REGIONAL EVALUATION,
MEASUREMENT & VERIFICATION FORUM

THE CHANGING EM&V PARADIGM

A Review of Key Trends and New Industry Developments, and Their Implications on Current and Future EM&V Practices

Prepared by DNV GL

A project of the Regional Evaluation, Measurement and Verification Forum

DECEMBER 2015

Facilitated and Managed by Northeast Energy Efficiency Partnerships

ACKNOWLEDGEMENTS

This report was prepared by the DNV GL team: Miriam Goldberg, Michelle Marean, Curt Puckett, Claude Godin, Wendy Todd, Shawn Bodmann, and Kristina Kelly.

The project was managed by Julie Michals, Regional EM&V Forum Director.

Thank you to the following people who provided peer review of the paper:

Jordana Cammarata, FirstFuel

Kevin Cooney, Navigant

Jessica Granderson, LBNL

Tim Guiterman, EnergySavvy

David Jacobson, Jacobson Energy Research

David Jump, Quest

Mike Kaplan, Retroficiency

Emily Kemper, ClearResult

Chris Long, Opower

Julie Michals, NEEP

Bill Miller, LBNL

Claire Miziolek, NEEP

Ralph Prahl, Prahl & Associates

Ethan Rogers, ACEEE

Elizabeth Titus, NEEP

Marisa Uchin, Opower

Daniel Violette, Navigant

Austin Whitman, First Fuel

Eric Winkler, ISO New England

Table of Contents

Table of Exhibits

1 EXECUTIVE SUMMARY

This report was prepared for the Regional Evaluation, Measurement and Verification Forum (EM&V Forum). The Forum, established in 2008, is a regional project facilitated and managed by Northeast Energy Efficiency Partnerships (NEEP) that in 2015 includes Connecticut, Delaware, the District of Columbia, Maryland, Massachusetts, New Hampshire, New York, Rhode Island, and Vermont. The Forum supports the development and use of consistent and transparent protocols to evaluate, measure, verify, and report the savings, costs, and emissions impacts of energy efficiency and other demand-side resources, and informs national EM&V protocols. Its work includes monitoring and informing Forum members of developments and opportunities to improve EM&V practices.

1.1 Background and Objectives

This paper characterizes key trends in the changing EM&V paradigm and the implications new industry developments have on current and future EM&V practices and activities. The findings and recommendations in this paper are provided to address key trends energy efficiency stakeholders are currently observing.

In particular, this paper addresses data collection and analysis enhancements related to new technology and computing capabilities. These advanced capabilities are often referred to as "M&V 2.0," and their incorporation into evaluation methods as "EM&V 2.0." Because of the potential ambiguities of both these catch phrases, this paper generally avoids the use of either term. Instead, this paper refers to specific types of enhancements, and their potential roles in evaluation. The broad categories of changes discussed are:

Figure 1-1: Two General Types of EM&V Advancements

Advanced Data Analytics	**Software:** The growing range of cloud-based software and platforms that process large volumes of data quickly using publically available or proprietary algorithms. These solutions present opportunities for analyzing the increased data collected in order to inform utility programs and EM&V efforts.
Improved Data Collection Tools and Increases in Data Availability	**Hardware:** The use of smart meters and smart thermostats, non-intrusive load monitoring and other tools that both collect energy usage data in new ways and present opportunities to increase the amount and type of data collected.

Under the first broad category, advanced data analytics, this paper focuses in particular on analytic tools and services that provide automated, ongoing analysis of energy consumption data. This process is referred to here as "automated M&V" or "auto M&V."

This paper characterizes the trends in data collection and analytics with the purpose of furthering stakeholders' understanding of how these approaches can be leveraged for EM&V, as well as discussing the limitations of these new tools and techniques. In particular, the research examines how and to what extent the enhanced data and new tools can help address stakeholders' concerns that standard EM&V procedures are costly and results take a long time to produce. Similarly, the paper examines whether these new capabilities can maintain or improve the accuracy and reliability of EM&V. Two case studies of experience to date with advanced data analytics and use of improved data collection tools are provided, and help to inform key findings and recommendations. Descriptions of several automated M&V products and vendors are also included.

1.1.1 Primary Research Questions

The paper explores how various elements of evaluation planning and reporting could be addressed by these new approaches. The primary questions, which guided the research, are shown in Table 1-1.

Table 1-1: The Primary Research Questions

Evolution of EM&V Practices	How are EM&V practices likely to evolve?
Program Enhancement with Real-Time Results	How has experience with automated M&V demonstrated the value of real-time or quick turnaround evaluation the supports a feedback loop that informs program changes in a timely manner?
Leveraging New Data	What data from emerging tools can by leveraged to meet EM&V needs and how?
Benefits and Challenges	What benefits of and challenges to the deployment of new tools and increasing data sources need to be considered in the context of improving EM&V practices for EE programs?
Standardization	How, where, and at what level would any standardization of practices associated with these trends benefit EM&V?
Program Preparation	What can and should administrators do to prepare for changes in EM&V?

To answer these questions, DNV GL conducted primary research on the new trends in data collection and analysis tools, including in-depth interviews with industry representatives, researchers (e.g., LBNL), and experts in the EM&V field, and consulted recent research and publications.

1.2 Broad Themes Emerging from This Review

The new tools, services, sources, and processes explored in this paper have a wide range of capabilities and potential implications for EM&V. The following broad themes emerge from this work:

1. **Advanced data collection and analysis tools and systems offer new opportunities for understanding and engaging customers, offering value to project and program delivery as well as to evaluation.**
    - On the delivery side, these capabilities currently offer value for project identification, program planning, and rapid and ongoing feedback to customers and programs for continuing engagement and problem solving.
    - On the evaluation side, the new capabilities can provide early indications of project and program effects, as well as enabling finer resolution analysis of customer loads, and new bases for customer segmentation using machine learning.

2. **There remain important evaluation challenges that are not solved by greater volumes or frequency of consumption data, or higher speeds of data processing.** These challenges include estimation of appropriate baselines, analysis of complex projects and processes, and assessing market conditions and program influence.
3. **Automated M&V could help shorten evaluation timelines in some cases, but will not necessarily accelerate the timing of final evaluation results.**
   - Automated M&V based on analysis of whole-premise consumption data could be used as the basis for final evaluation results for some types of programs; for many programs different or additional methods are needed.
   - Evaluation timelines are driven by many factors, including stakeholder processes, how far ahead evaluation planning is initiated, the degree of standardization versus customization at each step, and the data collection and analysis methods used.
   - Automated M&V is one tool that may help shorten evaluation timelines by accelerating early analysis, but in most cases will still require up to a year or more after the final program participation to provide final results.
4. **All evaluation methods have strengths and limitations that vary according to the characteristics of the programs and customers studied.** Evaluation planning involves selection of combinations of methods, balancing issues of accuracy, timeliness, and cost, according to the defined objectives and priorities for the work. Advanced data collection and analysis capabilities expand the toolkit of available evaluation techniques.
5. **The energy efficiency industry should continue to track emerging product and service offerings in this area, and explore the benefits and costs of incorporating the new capabilities into program delivery and evaluation**.
   - Establishing agreed protocols for product testing can support this process.
   - Cost-effectiveness should be assessed in terms of the range capabilities provided.

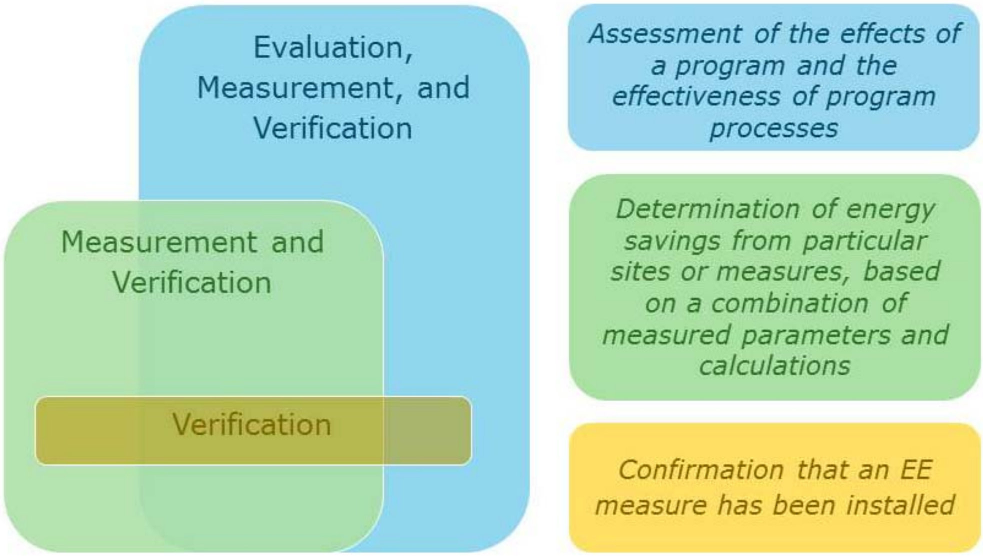
These themes are developed further below. Before presenting more fully the key findings and recommendations from this research, an overview of EM&V is provided as a framework.

1.3 EM&V Framework

1.3.1 Overview of EM&V Functions

In understanding how advanced data analytics and advanced data availability can help to improve current EM&V practices, a basic overview of EM&V is warranted. "EM&V" is often considered a single body of methods, but the three processes (Evaluation, Measurement, and Verification) have distinct roles, as illustrated in Figure 1-2. Distinguishing these three processes, and where and how advanced data analytics and data collection tools are relevant to each, is important to understanding the potential role for streamlining and improving on current EM&V practices. These opportunities and potential trade-offs are summarized in the Findings & Recommendations section below.

Figure 1-2: Relationship of Evaluation, Measurement, and Verification



Evaluation is only one purpose of M&V. M&V is one approach to developing evaluated savings. Verification is a component of an M&V process, and also can be conducted without measurements as part of program evaluation.

1.3.2 Evaluation Independence

Energy program evaluation is required to be conducted by an independent third party in almost all jurisdictions. For purposes of this paper, it is assumed that the principle and policy of independent evaluation is retained in the future. Absent such a requirement, program evaluation could be dramatically different, if it continued to exist at all, apart from the implications of new tools explored here.

1.3.3 Broad Categories of Evaluation Approach

Widely used EM&V guidance identifies three broad categories of evaluation approach[1]:

- M&V
- Deemed savings (with verification)
- Large-scale consumption data analysis (using a comparison group).

1.3.4 Baselines

A key question for impact evaluation (determination of program savings accomplishments) is the baseline. The baseline represents what would have occurred without the installation of the program measure (gross savings) or without the influence of the program (net savings). Different circumstances inform appropriate baselines.

Fully specifying the baseline for gross or net savings calculation requires a combination of technical information and assumptions together with policy guidance. The specific policies and assumptions for defining baselines vary by jurisdiction, but the combination of these considerations means that, even adjusting for changes in weather, direct comparison of participants' pre-installation with post-installation consumption does not provide savings consistent with common baseline definitions in most contexts.

1.3.4.1 Evolving Baseline Perspectives

Recently enacted legislation in California, AB802[2], could be interpreted to imply that the pre-installation condition is always the appropriate baseline, and that the weather-normalized change in consumption is the preferred measure of savings. However, in implementing the new law, the California Public utilities Commission is taking a more measured approach. Opportunities for to-code measures are currently being explored, and change in normalized consumption is recognized as a measure of energy savings, but not necessarily the only or preferred measure.

Other states may also widen the definition of savings that may be captured or counted by utility programs. It is still to be expected that baselines will often be defined not by the previously existing condition, but by some higher level of efficiency corresponding to natural replacement.

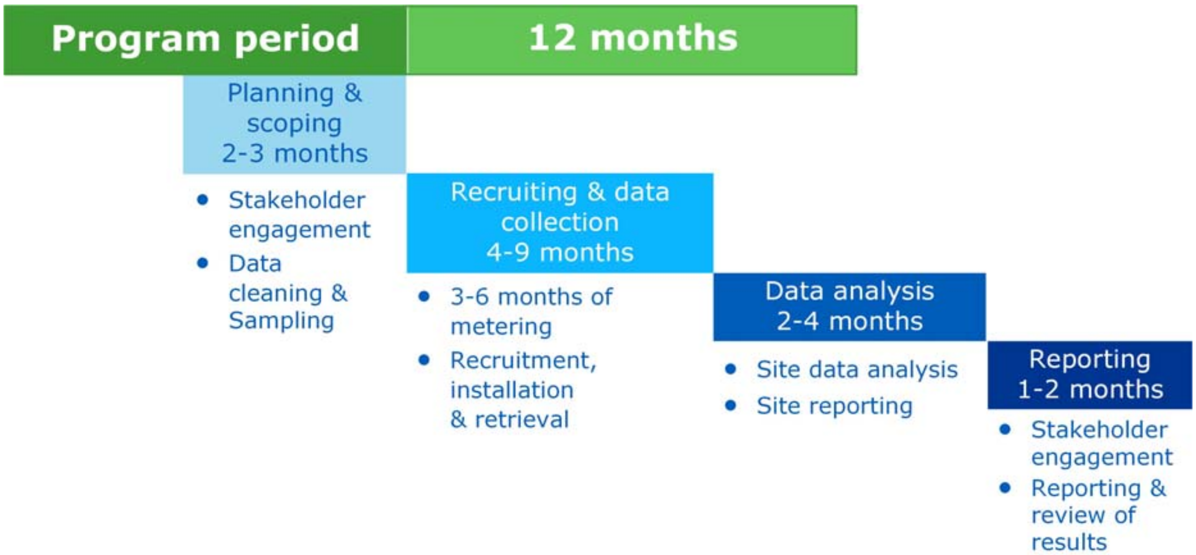
1.3.5 Evaluation Timeline and Potential for Streamlining

The time required to provide evaluation results currently includes:

- Planning time to understand programs and data available, and agree on research plans, instruments, and protocols with evaluation clients and stakeholders
- Data collection time, including developing sample designs, recruiting respondents for customer or vendor surveys for onsite data collection, and obtaining the consumption and tracking data themselves
- Data analysis time
- Reporting time, including responding to client and stakeholder feedback.

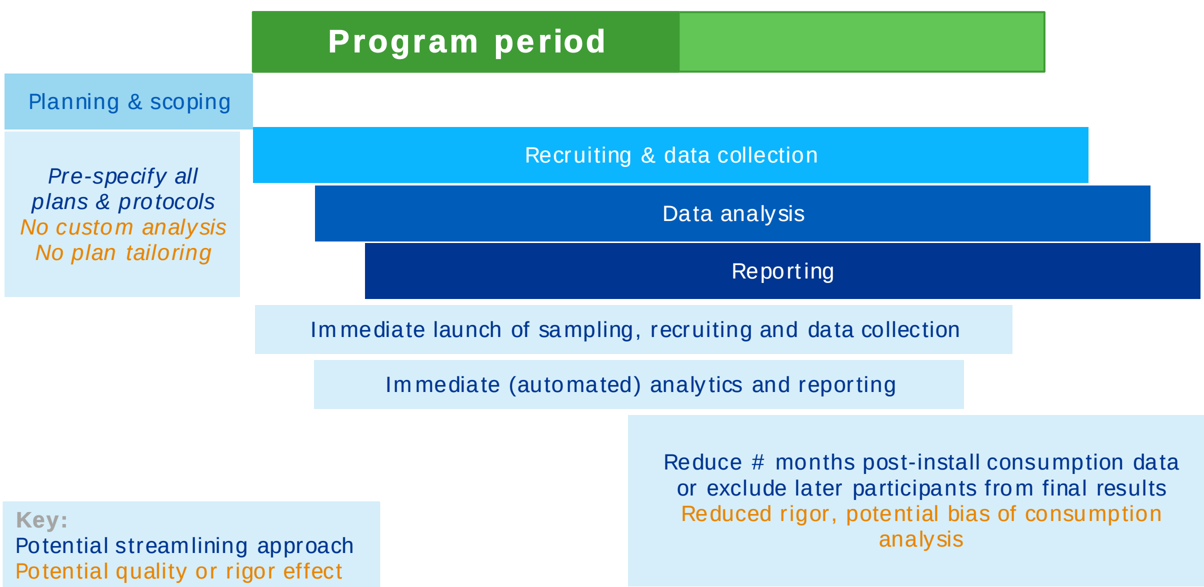
Figure 1-3 below presents an illustrative schedule for a typical impact evaluation. Depending on extent of stakeholder input to the draft and final evaluation results, as well as the types of data being collected and analyzed, the overall timeframe could go beyond the 18 months indicated in the figure.

Figure 1-3: Illustrative Evaluation Timeline



The total time to conduct an evaluation can be shortened by agreeing, prior to the program start, on fully specified research methods, instruments, and protocols, and by implementing some form of continuous sampling, analysis, and reporting. On the other hand, more comprehensive evaluation typically requires the flexibility to re-allocate resources based on issues that have emerged for the program, and to explore unusual cases. This opportunity is given up if a fully pre-specified approach is followed.

Figure 1-4: Illustrative Evaluation Streamlining Potential



There are other limits to how much the evaluation cycle can be compressed. If pre and post-installation energy consumption data analysis is to be the basis for evaluated savings, 9 to 12 months of post-installation consumption data is typically needed to capture all the seasonality of annual savings. If all program participants are to be included in the evaluated savings, this analysis cannot be finalized until close to a year after the program close, even with immediate data processing and reporting. In addition, many jurisdictions have broad stakeholder processes that entail substantial time for review at several stages of the evaluation. Timelines cannot be substantially compressed without curtailing these processes at some level.

One focus of this paper is to explore how advanced data analytics and data collection can help to shorten the evaluation process and provide timely information about program performance.

1.4 Key Findings and Recommendations

The findings and recommendations of this paper are organized to address the primary research questions under the two broad areas of **advanced data analytics** and **advanced data availability**:

1. Advanced data analytics/program enhancements with real-time results—using **automated M&V** in evaluation, to reduce the time and costs to develop evaluation results
2. Advanced data availability/leveraging new data—**improved data collection tools and data availability** to support evaluation

In addressing these broad areas, the paper discusses benefits, challenges, the potential role of standardization, and ways that program administrators and regulators can prepare for the coming changes.

The recommendations cover a range of topics. However many are focused on how EM&V stakeholders and practitioners can continue to assess the opportunities represented by the emerging tools and methods, and develop key protocols and principles for their application.

1.4.1 Automated M&V

1.4.1.1 Key Research Findings

Comparison of Automated M&V to Existing Evaluation Methods

- For calculation of savings, current automated M&V tools use weather normalization methods similar to those that have been used for program evaluation for many years, with similar performance.
- A key difference of the automated M&V applications is the rapid and ongoing analysis.
- Fully taking advantage of the opportunities offered by automated M&V, for both program delivery and evaluation, requires non-automated examination, exploration, and follow-up on the emerging results.

Current Uses

- Deployments of automated M&V are currently in the pilot stages. Most current applications are using monthly consumption data, especially for smaller customers.
- Primary uses currently being promoted for automated M&V are services provided to the program administrator, and are not used to support claimed savings reported to commissions. These services are:
    1. Program targeting and opportunity identification through remote assessments and audits.
    2. Rapid and continuous feedback to customers and programs on the changes in facility consumption after measure installation.

Capabilities

- Once an automated M&V installation is in place, automated analysis of energy consumption data can begin providing feedback to programs within one month of measure installation.
- Delivery of such automated analysis requires advance agreement and the implementation of methods for ongoing access to program and consumption data.
- Automated analytics vendors report excluding up to 20% of accounts based on data anomalies. This loss rate is not viewed by vendors as a problem for opportunity identification. However, if the automated tools are to be used for comprehensive program evaluation, the effect of such exclusions on savings estimates would need to be explored.
- A disadvantage of sole reliance on highly automated, rapid analysis for evaluation is the loss of opportunity to customize the analysis in response to changing program conditions or emerging findings. High quality evaluation typically requires the flexibility to re-allocate resources based on issues that have emerged for the program, and to explore unusual cases
- Automated analytics designed to be applied for large volumes of data in diverse contexts also can bring the advantage of more thoroughly tested and vetted analytic software.

Performance

- The performance of automated M&V tools has recently been assessed for their accuracy in predicting whole-building consumption based on historical consumption. Findings from these assessments are generally consistent with prior studies of monthly consumption data analysis using variable degree-day models.[3]
- Current research has not yet addressed the ability of these predictive models to measure energy savings *after measure installation*, but future research planned by LBNL will address this issue

Transparency

- Transparency is a key principle for M&V and for evaluation.[4] Under current practice, evaluators are required to provide full specifications and documentation of modeling methods, equations, assumptions, and tools used.
- The majority of vendors of automated M&V tools have stated that they will provide full transparency of the equations, and the process of constructing a comparison group where relevant.
- To a certain extent, the accuracy of proprietary tools can be tested via protocols such as those developed by LBNL in their recent and continuing work. However, empirical performance in a particular test environment may not by itself ensure confidence in the tool results for purposes of evaluation. The greater the transparency of methods and assumptions, the more readily a tool is likely to be accepted by regulators and stakeholders.

Applicability

- Whole-premise pre-post installation consumption analysis is an appropriate basis for evaluated savings in limited contexts, where the pre-program consumption pattern is the appropriate baseline for savings, and where other changes in facilities can be accounted for. These contexts include:

1. Retrofit programs for large homogeneous populations, with a valid comparison group capturing average changes for participants absent the program, or absent the program measures.
2. Commercial retrofit programs, using analysis of individual buildings where non-program-related changes are known and their effects on consumption can are reliably quantified.

Large commercial sites or industrial processes are not likely to be well addressed by these methods.

- Except in cases where comparison group methods provide net savings, determination of net-to-gross adjustment ratios, where required, will typically require other types of methods than consumption data analysis. Large scale rapid analytics is not likely to contribute substantially to expediting this research.

Cost Effectiveness

- The cost-effectiveness of these methods as part of overall program delivery or evaluation is not yet established.

1.4.1.2 Recommendations

1. **Programs should continue to explore the benefits of the early feedback made possible by automated M&V.** Estimated savings reductions from automated consumption data analysis can provide rapid feedback to programs whether or not this analysis is used as the final evaluated savings. Such rapid feedback is useful whether it is provided as part of program delivery or as part of evaluation.
2. **Where programs are using these tools on the delivery side to monitor project performance, programs and evaluators should explore the potential time and cost savings of evaluator's making use of the automated analysis.** Evaluators should take the automated M&V results into account in developing interim and final evaluations, including process as well as impact evaluation, to the extent practical. If evaluators are to make use of the automated M&V results themselves as a basis for evaluated savings, they should have access to the methods and results.
3. **Programs should continue to explore the benefits of using automated consumption data analysis for opportunity identification or to engage customers, in comparison to existing targeting and customer engagement processes.** Existing evaluation methods can be applied to support such assessment.
4. **Making use of automated consumption data analysis to provide final evaluated results would require advance agreement on the analysis methods, including data screening and cleaning, supplemental data to be collected and used in the analysis, procedures for adjustments for non-program-related changes, and comparison group construction if any**.

Evaluators and evaluation stakeholders should consider establishing automated consumption analysis as an accepted evaluation method that falls somewhere between customized M&V and a well vetted deemed savings process. Stakeholders should consider the situations where the benefits of cost and time savings would outweigh the disadvantages of less thorough or customized analysis.

5. **Evaluators and evaluation stakeholders should continue to explore and apply other means to reduce the total time required to complete evaluation of a program, and to provide early feedback.**

6. **For programs where pre-post analysis of consumption data is to be used as the basis for final evaluated results, a post-evaluation period long enough to capture all seasonal use patterns, if any, should be used.** In most cases this will require 9 to 12 months of post-program data, consistent with current evaluation practice.
7. Cost-effectiveness of automated EM&V tools should be assessed taking into account the full set of benefits for program delivery and evaluation.

1.4.2 Use of Improved Data Collection Tools and Data Availability to Support Evaluation

1.4.2.1 Key Research Findings

AMI Data

- The primary opportunity offered to evaluation by AMI is the potential to obtain whole-premise consumption data at daily, hourly, or higher frequency, for all customers. Daily data from AMI systems are currently used in evaluations for energy consumption analysis, and hourly or finer data for demand-response (DR) impact analysis and energy program load-shape impacts. M&V for both program support and evaluation can make use of hourly whole-premise load data for individual facilities to examine usage patterns and changes in usage patterns, and to calibrate simulation models.
- Two-way communications via AMI systems can interface with home or building control systems, for DR program implementation, and for program response monitoring and measurement. Very high frequency AMI data can be used with emerging Non-Intrusive Load Monitoring systems, to decompose whole-facility energy use into end uses.
- There are currently limitations for evaluation in the useful application of interval data from smart meters:
    - AMI systems are not deployed in all territories, and in some are deployed only to certain customer segments and geographies.
    - Where AMI systems are in place, systems are not necessarily capable of providing ready access to the data by programs, evaluators, or researchers. Restrictions due to regulatory and utility customer data protection policies can also be a factor in data availability.
    - Due to its volume, AMI data requires a significant increase in IT and data management capabilities for all stakeholders.
    - Hourly or daily data is typically not subject to the same routine data validation and cleaning as is typically used for monthly kWh consumption data and kW demand data from revenue accounting processes.
- Because of the current challenges of data processing, data management, data access, and data cleaning, use of AMI data is not necessarily a lower cost evaluation option compared to traditional evaluation approaches, even where AMI is already in place.

- If these limitations can be overcome, there are significant opportunities for using more granular AMI data, supporting better estimation of simple consumption models, plus the ability to fit more complex models. The use of daily AMI data rather than monthly data, and incorporation of supplemental premise information from existing databases, has the potential to extend the applicability of these methods to smaller proportions of savings. AMI data also presents the possibility of using similar analyses to determine peak or time-based impacts that monthly data cannot provide.
- AMI with machine learning or pattern detection and classification routines has the potential to be used to classify accounts based on load shape patterns. Once a set of load profile categories is developed, surveys can be designed to provide information linked to the load patterns and categories. Survey information could both validate conjectures about end uses and operations developed from the load patterns, and collect key parameters to be used in analysis of those uses. Thus, the AMI data can support both richer analysis of usage patterns and more informed survey design.

Other Granular Data

- Smart devices and non-intrusive monitoring have the ability to provide more granular consumption data, and in some cases, at a lower cost than traditional measurement approaches. Where data collection devices are already deployed, access to data can avoid the costs of on-site meter installation specifically for evaluation. A potential scenario is that the HEMS or smart thermostats can serve as a communications hub for the various smart devices in a home, and relay all of this information to third party providers or utilities. While these data sources are not at the stage where they can fully replace on-site metering and audits, they can provide useful information for calibrating results and supplanting data that is often difficult or costly to obtain in necessary quantities.
- Barriers exist in the ability for utilities to collect customer level data from smart devices. Many smart device vendors have strict data protection and customer privacy policies. However, most providers will supply aggregate data to the utilities. For some of the largest vendors, this is unlikely to change. These policies reduce the ability for using smart devices for demand response and behavior programs, but aggregate data is still very valuable for measuring some utility energy efficiency programs.
- For HEMS, other limitations to the use of smart devices for evaluation include:
    - Baseline conditions: If the program measure of interest is the installation of the HEMS itself, the system cannot provide data for the period prior to its installation, which would be the likely baseline of interest. (If the program measure of interest is occasional program-initiated control, the HEMS can provide data for periods with and without control events, as a basis for evaluation.)
    - Comparison group definition: If the program measure of interest is the installation of the HEMS itself, evaluation would require a comparison group of otherwise similar homes that did not install the HEMS.
    - While HEMS can be a valuable source of information, it should not be viewed as a tool that can replace standard EM&V. Information provided by HEMS and other smart devices can

enhance EM&V or reduce additional data collection requirements, but these tools do not provide enough information on their own to sufficiently estimate a program's impact.5

For NILM the challenges include the accuracy and predictability of end-uses captured. The technology is still in early stages of deployment.

1.4.2.2 Recommendations

1. **Stakeholders should continue to support and develop the potential for AMI data to expand EM&V techniques as the availability of more granular data continues to increase.**

2. **Stakeholders should consider establishing guidance for best practices for how AMI data is incorporated into EM&V.** Topics for consideration include:

   - Identifying contexts where whole-premise consumption data can be the basis for savings estimates using monthly and daily data
   - Development of hourly load impacts using interval data
   - Development of data screening procedures
   - Assessment of bias due to data loss and potential bias-mitigation strategies
   - improving simulation calibration using whole-premise interval data
   - Development of matched comparison groups using hourly load patterns
   - Identifying the contexts where AMI analysis is likely to be a lower cost evaluation method than alternatives, and the associated quality effects.

The EM&V community should continue working with smart device vendors, consumers, and the utilities should optimize smart device data uses and work to improve data ownership issues to improve data access. The more restricted access remains, the less likely this data will be incorporated into large scale EM&V.

2 INTRODUCTION

2.1 Project Overview

This paper characterizes key trends in the changing EM&V paradigm and the implications new industry developments have on current and future EM&V practices and activities. The findings and recommendations in this paper are provided to address key trends energy efficiency stakeholders are currently observing.

In particular, this paper addresses data collection and analysis enhancements related to new technology and computing capabilities. These advanced capabilities are often referred to as "M&V 2.0," and their incorporation into evaluation methods as "EM&V 2.0." Because of the potential ambiguities of both these catch phrases, this paper generally avoids the use of either term. Instead, this paper refers to specific types of enhancements, and their potential roles in evaluation. The broad categories of changes discussed are shown here.

Figure 2-1: Two General Types of EM&V Advancements

Advanced Data Analytics	**Software:** The growing range of cloud-based software and platforms that process large volumes of data quickly using publically available or proprietary algorithms. These solutions present opportunities for analyzing the increased data collected in order to inform utility programs and EM&V efforts.
Improved Data Collection Tools and Increases in Data Availability	**Hardware:** The use of smart meters and smart thermostats, non-intrusive load monitoring and other tools that both collect energy usage data in new ways and present opportunities to increase the amount and type of data collected.

Under the first broad category, advanced data analytics, this paper focuses in particular on analytic tools and services that provide automated, ongoing analysis of energy consumption data. This process is referred to here as "automated M&V" or "auto M&V."

This paper characterizes these trends with the purpose of furthering stakeholders' understanding of how these approaches can be leveraged for EM&V, as well as discussing the limitations of these new tools and techniques. In particular, it examines how and to what extent the enhanced data and new tools can help address stakeholders' concerns that standard EM&V procedures are costly and results take a long time to produce. Similarly, it examines whether these capabilities can maintain or improve the accuracy and reliability of EM&V.

2.2 Primary Research Questions

The paper explores how various elements of evaluation planning and reporting could be addressed by these new approaches. The primary questions, which guided our research objectives, are shown in Table 2-1.

Table 2-1: The Primary Research Questions

Evolution of EM&V Practices	How are EM&V practices likely to evolve?
Program Enhancement with Real-Time Results	How has experience with automated M&V demonstrated the value of real-time or quick turnaround evaluation the supports a feedback loop that informs program changes in a timely manner?
Leveraging New Data	What data from emerging tools can by leveraged to meet EM&V needs and how?
Benefits and Challenges	What benefits of and challenges to the deployment of new tools and increasing data sources need to be considered in the context of improving EM&V practices for EE programs?
Standardization	How, where, and at what level would any standardization of practices associated with these trends benefit EM&V?
Program Preparation	What can and should administrators do to prepare for changes in EM&V?

2.3 Summary of Approach

With these trends and questions in mind, several primary research efforts were undertaken. We held in-depth interviews with evaluation consultants and independent researchers. Five utility software providers spent many hours during the course of this project describing their products, their current and future markets, the nature of the utility programs they were supporting, and their observations about automated M&V and what “EM&V 2.0” means to them. These discussions were complemented by interviews and exchanges with smart thermostat manufacturers, implementers, and evaluators of thermostat-based demand response programs and other tools covered in this paper, including evaluators beginning to use AMI data in their current evaluation studies. Further, literature and papers were reviewed, which describe potential applications and implications to future evaluation practices influenced by the emerging tool and analytical techniques.

Lastly, this paper includes case studies of a software provider’s experience supporting utility and government program delivery and the use of AMI data to support customer engagement and demand response programs.

2.4 Organization of the Report

The remainder of this report is structured as follows:

- Section 3: Background: This section provides an overview of key terms and distinctions, to ensure a common understanding of EM&V in the context of the key trends and research questions listed above.
- Section 4: New Methods and Tools for Data Collection and Analysis: This section discusses the methods and tools that make up the landscape of interest. Information is provided to give the reader context when assessing the key trends and answering the primary research questions listed above.

- Section 5: Recommendations and Considerations for Emerging EM&V Practices: This section presents our conclusions and recommendations regarding the changing EM&V paradigm in the context of the research conducted.
- Section 6: Select Case Studies: This section presents cases studies that describe two example programs that utility Software as a Service (SaaS) providers are supporting and the use of AMI data in implementing and estimating results from demand response programs.

3 BACKGROUND

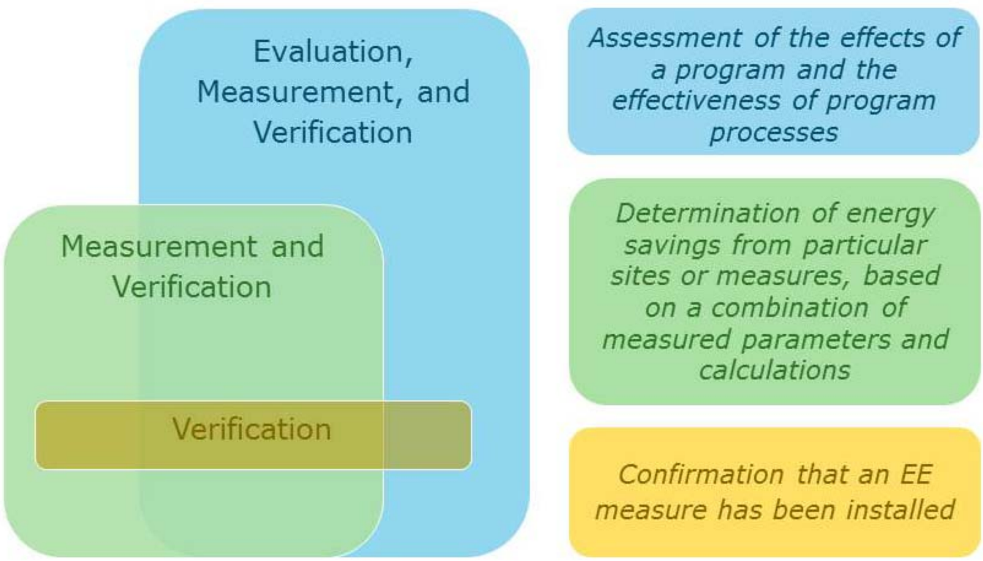
3.1 EM&V Framework

This chapter provides a foundation for our discussion of newer EM&V data collection and analysis tools by describing current EM&V methods and guidelines, and presenting the associated challenges and data management requirements.

3.1.1 General Terms

Often the terms evaluation, measurement, and verification are frequently, but inaccurately, used interchangeably. Figure 3-1 indicates commonly accepted definitions associated with EM&V and its components. For purposes of this paper, these are the definitions used when describing EM&V.

Figure 3-1: Common Definitions of Evaluation, Measurement, and Verification



Verification is a process of confirming that a measure has been installed as planned. Verification may be conducted for individual projects as part of broader site-level M&V, or may be conducted as part of evaluation without a corresponding measurement effort.

Measurement and Verification (M&V) in general is the determination of energy savings for an individual project or measure, using observations and measurements taken at the site together with engineering calculations. Techniques may include onsite inspection, data logging and on-site metering, and engineering review of projects. M&V can be used by service providers within or outside of programs, as a component of measure assessment, planning, or implementation. An M&V process can be used as part of building or project commissioning, to confirm that systems are working correctly and make corrections as needed. M&V is also used as a basis for determining savings in the context of performance contracts between facility owners and energy service providers. In the program evaluation context, M&V of individual sites is a basis on which program-level savings are developed.

Evaluation includes impact evaluation, quantifying program effects, and process evaluation, assessment of program processes and recommendations for improvement. An overall EM&V strategy may include research activities that do not directly contribute to impact or process findings, but are well aligned with these evaluation functions. In particular, market assessment, characterizing customers and suppliers related to adoption of technologies and practices, is often included in the scope of EM&V. In this paper, the evaluation focus is on impact evaluation.

3.1.2 Evaluation Independence

Energy program evaluation is required to be conducted by an independent third party in almost all jurisdictions. This requirement is to protect the credibility of evaluation results provided to regulators on behalf of ratepayers. While evaluators will have individual perspectives on programs and methods, they do not have a direct stake in the evaluation outcomes in the ways that program implementers and administrators do. The specific restrictions and roles of independent evaluation vary across jurisdictions. For purposes of this paper, it is assumed that the principle and policy of independent evaluation is retained in the future. Absent such a requirement, program evaluation could be dramatically different, if it continued to exist at all, apart from the implications of new tools explored here.

3.1.3 Broad Categories of EM&V Approaches

There are several approaches evaluators employ to conduct EM&V. The SEE Action Report identifies three broad categories of impact evaluation methods: [6]

- **Deemed Savings** use pre-determined unit savings values, together with tracked counts of the numbers of units implemented through the program. This approach is most suitable for common, prescriptive measures with well-established average savings per unit. Evaluation using deemed savings ordinarily includes validation that the deemed values have been appropriately applied, and some level of verification of the units installed.
- **M&V** when referring to a program evaluation approach means using project-level M&V for a sample or census of projects installed through the program. Program-level results are determined by statistical expansion of the sample if a census is not obtained. These evaluation approaches are most commonly used for large, custom, or complex measures or projects. These methods provide the most detailed information about individual projects, as trained field staff collect and review site-specific data. These methods also tend to be the most time consuming and costly of the evaluation approaches, often requiring months of data collection.
- **Billing analysis, or large-scale consumption data analysis with a control group,** uses analysis of consumption data taken from customer billing records from before and after installation of a program measure. The method typically utilizes a comparison group to control for non-program factors that may contribute to consumption changes. This method is most suitable for evaluation of programs with relatively large, homogeneous participant groups, availability of an appropriate

comparison group, and average savings on the order of 10% or more of pre-program consumption. How more granular AMI data may influence this threshold is discussed later in this paper.

For deemed and M&V methods, a separate net-to-gross adjustment is commonly but not always required, depending on the jurisdiction. Components of this adjustment may include free ridership, participant and nonparticipant spillover, and market effects. These adjustments are developed from market studies and customer and vendor surveys. When deemed savings are used, the net-to-gross adjustments may also be deemed. For billing analysis methods, the analysis may provide net savings directly, or may require a separate net-to-gross adjustment, depending on the structure of the comparison group and the analysis.

3.1.4 Measurement and Verification for Projects versus Programs

The International Performance Measurement and Verification Protocol (IPMVP) provides a well-defined standard for M&V at the individual project level. Evaluation regulatory requirements in several jurisdictions specifically require use of these protocols. However, the IPMVP by itself does not fully define evaluation methods at the program level, for several reasons.

1. The IPMVP defines several options for M&V methods that may be used between parties to a performance contract. It is up to these parties to define which method to use. Evaluation protocols for particular jurisdictions may define which IPMVP option must be applied for different program types.
2. In the performance contracting context, savings are measured from the perspective of the end customer. For measures installed at existing facilities, these savings typically are relative to the prior equipment. For many energy efficiency programs, the baseline for calculating savings is not the prior equipment, but new standard efficiency equipment, or a composite of prior equipment and standard new equipment.
3. The IPMVP defines methods for measuring savings for individual projects or facilities. Program-level savings are often developed by applying IPMVP methods to a sample of participating facilities, and statistically expanding these results to the full program participant population. Program-level savings using certain forms of billing data analysis may develop results only at the program level, and not for individual participants.
4. Impact evaluation in most jurisdictions requires determination of net savings, i.e. the savings that would not have happened without the program. Even if a “standard new” baseline is used, IPMVP methods provide savings with respect to the baseline as experienced by the customer, without regard to the program’s effect on the customer.
5. A full evaluation includes not only impact evaluation, determining the magnitude of savings attributable to the program, but also process evaluation, assessing how well the program was administered, if the program was as effective as it could be in light of other options, and how the program can be improved.

3.1.5 The Baseline Question

A key question for impact evaluation is the baseline. The baseline represents what would have occurred without the installation of the program measure (gross savings) or without the influence of the program (net savings). If equipment was going to be replaced regardless of the program, the baseline condition corresponds to the new unit that would otherwise have been installed. If other equipment or operating practices change at the same time as the installation of the program measure, the baseline corresponds to the presence of the other new equipment and operations, but without the program measure. This alternative configuration may be difficult to specify without some probing. If the equipment was going to be replaced in a year or two but the program accelerated the replacement, the baseline corresponds to the existing equipment up to the time it would have been replaced and new non-program equipment thereafter. In a new construction context, there is no pre-installation baseline to observe.

Fully specifying the baseline for gross or net savings calculation requires a combination of technical information and assumptions together with policy guidance. The specific policies and assumptions for defining baselines vary by jurisdiction, but the combination of these considerations means that, even adjusting for changes in weather, direct comparison of participants' pre-installation with post-installation consumption is not by itself a recommended evaluation approach in most contexts.

Recently enacted legislation in California, AB802[7], makes several changes to the landscape of energy efficiency policy in that state. Two key provisions are

1. Utilities programs may include actions to bring facilities up to code, as well as beyond code, rather than counting savings only relative to existing code.
2. Programs are to support improvements in existing buildings "taking into consideration the overall reduction in normalized metered energy consumption as a measure of energy savings."

These provisions could be interpreted to imply that the pre-installation condition is always the appropriate baseline, and that the weather-normalized change in consumption is the preferred measure of savings. However, in implementing the new law, the California Public utilities Commission is taking a more measured approach. Opportunities for to-code measures are currently being explored, and change in normalized consumption is recognized as **a** measure of energy savings, but not necessarily the only or preferred measure.

Other states may also widen the definition of savings that may be captured or counted by utility programs. It is still to be expected that baselines will often be defined not by the previously existing condition, but by some higher level of efficiency corresponding to natural replacement. EPA's draft EM&V Guidance for the Clean Power Plan[8] requires use of a "Common Practice Baseline" (CPB) representing the "default technology or condition that would have been in place at the time of project implementation absent the EE installation." For natural replacement situations, the CPB corresponds to market average efficiency or formal building or product standards, and for early replacement a dual baseline is specified.

3.2 Challenges and Limitations to Current EM&V Approaches

Preferred EM&V methods can differ depending on program characteristics, including the target population, measure types, delivery method, and program penetration. For example, billing analysis may be used for some types of residential or small/medium business programs, and on-site metering and custom project reviews are used more often for larger scale, non-residential projects. Regulatory requirements, ratepayer and shareholder interests, market participation, and the need to inform internal decision-making also influence the EM&V method selection process.

Within the broad categories of impact evaluation methods noted above, a combination of techniques are commonly used:

- Deemed savings are developed from prior evaluation results and from engineering analysis. Deemed savings most commonly provide gross savings only, but may also include deemed net-to-gross factors.
- Survey data are used to establish baseline and operating conditions for gross savings calculation for engineering analysis to determine gross savings, to develop inputs for consumption data regression analysis, as well as to determine net-to-gross factors.
- Engineering desk review by itself may be considered a form of verification, but may also be the foundation for measurement steps in an overall M&V process.
- End-use metering, data logging and on-site inspection are measurement methods as part of M&V.
- Billing analysis may provide gross or net savings (or something in between) depending on the structure of the comparison group, as described in the Uniform Methods Project chapter on Whole-Building Retrofit (which primarily addresses residential applications).[9]

Figure 3-2 presents many of the typical challenges for the common EM&V techniques.

Figure 3-2: Key Challenges for Common Evaluation Techniques

Deemed Savings
- Currency, relevance
- Comparability across programs and states

Engineering Desk Review
- Sources of assumptions
- Choice of formulas or models
- Comparability across programs and states

Surveys
- Response rates
- Length vs. survey fatigue
- Post survey data cleaning
- Reliability of responses
- Interpretation and scoring
- Program planning that may not facilitate surveys

End-use Metering
- Data quality
- Data access
- Customer access and scheduling
- Access to plans, schedules, info on baseline equipment

On-site Inspection
- Customer access and scheduling
- Access to plans, schedules, info on baseline equipment
- Availability of consumption data

Billing Analysis
- Data collection for the pre- and post-periods
- Data cleaning requirements
- Signal-to-noise ratio
- Comparison group specification

Developing a valid comparison group is challenging, and not necessarily possible for all program types. A variety of techniques exist to construct comparison groups that closely match the participants in terms of pre-participation consumption. However, these methods do not guarantee an appropriate comparison group for changes in consumption.[10] In particular, when program activity tends to be undertaken at times of other major changes at premises, either the comparison group must consist of non-participating premises with similar major changes, or the other changes need to be known and explicitly adjusted for. Matching on pre-participation consumption is a useful step, but insufficient to ensure the validity of the comparison group.

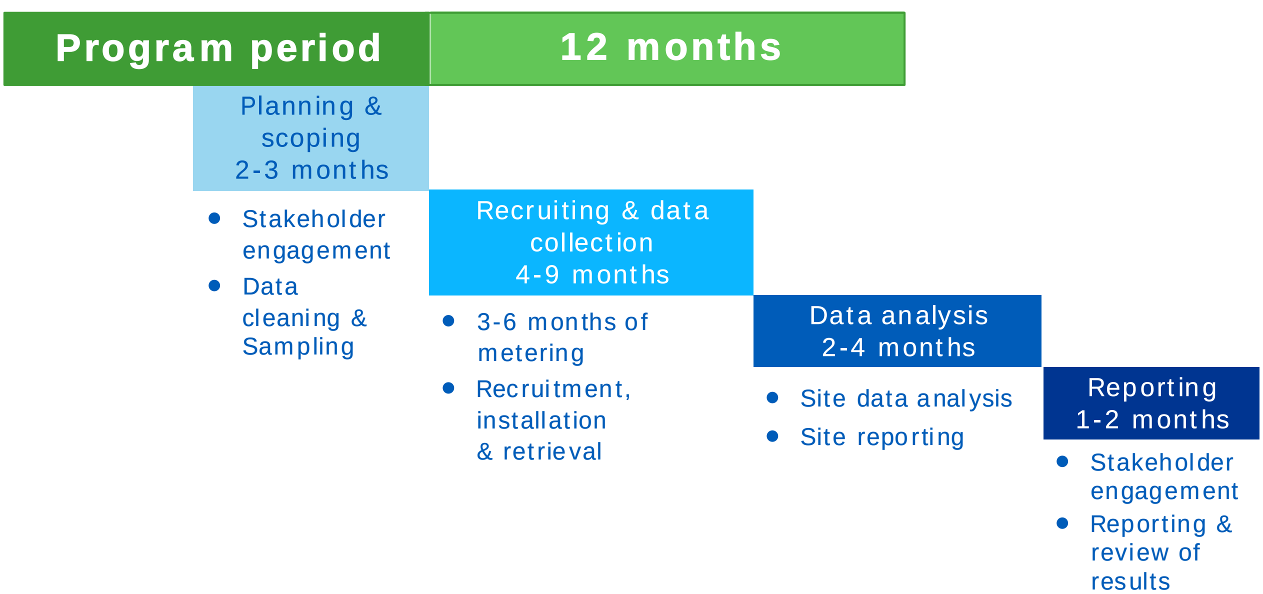
Determining appropriate baselines, as discussed, is a challenge for all impact evaluation approaches. For deemed savings, and engineering review, the baseline assumptions are among the factors that may be out of date, or assumptions may be in error. Baseline conditions are among the information surveys, onsite observation, and end-use metering attempt to collect or provide a basis for estimating. Comparison group specification is itself a means of adjusting to an appropriate baseline of a no-program alternative.

Baseline conditions are not necessarily directly observable, and may depend on the policy under which programs operate. Even when the baseline corresponds to a hypothetical standard rather than to observable conditions, information about the site characteristics and operations can be useful. This information supports construction of an appropriate estimate corresponding to the baseline definition, and accounting for post-installation premise characteristics and operations not related to the program.

As noted, different evaluation methods are appropriate for different situations. Methods that rely more on current, primary data collection and analysis specific to current participants, with higher sampling precision where applicable, and less on prior information or secondary data sources or population-level assumptions, are described as having higher rigor, meaning a greater level of effort is applied. Higher rigor evaluations are conducted to provide more accurate and reliable results. However, there are always trade-offs to be considered as to which combination of methods best mitigates the potential biases and uncertainties that are of concern for a particular program.

In general, methods involving primary data collection and analysis entail higher costs and often, but not always, more time. The primary drivers of the time and costs to perform EM&V are shown in the example timeline in Figure 3-3.

Figure 3-3: Typical Impact Evaluation Process and Timing



The number of stakeholders, in particular, has a major impact on EM&V time and cost. Stakeholders often have conflicting or disparate objectives. Consensus building is valuable and necessary, but it can also take a great deal of time, and can add months to an evaluation process.

3.3 Using Automated M&V to Reduce Impact Evaluation Effort and Timeline

One potential benefit of automated M&V is the potential for reducing impact evaluation timelines and effort described in Section 3.2. The potential for streamlining evaluation efforts at each of the key stages is discussed below. Automated M&V is discussed as one potential contributor to such streamlining. The broader issues of what would need to be agreed or accepted for such streamlining to be possible are also discussed.

3.3.1 Planning and Scoping

This first step includes review of program tracking data, and development and approval of evaluation plans. These plans include sample designs, data collection protocols and instruments, and data analysis plans. Typically, to ensure efficient deployment of evaluation resources, specific plans, and sample designs are developed based on review of the tracking data. The timeline for this stage of evaluation can be reduced if some or all of these plan elements are fully agreed at the outset. Such agreement gives up the potential for tailoring these steps to the current program conditions and participant characteristics, in the interests of streamlining evaluation timelines and costs. Use of automated retrieval and processing of program tracking data and whole-premise consumption data as a foundation for impact evaluation would require at a minimum:

- Determination that savings can be estimated using analysis of whole-premise consumption data (billing analysis)
- Agreement to a particular automated analysis as the basis for the billing analysis
- Advance specification of any supplemental customer data collection required for determination of gross savings baselines or of program attribution
- Acceptance of the automated data cleaning and screening procedures of the automated analysis
- Advance specification of a comparison group construction method, whether for determination of gross or net savings, or of a process for customer-specific adjustments.

3.3.2 Recruiting and Data Collection

This step involves customer recruitment and data collection via remote surveys (phone, internet, or mail) or via onsite visits. If no primary data is collected for evaluation, outside of compilation from existing utility and external sources, this step is eliminated. If sampling, recruitment, and data collection are conducted on a continuous basis as customers complete participation, according to a pre-agreed selection process, the timeline for recruitment and data collection can begin essentially with the start of the program, and typically be completed within one to three months after the close of the program period. Longer completion times are needed for more challenging recruitment and scheduling, such as for onsite measurements at complex sites.

A particular factor affecting the timing of data collection is the seasonality of energy use. If whole-facility energy analysis is used, current practice is that the post-installation period should include all seasons, to provide a basis for meaningful estimation of annual savings. In practical terms, this means that the evaluation typically cannot be finalized until close to a full year after the last installation to be included in the analysis, even if the entire analysis and reporting processes take minimal time. Depending on the timing of installation, a post-installation period as short as 6 months could capture winter, season, and shoulder months, but this would be true only for participants at certain times of the year.

One way to relax the requirement to span all seasons in the post-installation period would be simply to accept a level of bias, the nature of which would depend on the seasonality of participants' usage, together with the patterns of seasons covered across the consumption data. The other way would be to incorporate supplemental information or assumptions into the analysis, related to annual operating profiles and seasonal efficiency.

In effect, the second approach would shift the consumption data analysis to an estimation process based more on supplemental operational information, engineering assumptions and engineering analysis. This would be similar to the analysis conducted under an M&V approach, but relying on whole-premise data for metered energy use. Such an approach is not currently being advocated either by evaluators or by vendors of automated M&V tools.

When M&V is the primary evaluation method, the measurements are usually conducted for particular systems or equipment, and span only a few weeks. Annual savings are developed by applying information or assumptions on annual operating profiles. As a result, unless M&V is based on long-term post-installation metering, the need to capture seasonal patterns of use does not push out the completion of the evaluation cycle as much as does whole-facility analysis under current practice. For measures that are related to heating or cooling, data collection needs to occur during the corresponding seasons.

Customer surveys are used in impact evaluation to collect information on equipment operations, as well as on program influence. Completing such surveys quickly after program participation can improve the quality of information on program influence and intended operations, but will be less informative about actual post-installation operations.

3.3.3 Analysis and Reporting

This step is the development of site-level findings and program-level savings estimates from the compiled program data, other primary data, and external data. Time for the analysis step is shortened by using pre-established automated processing routines, including both processing of site-specific data and any sample expansion required.

Such pre-specification limits, or eliminates, the opportunity to explore unusual or anomalous cases or unexpected results; such investigation of the unexpected often provide the most important insights in any study. Such pre-specification also eliminates the ability to conduct customized analysis of large, complex, custom, or unusual projects and applications. Currently, customized data collection and analysis is commonly needed in particular for large commercial facilities and for industrial processes. It is not clear that pre-specified processing could be applied effectively in such cases.

Time for reporting is shortened by using pre-established reporting templates with minimal interpretation specific to the particular program period or findings.

Considering the potential for automated M&V in particular to reduce the analytic timeline, these tools currently consist of whole-premise consumption data analysis, mostly using monthly data. Whole-premise consumption analysis with comparison group is a recommended analysis method for large, homogeneous populations where a valid comparison group can be developed. This essentially limits its application to certain types of residential and small commercial programs.

Whole-premise consumption data analysis with site-specific custom adjustments is a recommended M&V method for individual projects or facilities. Such adjustments are not currently part of automated M&V.

3.3.4 Summary of Potential Streamlining

The potential time reductions using the streamlining procedures described above are summarized in Figure 3-4. Automated data collection tools and automated analysis tools can be employed in such streamlining. However, neither is essential to adoption of these streamlining methods.

Figure 3-4: Illustrative Evaluation Streamlining Potential

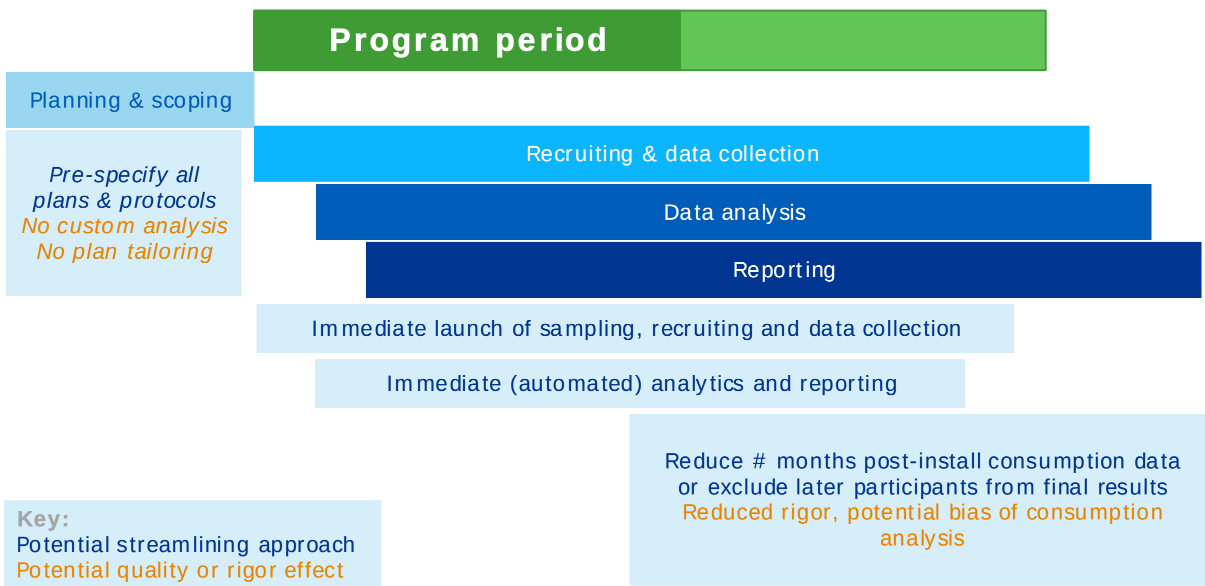


Table 3-1 illustrates the potential time to complete an evaluation after the close of a program, assuming fully pre-specified data collection and analysis processes, with automated sampling and analysis. The timeline depends on the types of primary data collection included. Also indicated in the table is an indication of the relative level of quality that would be associated with this level of pre-standardization. The quality level is not indicated as “high” for any of these highly constrained and automated evaluation structures; high quality would require the flexibility to re-allocate resources based on issues that have emerged for the program, and to explore unusual cases.

Table 3-1: Potential Time to Complete an Evaluation after the Close of a Program

Combination of Methods Used				Total Elapsed Months from Program Close			Relative Quality		
Remote Surveys	**Onsite Data Collection**	**Fully Automated Reporting Template**	**Whole Premise Consumption Analysis with Comparison Group**	**Complete Data Collection**	**Complete Analysis**	**Complete Reporting**	**Gross Savings Analysis**	**Attribution**	**Interpretation and Recommendations**
No	No	Yes	No	0	1	1	Low	Low	Low
Yes	No	Yes	No	1	2	2	Low	Medium	Low
Yes	Yes	Yes	No	2	3	3	Medium	Medium	Low
Yes	Yes	No	No	2	3	4	Medium	Medium	Medium
Yes	Yes	Yes	Yes	11	12	12	Medium	Medium	Low
Yes	Yes	No	Yes	11	12	13	Medium	Medium	Medium
Yes	Yes	Yes	Yes	6	7	7	Low	Low or Medium	Low
Yes	Yes	No	Yes	6	7	8	Low	Low or Medium	Medium

Of course, the evaluation timelines and quality, in terms of accuracy and thoroughness, depend on a wide range of factors. The table illustrates that

- Fully pre-specified plans and protocols with automated reporting could allow final results within a very short time after program close, depending on the data collection and analysis included.
- However, evaluation based on post-installation consumption analysis requires several months after the close of the program. Using less than a full year of post-installation data for the analysis reduces the quality of the results.

3.4 Requirements for Data Quality and Management

Compiling data needed to evaluate EE and demand reduction (DR) programs from consumption meters, program-tracking information, or on-site data collection is the first step in the EM&V process.

Data cleaning is the process of reconciling or removing data in a database that are incorrect, incomplete, improperly formatted, or duplicated; anomalies and outliers must be identified and resolved. The level of effort devoted to data cleaning depends on the intent of the EM&V activity and can consume 20 to 30% of the timeline. This data typically demands a significant amount of data preparation, which often presents significant challenges and adds time to the EM&V timeline. Some EM&V activities can meet their objectives with less effort and may be able to discard anomalies and outliers without compromising the EM&V objectives. Improving data quality at its source will likely result in significant time and cost savings for evaluators and program managers.

As noted, one of the ways that automated M&V can shorten evaluation timelines is by having automated compilation, screening, and processing of the data. Making this possible requires establishing automated data transfer processes, and establishing or accepting existing automated screening processes. Having established data transfer and screening mechanisms could shorten evaluation timelines with or without automated high volume processing. The trade-off is that the less automated data cleaning often surfaces issues that programs can address, and want to, and can result in a smaller proportion of customers being excluded from the analysis.

4 NEW METHODS AND TOOLS FOR DATA COLLECTION AND ANALYSIS

New methods that incorporate advances in technology and computing capabilities have the ability to enhance EM&V approaches and reduce the time and cost typically associated with evaluation. Figure 4-1 lists a variety of new methods and tools available to the evaluation community today. As discussed above, many of these tools and methods are currently being deployed for program delivery support, and not yet for program evaluation. In the following sections, an overview is provided of each method or tool and the potential implications it may have on EM&V.

Figure 4-1: Overview of New Methods and Data Collection Tools for Program Support and Evaluation

4.1 Data Analytics and the New Frontier of Technologies

4.2 Automated M&V Software

4.3 Auto M&V Software as a Service

4.4.1 Smart Meters and Advanced Metering Infrastructure

4.4.2 Smart Thermostats and Devices

4.4.3 Other Smart Devices and Home Energy Management Systems

4.4.4 Intrusive and Non-Intrusive Metering

4.1 Advanced Data Analytics

4.1.1 Data Analytics

Method Overview: Traditional analytical tools examine historical data and have produced business intelligence and operational insight for decades. Advanced analytics, on the other hand, are forward focused and predictive; they use statistical models on data about the past to predict the future.[11] The Internal Revenue Service and financial sector were early adopters of advanced analytical techniques.

Whether operating millions of smart thermostats, or performing near-real time M&V on very large samples of metered data, modeling at this level requires some element of machine learning.

Potential Implications for EM&V: Utilities and independent system operators take advantage of advanced analytics for grid optimization, operations management, and theft detection. Adopted by evaluators and program administrators, traditional and advanced analytics are becoming more common. Utilities and their stakeholders are seeking a deeper understanding of their customers' attributes, how those attributes influence future participation, and asking more complex questions answered with higher volumes of data, resulting from wider customer participation, smart meters, third party data, or all three simultaneously. Using modern platforms with large datasets is more accessible due to reduced costs of storage and processing.

Utilities have been hesitant to move "mission critical" operational systems like "SCADA"[12] or customer billing to cloud-based environments, but a growing number of utilities are now moving to cloud-based platforms for non-critical systems that do not require a high degree of two way, or real time (sub second) communication.

4.1.2 Machine Learning

Method Overview: Machine learning is a field of computer science that studies patterns in large datasets to make observations or predictions.

Potential Implications for EM&V: Machine-learning software can process billions of data points of disparate data within seconds or minutes. It is also possible to prepare large datasets for EM&V with machine-learning tools. Most data anomalies, including anomalies in utility data, fall into discrete categories and consistent patterns. High-speed analytics and machine learning can identify these patterns. In the case of consumption data machine learning may be helpful in detecting data anomalies. Rules-based analysis assists with identifying and reconciling consumption and program data, and flagging data that requires more extensive analysis using traditional (albeit human) intervention. In some cases, creating predictive models using machine-learning processes may improve utility forecasting over traditional methods when large amounts of data are available.

Machine learning puts the "smart" in smart thermostats by recognizing temperature and occupancy patterns during a training period and controls system run times to maintain indoor temperatures according to the occupants schedule. As occupants override settings or change schedules, the thermostats learn and adjust accordingly. Smart thermostats are discussed in more detail in later in this paper.

Opportunities for using machine learning uses EM&V are include the following.

1. Data cleaning, and anomaly detection, as described above.
2. Developing predicted energy usage absent the program. Machine learning can provide more informed predictions than a set of default engineering assumptions. On the other hand, predicted consumption based on pre-installation consumption is appropriately the basis for the "no-program" or "no project" baseline only in certain situations, as discussed in Chapter 3..
3. Customer segmentation based on machine learning pattern detection and classification. Such segmentation can strengthen the analysis for impact evaluation as well as for market assessments and planning.

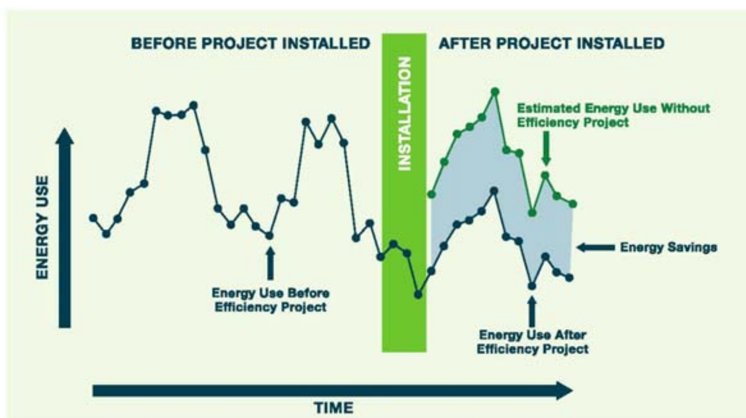
When vendors are using machine learning to help deliver programs, evaluation may be able to make use of the results of the machine learning, to the extent these results are made available to evaluation. For example, analysis of a smart thermostat program can benefit from using a physically based model that takes into account the details of the control strategy. However, but if these details are proprietary, the analysis will be less informed, and may be similar to what would be done for a more conventional control strategy.

An important caveat is that machine learning on any volume of data cannot overcome the fundamental baseline challenges described in Chapter 3: pre-installation usage by itself cannot predict no-program usage—i.e., does not define the appropriate baseline--in cases of natural replacement, early replacement, new construction, or where other major changes are occurring at the same time. The machine learning may be able to flag cases where there appear to be other changes in the facility, but by itself the automated process is unlikely to be able to determine what adjustments are appropriate for those. For example, machine learning without supplemental information from the customer cannot sayto what extent the other changes are lasting versus ephemeral, stable versus varying with other conditions or season, or related to the measure of interest as interactive effects, takeback, or spillover.

4.2 Automated M&V

Method Overview. Automated measurement and verification software takes advantage of automated data processing, and sometimes machine learning processes, to produce building energy profiles, estimate savings potential, or estimate whole-building energy savings in near real time. Using AMI and monthly data, automated M&V methods are similar to traditional degree-day billing analyses. The software estimates whole building energy savings by creating an adjusted baseline from metered consumption in the pre-program period to model buidling energy use into the future absent the energy efficiency measure. Savings are reported as the difference between the adjusted baseline and metered consumption. In Figure 4-2, baseline energy use is shown on the left and the adjusted baseline is shown as the green line on the right.

Figure 4-2: Baseline (left) and adjusted baseline (right) (Source: EPA)



Historically, metered data has only been available in monthly intervals. As has been recognized for years in the evaluation context, billing analysis methods are most reliable when they use close to 12 months of consumption data for the pre- and the post-implementation period if consumption data is the basis for evaluating measures with seasonal effects. Six to nine months may suffice if all seasons are included, but accuracy will be lower than a 12-month analysis period. Many believe that AMI data at daily or hourly intervals will shorten the amount of time needed to calculate the adjusted baseline post-implementation as compared to the traditional 9–12 month analysis period. Certainly using daily rather than monthly data allows a wider range of weather conditions to be observed within a given time frame. To the extent that there are seasonal usage patterns not fully explained by weather or other explicit model terms, using less than a year of data would still leave some bias in the estimate of annual savings. Even with daily data, a 6-month analysis window will only sometimes capture the extremes of both summer and winter, along with the extended behavioral effects surrounding those extremes. More experience using AMI data for evaluation, together with performance assessments such as the recent and future research by LBNL,

When automated methods are used to track building or program performance on an on-going basis, analysis and reporting of results can begin immediately after measure installation. These reported results, can be informative to the customer and to the program even if they do not use the evaluation baseline as the basis for calculating savings, and do not represent final evaluated results.

The private sector has embraced automated M&V for many years (in the form of energy management information systems) to measure impacts from energy conservation investments, to support building commissioning, and as a maintenance management strategy for monitoring building operations.

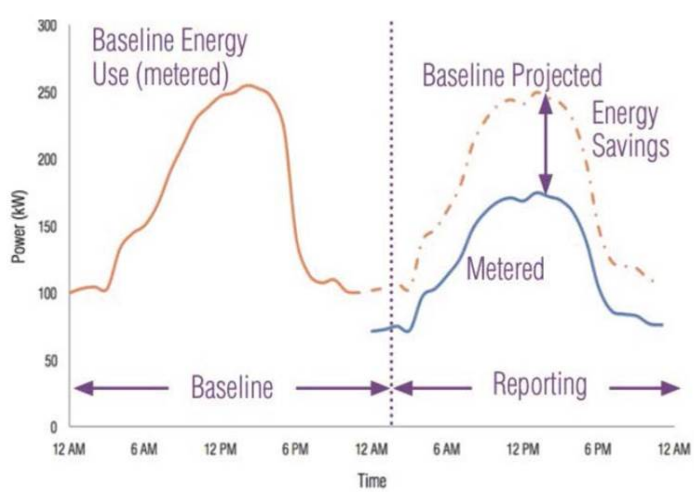
Currently, most of the algorithms and methodologies of automated M&V vendors are proprietary if automated M&V is to be used directly as evaluated savings, transparency requirements in the utility sector M&V or evaluation context will require that vendors publish their methodologies and the equations used to estimated energy savings. If the automated M&V is used to inform the evaluation, with a retrospective adjustment factor determined by other evaluation analysis, detail on the automated M&V methods is helpful to evaluation but not essential. If the adjustment factor is to be used prospectively, there needs to be a way to confirm that the future automated processes are consistent with past processes that were the basis for developing the empirical adjustment.

Depending on the application, some automated models may incorporate third party information beyond weather data, such as property assessments, neighborhood census data, occupancy data, operations schedules, and product throughput. The use of third party data in EM&V in general has become less expensive because the public sector, academia, and private companies make large amounts of data available through data markets and open source communities. Automated M&V requires continuous and systematic streams of meter data which is challenging for some utilities.

Ongoing Research. Lawrence Berkeley National Lab (LBNL) compared how well data-driven automated baseline models predicted commercial energy consumption from high interval metered data. LBNL tested public and proprietary automated model specifications from three, six, nine and twelve months of consumption (the training period), shown in Figure 4-3. Predictions from each model were compared to metered data in the post period, and each other, using standard statistical tests.

In the studies of automated M&V models for commercial buildings, performance across various models was similar. The most recent study, which included proprietary and public models, found that for most buildings in a large data set (500+ buildings), the models are likely to meet ASHRAE guidelines for how well baseline models should fit the pre-measure data to be used for EM&V. In addition, the

Figure 4-3: Model training period (left), predicted model and metered consumption (right) (LBNL)



study found that for half of the buildings, annual whole building energy consumption could be predicted with errors less than 5%. The accuracies achieved in this study were for a fully automated case. In practice, errors can be further reduced with the oversight of an engineer to conduct non-routine adjustments where necessary. Errors are also further reduced when individual buildings are grouped and considered at the program level.[13]

Future LBNL research will focus on demonstrating the automated M&V approach in partnership with utilities and program implementers. Gross savings from sets of program data will be calculated using interval data baseline models, along with the associated uncertainty and confidence in the savings results. Where possible, the time savings with respect to more conventional EM&V methods will also be investigated.

Potential Implications for EM&V. Automated M&V supports program delivery by providing whole building feedback at the site, project, and program level to program administrators and participants early and throughout the post-implementation period. Ongoing automated analysis could be useful in an evaluation context also. Evaluation use of automated M&V would require agreement on the automated analysis and acceptance of the data screening process and data loss. Rapid and ongoing feedback to the program using automated M&V would be a useful evaluation function if not already provided by the program delivery.

The main time and potentially cost savings from evaluation use of automated analyses would come from establishing a fixed analysis protocol and not deviating from it. Common data preparation formats and reporting protocols will assist in realizing the savings opportunities associated with automated analyses.

Thus, automated methods could reduce the turn-around time to the delivery of evaluated results, if there is up front agreement on the use of the methods, with no extended exploration of the results afterward.

When automated M&V is used in program implementation, the program has access to the early and ongoing results and can explore reasons for individual anomalies or program-level trends. If savings are different from what is expected, reasons could be explored early and possible changes to the program or to the analytic method considered. There may also be a role for evaluators in reviewing and interpreting the ongoing feedback from the automated M&V.

For program managers and DSM planners, automated M&V can provide timely and actionable feedback allowing for enhanced customer engagement, continuous improvement, and program optimization.

To the extent the automated M&V results from program delivery can feed directly into evaluation, the automation may reduce evaluation data processing time. Acceptance of existing automated M&V as evaluated results would require acceptance of the basic analytic framework, as well as acceptance of the data screening processes. Tolerances for data losses will vary by type of data and the type of evaluation. Therefore, automated data preparation is not a one size fits all solution. If data cleaning is performed by the automated M&V vendors prior to a third party evaluation, data sharing may reduce data preparation time even further if data protocols are specified in advance.

The main challenge for use of automated M&V as it currently exists for program evaluation is the definition of savings itself. As described in Section 3, whole-premise pre-post installation consumption analysis is an appropriate basis for evaluated savings in limited contexts, where the pre-program consumption pattern is the appropriate baseline for savings, and where other changes in facilities can be accounted for. These contexts include:

- Retrofit programs for large homogeneous populations, with a valid comparison group capturing average changes for participants absent the program, or absent the program measures.
- Commercial retrofit programs, using analysis of individual buildings where non-program-related changes are known and their effects on consumption can are reliably quantified.

In cases of natural replacement, the basis for establishing the baseline for savings is the “standard” efficiency equipment that would otherwise have been installed. In the case of accelerated replacement, the baseline is the prior condition up to the time the equipment would have been replaced and standard efficiency thereafter. Currently, the savings reported by the automated M&V tools is relative to a baseline defined by pre-program consumption, and often assumes no other major changes at the facility beyond the program measures and weather. In some cases, additional variables such as occupancy, hours of operation, or throughput are added to account for levels of commercial activity. Thus, additional or different analysis would be needed to translate the automated M&V savings relative to the prior condition to the appropriate baseline.

Also as noted in Section 3, developing a valid comparison group is challenging and not possible for all programs, even those that serve mass markets. For programs where a particular comparison group definition is considered valid, it may be possible to automate the comparison group development. To make the establishment of a comparison group as automated as other aspects of the analysis, it would be necessary to agree in advance on a protocol, and a basis for testing its validity. The application of this protocol to creating the comparison group should be overseen by an independent third-party evaluator. This approach of a comparison group selection process agreed in advance and overseen (or in some cases implemented) by the evaluator is common today for residential behavioral programs using a randomized assignment as the basis for evaluation.

In contexts where a direct comparison group is not available, regression analysis is sometimes used in evaluation to control for the underlying non-program differences between participants and the comparison group. This type of analysis would be extremely challenging to automate.

As noted, even if a valid comparison group is available, the pre and post analysis in general provides program impacts only in the context of energy conservation measures with homogeneous participant groups. For other contexts, including commercial/industrial programs, natural replacement, accelerated replacement, and new construction, a post-only automated analysis could be used as part of a calibration process. While this work could, in principle, begin earlier than 12 months after installation, there may be limited value to the preliminary savings estimates beyond the benefits of the pre-post automated M&V as currently provided.

For use of whole-premise automated analytics for evaluation, a first step would be to validate the predictive accuracy of the tools using a protocol similar to LBNL. Ideally, such validation would be conducted for different customer types, within multiple regions, and over multiple years. The protocol should also address data screening rules and transparency of these rules.

Finally, while these modeling and analytical approaches do have the potential to reduce the time and cost associated with evaluation calculation of gross savings, they do not however address other, often-required aspects of evaluation. In particular, determination of net savings factors including free ridership, spillover, and market effects require other types of methods. These are based on surveys, market studies, expert interviews, case studies, and econometric regressions. While all of these methods could be expedited by establishing pre-agreed detailed research specifications, large scale rapid analytics is not likely to contribute substantially to expediting this research.

4.3 Utility Software as a Service (SaaS)

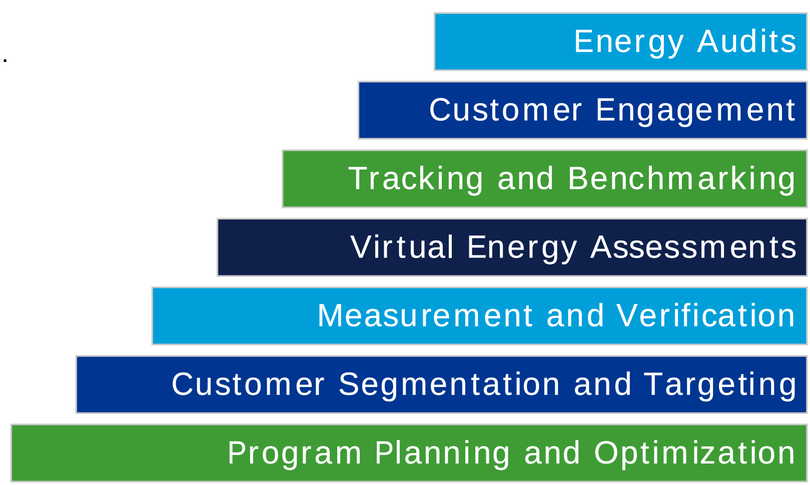
4.3.1 SaaS Description

As described above, most current applications of automated M&V are for program delivery support, not for program evaluation. The software-as-a-service (SaaS) model and the automated M&V vendors are, at this time, primarily providing program support.

Where such a model is in place for some combination of the program benefits described here, evaluation could be another use of the service, as described in Section 4.2. In principle, even absent a program delivery role for automated M&V, evaluations could also provide analytics to programs on a similar automated, rapid feedback basis. However, many of the benefits of automated M&V, which help justify the investments, fall on the program delivery side. Research for this paper did not identify any instances of SaaS offerings strictly for evaluation.

Automated M&V software is being licensed to utilities in a SaaS model to support program delivery, customer engagement, marketing, and to track savings (relative to the pre-program condition) from energy efficiency measures. A comprehensive list of service offering is shown in Figure 4-5.

Figure 4-4: Typical Service Offerings of Automated M&V SaaS Vendors



Using the automated M&V techniques described in the previous section, the utilities purchase customized software that accesses the vendor's cloud-based analytic platforms, and leverage the engineering expertise of the vendors. Although outside the context of EM&V, there are several value added services of SaaS vendors for utility program delivery:

- The SaaS vendors provide custom software tools that offer multiple customer communication channels to increase customer engagement throughout the program cycle
- They give the program managers the ability to conduct in-house analysis of project and program performance and monitor program savings on an ongoing basis
- It enables utility customers direct access to continuous tracking and analysis
- It increases scalability of services through automation of analytics
- Provides new services for data aggregation and management

The range of services offered by the utility SaaS market is increasing as the utilities gain experience and request a broader range of services and as new vendors enter the market. SaaS providers often form partnerships with program implementers and engineering firms to expand their services across the program cycle. Although by no means a complete list, several high profile SaaS vendors are listed in Figure 4-5. One case study of an automated M&V SaaS vendor is included in Section 6.

Figure 4-5: Selected SaaS Providers of Automated M&V

Virtual Audits, Remote Audits, and Virtual Assessment

The SaaS products for virtual audits, remote audits, and virtual assessment are a subset of automated M&V functions beyond tracking savings. These tools identify candidate C&I customers and engage customers as they assess investments in energy conservation measures or are pursuing maintenance and operational changes to improve energy efficiency. The remote assessment allows the utility to use customer specific data in targeted marketing and customer engagement campaigns. A traditional audit requires a site visit to assess energy savings potential, usually at significant expense to the customer. Remote assessments model consumption, building characteristics and operations, without a site visit. Remote assessment often precedes an on-site audit. Vendors report that savings estimates are produced on a by building-by-building basis using industry accepted simulation models.

If a small percentage of C&I customers account for the majority of savings, remote assessments function as a first step to engage the largest potential savers. Bringing the most fruitful projects into the program early increases the speed and scale at which projects are executed across portfolios. Remote assessments are a predictive EM&V tool used to speed up the program delivery cycle. Remote assessments are sometimes conducted in highly specific geographical areas (geo-targeting) to reduce loads on locally constrained grids. One application to Con Edison's geo-targeting program is presented as a case study later in this paper.

Vendor Perspectives on Barriers to Implementation of Automated M&V Software in the Utility Sector

Several of the SaaS vendors have identified barriers to implementation of automated M&V software in the utility sector. From a regulatory perspective, some vendors speculate that the utilities are wary of being penalized if the reported savings from whole-building automated methods differ from either expected savings and/or from results from previous evaluations. The focus on measure-based evaluation as opposed to whole building analysis, and the stipulation for elective and manual audits and inspections to validate savings, also limits the opportunities of the SaaS vendors and the use of their automated tools for EM&V.

Customer analytics and new software tools provide benefits to multiple departments within a utility. This can be a challenge to vendors and evaluators, as historically utilities do not fund products or projects across departments. Additionally, demand-side management (DSM) programs and their evaluation managers operate under highly constrained budgets. Evaluation managers and the organizations that provide evaluation oversight may recognize benefits, but unless new tools are replacing existing evaluation processes, utilities may be reluctant to allocate funds, especially for evaluation activities conducted by a utility vendor (as opposed to an independent third party).

Specific recommendations include:

- Define a transition path that assures future changes to EM&V methods will not jeopardize savings credited in the past.
- Revise EM&V requirements and move away from measure-based savings analysis to facilitate the use of automated analytics to complement existing methods.
- Encourage continuous monitoring of energy use and strategies for re-engaging with customers to ensure desired savings are achieved and to find new savings opportunities.

One way to promote the integration of automated methods into standard utility EM&V practice is for utilities and their third party evaluators to work collaboratively with the SaaS providers to understand their respective methods and explore how the automated tools can enhance and improve existing processes. “Integrating analytics into the array of techniques used for EM&V across the country will require stakeholder input and synthesis, and cooperation between utilities and third-party providers of technologies and services.”

4.3.2 Summary of SaaS Implications for EM&V

Table 4-1: Key Characteristics of SaaS

Function	With SaaS, mature business intelligence analytics has migrated from the private sector to the utility sector. Providers use cloud-based, internet enabled software and very large data sets to support decision for both utilities and their customers.
Current Uses/Value	Uses include measurement and verification, benchmarking, customer segmentation/targeting services, remote energy assessments, energy audits, and customer engagement programs. Results inform customers and program administrators in near-real time allowing for more agile program management and enhanced customer engagement.

Function	With SaaS, mature business intelligence analytics has migrated from the private sector to the utility sector. Providers use cloud-based, internet enabled software and very large data sets to support decision for both utilities and their customers.
Challenges for Utility	Monthly billing data rather than hourly usage is still the most common interval for analysis. Monthly consumption data is sufficient to provide results in many applications.
Impact on Project and Program - M&V	• The time to deliver consumption profiles is shortened, resulting in agile program management. Consumption and tracking data from December is reported in mid-January. • Data cleaning is rule-based and automated
Impact and Challenges for Evaluation	• If SaaS of automated M&V is in place for program delivery support, evaluation can be one of the users. • Evaluation processes can be expedited by ongoing data availability and pre-specified automated analysis. • Baselines for evaluated savings correspond to pre-program usage patterns only in retrofit contexts with no non-program changes taking place. • For mass market retrofit programs, comparison groups can sometimes account for overall non-program changes, but valid comparison group construction is not always possible. • For commercial/industrial customers, non-program-related changes need to be accounted for explicitly if whole-premise consumption data is the basis for savings. • Data attrition from automated screening processes may be more severe than is currently accepted or expected by evaluation • If automated comparison group construction and automated consumption data analysis are to be used as the basis of final evaluated results for a particular program, evaluation needs to be able to vet the methods and confirm they are being consistently applied.
Ethical/Regulatory Issues	PII management with third party data providers

4.4 Improved Data Collection Tools and Enhanced Data Availability

4.4.1 Smart Meters and Advanced Metering Infrastructure

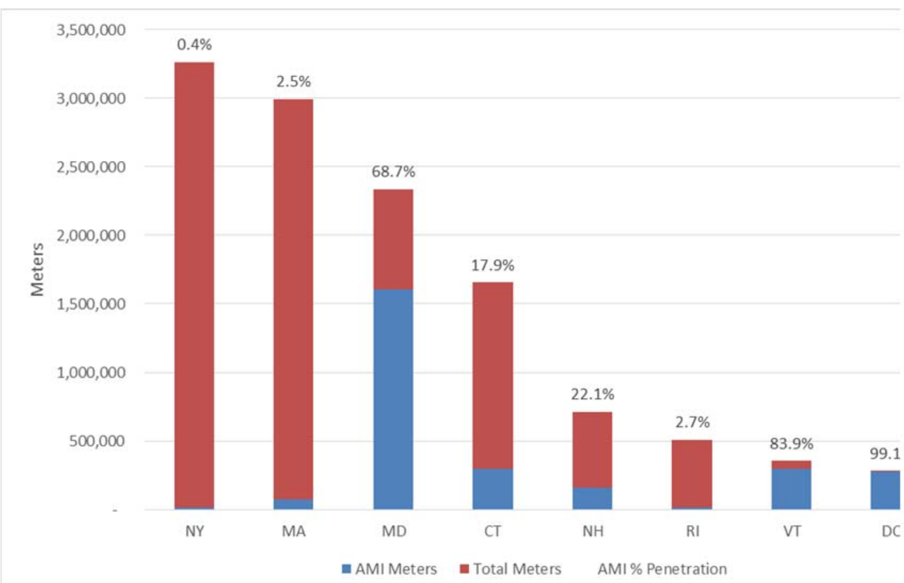
Technology Overview: Smart meters range from basic hourly interval meters to real-time meters with built-in two-way communication, capable of recording and transmitting instantaneous data.

Smart meters produce sub-daily consumption data, which is useful for some analyses. AMI data facilitates dynamic load profiling and time-differentiated impact estimation for EE programs, demand reduction, short-term forecasting, and can inform rates. Most states require consumer access to data as a condition of smart meter deployment. Some argue that the precedent for utility ownership of meter data on the “utility side” applies to old technology and not to AMI meters, as they give visibility to schedules and habits. Although not perceived as a threat to deployment, the topic is receiving some attention.

Utilities who implement AMI must either increase IT infrastructure or outsource data management, which is seen as undesirable. Data silos are the status quo within utilities and it is often difficult to transfer data between operations and EM&V functions. In the face of an already difficult process, AMI data collection and preparation can be laborious for both the utility staff and EM&V practitioners. Smart meters produce data with sufficient reliability, although a 20% data loss is common due to quality and data interruptions.

Since 2009, the U.S. Department of Energy (DOE), utilities, and other organizations have invested over $7.9 billion in 99 Smart Grid Investment Grant (SGIG) projects. By 2013, 14.2 million smart meters were installed as part of SGIG, representing 32% of all smart meters in the United States. As of December 2014, there were approximately 50 million installed AMI meters, representing 37% of all electric meters. Figure 4-6 shows the number of meters installed and the penetration of AMI in the NEEP Forum states.[14]

Figure 4-6: Total Meters, AMI Meters, and Percent Penetration Rates in NEEP Forum States (EIA)



Statistical analysis of more granular data from smart meters may enhance the ability to separate heating, cooling, and base loads with increased precision compared to analysis of monthly consumption data. With increased granularity new variables can be included in consumption models, such as day types (weekends, holidays), peak hours, and consecutive hot or cold days. These expanded model specifications enhance the ability to analyze and predict building

AMI data, at 15-minute or hourly intervals, has the capacity to measure demand at the time it occurs and may yield greater visibility to smaller loads.

performance. Combining daily meter and temperature data reveals the relationship between degree-days and HVAC loads when they occur, rather than averaged over 30 days. Finer interval data also can reveal insight into differences in building performance in mild versus extreme conditions. However, finer resolution does not improve the ability to answer many questions, such as estimating the persistence of energy savings.

AMI data can be useful for grouping program participants or market segments by usage profiles. DNV GL recently completed a study that compared monthly to hourly billing analysis data in a whole house retrofit program evaluation in California.[15] The study assessed whether hourly data improved the valuation of energy efficiency, using methods acceptable for impact evaluations for forward capacity markets and avoided generation and emissions.

The results from the AMI data were more consistent with savings expectations based on energy conservation measures and climate than were the results from the monthly analyses. The study concluded that combining account-level AMI data for residential and small commercial buildings with detailed engineering, statistical, and building modeling methods enables estimation of some HVAC loads with a higher degree of precision at a much lower cost than conventional end-use metering approaches. In this study, hourly data produced 730 times the volume of the monthly data.

Potential Implications for EM&V: AMI data are used for daily consumption analysis, and for DR impact analysis and energy program load shape impacts using hourly data. M&V for both program support and evaluation can make use of hourly whole-premise load data for individual facilities to examine usage patterns and changes in usage patterns, and to calibrate simulation models.

Key issues for wide-scale use of hourly data include data validation, estimation, editing (VEE, or data cleaning) and data volume. More work is needed to determine best practices for using daily and hourly AMI data for evaluation. However, combining survey methods and site visits remains important whether using AMI or monthly data.

Monthly consumption analysis is generally recognized as an appropriate evaluation method for certain types of programs—in particular, those with relatively homogeneous participant groups and with savings a relatively large fraction of pre-installation load. Use of daily rather than monthly data via AMI systems, and incorporation of supplemental premise information from existing utility or third party databases, have the potential to extend the applicability of these methods to smaller proportions of savings. AMI data also presents the possibility of using similar analysis to determine peak impacts. The need for meaningful comparison groups and relatively homogeneous populations, as discussed in Section 3.1, is not mitigated, however.

AMI data is not without drawbacks, notably the volume. Historically, AMI data may be the largest volume of data collected and managed by utilities.[16] For example, a small study using two years of monthly data and analyzing six variables for 10,000 premises produces 1.5 million data points. The same analysis using hourly data produces 1 billion data points. Due to the volume of data, many analyses require cloud-based computer platforms and off-site data storage. Many utilities are not accustomed to placing data repositories outside of utility direct physical control.

The key characteristics of smart meters and AMI related to customer consumption data analysis are listed in Table 4-2. AMI systems also provide key metering data and communications related to utility grid operations.

Table 4-2: Key Characteristics of Smart Meters and AMI

Function	Smart meters range from basic hourly or finer interval meters to real-time meters with built-in two-way communication capable of recording and transmitting instantaneous data.
Significance	They make hourly or 15-minute consumption data possible for large volumes of customers.
Current Uses/Value	They facilitate dynamic load profiling, time-differentiated impact estimation for EE programs, short-term forecasting, and can inform rates.
Challenges for Utility	Challenges for utilities include data transmission, data storage, and data processing. Utilities will have to increase IT infrastructure or outsource data management, which is seen as undesirable.
Impact on Project and Program M&V	Most programs are not equipped to handle the data volumes from AMI data unless they use outsourced services. When available and analyzed, AMI data can increase the visibility of EE behavior. They offer the ability to estimate overall changes in kWh use by time of day.
Impact and Challenges for Evaluation	Evaluation can use AMI data to produce savings estimates by time of day, and to develop higher resolution energy savings models. Evaluation challenges include data access and data quality. Data silos are the status quo within utilities; it is often difficult to transfer data between grid operations and program M&V or evaluation functions. There can be data losses in transmission from meter to collection point.
Ethical/Regulatory Issues	Privacy and data security.

4.4.2 Smart Thermostats and Devices

Technology Overview:

Customer Services: New smart devices are coming to the utility and consumer market at an amazing pace. Some of these devices, such as the NEST smart thermostat, are moving into the mainstream but the market is still serving early adopters. These devices help consumers in numerous ways, including, but not limited to, security, comfort, and safety.

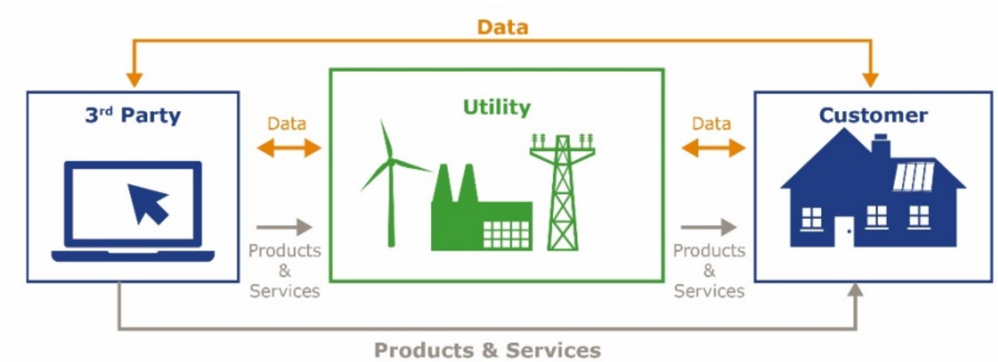
Figure 4-7: Ecobee Smart Thermostat

Smart thermostats are an important part of the “customer experience” market for Apple and Google. Smaller but worthy competitors are Honeywell, Ecobee, and EnergyHub. Researchers have long known that consumers do not operate manual programmable thermostats correctly and they were not an effective energy saving strategy. As a result, the Environmental Protection Agency (EPA) suspended them from the ENERGY STAR® program in 2009,[17] however; it is currently developing a new specification for connected thermostats.[18] Using machine-learning techniques, smart thermostats replace manual controls with a customized, agile, and automated system. The thermostat generates savings if it reduces household energy consumption compared to household consumption before the thermostat was installed. Algorithms trained to learn customer preferences control HVAC equipment.

Sensors relay data to third party providers over Wi-Fi. Multiple temperature and occupancy sensors learn household schedules, monitor indoor and outdoor conditions, and set back thermostats accordingly. They know when occupants return home and set the indoor temperature to their comfort zone. Smart thermostats use input from weather forecasts (e.g., they know when the polar vortex will arrive and will likely skip the setback that evening). Users can program them from smartphones and tablets, and review their energy history even when they are away from home. The system will also "learn" new routines, such as changes in users' schedules or the need to have the house a few degrees warmer.

Program Services: These devices also support customer-enabled load reduction programs, such as dispatchable demand response programs. A common design for demand response programs employs direct load control using the smart thermostat as the control device. The utility contracts with third party providers (i.e., EcoFactor, EarthNetworks, Ecobee, Honeywell, and Nest) who signal and control thermal loads via the thermostat following a signal from the utility. Figure 4-8 shows the smart device data flow between customers, third parties, and utilities.

Figure 4-8: Smart Device Data Flow (NARUC)[19]



Although smart thermostats have been used in various demand response programs for years, the two-way communicating and web/smart phone-enabled thermostat is creating a more robust interactive vehicle for utility (or third party) customer engagement. These devices receive and transmit customer and equipment data 24 hours a day, seven days a week. Gathered intelligence from the thermostat or other smart device creates new outreach opportunities for various parties. Utility rebate programs are pushing smart thermostats beyond the coolness factor and accelerating adoption across more diverse customer segments for direct control for demand reduction programs.

Potential Implications for EM&V: While there are expectations that smart devices and their associated data are valuable EM&V tools due to their capabilities and the data collection potential, significant barriers exist today that limit their potential. While the ability for utilities or other third parties to tap into the smart device communication stream to garner meaningful and insightful information would be a valuable asset, customer privacy, data ownership, and right to access the data are key issues.

From the utility standpoint, EM&V of programs utilizing smart devices could be vastly simplified and significantly less expensive if the utilities had access to customer-identified site level data produced from the thermostats or smart device. This data could answer questions about customer overrides, days or time of days with fewer or more overrides, and changes in interior temperature based on curtailment. This data is valuable to both the utility and the third party provider, and both parties have a stake in maintaining control, or at least, having access to the data. A current lack of standard communication protocols, which would allow for cross communication among devices, i.e., the Internet of Things, however, limits the ability of sharing information between devices.

Currently, utility access to customer-level data produced from smart devices varies across vendors. NEST's privacy policy states, "We may also share *non-personal information* with our partners, for example, if they are interested *in providing demand-response services or other incentive programs*." The key phrase in this statement is "non-personal information."[20] NEST and the other thermostat manufacturers are motivated to collaborate with utilities because it accelerates their market penetration and will earn them small fees from transmitting the curtailment signals. However, they consider those payments compensation for access and are not relying on utility fees as a revenue stream. Smart thermostat characteristics to consider are provided in Table 4-3.

Table 4-3: Key Characteristics of Smart Thermostats

Function	Smart devices can include ANY energy consuming device. They interface with a micro-computer or hub and communicate in multiple (N) directions. Some devices can be electrified like deadbolt locks for houses/cars.
Current Uses/Value	They offer centralized/easy control of household devices and "coolness" or novelty factor.
Current Challenges	Main challenges are the lack of common communication protocols and the inability to access customer level data from some smart thermostat vendors.
Challenges for Utility	Challenges include lack of standard communication protocols, data access, that data is not actively collected or coordinated, data volume, and data consistency. Efficiency is an afterthought—consumers purchase for convenience, comfort, control, security, and "coolness" rather than efficiency.
Impact on Program and Project M&V	Interconnectivity of all devices in a home will allow live experiments such as push intervention to specific households and monitoring data feeds to verify action.
Impacts and Challenges for Evaluation	Traditional methods and analytics might not handle data volume and complexity; we need to figure out how to leverage smart phones as a communication vector.
Ethical/Regulatory Issues	The main regulatory issues are data ownership, similar to HIPAA challenges regarding data privacy.

4.4.3 Other Smart Devices and Home Energy Management Systems

Technology Overview: Home energy monitoring or management systems and/or solutions (collectively HEMS) are defined as "any hardware and/or software system that can monitor and provide feedback about a home's energy usage, and/or enable advanced control of major and minor energy loads in the home."

HEMS present a unique opportunity for the customer, the program administrator, and implementers by collecting and reporting real time feedback on site level (and potentially population's) energy use. In addition, while it has not been a primary focus for the HEMS industry, more products, such as smart thermostats and their vendors, are incorporating automated M&V into products and/or technologies for feedback to both customers and program administrators.[21] For example, customers can be provided potential energy savings information in near real time by the HEMS system using deemed savings and measured run times for controlled products.

The potential goes beyond whole house energy management systems and smart thermostats. Manufacturers are adding communications and control ("smart") technology to a wide array of powered devices (e.g., appliances, LED lamps, electronics) and devices not typically thought of in this space (e.g., wall switches, deadbolts). As more devices are made smart, each device gains the potential to self-report its own energy use and operations. A potential scenario is that the HEMS or smart thermostats can serve as a communications hub for the various smart devices in a home, and relay all of this information to third party providers or utilities.

Some smart device providers will only release customer data in the aggregate which is perceived as a lost opportunity for replacing or supplementing expensive, direct metering.

Potential Implications for EM&V: While the case for leveraging HEMS technology for EM&V is clear, the opportunity to leverage the technology for independent, third party evaluation studies is murky. The HEMS technology certainly presents the opportunity to reduce the data collection burden by transmitting the data at the whole house or widget level in real (or near real) time, but a number of barriers exist that could limit the usefulness of the data for both program implementers and evaluators. These barriers include:

- Quality control: The data stream from HEMS does not undergo any significant quality checks or controls. HEMS data will require significant quality control procedures and data preparation, which will be time consuming due to the sheer volume of data.
- Access to data: In many cases, data collected by third party HEMS providers is held by the third party providers and is not readily available to program administrators to be used for EM&V. HEMS, as well as smart thermostats and devices, are vehicles to create an improved customer experience. As such, many providers are not inclined to provide customer level data to the utilities and this is not likely to change.
- Breadth of data: It is likely that smart devices can and will record much broader data streams than utilities want or need. It will be important for utilities to work with device vendors to find solutions to acquire only the data fields that are necessary for EM&V.

- Baseline conditions: If the program measure of interest is occasional program-initiated control, the HEMS can provide data for periods with and without control events, as a basis for evaluation. If the program measure of interest is the installation of the HEMS itself, the system cannot provide data for the period prior to its installation, which would be the likely baseline of interest.
- Comparison group definition: If the program measure of interest is the installation of the HEMS itself, evaluation would require a comparison group of otherwise similar homes that did not install the HEMS.

While HEMS can be a valuable source of information, it should not be viewed as a tool that can replace standard EM&V. Information provided by HEMS and other smart devices can enhance EM&V or reduce additional data collection requirements, but these tools do not provide enough information on their own to sufficiently estimate a program's impact.[22] Table 4-4 and Table 4-5 list the key characteristics of smart devices and HEMS.

Table 4-4: Key Characteristics of Smart Devices

Function	**Smart devices can include ANY energy consuming device. They interface with a micro-computer or hub and communicate in multiple (N) directions. Some devices can be electrified like deadbolt locks for houses/cars.**
Current Uses/Value	They offer centralized/easy control of household devices and "coolness" or novelty factor.
Current Challenges	Main challenges are lack of common communication protocols.
Challenges for Utility	Challenges include lack of standard communication protocols, data access, that data is not actively collected or coordinated, data volume, and data consistency. Efficiency is an afterthought—consumers purchase for convenience, comfort, control, security, and "coolness" rather than efficiency.
Impact on Program and Project M&V	Interconnectivity of all devices in a home will allow live experiments such as push intervention to specific households and monitoring data feeds to verify action.
Impacts and Challenges for Evaluation	Traditional methods and analytics might not handle data volume and complexity; we need to figure out how to leverage smart phones as a communication vector.
Ethical/Regulatory Issues	The main regulatory issues are data ownership, similar to HIPAA challenges regarding data privacy and ownership.

Table 4-5: Key Characteristics of HEMS

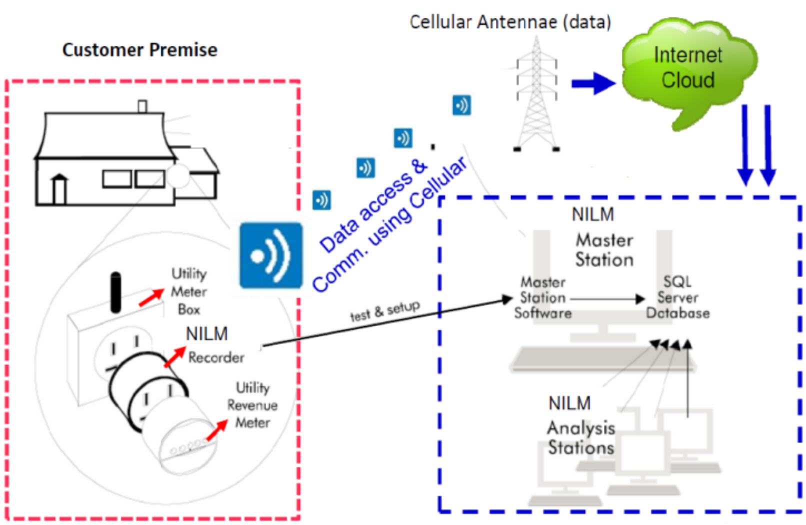
Function	**HEMS are hardware and/or software systems that can monitor and provide feedback about a home's energy usage and/or enable advanced control of energy-using systems and devices in a home.**
Significance	They offer energy efficiency potential, demand response potential, M&V potential, and distributed-energy resources control potential.
Current Uses/Value	They are fun gadget for early adopters, offer comfort (smart Thermostat) and security (installed through home security system), and some energy savings.

Challenges for Utility	Challenges include validation of savings, emerging technology, data access, and protecting consumer data (not an actual challenge, but public perception is concerned about it).
Impact on M&V	HEMS require interval data collectors to confirm the savings of other measure in a home. Systems can interact with and report on a wide range of measures (e.g., lighting, plug load, appliances, and HVAC) at no added cost.
Impacts and Challenges for Evaluation	Limited products currently ready to be deployed for M&V, limited customer penetration, new product area, control devices and thus not inherently efficient, making deemed savings a challenge. Baseline and comparison group issues for assessing overall energy impacts of device installation
Ethical/Regulatory Issues	Concerns over homeowner data privacy and protection are the main issues.

4.4.4 Intrusive and Non-Intrusive Metering

Method Overview: Intrusive methods of direct load measurement place data loggers, meters, and sensors on building systems and household appliances to capture energy consumption and customer demand on a high frequency basis. End uses can also be measured at the electric panel if it is isolated on a single circuit. However, traditional customer-facing EM&V using direct measurement is time consuming, expensive (between $10k and $15k per household), intrusive, and requires a high degree of cooperation from the customer.

Figure 4-9: Utility Grade NILM Data Collection and Data Analysis Infrastructure



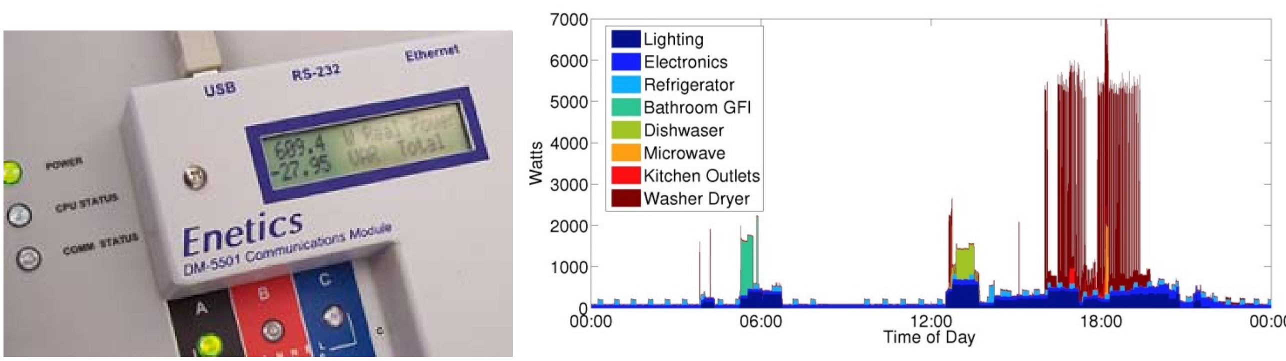
Alternatively, non-intrusive load metering (NILM) is a process for analyzing changes in the voltage and current going into a building or the run times of in-house systems, and deducing what appliances or equipment are in use and then measuring their respective energy consumption. NILMs are meters that record high frequency[23] signals at the building level, compute granular[24] level load profiles, and identify time-stamped transitions in energy consumption.[25] These load profiles and load transitions are stored in the NILM and either transmitted to a central system using some form of IP-based communication periodically[26] or collected with a laptop computer. By definition, utility researchers use NILMs because they don't want to "intrude" on the customer. To compensate for this, utilities may release detailed load shape libraries that are calibrated for a given region or household type rather than a specific customer.

There are two types of NILMs. The first are consumer products, which provide customers with a fuller understanding of their energy consumption. Most consumer NILMs operate over the internet and can include the aforementioned smart thermostats, smart devices, and HEMS. In theory, these devices could transmit customer energy data back to the utility or vendor, but as discussed above there are currently numerous issues preventing this transfer.

At the other end of the spectrum are utility grade hardware and analytical software systems, designed with the utility researcher in mind, providing quality end-use data at a reasonable cost. The process used by consumer grade NILMs is shown in Figure 4-10. Utility grade NILMs have been studied for many years but are the least mature of all the technologies discussed in this paper, but there are efforts to begin piloting NILMs in the near future.

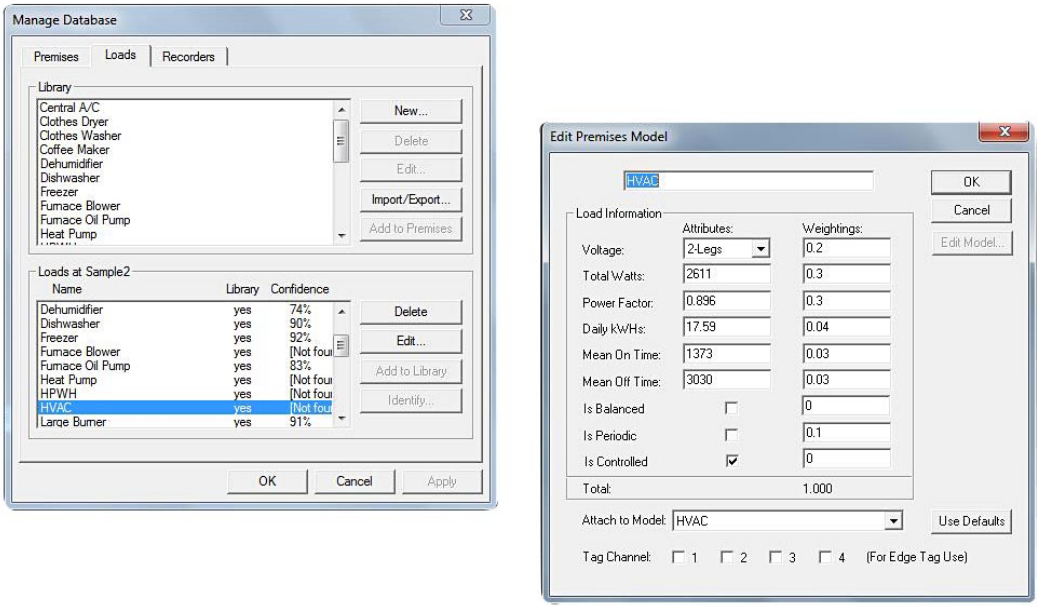
Consumer technologies have an edge in labeling loads as most allow the end user to train the system by turning loads on and off to label the loads directly. An example of a training procedure is shown in Figure 4-10 (right). Utility grade systems do not have this luxury, though some are using direct measurement (intrusive sensors and loggers) to capture loads to reduce the uncertainty and create regional load libraries.

Figure 4-10: Utility-grade NILMs and Sample Output with Appliance Load Disaggregation (Right)



An emerging technology in the NILM category uses harmonic signatures to identify end uses. Although promising, it is intrusive because individual loads must be turned on and off during the labeling (training) process.

Figure 4-11: NILM Training Period Procedure



Equipment and installation cost for a consumer device along with the required cloud-based data storage and mobile applications range from a one-time fee of $200 to $300 to monthly fees based on functionality. Utility-grade systems are much costlier, costing around $1,500 for the measurement device and additional licensing fees for the vendor's data collection and analytical software. Installation costs are dependent on whether the device is installed at the customer panel or at the utility meter where the measurement device sits between the utility meter and the customer service panel.

Potential Implications for EM&V: Utility grade NILM vendors using proprietary algorithms process the data and convert it into appliance level load shapes. The output available from the various technologies depends on the sophistication of the measurement device, the intricacy of the processing algorithms, and the ability to correctly "label" a load. However, if the M&V task can sacrifice accuracy in exchange for reduced costs and burden on the customer, the NILM will be significantly less expensive once the technology matures.

The key characteristics of NILM are listed in Table 4-6.

Table 4-6: Key Characteristics of NILM

Function	**This technique can identify loads from specific devices inside building without installing separate meters on those devices.**
Current Uses/Value	It is used in load research, DSM program design, and grid stability models.
Challenges for Utility	Data transmission, installation, and coordination with metering department for some products are some of the potential challenges for utilities.
Impact on Project or Program M&V	Multiple loads could be captured from a single device, for assessment of multi-measure projects or programs, at lower cost.

Function	This technique can identify loads from specific devices inside building without installing separate meters on those devices.
Challenges for Evaluation	Challenges include accuracy and predictability of end-uses captured. Technology is still in early stages of deployment.
Ethical / Regulatory Issues	The main issues are privacy and liability if installed on the customer side.

5 SUMMARY AND RECOMMENDATIONS FOR EMERGING EM&V PRACTICES

This section presents findings recommendations and considerations related to the key trends in the changing EM&V paradigm and the implications new industry developments have on current and future EM&V practices and activities.

Overall, the trends in new methods and tools for data collection and analysis hold great promise for program management and optimization, M&V and EM&V. Emerging software, hardware, and data availability are still evolving and should continue to be monitored by EM&V industry stakeholders. That said, our research indicates that the new tools and methods will enhance data collection, data availability and data processing efforts.

The recommendations and findings characterize the trends in data collection and analytics with the purpose of furthering stakeholders' understanding of how these approaches can be leveraged for EM&V, as well as discussing the limitations of these new tools and techniques. In particular, the research examines how and to what extent the enhanced data and new tools can help address stakeholders' concerns that standard EM&V procedures are costly and results take a long time to produce.

5.1 Broad Themes Emerging from This Review

The new tools, services, sources, and processes explored in this paper have a wide range of capabilities and potential implications for EM&V. Certain broad themes emerge from this work:

1. Advanced data collection and analysis tools and systems offer new opportunities for understanding and engaging customers. These capabilities can offer value for project identification and delivery, program planning and operations, and evaluation.
2. There remain important evaluation challenges that are not solved by greater volumes or frequency of consumption data, or higher speeds of data processing. These challenges include estimation of appropriate baselines, analysis of complex projects and processes, and assessing market conditions and program influence.
3. Tools such as automated M&V can provide valuable early and ongoing feedback to customers and programs; however, their ability to reduce overall evaluation timelines is limited, and their cost effectiveness is currently unknown.

These themes are developed further in the remainder of this section, and more fully in the main text.

5.2 Framing Issues

5.2.1 Evaluation Independence

Energy program evaluation is required to be conducted by an independent third party in almost all jurisdictions. This requirement is to protect the credibility of evaluation results. While evaluators will have individual perspectives on programs and methods, they do not have a direct stake in the evaluation outcomes in the ways that program implementers and administrators do. The specific restrictions and roles of independent evaluation vary across jurisdictions. For purposes of this paper, it is assumed that the principle and policy of independent evaluation is retained in the future. Absent such a requirement, program evaluation could be dramatically different, if it continued to exist at all, apart from the implications of new tools explored here.

5.2.2 Baselines

5.2.2.1 Baselines for Different Contexts

A key question for impact evaluation (determination of program savings accomplishments) is the baseline. The baseline represents what would have occurred without the installation of the program measure (gross savings) or without the influence of the program (net savings). Different circumstances inform appropriate baselines, as described below:

- In a retrofit context, where the existing equipment might have stayed in place indefinitely, and no other major changes, the baseline corresponds to the *existing equipment.*
- If equipment was going to be replaced regardless of the program, the baseline condition corresponds to *the new unit that would otherwise have been installed.*
- If other equipment or operating practices change at the same time as the installation of the program measure, the baseline corresponds to *the presence of the other new equipment and operations, but without the program measure.*
- If the equipment was going to be replaced in a year or two but the program accelerated the replacement, the baseline corresponds to the *existing equipment up to the time it would have been replaced and new non-program equipment thereafter* (commonly referred to as a "dual baseline."
- In a new construction context, the baseline corresponds to the non-qualifying premise that would otherwise have been built.

Only in the first case of retrofit installation with no other major changes to the facility and its operations, is the baseline observable based on pre-installation conditions. For new construction there is clearly no pre-installation baseline to observe. However, even when pre-installation data exist, these do not define the baseline for most contexts.

Fully specifying the baseline for gross or net savings calculation requires a combination of technical information and assumptions together with policy guidance. The specific policies and assumptions for defining baselines vary by jurisdiction, but the combination of these considerations means that, even adjusting for changes in weather, direct comparison of participants' pre-installation with post-installation consumption does not provide savings consistent with common baseline definitions in most contexts.

M&V is designed to estimate the effect of installing the program-supported measure within the facility, often referred to as "gross savings." Net savings adjustments, including free ridership, spillover, or market effects are not accounted for in facility-level M&V. Program-level impact evaluation in many jurisdictions and contexts requires determination of net as well as gross savings. Large-scale consumption data analysis using a comparison group can produce gross or net savings, or something in between, depending on the composition of the comparison group.[27] In the context of individual facility M&V using whole-facility consumption analysis, standard M&V practice includes custom adjustment ("non-routine adjustment" [28]) as needed for other conditions that changed at the premise. Thus, if the pre-installation condition alone does not provide the appropriate baseline post-installation, an explicit adjustment based on engineering information and principles may be applied.

In the context of program-level evaluation using large-scale consumption data analysis, the recommended evaluation practice uses a comparison group to account for typical non-program-related changes occurring in a similar population. This approach requires establishing a comparison group of customers that are similar in terms of relevant consumption drivers, but have not participated in the program. To the extent the comparison group used is not a good representation of how participants would have used energy absent the program or program measures, the resulting savings estimate will be biased. Defining a comparison group that will be recognized conceptually as being that good representation, and can be identified in practice is one of the challenges for applying this method.

5.2.2.2 Evolving Baseline Perspectives

Recently enacted legislation in California, AB802[29], makes several changes to the landscape of energy efficiency policy in that state. Two key provisions are:

1. Utilities programs may include actions to bring facilities up to code, as well as beyond code, rather than counting savings only relative to existing code.
2. Programs are to support improvements in existing buildings "taking into consideration the overall reduction in normalized metered energy consumption as a measure of energy savings."

These provisions could be interpreted to imply that the pre-installation condition is always the appropriate baseline, and that the weather-normalized change in consumption is the preferred measure of savings. However, in implementing the new law, the California Public utilities Commission is taking a more measured approach. Opportunities for to-code measures are currently being explored, and change in normalized consumption is recognized as a measure of energy savings, but not necessarily the only or preferred measure.

Other states may also widen the definition of savings that may be captured or counted by utility programs. It is still to be expected that baselines will often be defined not by the previously existing condition, but by some higher level of efficiency corresponding to natural replacement. EPA's draft EM&V Guidance for the Clean Power Plan[30] requires use of a "Common Practice Baseline" (CPB) representing the "default technology or condition that would have been in place at the time of project implementation absent the EE installation." For natural replacement situations, the CPB corresponds to market average efficiency or formal building or product standards, and for early replacement a dual baseline is specified.

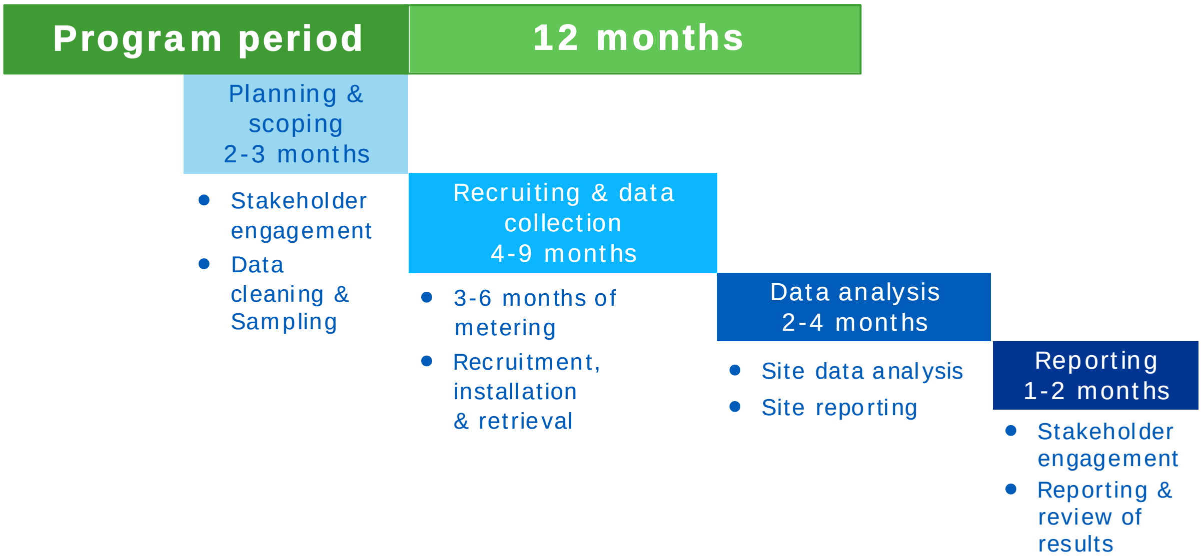
5.2.3 Evaluation Timeline and Potential for Streamlining

The time required to provide evaluation results currently includes:

- Planning time to understand programs and data available, and agree on research plans, instruments, and protocols with evaluation clients and stakeholders
- Data collection time, including developing sample designs, recruiting respondents for customer or vendor surveys for onsite data collection, and obtaining the consumption data itself
- Data analysis time
- Reporting time, including responding to client and stakeholder feedback.

Figure 5-1 below presents an illustrative schedule for a typical impact evaluation. Depending on extent of stakeholder input to the draft and final evaluation results, as well as the types of data being collected and analyzed, the overall timeframe could go beyond the 18 months indicated in the figure.

Figure 5-1: Illustrative Evaluation Timeline

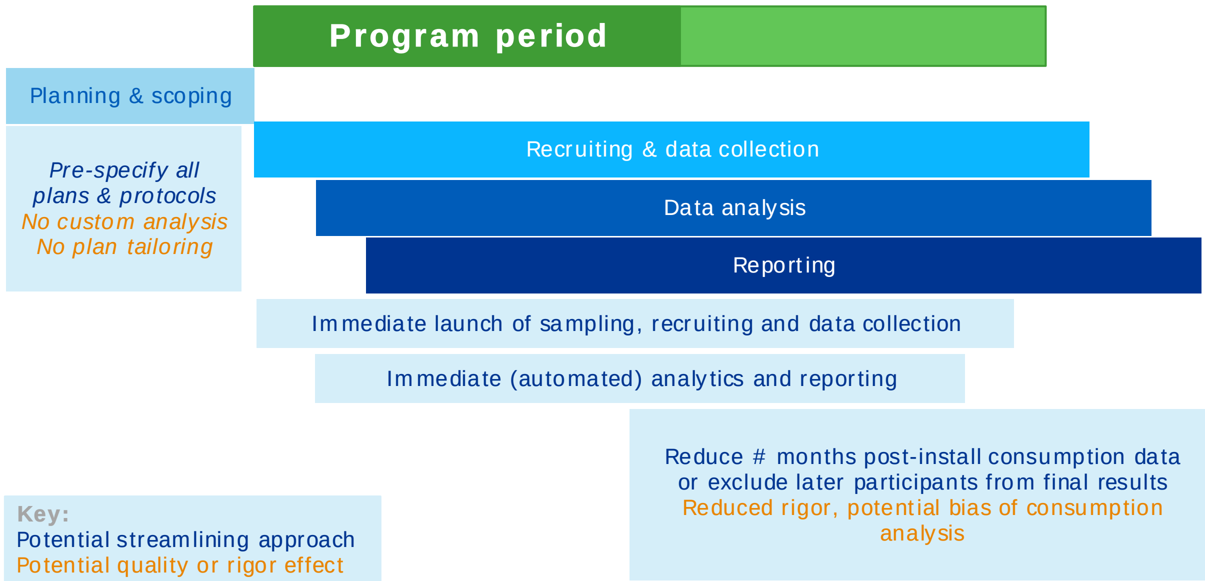


The total time to conduct an evaluation can be shortened by agreeing, prior to the program start, on fully specified research methods, instruments, and protocols, and by implementing some form of continuous sampling, analysis, and reporting. As an example, some jurisdictions have established questionnaires and analysis frameworks for free ridership and spillover assessment that are implemented on a regular basis and updated periodically. On the other hand, more comprehensive evaluations typically require the flexibility to re-allocate resources based on issues that have emerged for the program, and to explore unusual cases. This opportunity is given up if a fully pre-specified approach is followed.

There are other limits to how much the evaluation cycle can be compressed. If pre and post-installation energy consumption data analysis is to be the basis for evaluated savings, 9 to 12 months of post-installation consumption data is typically needed to capture all the seasonality of annual savings. If all program participants are to be included in the evaluated savings, this analysis cannot be finalized until close to a year after the program close, even with immediate data processing and reporting. In addition, many jurisdictions have broad stakeholder processes that entail substantial time for review at several stages of the evaluation. Timelines cannot be substantially compressed without curtailing these processes at some level.

The potential time reductions using streamlining procedures are illustrated in Figure 5-2. Automated data collection tools and automated analysis tools can be employed in such streamlining. However, neither is essential to adoption of many of these streamlining methods.

Figure 5-2: Illustrative Evaluation Streamlining Potential



5.3 Key Findings and Recommendations

5.3.1 Potential Roles of New Tools and Methods

As described in Section 3.1, the current EM&V framework involves various methods, in some cases complicated baselines, and evaluation timelines that can take multiple years to complete. There are several potential roles for automated M&V and advanced data collection methods to improve and streamline current EM&V practices. These opportunities are summarized below, with specific findings and recommendations following:

- Automated energy consumption data analysis for an individual facility could be used as a form of M&V, consistent with whole-facility analysis under IPMVP option C.
- Automated energy consumption data analysis for a group of facilities with an appropriately defined comparison group could be used as a form of large-scale consumption data analysis.
- Automated Metering Infrastructure (AMI) whole-premise metering at higher frequency than monthly can be incorporated into whole-facility M&V, or into large-scale consumption data analysis.
- End-use energy or demand metering, or estimated end-use decomposition of AMI data can be incorporated into end-use M&V.
- Metering of other parameters such as temperatures or flows from advanced control or monitoring systems can also be incorporated into M&V.

The findings and recommendations are organized to address the primary research questions under the two broad areas of **advanced data analytics** and **advanced data availability**:

1. Advanced data analytics/program enhancements with real-time results—using **automated M&V** in evaluation, to reduce the time and costs to develop evaluation results
2. Advanced data availability/leveraging new data—**improved data collection tools and data availability** to support evaluation

In addressing these broad areas, the paper discusses benefits, challenges, the potential role of standardization, and ways that program administrators and regulators can prepare for the coming changes.

The recommendations cover a range of topics. However many are focused on how EM&V stakeholders and practitioners can continue to assess the opportunities represented by the emerging tools and methods, and develop key protocols and principles for their application. The key findings and recommendations are presented next.

5.3.2 Automated M&V

5.3.2.1 Key Research Findings

Comparison of Automated M&V to Existing Evaluation Methods

For calculation of savings, current automated M&V tools use weather normalization methods similar to those that have been used for program evaluation for many years, with similar performance. The automated tools can process all program participants, other than those dropped due to data anomalies. Comprehensive analysis of a full program and a corresponding comparison group is also frequently done in traditional evaluations.

A key difference of the automated applications is the rapid and ongoing analysis. On the program delivery side, this early and continuing information provides opportunity identification, early warning of systems not working as anticipated, with the potential for correction, as well as a means of ongoing customer engagement with the implemented measure. On the evaluation side, fast feedback is an opportunity to share and explore information with the program, and to investigate unusual situations from and evaluation perspective, prior to the final stages of analysis. Thus, fully taking advantage of the opportunities offered by automated M&V requires non-automated examination, exploration, and follow-up on the emerging results.

Current Uses

Deployments of automated M&V are currently in the pilot stages. Most current applications are operating in environments where AMI is not deployed and hourly data are not available for most customers. These applications are using monthly consumption data, especially for smaller customers.

Primary uses currently being promoted for automated M&V are:

1. Program targeting and opportunity identification through remote assessments and audits. This is accomplished by combining information from other sources with the consumption analysis, to identify high usage relative to building type and size.
2. Rapid and continuous feedback to customers and programs on the changes in facility consumption after measure installation.

In both types of current applications, these services are provided to the program administrator, and are not used to support claimed savings reported to commissions.

Capabilities

Once an automated M&V installation is in place, automated analysis of energy consumption data can begin providing feedback to programs within one month of measure installation. Delivery of such automated analysis requires advance agreement and the implementation of methods for ongoing access to program and consumption data.

Automated analytics vendors report excluding up to 20% of accounts based on data anomalies. This loss rate is not viewed by vendors as a problem for opportunity identification, since there remain are large numbers of cases that can be fruitfully targeted, and for which automated post-installation reporting can provide customers and programs with useful information. If the automated M&V tools are to be used for comprehensive program evaluation, the effect of such exclusions on savings estimates would need to be explored.

A disadvantage noted above of highly automated, rapid analysis for evaluation is the loss of opportunity to customize the analysis in response to changing program conditions or emerging findings. High quality evaluation typically requires the flexibility to re-allocate resources based on issues that have emerged for the program, and to explore unusual cases. With feedback beginning with the first program participants and continuing throughout the program period, there is time to investigate and understand anomalous situations as they arise, provided the evaluation isn't limited to the automated analytics. Automated analytics designed to be applied for large volumes of data in diverse contexts also can bring the advantage of more thoroughly tested and vetted analytic software.

Performance

The performance of automated M&V tools has recently been assessed for their accuracy in predicting whole-building consumption based on historical consumption. Findings from these assessments are generally consistent with prior studies of monthly consumption data analysis using variable degree-day models.[31] In particular, twelve months of pre-installation consumption data provides somewhat better accuracy than nine months, but nine months or even six may provide unbiased results as long as all seasons are represented in the data, or if the facility has no seasonality of consumption. For individual customers, annual prediction accuracy based on 12 monthly observations is on the order of ±3 to 7% at 90% confidence. This means that a pre-post change on the order of 5 to 10% would be considered statistically significant for an individual building. Because consumption and consumption changes are themselves highly variable across customers, the average level of savings that can be detected for a population cannot be directly inferred from these metrics. . Studies involving hundreds of premises have provided statistically significant estimates for savings on the order of 5 percent of consumption.

Current research has not yet addressed the ability of these predictive models to measure energy savings *after measure installation*, but future research planned by LBNL will address this issue. These two research efforts address very different capabilities as the accuracy of predicting whole building energy consumption is not the same things as accuracy of *measuring savings* for several reasons:

- Other changes typically occur at the same time.

- If both pre and post consumption data are weather-normalized, the savings accuracy must reflect statistical error in both. If only pre-installation data are weather normalized, the measured savings represents only the first year, not the long term savings.
- If savings are being determined for a population based on calculations for a sample, the variability in consumption and change in consumption over the population affects the accuracy of the estimated program average.

Transparency

Transparency is a key principle for M&V and for evaluation.[32] Under current practice, evaluators are required to provide full specifications and documentation of modeling methods, equations, assumptions, and tools used.

Many details of the software implementations for commercially developed tools are proprietary because the tools are in an early stage of implementation. However, the majority of vendors of automated M&V tools have stated that they will provide full transparency of the equations, and the process of constructing a comparison group where relevant. At the same time, vendors wish to be able to retain some proprietary features of their tools, including some targeting and outreach analytics, and software coding details.

To a certain extent, the accuracy of proprietary tools can be tested via protocols such as those developed by LBNL in their recent and continuing work. A tool may be determined to be acceptable as a basis for providing final evaluated results based on results of such a performance test. However, empirical performance in a particular test environment may not by itself ensure confidence in the tool results for purposes of evaluation. One consideration is that “black boxes” are not easily trusted regardless of performance metrics. The greater the transparency of methods and assumptions, the more readily a tool is likely to be accepted by regulators and stakeholders. A second issue is that a performance test is most convincing if results are available specific to the conditions where the tool will be applied. Finally, some process may be needed to ensure that the version of the tool being used in particular context is substantively the same as the tool that was assessed.

Applicability

As described in Section 1.4.3, whole-premise pre-post installation consumption analysis is an appropriate basis for evaluated savings in limited contexts, where the pre-program consumption pattern is the appropriate baseline for savings, and where other changes in facilities can be accounted for. These contexts include:

1. Retrofit programs for large homogeneous populations, with a valid comparison group capturing average changes for participants absent the program, or absent the program measures.
2. Commercial retrofit programs, using analysis of individual buildings where non-program-related changes are known and their effects on consumption can are reliably quantified.

Large commercial sites or industrial processes are not likely to be well addressed by these methods.

Except in cases where comparison group methods provide net savings, determination of net-to-gross adjustment ratios, where required, will typically require other types of methods than consumption data analysis. These methods are based on surveys, market studies, expert interviews, case studies, and econometric regressions. While all of these methods could be expedited by establishing pre-agreed detailed research specifications, large scale rapid analytics is not likely to contribute substantially to expediting this research.

Cost Effectiveness

The cost-effectiveness of these methods as part of overall program delivery or evaluation is not yet established. These offerings are still in the early stages of development, and most applications are being conducted on relatively small scales.

5.3.2.2 Recommendations

Programs should continue to explore the benefits of the early feedback made possible by automated M&V. Estimated savings reductions from automated consumption data analysis can provide rapid feedback to programs whether or not this analysis is used as the final evaluated savings. Such rapid feedback is useful whether it is provided as part of program delivery or as part of evaluation.

Where programs are using these tools on the delivery side to monitor project performance, programs and evaluators should explore the potential time and cost savings of evaluator's making use of the automated analysis. Evaluators should take the automated M&V results into account in developing interim and final evaluations, including process as well as impact evaluation, to the extent practical. If evaluators are to make use of the automated M&V results themselves as a basis for evaluated savings, they should have access to the methods and results.

Programs should continue to explore the benefits of using automated consumption data analysis for opportunity identification or to engage customers, in comparison to existing targeting and customer engagement processes. Existing evaluation methods can be applied to support such assessment.

Making use of automated consumption data analysis to provide final evaluated results would require advance agreement on the analysis methods, including data screening and cleaning, supplemental data to be collected and used in the analysis, procedures for adjustments for non-program-related changes, and comparison group construction if any.

Evaluators and evaluation stakeholders should consider establishing automated consumption analysis as an accepted evaluation method that falls somewhere between customized M&V and a well vetted deemed savings process. Similar to deemed savings, automated analytics and pre-specified protocols have the potential to reduce evaluation costs and timelines, at the expense of less thorough or customized analysis. Less customized analysis means reduced capacity for adapting to changing program and market conditions, or to new methodological insights; reduced capacity to explore and learn from anomalous situations; and reduced ability to conduct tailored analysis where indicated by the results. Stakeholders should therefore consider the situations where the benefits of cost and time savings would outweigh these disadvantages.

Incorporating automated M&V tools into evaluation would require:

- Independent empirical assessment of the performance of the calculation tools in regional and market contexts similar to their planned applications. These assessments could be conducted using an agreed test protocol similar to those used in recent studies.
- Independent empirical and theoretical assessment of the accuracy and validity of any comparison group methodology

- Independent methodological assessment of the calculation methods, including any adjustments for non-program-related changes, and the implications of the automated data screening procedures for potential bias.
- Establishment of ongoing access procedures by the automated analysis software to consumption data and program tracking data
- Identification of the programs and contexts where this tool will be considered an acceptable evaluation method
- An agreed process for periodic re-assessment

Evaluators and evaluation stakeholders should continue to explore and apply other means to reduce the total time required to complete evaluation of a program, and to provide early feedback. These methods include early agreement on and ongoing deployment of fully specified data collection instruments, sample designs or continuous sampling protocols, analysis steps and programs, and reporting tables and templates.

For programs where pre-post analysis of consumption data is to be used as the basis for final evaluated results, a post-evaluation period long enough to capture all seasonal use patterns, if any, should be used. In most cases this will require 9 to 12 months of post-program data, consistent with current evaluation practice. As indicated above, standardizing other aspects of the evaluation process, and providing continuous analysis and reporting that begins shortly after a program launch, could still reduce the overall evaluation timelines and costs, and provide valuable early feedback to programs. Some aspects of these early feedback mechanisms are already incorporated into many evaluations, without automated M&V.

Cost-effectiveness of automated EM&V tools should be assessed taking into account the full set of benefits for program delivery and evaluation. Potential benefits include program targeting, opportunity identification, ongoing customer engagement, early program feedback, and evaluation applications.

5.3.3 Use of Improved Data Collection Tools and Data Availability to Support Evaluation

5.3.3.1 Key Research Findings

AMI Data

The primary opportunity offered to evaluation by AMI is the potential to obtain whole-premise consumption data at daily, hourly, or higher frequency, for all customers. Daily data from AMI systems are currently used in evaluations for energy consumption analysis, and hourly or finer data for demand-response (DR) impact analysis and energy program load-shape impacts. M&V for both program support and evaluation can make use of hourly whole-premise load data for individual facilities to examine usage patterns and changes in usage patterns, and to calibrate simulation models.

Two-way communications via AMI systems can interface with home or building control systems, for DR program implementation, and for program response monitoring and measurement. Very high frequency AMI data can be used with emerging Non-Intrusive Load Monitoring systems, to decompose whole-facility energy use into end uses.

Despite the opportunities and established uses, there are currently limitations for evaluation in the useful application of interval data from smart meters. A first limitation is that AMI systems are not deployed in all territories, and in some are deployed only to certain customer segments and geographies. In The Northeast region, AMI penetration is lowest among the largest states, New York and Massachusetts, and in New England, only Vermont and Maine have AMI in place. Many states across the country implemented AMI in recent years using ARRA grants. A similar source of funding is not on the horizon. Some states will pursue AMI; others are proposing implementation of "smart metering capability," as a substitute for AMI. In some cases states have not captured all of the depreciation on their existing metering infrastructure, and with the rate of technological innovation, there is some risk of investing in hardware that may be obsolete in three years.

Where AMI systems are in place, systems are not necessarily capable of providing ready access to the data by programs, evaluators, or researchers. Restrictions due to regulatory and utility customer data protection policies can also be a factor in data availability.

Due to its volume, AMI data requires a significant increase in IT and data management capabilities for all stakeholders. These include increases in data management requirements for utilities and evaluators including validation, estimation, and editing.

A challenge cited by evaluators and other researchers is that even where utilities have AMI data, the limitation exists in utility usage and approaches to get value out of the data. Specifically, the hourly or daily data is typically not subject to the same routine data validation and cleaning as is typically used for monthly kWh consumption data and kW demand data from revenue accounting processes. Smart meters have multiple data output and input channels and in many jurisdictions, the data stream used for account billing is not the same data stream used for EM&V.

Because of the current challenges of data management, data access, and data cleaning, use of AMI data is not necessarily a lower cost evaluation option compared to traditional evaluation approaches, even where AMI is already in place.

If these limitations can be overcome, there are significant opportunities for using more granular AMI data, supporting better estimation of simple consumption models, plus the ability to fit more complex models. The use of daily AMI data rather than monthly data, and incorporation of supplemental premise information from existing databases, has the potential to extend the applicability of these methods to smaller proportions of savings. AMI data also presents the possibility of using similar analyses to determine peak or time-based impacts that monthly data cannot provide.

Even without a full NILM deployment, AMI with machine learning or pattern detection and classification routines has the potential to be used to classify accounts based on load shape patterns. Once a set of load profile categories is developed, surveys can be designed to provide information linked to the load patterns and categories. Survey information could both validate conjectures about end uses and operations developed from the load patterns, and collect key parameters to be used in analysis of those uses. Thus, the AMI data can support both richer analysis of usage patterns and more informed survey design.

Other Granular Data

Smart devices and non-intrusive monitoring have the ability to provide more granular consumption data, and in some cases, at a lower cost than traditional measurement approaches. Where data collection devices are already deployed, access to data can avoid the costs of on-site meter installation specifically for evaluation. A potential scenario is that the HEMS or smart thermostats can serve as a communications hub for the various smart devices in a home, and relay all of this information to third party providers or utilities.

While these data sources are not at the stage where they can fully replace on-site metering and audits, they can provide useful information for calibrating results and supplanting data that is often difficult or costly to obtain in necessary quantities.

Barriers exist in the ability for utilities to collect customer level data from smart devices. Many smart device vendors have strict data protection and customer privacy policies. However, most providers will supply aggregate data to the utilities. For some of the largest vendors, this is unlikely to change. These policies reduce the ability for using smart devices for demand response and behavior programs, but aggregate data is still very valuable for measuring some utility energy efficiency programs.

For HEMS, other limitations to the use of smart devices for evaluation include:

- Baseline conditions: If the program measure of interest is the installation of the HEMS itself, the system cannot provide data for the period prior to its installation, which would be the likely baseline of interest. (If the program measure of interest is occasional program-initiated control, the HEMS can provide data for periods with and without control events, as a basis for evaluation.)
- Comparison group definition: If the program measure of interest is the installation of the HEMS itself, evaluation would require a comparison group of otherwise similar homes that did not install the HEMS.
- While HEMS can be a valuable source of information, it should not be viewed as a tool that can replace standard EM&V. Information provided by HEMS and other smart devices can enhance EM&V or reduce additional data collection requirements, but these tools do not provide enough information on their own to sufficiently estimate a program's impact.33

For NILM the challenges include the accuracy and predictability of end-uses captured. The technology is still in early stages of deployment.

5.3.3.2 Recommendations

Stakeholders should continue to support and develop the potential for AMI data to expand EM&V techniques as the availability of more granular data continues to increase.

Stakeholders should consider establishing guidance for best practices for how AMI data is incorporated into EM&V. Topics for consideration include:

- Identifying contexts where whole-premise consumption data can be the basis for savings estimates using monthly and daily data
- Development of hourly load impacts using interval data
- Development of data screening procedures
- Assessment of bias due to data loss and potential bias-mitigation strategies
- improving simulation calibration using whole-premise interval data
- Development of matched comparison groups using hourly load patterns
- Identifying the contexts where AMI analysis is likely to be a lower cost evaluation method than alternatives, and the associated quality effects.

The EM&V community should continue working with smart device vendors, consumers, and the utilities should optimize smart device data uses and work to improve data ownership issues to improve data access. The more restricted access remains, the less likely this data will be incorporated into large scale EM&V.

6 CASE STUDIES

Two case studies are presented here. Although one is commercial and the other residential, both fall into the category of utility or government program support. The first is an example of an automated M&V software application and the second illustrates the value of AMI data. Implemented by Retroficiency, case one describes two commercial programs: a geo-targeting remote assessment and customer engagement program for Con Edison, and a combined remote and on-site assessment program for an energy efficiency program in three Montgomery County Maryland Schools.

In the second case, Opower describes a residential demand response pilot at Glendale Water and Power that tested the effectiveness of behavioral demand response programs when implemented with and without home energy reports.

In each example, data analytics, automated M&V, early and continuous feedback and/or customized reports were used to engage customers to take actions to improve energy efficiency and to keep them engaged in energy efficiency actions and behaviors over time.

Company descriptions for the software vendors EnergySavvy and FirstFuel are also included. Both vendors are fully engaged with multiple utilities but it at the time of this writing, it was not possible to describe specific deployments in detail.

6.1 Retroficiency

This case study addresses two separate deployments of Retroficiency's software: Con Edison's "Targeted Demand Side Management (DSM) Program" and Maryland Energy Administration's "On Ramp Pilot."

Company and Product Overview

Bennett Fisher and Bryan Long co-founded Retroficiency in 2009 with "Automated Energy Audit" (AEA). AEA is an on-site energy audit software marketed primarily to energy service providers. In 2012, Retroficiency released "Virtual Energy Assessment" (VEA). VEA uses meter data to disaggregate end uses (i.e., HVAC, lighting, refrigeration, and other heavy equipment) to identify candidate buildings and systems for efficiency improvements. Retroficiency is a software and service provider to utilities and program implementers, and provides end-customer engagement services on behalf of the utility. "Efficiency Track" is Retroficiency's automated M&V software. Although it is a standalone product, it is often integrated with VEA to verify measure savings. Each product is customer facing and promotes customer engagement by identifying and encouraging energy efficient behaviors (http://www.retroficiency.com/).

Consolidated Edison

Con Edison has used geo-targeting strategies for many years. A description of Con Edison's long-standing program, which ends just before Retroficiency joins the project, is detailed in a prior NEEP report, "Energy Efficiency as a T&D resource."[34] In the example below, the objective was to determine if EE savings or DR programs could achieve more savings in less time and at reduced costs across a portfolio faster than with non-automated approaches. Con Edison's "Targeted DSM Program" focuses on reducing demand during peak periods in key constrained areas of the grid in Brooklyn and Queens (NY). Con Edison's objective was to identify customers with significant potential for kW reduction (as opposed to only kWh savings) using a multi-channel customer engagement effort.

Con Edison deployed Retroficiency's VEA software to find business customers with high savings potential and engage those customers with building-specific EE opportunities. Retroficiency's VEA software produced customer-facing reports to attract C&I program participation. These reports included estimates of energy savings potential, load disaggregation, and actionable capital and operational recommendations to achieve energy savings and reduce operating costs. It is believed that in some cases, information barriers prevent C&I customers from acting on energy efficiency opportunities. Virtual assessments (remote audits) are one way to overcome this information barrier without a site visit.

The first phase of the Con Edison project included more than 1000 accounts. The assessments also allowed Con Edison to rank buildings by demand reduction potential. In total, Retroficiency identified 23 MW of demand reduction potential during peak network periods.[35]

Retroficiency's assessment reports were delivered to customers by Con Edison, and those interested in pursuing projects worked with Con Edison engineers and/or local contractors to scope potential projects. Con Edison's internal deemed savings values and/or custom measurement approaches were used to determine savings from implemented energy conservation measures. Using this analytics-based engagement approach, Con Edison experienced a four-fold increase in project interest relative to their traditional sales and marketing results. VEA identified the highest potential projects first, which increased the pace and scale of demand reduction and energy savings for Con Edison.

Maryland Energy Administration

The "On Ramp Pilot" also used Retroficiency's "Virtual Energy Assessment" software but with some key differences in program strategy, design, approach, and goals. The pilot, conducted on behalf of PEPCO, by the Maryland Energy Administration (MEA), was focused on buildings with savings opportunities related to operational measures only (as opposed to capital). MEA believed that although low- or no-cost measures represented a significant opportunity for energy savings in the commercial sector, savings have been difficult to estimate, in part due to a lack of standardized programmatic EM&V protocol. Retroficiency has adopted IPMVP Option C and uses 15-minute or hourly-interval meter data when it is available.

Objectives

MEA's objective entering the pilot was to understand whether energy savings resulting from operational measures with limited economic hurdles in the schools could help drive more savings throughout the school system and across other segments of commercial customers. Example no-cost energy conservation measures included improving building system start and stop times, and optimizing temperature set points for the whole building level, specific spaces, and for different times (seasonal, occupied/unoccupied).

Study

Retroficiency worked with Montgomery County Public Schools (MCPS) and its energy management team that oversees more than 200 schools in total. The pilot began by analyzing energy consumption from eight schools to identify three with the best no-cost operational opportunities, determined by differences in operations or existing practices. The buildings were identified based on VEA results, which determined the that schools weren't operating as well as they could be. VEA's findings were presented and disussed with building operators.

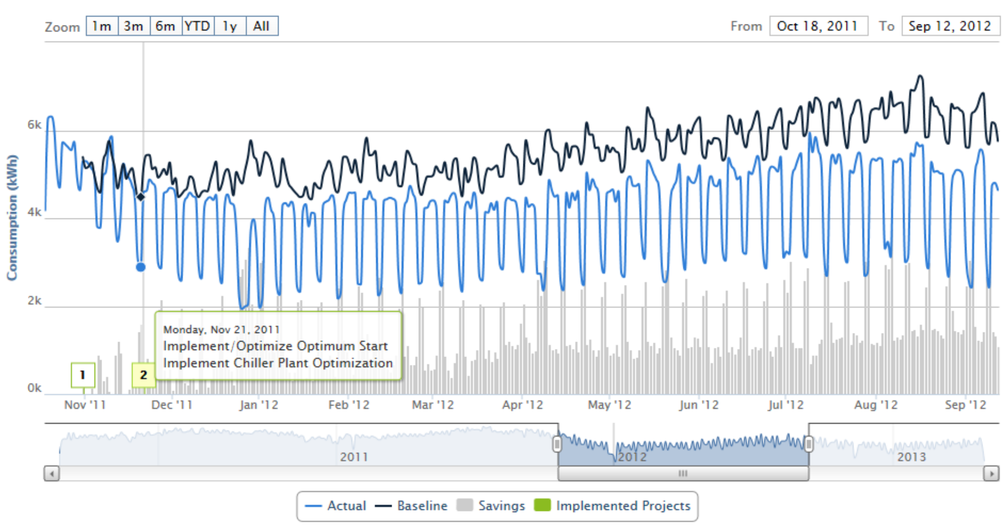
Three buildings (an elementary, middle, and high school) were identified as the best candidates to adjust their operations or existing maintenance practices. Retroficiency engineers worked with MCPS staff to align and refine the results of the assessment with each building's operational requirements. This process included phone conversations with both central and localized building management and on-site walk throughs for a more detailed review of building systems. These assessments led to specific operational recommendations across each of these schools. For example, analytics reports and discussions with building staff revealed that after one-off nighttime events, school operations were not always quickly set back to their optimal control settings for typical usage.

MCPS implemented the first batch of applicable schedule change measures at the onset of the summer. Once implemented, Retroficiency began estimating realized savings with its "Efficiency Track" automated M&V software. Efficiency Track employs a proprietary algorithm based on IPMVP Option C baseline approaches to automatically generate a weather- and occupant-normalized consumption baseline, which was compared to post-implementation metered consumption. Figure 6-1 includes a screen shot from one of Efficiency Tracks modules.

Results

Retroficiency's Efficiency Track software measured savings of 23%, 15%, and 1% percent respectively for the three pilot buildings. These results demonstrate the potential differences in estimated savings even though the buildings implemented the same energy savings measure. There are several potential reasons for this difference. One building (with the lowest reported energy savings on a percentage basis) was undergoing construction during a portion of the measurement period, potentially creating a one-time increase in energy consumption, and it is well known that billing analysis regression models perform best under stable operating conditions. This underscores the need for standardized methods for documenting and accounting for observed events, such as implemented measures or baseline adjustments, when using interval-analytics for building savings estimation and is a good example of where automated M&V can precede the more traditional investigation of program results.

Figure 6-1: "Efficiency Track" Example Screen of "Energy Track" Software



Shela Plank, MCPS Energy Program Manager, noted "MCPS invests significant time and resources to effectively manage its energy usage and has been successful in doing so, but it is always challenging when dealing with hundreds of facilities with ever-changing needs. Retroficiency's ability to rapidly analyze hourly meter data and make actionable recommendations is a very unique solution that highlighted further operational improvements and savings for MCPS with no capital investment required."

6.2 Opower

Company and Product Overview

Opower, founded in 2007 by Dan Yates and Alex Laskey, is an enterprise software company that helps utilities enhance customer engagement. Energy providers use Opower's customer engagement platform to deliver digital communications with the goal of raising customer satisfaction, manage energy demand, and lower service costs. Opower's software is deployed to more than 95 utilities worldwide and reaches more than 50 million homes and businesses (https://opower.com/).

The Opower programs referenced in the case study below are Home Energy Reports (HERs) and Behavioral Demand Response (BDR) as implemented by Glendale Water and Power.

HERs Programs

HERs programs use monthly or AMI data to produce customized energy profiles. Each enrolled customer receives his or her individualized recent and historical consumption. The report provides recommendations and where applicable, promotions for utility energy efficiency programs. Customer are compared (or ranked) against their neighbors with similar energy profiles.

Behavioral Demand Response

Behavioral Demand Response (BDR) is a software-only approach to demand response that uses AMI data to drive peak reduction at scale. BDR participants receive information about the program and recommendations on how to reduce load during peak load events. The day before a peak event, when the utility seeks system-wide demand reduction, BDR information is delivered via the customer's preferred communication channel (email, text, phone) to encourage customers to take actions such as raising the temperature on their thermostat by a few degrees. Each customer then receives timely post-event feedback about his or her individual demand savings within two to three days. These post-event communications reinforce the value of program participation and keep the customer engaged for future events.

Case Study–Glendale Water & Power

This case study focuses on a 2014 BDR pilot program at Glendale Water and Power (GWP). GWP is an innovative municipal utility in Southern California. GWP wanted to enhance energy savings by adding a BDR program to its previously deployed Opower HER program. This BDR pilot engaged 39,331 participating households with the technology. Over the course of three peak day events, this pilot achieved an average of 4% savings. GWP saw savings as high as 5%. Because these savings were achieved without installing a device or implementing a price signal, they are less expensive than traditional forms of demand response.[36]

The pilot will be evaluated again using 2015 data. (Opower has since expanded BDR to 1.5 million homes in the summer of 2015 including to customers at Commonwealth Edison, Baltimore Gas and Electric, Glendale Water and Power, and other large IOUs in the Midwest and California.)

Program Design

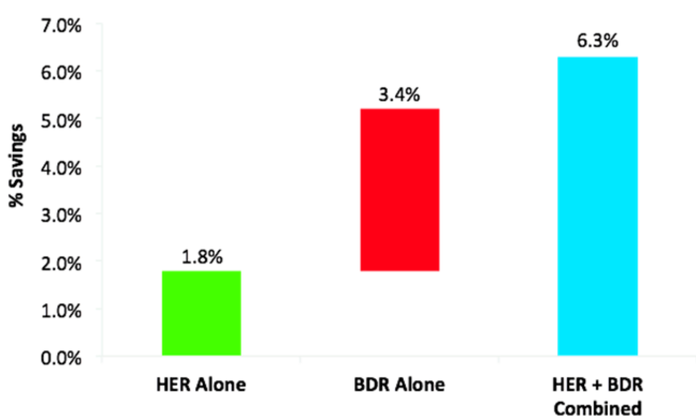
GWP tested BDR across two customer groups: customers that were already receiving Opower's HERs, and customers that were not. Both the BDR and HER programs were implemented as randomized controlled trials (RCTs), which allowed Opower to rigorously measure the impact of its behavioral energy efficiency and peak solutions separately and in combination.

Households were selected from three customer groups: (1) those who had been receiving monthly HERs from Opower since November 2009, (2) those who had been receiving monthly HERs from Opower since July 2011, and (3) those who have not received prior communications from Opower. This design allowed Opower to test whether the BDR program results were impacted by the HER program. The three treatment groups were randomly assigned to receive the BDR messages across a sample of 63,996 households. Messages were delivered via email, phone (interactive voice response), and text.

Results

To measure the impact of BDR on hourly energy consumption, Opower measured the difference in hourly energy usage between treated households (customers receiving the BDR messages) and control households (those who did not receive any communication from Opower). In its analysis, Opower estimated the impact of the BDR program on hourly consumption over the peak events. The results show that treated households used less energy during peak hours on peak event days than households in the control group.

Figure 6-2: Percent Reduction in Peak Energy Use Glendale Water and Power Case (Source: Opower)



The experiment found that households across the treatment groups used between 3.3–5.4% less energy during peak periods across peak event days than those in the control group. In addition, each treatment group achieved significant demand savings across all of the three peak days, with the greatest demand savings occurring on the final peak day. This demonstrates the persistence of the program over consecutive hot days during the experiment. Understanding customer willingness to maintain DR efforts over consecutive hot days has been a topic of interest for several years as this factor can differ between utilities and programs.

The results, shown in Figure 5-3, indicate that the reduction in peak energy use among customers who received both BDR messages and HERs was 1.9 times as large as those who only received BDR messages. Additionally, the peak reduction from the combined HER/BDR program was approximately the sum of the peak reduction from HERs alone and the peak reduction of BDR alone. This demonstrates that behavioral energy efficiency and behavioral DR programs can be integrated to create additional value for customers.[37]

When surveyed, 85% of GWP's customers expressed satisfaction with the BDR program, a high percentage for utility programs.[38] Connecting the customers with direct feedback of their energy use increased GWP's customer satisfaction.

This pilot was an example of the potential for more granular AMI data to make BDR programs possible. Without AMI, it would be imprecise to measure results (at the hourly or daily level) and generate personalized customer information. This level of information and direct communication makes BDR a viable utility DR strategy that can be accomplished with or without a price signal or device in the home.

Opowers Glendale pilot is a good example of how AMI data has the potential to increase the portfolio of utility DR and energy efficiency programs.

Studying Customer Load

Residential customers use different amounts of energy at different times of the day, but research has shown that individual load profiles and total consumption tend to be stable over time absent changes in occupancy, increases or decreases in building size, or major life events.

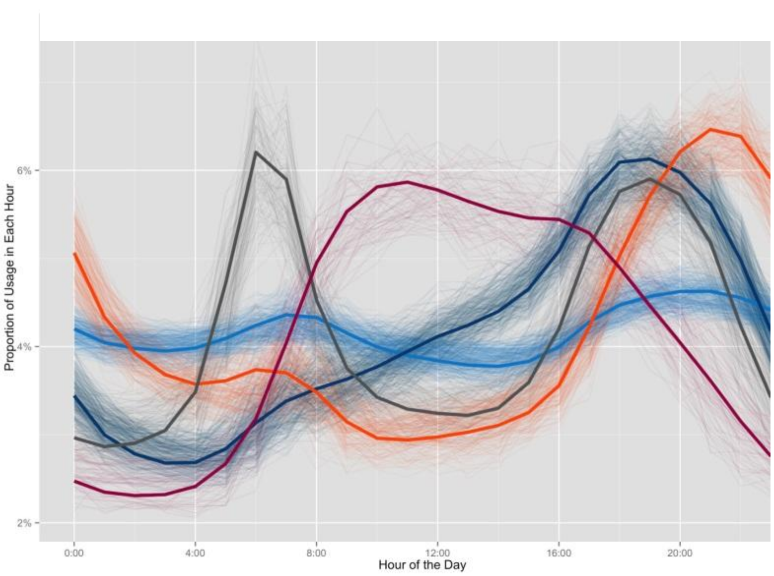
Utilities representing 812,000 residential electric customers in three major U.S. metropolitan regions collaborated with Opower to analyze load profiles and identify customers with the highest potential contribution for kW reduction during peak events. The utilities were interested in identifying these customers to offer them targeted, customized engagement strategies and program offerings, by combining AMI data and other variables such as time, geography, and weather.

Based on statistical clustering, five distinct hourly electric load profiles emerged. Each of these well-defined patterns form a particular load shape or “load archetype.” This includes customers who peak in the morning and evening (“twin peaks,” black curve in Figure 6-3) and those who peak around five o'clock (“evening peakers,” dark blue curve).

For example, if a utility identifies the customers who most closely fall into “evening peakers,” then the utility can take a more targeted and direct approach to delivering peak reduction programs like behavioral demand response or smart thermostat management.

Opower further analyzed the archetype distribution among BDR customers using data from three utility deployments across nine peak load events. The results show that customers whose usage remain relatively flat and stable all day (light blue curve) reduce consumption at lower rates than other archetypes, while the other archetypes save at roughly the same rate. In such a scenario, a utility saves time and money by focusing its demand reduction efforts on customers who have the potential to shed the most load, and customers are happier because they receive offers that are most relevant to them, reducing information fatigue or information overload.

Figure 6-3: AMI Data from 812,000 Utility Customers and Associated Load Patterns (Source: Opower)



Alternatively, utilities could identify customers with load profiles that aligned with solar photovoltaic output (i.e., daytime peak usage) and target those customers with offerings for solar solutions.

The ability to perform this analysis and similar predictive analytics for customer profiling and utility decision making would not have been possible with monthly data.

6.3 Selected Vendor Descriptions

6.3.1 Energy Savvy

Company and Product Overview

Energy Savvy licenses Optix Quantify, a cloud-based software tool for the electric and gas utility market that models consumption and weather data to create predictive models of premise level weather-normalized energy usage. This automated analysis can be completed with both monthly, bi-monthly, or more granular AMI data. The predicted consumption is compared with the actual usage data, to assess the increase or decrease in energy consumption for each project or for the program as a whole. User's access results such as summaries of program performance or realization rates through a web-accessible dashboard. Their target markets are utilities, implementers, evaluators, or other stakeholders. EnergySavvy supports energy efficiency programs, customer engagement programs, calculates potential high savers in constrained load pockets and can support utility operations by monitoring accounts for possible high bill instances.

Optix Quantify employs open industry protocols, such as IPMVP Option C, ASHRAE Guideline 14-2002, Uniform Methods Project (UMP) to build models of energy usage for each treated premise and an associated comparison group. EnergySavvy's specific methodology is available for review.

Current Applications

Currently, one of the most common applications is automated M&V for ongoing measurement of utility-funded residential and small/medium commercial energy efficiency programs. EnergySavvy reports that utility customers are utilizing Quantify primarily for program optimization, including feedback on contractor performance, contractor ranking, remote QA/QC and continuous assessment of factors influencing savings. Utilities are also analyzing demographic and propensity data from past participants for target marketing.

https://www.energysavvy.com/

6.3.2 FirstFuel

Company and Product Overview

FirstFuel licenses two software-as-a-service products to utilities and energy retailers: FirstAdvisor and FirstEngage. FirstAdvisor provides building-level analysis of customer meter data on monthly, or AMI data. The purpose of FirstAdvisor is to identify and track energy savings opportunities, based on technical potential and propensity to save, and to help utilities increase program participation.

FirstEngage is a customer engagement/customer service tool. Both products can be used to track meter-level energy consumption and the FirstAdvisor product can be used to perform automated M&V, but is not restricted to automated analysis.

FirstFuel utilizes machine-learning algorithms and customer meter data to quantify the reductions in load and consumption that result from energy efficiency measures. FirstFuel's models generate a metered consumption baseline model. The product estimates the increase or decrease in energy by subtracting metered consumption from the predicted model. FirstFuel's error metrics are compliant with the International Performance Measurement and Verification Protocol (IPMVP) Option C and ASHRAE Guideline 14.

Current Applications

Utilities are using the information created from FirstFuel's analysis (comparing predicted energy usage to actual energy usage) for assessing the effectiveness of EEMs. They are also using the EEM potential analysis to target energy efficiency programs and reduce customer load, in order to reduce capital investment needs. At a planning level, utilities are using the automated M&V capabilities to assess the distribution system impact of EEMs. FirstFuel also markets its platform as a "customer engagement" tool. As such, utilities are using it internally to reduce costs of handling customer inquiries, target their marketing of specific energy efficiency program offers based on customer information, track customer-specific actions, and follow up with customers to pursue deeper savings and monitor savings persistence.

http://www.firstfuel.com/

End Notes

---

[1] As set forth in the US DOE/US EPA State Energy Efficiency Network Action (SEE Action) Impact Evaluation Guide (2012)

[2] https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201520160AB802

[3] M. Fels, ed. "Measuring Energy Savings: The Scorekeeping Approach," *Special PRISM Issue of Energy and Buildings.* (1986): 9 (1–2) [sixteen papers describing background on PRISM and sample applications; primary background reference for PRISM users]; Eto, Joseph H. 1988. "On using degree-days to account for the effects of weather on annual energy use in office buildings". *Energy and Buildings*. 12 (2): 113-127; Northwest Energy Efficiency Alliance, E. Crowe. A. Reed, H. Kramer, J. Effinger, E. Kemper, M. Hinkle. 2015. “Baseline Energy Modeling Approach for Residential M&V Applications.” Report #E15-288.

[4] Eg., the American Evaluation Association’s Guiding Principles for Evaluators include, “Communicate the approaches, methods, and limitations of the evaluation accurately and in sufficient detail to allow others to understand, interpret, and critique their work.”

[5] NEEP, Home Energy Management Systems: Opportunities for HEMS in Advancing Residential Energy Efficiency Programs, August 2015.

[6] State and Local Energy Efficiency Action Network. 2012. ‘Energy Efficiency Program Impact Evaluation Guide.’ Prepared by Steven R. Schiller, Schiller Consulting, Inc., www.seeaction.energy.gov

[7] https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201520160AB802

[8] U.S. Environmental Protection Agency, *Evaluation Measurement and Verification (EM&V) Guidance for Demand-Side Energy Efficiency (EE) DRAFT FOR PUBLIC INPUT,* August 3, 2015

[9] Agnew K, Goldberg M. 2013. Chapter 8: Whole-Building Retrofit with Consumption Data Analysis Evaluation Protocol.

[10] Agnew K. Goldberg, M. DNV-GL. 2014.Whitepaper: Evaluating Opt-In Behavior Programs: Issues, Challenges and Recommendations prepared for the California Public Utilities Commission – Energy Division. CPU0088.01 Rev. V1.

[11] Davenport, Thomas H. 2014. Big data at work: dispelling the myths, uncovering the opportunities. Boston, Massachusetts: Harvard Business Review Press.

[12] SCADA, or supervisory control and data acquisition, is a system that operates over communication channels to control remote utility equipment.

[13] Jessica Granderson contributed substantially to this section. Granderson, J, Price, PN, Jump, D, Addy, N, Sohn, M., “Automated measurement and verification: Performance of public domain whole-building electric baseline models,” Applied Energy 144: (2015) 106-113; Granderson, J, Touzani, S, Custodio, C, Fernandes, S, Sohn, M, Jump, D, “Assessment of automated measurement and verification (M&V) methods,” Lawrence Berkeley National Laboratory, July 2015, LBNL#-187225.

[14] U.S. Energy Information Administration, “Electric power sales, revenue, and energy efficiency Form EIA-861,” August 11, 2015 for Early Release of 2014 data;

[15] J. Metoyer, DNV GL, M. Dzvova, California Public Utilities Commission. 2014. "Expanding the Value of AMI Data for Energy Efficiency Savings Estimation in California." ACEEE Summer Study in Energy Efficiency, Asilomar, CA.

[16] C. Holmes, K. Gomatom, A. Chuang. 2014. "Unlocking Customer Insights on Energy Savings and Behavior through the Use of AMI Metering Electric Power Research Institute." Proceedings from ACEEE, August 2014, Asilomar, Ca.

[17] Notice of suspension of the programmable thermostat from the ENERGY STAR® Program, Office of Air and Radiation, United States Environmental Protection Agency, Washington, D.C. 20460, May 2009

[18] EPA. Connected Thermostats Specification V1.0. 2015. https://www.energystar.gov/products/spec/connected_thermostats_specification_v1_0_pd

[19] National Association of Utility Regulatory Commissioners, "Value of Customer Data Access, Market Trends, Challenges and Opportunities," prepared for the National Association of Regulatory Utility Commissioners, by Navigant Consulting, Inc. (with permission) April 2015.

[20] We may share non-personal information (for example, aggregated or anonymized customer data) publicly and with our partners. For example, we may publish trends about energy use or elevated carbon monoxide levels in the home. This information may also be shared with other users to help them better understand their energy usage compared to others in the Nest community, raise awareness about safety issues, or help us generally improve our system. We may also share non-personal information with our partners, for example, if they are interested in providing demand-response services or other incentive programs. We take steps to keep this non-personal information from being associated with you and we require our partners to do the same. Privacy Policy for Nest Web Sites, https://nest.com/legal/privacy-policy-for-nest-web-sites/

[21] Northwest Energy Efficiency Alliance (NEEA), "Baseline Energy Modeling Approach for Residential M&V Applications," May 2015.

[22] NEEP, Home Energy Management Systems: Opportunities for HEMS in Advancing Residential Energy Efficiency Programs, August 2015.

[23] Some systems sample voltage and current at a rate of over 1,900 time per second

[24] 5 minute or shorter

[25] Typically 150 watts or lower

[26] Daily is standard

[27] Agnew K, Goldberg M. 2013. Chapter 8: Whole-Building Retrofit with Consumption Data Analysis Evaluation Protocol. The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures. National Renewables Energy Laboratory. http://energy.gov/sites/prod/files/2013/11/f5/53827-8.pdf

[28] Per the International Performance Measurement and Verification Protocols (IPMVP), Option C.

[29] https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201520160AB802

[30] U.S. Environmental Protection Agency, Evaluation Measurement and Verification (EM&V) Guidance for Demand-Side Energy Efficiency (EE) DRAFT FOR PUBLIC INPUT, August 3, 2015

[31] M. Fels, ed. "Measuring Energy Savings: The Scorekeeping Approach," *Special PRISM Issue of Energy and Buildings.* (1986): 9 (1–2) [sixteen papers describing background on PRISM and sample applications; primary background reference for PRISM users]; Eto, Joseph H. 1988. "On using degree-days to account for the effects of weather on annual energy use in office buildings". *Energy and Buildings*. 12 (2): 113-127; Northwest Energy Efficiency Alliance, E. Crowe. A. Reed, H. Kramer, J. Effinger, E. Kemper, M. Hinkle. 2015. “Baseline Energy Modeling Approach for Residential M&V Applications.” Report #E15-288.

[32] Eg., the American Evaluation Association’s Guiding Principles for Evaluators include, “Communicate the approaches, methods, and limitations of the evaluation accurately and in sufficient detail to allow others to understand, interpret, and critique their work.”

[33] NEEP, Home Energy Management Systems: Opportunities for HEMS in Advancing Residential Energy Efficiency Programs, August 2015.

[34] Energy Efficiency as a T&D Resource: Lessons from Recent U.S. Efforts to Use Geographically Targeted Efficiency Programs to Defer T&D Investments. 2015. C. Neme and J. Grevatt, Energy Futures Group, for NEEP and the Regional EM&V Forum.

[35] R. Craft, Con Edison, B. Fisher, Retroficiency. Energy-efficient Buildings, Analytics and Con Edison. *Electric Light and Power*, vol. 92, no. 4, August 2014. http://www.elp.com/articles/print/volume-92/issue-4/sections/energy-efficiency-demand-response/energy-efficient-buildings-analytics-and-con-edison.html

[36] A. Brandon, J. List, R. Metcalfe, M. Price “The Impact of the 2014 Opower Summer Behavioral Demand Response Campaigns on Peak, Time Energy Consumption.” (Forthcoming), the University of Chicago and Georgia State University

[37] Percent savings calculated by Opower and are in line with the findings in the paper by Brandon, et al.

[38] Opower survey results