for EM&V of residential programs. It is accessed via Jupyter Notebook and does not include a user-interface. The tool accepts 15-minute utility data to create either change-point models using custom degree days based on daily or monthly data, or modified TOWT_OpenEE models based on hourly data. OpenEE meter calculates avoided energy consumption.

---

[24] Automated model development routines may lack the acuity of custom analyses due to simplifying assumptions.
[25] Developed by Bill Koran across multiple organizations.
[26] More information on the GBM model is available through Berkeley Lab and GitHub.

OpenEE Meter implements the 'CalTRACK Methods' via Python code, including a modified TOWT approach (TOWT_OpenEE) that uses hourly data to create 12 weighted 'monthly' TOWT models, rather than the typical annual modeling approach.

UT3 M&V Module. Added to the PG&E UT3 (Universal Translator) by Quantum Energy Services & Technologies, Inc. (QuEST), this free tool (not open-source) is appropriate for M&V of commercial projects. It is accessed via the UT3 tool's user-interface. The tool accepts 15-minute utility data to create either change-point models based on daily data, or modified TOWT_UT3 models based on hourly or daily data. UT3 M&V Module calculates both avoided energy consumption and normalized savings.

The UT3 M&V Module is part of the UT3 data analysis tool and allows for the filtering of data based on time-of-day or week schedules, or different building operation modes (e.g., Holidays, Summer school). The change-point and modified TOWT (TOWT_UT3) algorithms can be used with sub-hourly, hourly, or daily data and may be modified to produce time-of week only (TOW_UT3) or temperature-only models. Models created for each schedule are combined using a 'Model Assembler'.

NMECR: Developed by kW Engineering, this open-source tool is appropriate for M&V of commercial projects. It is accessed via R-Studio and does not include a user-interface. The tool accepts 15-minute utility data and allows the creation of change-point models based on daily or monthly data, or a modified TOWT model (TOWT_NMECR) using hourly or daily data. NMECR calculates both avoided energy consumption and normalized savings.

Released in November 2019, NMECR provides scripts coded in R to create energy models. NMECR uses indicator variables to describe different operation modes in buildings. The TOWT/TOW_NMECR models allow for the inclusion of additional day-types (e.g., holidays, summer-school), and the weighting factor that was included in RMV2.0 for demand response analysis can be disabled.

These free and mostly open-source AM&V tools have a variety of features that help complete meter-based savings analyses. Key features that vary between tools include:

- The user interfaces to facilitate analyses and level of user guidance documents;
- Data management and visualization tools;
- Automated analyses of load shapes;
- Model types included, the complexity of models and the variables/inputs included;
- Expertise, level judgement, and effort required;
- Level of automation possible and level of sophistication in automated modeling strategies;
- Statistical reporting and ease of comparing models;
- Detail provided on the calculations themselves (the equations);

» Calculation of both avoided energy and normalized savings;

» Automated retrieval of ambient temperature data for use in adjustments;

» Savings tracking capabilities;

» Identification of periods with unexpected performance and potential non-routine events;

» The tracking of non-routine events.

Savings Uncertainty

The only error that is typically quantified in meter-based M&V methods is the error from the empirical energy model(s), hence the intense focus placed upon proper model assessment. Measurement errors are not usually applied to meter-based methods that use revenue-grade utility meters, and AMI data is considered free of measurement errors once validated by the utility; sampling error would only apply in an evaluation study.

In reality, of course, other sources of error exist and include:

» Missing or irregular energy data;

» Flaws in independent variable data such as the source for local weather data;

» Methods used for addressing missing/anomalous data;

» Extrapolations beyond model limits;

» Model misspecification (e.g., specifying a 3-parameter versus 5-parameter change-point model, omitting an important production variable, overfitting from too little data, or leaving an unexplained residual trend);

» Dates selected for baseline and performance periods; and non-routine events and any subsequent adjustments.

One of the key benefits of meter-based methods over other M&V methods has been the ability to compute the uncertainty of the savings estimates based on the statistics from the energy model(s), often using the popular error metric "Fractional Savings Uncertainty" (FSU), or the relative precision of the model[27]. FSU quantifies savings uncertainty for models that are essentially valid (i.e., models that are not afflicted by the issues noted in the previous paragraph). Unfortunately, current FSU calculations are not reliable when using hourly or daily energy use data and tend to underestimate uncertainty[28].

---

[27] At a specific confidence interval, the precision can be determined from the standard error of the regression and provide a range of savings, (e.g., at 90% confidence avoided gas use was between 49,000 to 56,400 therms). See IPMVP Application Guide on Uncertainty Assessment for more details.

[28] Autocorrelation issues arise with use of frequent-interval energy data result underestimating the uncertainty in savings calculated for the reporting period. See Evaluation of Methods to Assess the Uncertainty in Estimated Energy Savings, 2019.

ASHRAE Guideline 14 published the original FSU calculation in 2002 based on monthly energy data and includes simplifications to account for model errors. With higher frequency energy use data, the model residuals are ‘auto-correlated’ confounding the estimation of the savings uncertainty. Correction factors to fractional savings uncertainty for hourly and daily data have been developed but are not sufficient, and uncertainty is still underestimated, especially for hourly models (additional testing is needed). Work is ongoing to develop other methods to estimate error for hourly and daily models.

Model Accuracy

This gap in uncertainty metrics can be mitigated by ensuring the models are as accurate as possible. Maximizing the accuracy of the models, without overfitting, will minimize uncertainty in savings. Often, a more accurate model is technically achievable, but it has not been optimized through additional analyses and customization. Evaluating additional model types, data increments, and independent variables for model improvements is best-practice but can add time. Specific consideration should be given to components affecting uncertainty – coefficient of variation of root mean squared error (Cv(RMSE)), number of points used, level of savings, and degree of autocorrelation.

Assessing the goodness-of-fit criteria for a baseline energy model alongside the percentage reduction in energy use by the project can indicate if subsequent savings estimates are likely to be ‘lost in the noise’. Ensuring that savings will be measurable seasonally, however, is often overlooked. Evaluating the overall percent savings achieved annually by fuel is not sufficient if multiple measures are implemented whose savings occur at different times (e.g., reduced electric heating loads for outside air and new chillers). At a minimum, the type of energy reduction measures planned should be considered.

Uncertainty in Avoided Energy vs. Normalized Savings

Savings uncertainty is primarily driven by the baseline model’s goodness of fit, but it also affected by the calculation approach: *avoided energy consumption* or *normalized savings*. Savings error in avoided energy consumption is lower than the error in normalized savings.

Avoided energy consumption is the simplest, and most intuitive approach, and is referenced in ASHRAE G14 as “actual savings.” As shown in Figure 1, a weather-adjusted baseline paints a picture of what the energy use would have been during the reporting period without the energy project, current energy use is subtracted from the adjusted baseline energy, and the estimated savings align with reduced energy costs. In this case, the quantifiable uncertainty is the prediction uncertainty incurred when the baseline model is adjusted to reporting period conditions to estimate the weather-adjusted baseline energy.

Normalized Savings contain increased modeling error over avoided energy consumption because an additional model for the reporting period energy is developed and adjusted along with the baseline model (see IPMVP Core Concepts equation 3 and 7 in the Appendix). Assuming non-routine adjustments are not required, the error in the savings includes the errors in both the reporting period model and the baseline model, increasing saving uncertainty.

Normalized savings are intended to represent what the savings would have been during an average or ‘normal’ year, often used for long-term planning, comparing year-to-year savings for buildings within a climate zone, or to minimize financial risk. Sometimes, avoided energy can be much different than normalized savings, especially during an extreme year, which can cause confusion. Questions regarding the reliability of long-term past weather patterns (e.g., TMY3 weather data[29]) to represent a contemporary "average year" has made using avoided energy calculations preferred in some cases.

COMPARATIVE AM&V TOOL TESTING

Comparing the performance of advanced M&V tools is challenging since most of the tools are proprietary, and data from multiple sites is not readily available. To address this concern, Berkeley Lab developed a method[30] to compare advanced M&V tools based on their predictive capabilities. Rather than assessing baseline model results, data from stable buildings without energy projects or known NREs is used to assess the models' actual performance.

Launched as EVO's *Advanced M&V Testing Portal*, registrants develop models using one year of energy and ambient temperature data for a portfolio of commercial buildings across multiple regions. The models are used with ambient temperature data from the subsequent year to generate 'adjusted baseline' predictions and compare to actual usage. Calculated metrics characterize the errors between predictions and actual use for each tool; individual results are published and posted graphically for comparison.

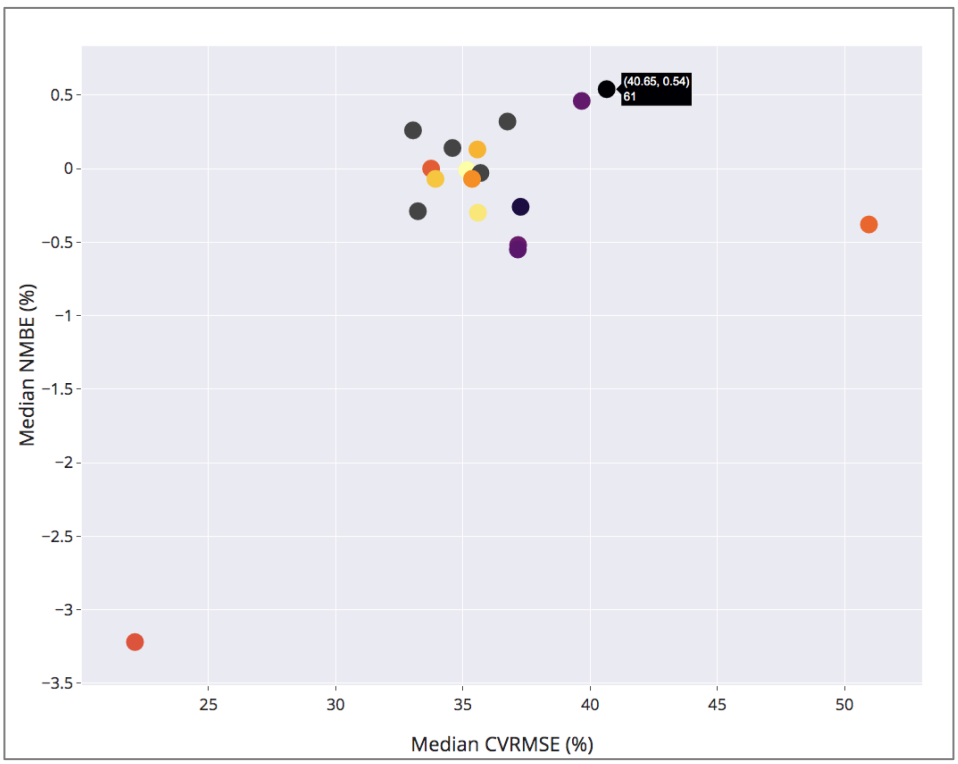

Figure 3: EVO's Advanced M&V Tool Testing Portal, Results in January 2020

[29]Typical meteorological year (TMY) data development is unique and does not include weather extremes as discussed in An Assessment of Typical Weather Year Data Impacts vs. Multi-year Weather Data on Net-Zero Energy Simulations

[30] See Lawrence Berkeley Lab's paper Assessment of Automated Measurement and Verification Methods detailing the M&V tool testing methods which use out-of-sample testing.

As shown in Figure 3, the two key statistics used are the coefficient of variation of the root mean square error (Cv(RMSE)) and the net mean bias error (NMBE). Results compare the median Cv(RMSE) vs NMBE from testing each tool, each depicted by a point (e.g., Tool 61 has a median Cv(RMSE) of 40.65% and a corresponding NMBE of 0.54%).

The median values for the population are displayed in the portal graphics, and the 25th and 75th quartile values are published. The overall results are based on a wide range of commercial buildings and provide a comparative snapshot. A considerable range of variation is seen in these median values across the different tools, with greater variations in Cv(RMSE) than in the median NMBE.

The EVO Tool Testing Portal was designed by Berkeley Lab to compare the predictive accuracy of any tool or model, independent of whether it is open source or proprietary. This active platform is in the public domain and could be leveraged by utilities to competitively screen tools. In future, conducting a custom test could indicate the most effective tool for a given region and market sector.

Results from new methods are continuing to be posted in the portal. For example, results from the winner of the Great Energy Predictor III[31], a short-term modeling contest that concluded in December 2019, will be added to the Portal. This is the third such contest following the original in 1993[32]; Sponsored by Kaggle and ASHRAE, over 3,600 teams competed for a $25,000 cash award for developing the best predictive energy model. The dataset included hourly meter readings and weather data from over one-thousand buildings at several different sites around the world. Kaggle selected root mean squared log error (RMSLE)[33] as the single metric in evaluating the results, which is not typically used within the energy industry.

NON-ROUTINE EVENTS (NREs)

Non-routine events (NREs) are changes in energy use due to changes in site characteristics or to "static factors" which are not used in the empirical energy models or related to the energy project. Typical changes to *static factors* at a site include significant changes in the number of occupants and occupancy schedules, significant operational changes, equipment shut-downs or removal, maintenance periods, modifications to tenant spaces, the addition of solar panels, or even changes in facility size. These unexpected changes in energy use are the most significant complication faced by meter-based M&V approaches. The IPMVP Core Concepts 2016 defines static factors as follow:

> *Those characteristics of a facility which affect Energy Consumption and Demand, within the defined Measurement Boundary, that are not expected to change, and were therefore not included as independent variables. If they change, Non-routine Adjustments need to be calculated to account for these changes. Note: Those characteristics may include fixed, environmental, operational and maintenance characteristics.*[34]

---

[31] See Kaggle/ASHRAE's Great Energy Modeling Shoot-out III
[32] Jeff Haberl of Texas A&M led these earlier 'Shootout' efforts.
[33] See https://www.kaggle.com/c/ashrae-energy-prediction/discussion/113064
[34] IPMVP Core Concepts 2016.

The two primary methods of identifying NREs are 1) analytical and 2) field data. Advanced M&V applications may use both approaches.

Using analytical approaches to identify NREs is an active area of investigation. Current strategies being explored include specific data visualization methods as well as analytical approaches such as the analyses of model residuals and the use of specific dissimilarity indices[35] to flag irregularities. Note the last two NRE identification approaches require a model of the reporting period energy use to establish post-project norms, which cannot be made until several seasons of data is available.

Notifications from the site staff of changes remain the easiest approach to identify unplanned changes at a project site. However, some resources to identify NREs are included in advanced M&V tools. Berkeley Lab has explored statistical time series analytics to identify non-routine events[36] and has included some of these techniques into the RM&V2.0 tool. ECAM includes assessments of long-term trends in energy use and flags changes that are statistically significant. The use of more granular data allows subtle changes to be detected but is limited by the predictive quality of the reporting period model.

Once a potential NRE has been identified, what action is warranted? That depends on when the 'event' happened, its duration, and level of impact on energy use. A phone call to the site to inquire about operations can save time deliberating its source. Minor, short-term anomalies are less concerning than significant lasting changes. Similarly, NREs that occur during the baseline period are more readily addressed (if identified during baseline model development) than changes occurring during the performance period. If a potential non-routine event has a '*significant-enough'* impact, a non-routine adjustment may be warranted. What warrants '*significant-enough'* will vary by project.

The significance of an NRE on energy savings can be gauged in many ways. Making a direct estimate from predicted values via a performance-period model is the most straightforward approach, but an accurate performance period model generally requires 3 to 9 months of data[37] and additional analyses. However, once developed the difference between the metered energy use and the performance-period model should be roughly equivalent to the impact of the event.

If the impact on the savings warrants action, the root cause of the change should be identified. The 'event' must be unrelated to the project being measured to justify an adjustment, which must be evaluated on a case-by-case basis.

NON-ROUTINE ADJUSTMENTS (NRAS)

Although non-routine adjustment strategies are outlined in IPMVP Option C guidance, this remains a big open issue. Submetering or custom engineering calculations have always been required for NRAs, but explicit examples are somewhat limited. With the wide-spread adoption of advanced meter-based methods, there are impressive new opportunities.

---

[35] See *Potential Analytics for Non-Routine Adjustments* by SBW Consulting for Bonneville Power Administration, 2018.
[36] See Statistical Change Detection of Building Energy Consumption: Application to Savings Estimation, 2019.
[37] See Automated M&V: Performance of Public Domain Whole-Building Electric Baseline Models, 2015.

The industry is very interested in more detailed guidance as to how to characterize and manage specific NREs and NRAs: How do you identify common NREs using AM&V? When do you need to make non-routine adjustments to the baseline or reporting period energy usage, and how is that done? What are the calculation options when using Option C methods? Can the additional uncertainty introduced by non-routine adjustments to savings be quantified? These are significant open issues and are a current focus of research. The upcoming IPMVP *Application Guide on AM&V Approaches* will address these issues.

When using meter-based methods, key pieces of information needed to manage NRAs are the date the change occurred, if the change is ongoing, or the date it ended. The form of a non-routine adjustment (NRA) will vary from simple to complex and may require sub-metered data and engineering calculations. The impacts of a temporary event are easier to manage as they may be quantified, and the NRA applied as a single value, whereas ongoing changes require further analyses to incorporate them into the ongoing savings calculations.

4. KEY OPEN ISSUES

Advanced M&V users are facing several key challenges inherent to meter-based approaches:

1. **Savings uncertainty.** Currently, error metrics are not reliable from models based on hourly data. 'Fractional Savings Uncertainty' is heavily relied upon but underestimates the level of error. This is especially concerning for aggregated methods relying on this calculation to validate portfolio level savings.

2. **Aggregated methods.** The efficacy of these methods is not yet established, and several issues remain unresolved. Open topics include: proving the reliability of fractional saving uncertainty calculations; establishing appropriate thresholds for baseline model goodness of fit that will help ensure large errors will balance-out across a portfolio; demonstrating effective methods within the portfolio to manage non-routine events and their impacts; incorporating methods to calculate weather normalized savings, especially in timeframes less than a year; and ensuring the savings for all participants in the portfolio can in fact be seen at the site level.

3. **"Bad buildings"**. Not all buildings and projects are suited for meter-based approaches because the energy use cannot be accurately predicted. Results from past studies show a portion[38] of commercial buildings do not obtain acceptable goodness of fit for AM&V. The pervasiveness of these "bad buildings" varies by customer sector, climate, model type, modeling strategies applied, and the model acceptance criteria used. The use of improved modeling strategies and targeted testing will be instrumental in making progress in this area.

4. **Site-level changes**. Unexpected energy changes at a site can be flagged as potential non-routine events (NREs), but determining if the site level changes warrant a non-routine adjustment (NRA) to the savings estimate, and how to best calculate that adjustment, is an open question. No automated silver bullet exists, and project level details and professional judgment are required. Research into NREs has brought to light the need to expand the consideration of NREs to events that:

   » Increase or decrease calculated savings.

   » Occur during the reporting or baseline period.

   Examples of the most common non-routine events and direction on handling these events will be helpful in mitigating these unavoidable changes.

[38] Berkeley Lab has published and unpublished screening results from several utility studies; EM&V studies include similar findings (*ComEd Commercial and Industrial (C&I) Custom SMART Screening Pilot*, Navigant Consulting June 2019).

5. IPMVP and Advanced M&V

IPMVP is an M&V protocol recognized around the world, translated into nine languages, and used in at least as many countries. IPMVP's mission is to ensure that savings and impacts from energy efficiency and sustainability projects are accurately measured and verified. The protocol provides common principles, terms, and methods for M&V of energy, demand and water savings.

IPMVP establishes the basis of meter-based M&V methods in 'Option C: Whole Facility,' last updated in 2016 it includes discussion on advanced meter-based methods, including system-level sub-metering (e.g., chiller plant). The IPMVP Option C method and guidance remain relevant.

EVO has worked for more than 20 years to create alignment in M&V by establishing a common vocabulary through the IPMVP and related CMVP trainings. Key terms have been long-standing and are explicitly defined by IPMVP to avoid confusion, allow transparency, and promote best practices. Some of these terms may not be ideal (e.g., static factors) and limiting (e.g., baseline adjustments), but changes are best made cautiously and with consensus. The established IPMVP language will prove invaluable as the industry moves towards more complex scenarios to account for energy efficiency, distributed energy, and demand response impacts from metered energy data.

Despite the ubiquitous references to IPMVP found in most utility program plans and EM&V reports, industry publications are seeming to increase in inconsistencies. Table 4 below includes some key IPMVP terms and their 'twin' expressions, which could benefit from being trued-up in some applications.

Table 4: Industry Vocabulary Check-up List

IPMVP's Terminology	Also Known As
Baseline model	*Forecast model* *Training period model* *Adjustment model*
Adjusted baseline model Adjusted baseline energy	*Counterfactual model* *Baseline projection* *Projected Baseline* *Predicted baseline*
Baseline period	*Pre-period* *Pre-retrofit*
Reporting period	*Post-period* *Performance period* *Achievement period*
Measurement & Verification (M&V)	*Monitoring & verification* *Monitoring targeting & reporting (MT&R)*
Advanced M&V (AM&V)	*Automated M&V* *M&V 2.0*
Avoided energy consumption	*Avoided savings*
Meter-based energy savings	*Metered savings* *Measured savings*

Not using the same terms can confuse even the most experienced M&V practitioner. Prevalence of some expressions may be spilling over from other industries, EM&V studies, and from specific program applications (e.g., SEM). There are many variations of some terms, and while it is not necessarily critical to avoid other terms, some degree of standardization of terminology is desirable.

IPMVP Principles

In addition to defining terminology, M&V options, and process requirements, the IPMVP establishes key principles that provide a framework for adherence to the M&V process. Here is a quick review to following the six best-practice M&V principles when working with an advanced M&V approach: *Complete, Consistent, Transparent, Relevant, Accurate, Conservative.*

Complete, Consistent, Transparent:

Empirical models used in advanced M&V require unique considerations and specific information to ensure advanced M&V approaches meet the thresholds of *Complete, Consistent, and Transparent*. The essence of these can be summarized as requiring full disclosure of data, techniques, and analyses.

"Full disclosure" requires sufficient detail for another practitioner to understand and evaluate the results fully and be able to replicate them if needed. Necessary information includes the measures implemented, key dates, the expected level of savings, details on the meters included, software & version of tools used, model selection process, description of the model type(s) used, explanatory variables considered and selected, data sources, as well as details of any data cleaning, other judgments and assumptions.

Relevant, Accurate, Conservative:

In addition to the full disclosure called for above, following the IPMVP principles of *Relevant, Accurate, and Conservative* ensure viable energy savings estimates. For advanced M&V, fulfilling these principles inherently demands best-practice modeling techniques be followed.

The technical accuracy of the mathematical models should be validated in consideration of the expected level of savings. Rigorous regression analysis procedures should be followed to ensure all relevant variables and factors are considered, the models are validated as free from critical errors, and current data is utilized. In general, models should be optimized so the calculated savings will contain fewer errors and be more accurate. In some cases, this may mean contacting the facility to confirm engineering details relevant to the savings calculation.

Providing key details, including the actions taken to ensure the accuracy of the model(s), can provide the needed technical insights into both proprietary and open-source software models to ensure the support of *Complete, Consistent, Transparent, Relevant, Accurate, Conservative*.

Although there is a preference for open-source methods, fully open-source AM&V solutions are limited but expanding. The current snapshot of the advanced M&V tool market discussed earlier shows only four of the

19 tools examined are open-source. Many of the private EMIS software tools have been long established, and often use the same M&V models as open-source methods. Open-source methods are not inherently better or necessarily based on best-practices, but options are expanding.

Both proprietary and open-source advanced M&V software developers are limited by business realities and driven by market opportunity. Rather than mandating the adoption of open-source methods, conducting comparative testing of methods may be more effective in driving advancements. If a project sponsor wishes to have third-party oversight or an 'under the hood' understanding of proprietary software, a non-disclosure agreement should be sufficient to facilitate the needed technical engagement.

FUTURE DIRECTIONS FOR IPMVP

Industry context is evolving rapidly with implications for future applications of advanced M&V methods. Beyond the present drivers for reporting accurate time-of-use energy savings, the brisk addition of demand-response (DR) efforts and new distributed generation (DG) resources (e.g., electric vehicles) will complicate known methods. Meter-based energy use is core to all of these efforts and will require the coordination of multiple baselines. Inevitably the need for 'integrated M&V' to delineate savings from EE, DR, and DG will require M&V approaches to evolve.

Option C methods using monthly data continue to be popular for natural gas and other fuels. Fuel use data is generally limited in granularity and frequency of collection but is improving over time. Strategies may evolve as natural gas metering advances, but the direction will likely be influenced by emissions accounting and efforts to de-carbonize buildings.

IPMVP will release an *Application Guide on Advanced M&V Approaches* in Spring 2020, and updates to the IPMVP Core Concepts will follow in 2021. The *Application Guide* will provide more specific guidance on issues related to advanced methods for energy efficiency applications. Direction on identifying and characterizing non-routine changes in energy use, quantifying their impacts, making necessary non-routine adjustments in savings, and managing savings uncertainty will be included. Since these issues also affect monthly data approaches, both will be covered.

USE MODEL ACCEPTANCE CRITERIA TO MAXIMIZE SAVINGS

Fundamentally, energy-use models must be sufficiently accurate for a given project for all of the energy savings to be discernable and tallied. If projects or programs use lenient accuracy thresholds, they miss some of the benefits of AM&V and can result in lower verified savings, effectively leaving savings on the table. Although perhaps not intuitive, using stringent acceptance criteria for baseline models can increase program-level savings in several ways:

- A more accurate model will measure lower levels of savings, e.g., ~5%, thereby capturing more of the savings achieved at a customer's site.
- More accurate models result in lower savings uncertainty, thereby allowing a larger portion of the estimated savings to be recognized by evaluators, increasing savings attributed to the program.

» Using better models can expand program participation by including customers with lower levels of whole building savings to participate.

Acceptance criteria may be customized but should meet or exceed industry guidelines such as those shown in Table 5, which include ASHRAE's Guideline 14, DOE's Superior Energy Protocol, and BPA's Regression for M&V Reference Guide. As an overarching protocol, IPMVP CORE 2016[39] does not provide rules of thumb for model fitness.

Table 5: Baseline Model Acceptance Criteria - Industry Guidance

Industry Guideline	Model Fit Criteria			
	CV(RMSE)	R^2	NMBE	Other Requirements[40]
ASHRAE G14 - Whole Building Performance Path	Varies. See FSU	None	< 0.005%	✓ Fractional Savings Uncertainty (FSU) < 50% annual savings at 68% confidence level[41] Note: FSU ~ f(Cv(RMSE), % savings, # baseline & reporting period points)
ASHRAE G14 - Whole Building Prescriptive Path	<25%	None	< 0.005%	✓ Expected savings > 10% ✓ Daily data is minimum interval ✓ Baseline model uncertainty, depends on length of reporting period: Energy < 20 – 30%, Demand < 30 – 40%
Superior Energy Performance (SEP) M&V Protocol	None	> 0.50	None	✓ F-test for overall model fit must have a p-value < 0.1 (i.e., the overall fit of the model is greater than the 10% significance level). ✓ All included relevant variables in the model shall have a p-value of less than 0.20. ✓ At least one of the relevant variables in the model shall have a p-value of less than 0.10.
BPA Regression for M&V: Reference Guide	A low value is desirable (often interpreted as 10% or 15%)	> 0.75*	< 0.005%	✓ p-value for independent variables <0.10 to 0.01 ✓ t-statistic for independent variables >1.96 (95% confidence level) ✓ F-statistic (used for entire model instead of individual variables; Larger the better.) ✓ Adjusted R-squared for multiple regression models. ✓ A low R^2 does not indicate a poor model; Professional judgment should be applied (*) This is a rule of thumb value

[39] Previous versions of IPMVP, including IPMVP 2012, included suggested statistical metrics.
[40] The intermediate statistics of F-statistic and p-value assume linear models. NMBE and Cv(RMSE) of the models apply other model forms.
[41] This is a modest FSU threshold; Confidence levels of 80% to 90% are typical.

Ideally, we would like to evaluate the effectiveness of a baseline model in terms of fractional savings uncertainty. However, since FSU estimates are not always reliable, practitioners currently rely on several related statistical metrics, which provide insights into different aspects of model accuracy. Relying on just one metric is usually not sufficient to fully understand the weakness and strengths of a specific baseline model. Note the 'other' requirements of these guidelines clarify the need for modeling expertise, as statistics must be evaluated during model development.

The primary metrics used, described below, are the coefficient of determination, or R-squared (R^2), the coefficient of variation of the root mean squared error (CV(RMSE)), and the normalized mean bias error (NMBE); mean average percentage error (MAPE) is also explained. In a real-world modeling context, these metrics are all calculated for 'in-sample' errors (i.e., actual and predicted values all refer to the data used in the baseline model). These three metrics provide complementary views of model performance for M&V applications.

- **R^2 (coefficient of determination[42]).** R^2, or Adjusted R^2 for more than one variable, range from 0 to 1 (higher is better) and should be used as an initial check of model quality, not as a pass-fail metric. Low R^2 can indicate missing variables, low variability in energy use (i.e., flat slope), or improper model form.

- **Cv(RMSE) (coefficient of variation of root mean squared error)**. Cv(RMSE) is a key metric for model evaluation and an indicator of random error. Calculated as the RMSE divided by the average energy consumption, it quantifies the typical prediction error as a percentage (expressed as a percentage, lower is better). Cv(RMSE) reports the model's ability to predict the energy use and is the basis for fractional savings uncertainty calculations.

- **NMBE (net mean bias error)**. NMBE is a measurement of bias error and should be very close to zero; NMBE is the total difference between model-predicted energy use and actual metered energy use given as a percentage (ranges from 0 – 100%, 0 is the target). If the value of NMBE is positive, it means that the model's prediction is lower than the measured value; a negative NMBE means that the prediction is higher.

- **MAPE (mean average percentage error).** MAPE is a best-practice error metric that is not included in these guidelines but provides a closer examination for bias. Bias can occur in linear models due to 'overfitting' and is of particular concern in non-linear model forms. Calculating MAPE for each month will indicate if there is a seasonal bias that may not show up when evaluation the entire period using NMBE.

Most applications and programs meet the industry guidance thresholds shown above, but it is recommended that stringent baseline model criteria be used to mitigate the risk of potentially faulty error metrics such as the current estimates of fractional saving uncertainty.

[42] EVO M&V Focus October 2019 [Why R^2 Doesn't Matter]

CONCLUSIONS & NEXT STEPS

The last five years have seen advanced M&V shift from promising but limited techniques and offerings using newly available smart-meter data, to an array of strategies and tools ready for adoption in mainstream applications. Utilities, regulators, and M&V practitioners now have several new resources to support these methods. These include research findings, regulatory actions, utility pilot program examples, and open-source software tools.

Technical developments in modeling methods and software tools are improving the accuracy of energy models, including the introduction of new and updated open-source methods and the launch of EVO's tool-testing portal for objectively comparing advanced M&V tools. Technical guidance, pilot program case studies, and regulatory language examples can provide direction to those looking to incorporate advanced M&V into their portfolio of projects or programs.

It remains critical, however, to ensure that tools and methods are applied using a rigorous process, and not to simply trust that a 'good' M&V tool is guaranteed to give an accurate result. For large-scale deployments, comparative tool testing is recommended to select the most technically accurate solution for a given population. Although not required for IPMVP adherence, several open-source AM&V tools are now available, with several focused on commercial applications.

As work is still needed to determine accurate uncertainty methods with interval energy data, practitioners can manage risks by using strict model fit metrics and paying careful attention to model details such as ensuring expected savings will be clearly and consistently discernable year-round at every site.

The potential benefits of advanced M&V over other savings estimation methods include:

» Discerning savings quickly from metered energy data.

» Measuring lower levels of whole-building savings due to improved model accuracy over monthly models.

» Capturing comprehensive savings from multi-measure projects with significant interactive effects.

» Reporting time-disaggregated savings, e.g., hourly, by season, etc., for grid management and emissions reporting.

» Detecting anomalies in energy use or in the avoided energy savings which could be non-routine events.

» Documenting baselines with fully measured data rather than system-level calculations.

» Verifying site-level issues via telephone engineering reviews to supplement and confirm M&V model findings, potentially avoiding an onsite visit.

As with all savings estimation methods, it is essential to understand not only the benefits but also the limitations, which for advanced M&V include:

- Current methods to quantify savings uncertainty have been found to underestimate error in hourly and daily methods, including the ASHRAE Fractional Savings Uncertainty (FSU) formula with autocorrelation adjustment. The FSU calculations should not be relied upon, although they may still be informative. Despite this, methods using more granular data can produce reliable models.
- Aggregated methods have many open issues and still need to be validated.
- Challenges in identifying, validating, and adjusting for *non-routine events* affecting savings exist.
- Avoided energy consumption estimates are the most accurate since errors are only introduced from the baseline model and can vary substantially from normalized savings in extreme weather years.
- Low levels of whole-building savings may not always be measurable with meter-based methods, as a high level of model accuracy is required (other M&V options may be more suitable).
- Incompatibility of some buildings with a meter-based method due to inconsistent energy use profiles or incompatibility with the specified model type.

Although site level verification requirements may be reduced, they are not eliminated and can be more critical than when using other M&V approaches (e.g., retrofit isolation). Practitioners evaluating savings with advanced M&V require project-level details such as the date(s) of project installation, measure types installed, expected savings, and a point of contact at the site, which is an essential resource to evaluate non-routine events.

IPMVP's upcoming *Application Guide on Advanced M&V* will provide necessary guidance on navigating the nuances when executing advanced meter-based M&V methods, and IPMVP Core Concepts will be updated in 2021. The evolving market and industry context that includes EE, DR, and DG will require M&V approaches to continue to evolve and foreshadows the need for 'integrated M&V' to delineate savings at the most advanced project sites. These ongoing changes keep IPMVP relevant and underscore the need for a unified vocabulary to discuss increasingly complex measurement and verification applications.

APPENDIX

Equations for Avoided Energy Consumption ***AND*** *Normalized Savings*

IPMVP CORE CONCEPTS 2016 – EQUATION 3

Avoided Energy Consumption =		(Baseline Period Energy
	±	Routine Adjustments to Reporting Period Conditions
	±	Non-Routine Adjustments to Reporting Period Conditions)
	–	Reporting Period Energy

IPMVP CORE CONCEPTS 2016 – EQUATION 7

Normalized Savings =		(Baseline Period Energy
	±	Routine Adjustments to Fixed Conditions
	±	Non-Routine Adjustments to Fixed Conditions)
	–	(Reporting Period Energy
	±	Routine Adjustments to Fixed Conditions
	±	Non-Routine Adjustments to Fixed Conditions)

Table 6: Selected Regulatory Actions Driving Advanced Meter-Based M&V Approaches, 2010 to 2019

State	Regulator	Mandate	Ruling	Date
WA, ID, OR	NW Power and Conservation Council	6th Northwest Conservation and Electric Power Plan	Permitted the launch of industrial SEM programs, acknowledged use of interval data for improved savings estimates	2010
CA	CA State Senate	SB 350 - Clean Energy and Pollution Reduction Act	Increased state efficiency and renewable energy goals, modified demand forecast methods with emphasis on metered savings	2015
CA	CA State House	Assembly Bill No. 802	Authorized utility programs to use *normalized metered energy consumption (NMEC)* as a measure of energy saving to estimate savings, allowing existing conditions baselines	2015
CA	CA PUC	Ruling Regarding High Opportunity Energy Efficiency Programs and Projects (HOPPs)	Authorized selected utility programs to pilot metered savings using existing conditions baseline rather than code-baseline (NMEC)	2015
NY	NY Governor	Executive Order 166	Mandated use of advanced M&V techniques for emissions tracking	2015
CT	CT General Assembly	General Statutes - 16-245m(d) - 2019-2021 Conservation & Load Management Plan	The PUC shall add approaches documenting savings achieved to include measurement methods through metering, with appropriate adjustment for weather normalization and other factors	2015
OR	OR Governor	Executive Order 17-20: Accelerating Efficiency in Oregon's Built Environment to Reduce Greenhouse Gas Emissions and Address Climate Change	Mandated meter-based savings pilot programs, including pay-for-performance pilots for all customer sectors	2017
NY	NY PUC	Order Adopting Regulatory Policy Framework and Implementation Plan	Encouraged Pay-for-performance programs to use advanced M&V with existing conditions baseline; EM&V should use advanced M&V to lower costs	2017
MO	MO Dept Economic Development	N/A	Published guidebook for the adoption of M&V 2.0 and EM&V 2.0, including the development of Technical Reference Manual (TRM)	2017
VA	Virginia PUC	20VAC5-318-40 Rules Governing Utility DSM Program EM&V	Utilities to consider the use of "advanced measurement and verification" or "evaluation, measurement and verification 2.0" when appropriate and cost-effective	2018
CA	CA PUC	Rulebook for Custom Program and Projects Based on Normalized Metered Energy Consumption (NMEC) - V1 & V2 (relates to mandate in AB 802)	Normalized meter-based savings (NMEC) guidelines for CA projects and programs	2018 - 2019

Table 7: Details on Selected Advanced M&V Tools (an unabridged version of this assessment is available at Facility Energy Solutions.com)

Tool Overview						Savings Type		Energy Data Used				Details	
Tool	Model Type(s)	Variables & Inputs Used	User Interface	Level of User Adjustments	Equations of Model(s) Shown	Avoided Energy	Normalized Savings	Interval Data Accepted	Hourly	Daily	Monthly	Version #, Date	Notes
ECAM	Average	None	**Yes**	High	**Yes**	**Yes**	**Yes**	**Yes**	■	■	■	V6r5, 2018	[1] Coefficients are given, but equation form is published elsewhere for change-point models
	Linear	OA Temperature (or other independent variable); Daily & Annual Schedules							■	■	■		
	Change-point (3P to 6P)				**Yes [note 1]**				■	■	■		
RMV2.0	TOWT	OA Temperature	**Yes**	Low	No	**Yes**	No	No	■			V1, 2016	
	GBM	OA Temperature; Holidays		Medium				No	■			V1, 2017	
OpenEE Meter	Average, linear, and change-point (3-P to 5-P)	OA Temperature	No	High(via code)	**Yes [note 1]**	**Yes**	No	**Yes**			■	v2.8.5, 11/21/2019	[1] Coefficients are given, but equation form is published elsewhere for change-point models; [2] TOWT_OpenEE modifications use CalTRACK methods and create 12 'monthly' TOWT models with data weighting.
										■			
	TOWT_OpenEE [note 2]								■				
UT3 M&V Module	Average	None	**Yes**	High(via code)	**Yes**	**Yes**	**Yes**	**Yes**	■	■	■	Universal Translator 3 M&V Module v1, 2014	[1] Coefficients are given, but equation form is published elsewhere for change-point models; [3] TOWT_UT3 modifications allow for filtering based on load profiles, e.g., Holidays, Summer school; TOWT model can be used with daily or hourly data; Models are combined using a 'Model Assembler'
	Linear	OA Temperature (or other independent variable)			**Yes**					■			
	Change-point (3-P to 6-P)	OA Temperature; Daily & Annual Schedules			**Yes [note 1]**					■			
	TOWT_UT3 [note 3]				No				■	■			
	TOW_UT3 [note 3]				No				■	■			
NMECR	HDD/CDD	OA Temperature, optional second independent variable	No	High(via code)	**Yes**	**Yes**	**Yes**	**Yes**			■	Version: 1.0.1, 11/27/2019	[4] TOWT_NMECR modifications allow for filtering based on load profiles (e.g. Holidays, Summer school); Weighting factor for demand response event can be disabled; Automated determination of operating schedules for day-typing
	Linear	OA Temperature (or other independent variable); Daily & Annual Schedules							■	■	■		
	Change-point (3-P to 5-P)	OA Temperature								■			
	TOWT_NMECR [note 4]	OA Temperature; Annual Schedule (additional day-types, e.g., Holidays)							■	■			
	TOW_NMECR [note 4]								■	■			

Efficiency Valuation Organization

1629 K Street NW, Suite 300
Washington, DC 20006, USA

EVO.Central@EVO-world.org

www.evo-world.org