Using Data Analytics to Bridge the Gap between M&V 1.0 and 2.0

Isaac Wainstein, Patrick Hewlett, and Paul Dobrowsky – ERS

Session 5B: The Shortest Distance Between Two Points

Abstract

Summary

Successful implementation of energy efficiency as a grid resource requires demonstration that installed measures confidently provide capacity relief during the network peak. Measurement and verification (M&V), often the means of defending these programs, is at a crossroads, with the debate intensifying between traditional M&V approaches and the advanced metering infrastructure (AMI)-based approach championed by "M&V 2.0." Traditional equipment-specific M&V is the gold standard for high-certainty results required by jurisdictional evaluations, but it can be relatively cost- and time-intensive, due to the numerous on-site inspections, metering, and analyses required for each new initiative. As a result, the industry has looked to M&V 2.0 as an alternative, with real-time, building-level data from AMI, which reduces project costs and timelines while providing robust data to assess hourly operation by account. However, the "black box" methodology of characterizing measure-level operations from building-level data leaves many skeptical of the ability of M&V 2.0 to support energy efficiency as a resource as accurately and defensibly as traditional M&V. This paper presents the lessons learned from a recent project utilizing the EDGE M&V methodology that borrows successful aspects of both traditional and 2.0 approaches, maximizing accuracy and minimizing costs by using predictive analytics on previously collected end-use metered data. While no M&V approach can ever be one-size-fits-all, this hybrid model shows that the best M&V solution is often a tailored combination of both approaches.

Results

New York's Con Edison has committed to providing more than 52 MW of capacity relief over a 12-hour period using targeted energy efficiency upgrade programs to permanently reduce demand in a network experiencing rapid load growth. The authors of this paper and Con Edison have embarked on a pioneering metering and market characterization effort, installing over 3,000 metering devices across 200+ facilities, to quantify how customers consume energy at the equipment level throughout all hours of the day. Instead of using deemed values or aggregate sector-level load shapes to estimate program savings, this metered data is used to develop an intricate analytical model of demand reduction at each hour not only in the studied population but for other segments as well.

This paper will detail how this method benefits from the real-time, low-cost analytic approach of M&V 2.0, while not losing the granular, measure-level accuracy of traditional M&V. Additionally, it will highlight the improvements in cost, accuracy, and the ability to produce real-time results as compared to traditional M&V and M&V 2.0 methods.

Key Takeaways

Using this hybrid M&V approach, Con Edison is empowered to make data-driven decisions about which programs and measures provide cost-effective and defensible capacity relief during the peak hours in order to achieve the 52 MW goal. During program implementation, administrators can monitor progress dashboards in real time to analyze trends and make any adjustments (e.g., neighborhood targets, incentive levels) to optimize the forecasted demand reduction at the hours of need. The M&V method described in this paper provides confidence and granularity of the results needed for Con Edison to enact similar programs at significantly lower costs in other networks throughout their service territory and defend energy efficiency as a grid resource.

Overview

Why do we need more M&V? This question is regularly asked by program implementers, and for good reason. The case for M&V is that it allows program planners to quantify the influence and cost-effectiveness of a program, pay for actual performance, improve implementation practices, and help forecast and defend reductions to establish energy efficiency as a viable grid resource. Yet when M&V results come long after a program is complete, they do not necessarily resonate with the current version of the program, which may have undergone multiple redesigns since the M&V timeframe. In addition, when the results comport with accuracy standards that are less stringent than those for grid operation, the concept of energy efficiency as a grid resource becomes harder to defend.

To assess the reliable contribution of this and other efficiency-based capacity-relief programs, substation planners are concerned about accurate projections of hourly performance. The switch from programs based on achieved kWh goals to those that deliver customer-side demand reductions as a reliable grid resource requires the following M&V innovations:

- **Reliability of results –** The industry's evaluation-standard accuracy of 10% relative precision at the 90% confidence interval is not sufficient for grid reliability. Hence, a program's M&V results must become more accurate, or forecasted demand reductions from customer-side solutions will be heavily discounted, thereby significantly reducing the program's overall cost-effectiveness.
- **Temporal kW forecasting and reporting –** Program planners need to confidently predict which measures will achieve the reduction, and by how much, at the specific hours of need. In a capacity-relief context, measures are no longer primarily valued by their lifetime kWh savings but by their ability to provide reliable demand reduction when needed.
- **Real-time results –** Timely results are needed to fine-tune the program before it is redesigned or ended. Learning after a program's lifecycle that reliability was low, or that the measure achieved the savings outside the hours of need, will result in energy efficiency

being multiple steps behind other grid resource solutions. Furthermore, emphasizing real-time results enables pay-for-performance based approaches that better align the motivation of program managers and implementation teams.

- **Insights for program stakeholders –** In order to maximize the value of M&V, results must go beyond only reporting what was saved. What measures are providing reliable savings? What customer types provide best opportunity for demand reduction at the hours of need? How can programs be redesigned and optimized to improve cost-effectiveness? M&V teams must leverage their data collection to help answer these types of questions.

The explosion of hourly data from advanced metering infrastructure (AMI) is often touted as a solution to the M&V challenges described above. Instead of a status quo of irregular, retroactive feedback of customer energy use through monthly meter readings, AMI provides a consistent, 15-minute stream of energy use for all participating customers. This granular data set can enhance the reliability, timeliness, and flexibility of customer-side analysis to quantify a program's impact on the grid.

A billing analysis with AMI involves modeling the energy use in the pre- and post-installation cases, with the difference equating to the influence of the measure or program. Yet there are significant concerns with the AMI method. Can we tease out the influence of a single measure from the whole building's usage, particularly if the program effect is small or if other measures were implemented as well? What about other economic, behavioral, or operational factors? Can these parallel changes in energy use be reliably separated from the influence of the measure?

This paper presents an alternate approach, termed Energy-focused Deep Granular Evaluation (EDGE) M&V, that leverages the explosion of whole meter-level data to provide dynamic, real-time assessment of grid impact through *modeling the measure operation* in order to improve the accuracy of results. In addition to AMI data, the presented approach also borrows additional data from edge devices, program-tracking databases, equipment load shapes, and demographic surveys to characterize the grid impact of each measure in the program. And with the increasing prevalence of smart devices and the ability to store complex streams of data, this measure-based approach will become more valuable to energy efficiency as grid technology matures.

Current and New M&V Approaches

This section will compare three M&V approaches presented in Table 1. The subsequent sections provide detailed descriptions of the three methods, outline when and why they are used, and identify their shortcomings.

Table 1. Comparison of M&V Approaches

	Description	Pros	Cons
EDGE M&V	Using multiple data sources (e.g., AMI billing, edge devices, equipment-specific M&V, demographic data) to model the respective measure(s)	• Real time • Flexible • In-depth insights from rich data sets to improve program planning and implementation • Rigorous site-level M&V NOT required	• Varying, and often complex, analysis approaches • Customization required for different territories/programs
Traditional M&V (IPMVP Options A or B)	Rigorous pre/post-installation M&V at a representative sample of sites	• Evaluation-standard approach • Sampling techniques allow evaluation at low cost, compared with overall program expenditures	• Results provided after program completion • Rigorous site-level, often customized M&V required • Insights limited to data from just a sample of projects
AMI Billing Analysis (M&V 2.0) [1]	Enhanced billing analysis using AMI data	• Real-time • Flexible • Single analysis approach • Requires no on-site presence	• Prone to bias and accuracy issues of modeling measure from full building data • Insights are limited to what can be inferred from whole-meter data

EDGE M&V

Energy-focused **D**eep **G**ranular **E**valuation M&V and Analytics seeks to retain the rigor of traditional M&V while also leveraging the influx of rich interval data used in AMI billing analyses. The key aspect of this method is that it characterizes the specific performance of a measure instead of the whole site or billing account.

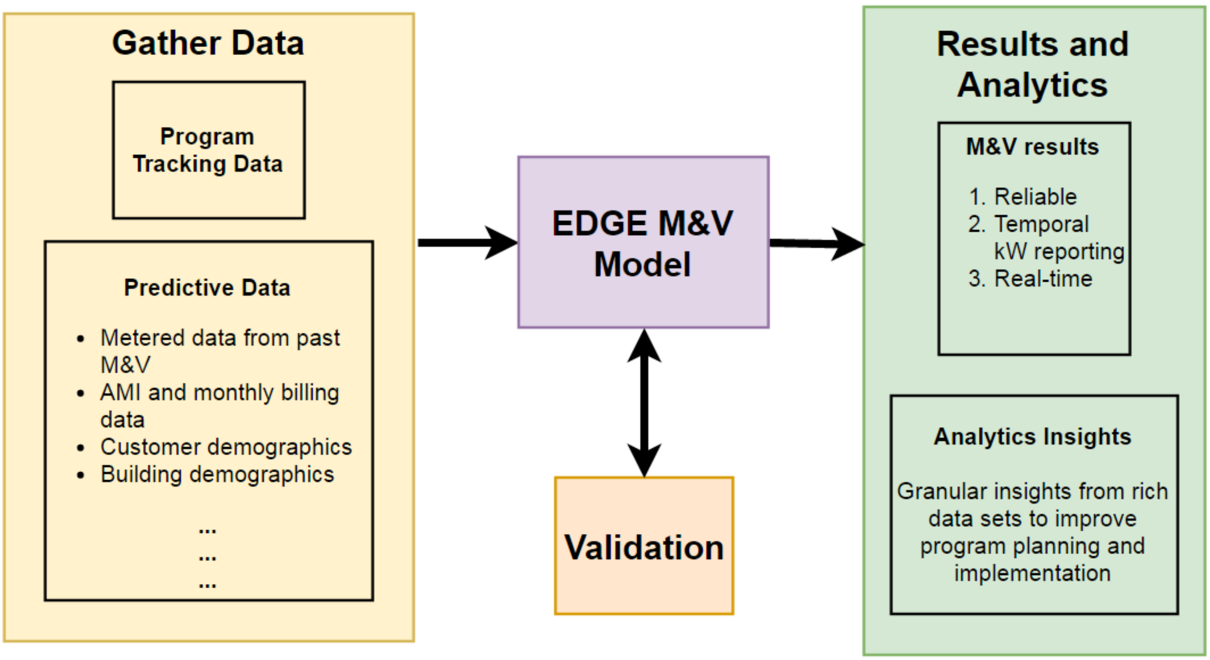
This method builds off of the benefits of the M&V 2.0 model's census-based, real-time, and flexible nature. These aspects allow programs to report performance closer to real time, pay customers accordingly, continuously improve program implementation, and better forecast and plan for future program iterations. EDGE combines these aspects with the improved accuracy of

[1] The energy industry has yet to settle on the definition of this and related terms. Sometimes M&V 2.0 and other terms such as Advanced M&V, Automated M&V, and Edge M&V are meant to include other forms of advanced data analytics such as equipment-specific interval data, on-board analytics, and other advanced elements. The narrower definition of M&V 2.0 used in this paper also is used by the Rocky Mountain Institute (see http://www.rmi.org/elab_accelerator_2016_m-v_2.0) and sometimes by the Lawrence Berkeley National Laboratory (see http://energy.gov/eere/buildings/downloads/assessment-advanced-measurement-and-verification-methods-mv-20).

traditional M&V, resulting in an approach that can be better defended by stakeholders when using energy efficiency as a grid resource.

Admittedly, M&V that is both highly accurate and produced in real time adds layers of complexity. However, the EDGE method involves three primary components that allow flexibility in how their data is collected and processed. The three phases include: 1) gathering data, 2) a continuous modeling approach, and 3) validating the model. The relation of these three analysis components are outlined in Figure 1.

Figure 1. Components of EDGE M&V



Gathering Data

Data is gathered for EDGE M&V with the purpose of building predictive tools. In this context, data collection is no longer oriented under the narrow view of analyzing a single site or measure. This data can come from multiple sources, such as demographics, building characteristics, equipment metered data, AMI, monthly utility bills, and more. One example is using business hours to determine cooling operation. A business that closes at 5:00 p.m. is less likely to cool their space overnight than a business that is open 24/7. Another example uses AMI data to create the feature of percentage of maximum meter load at certain hours of the day.

This approach does not rely on a specific data type for modeling; in fact, the independent variables that are utilized in the model will likely be uniquely chosen for each project. As a result of continuous iterations, M&V providers may learn which demographic, equipment, building, or other factors contain the most predictive power. This actionable information creates an incentive for communication between M&V and implementation teams, who compile the tracking data.

Modeling Approach

The modeling approach is dependent on the available data, and therefore there is no standard modeling approach for EDGE M&V. For example, one approach to predict the lighting operation at a certain hour is to correlate metered hourly coincidence factors (CFs) – the percentage likelihood of the light fixture operating at that hour – with predictive features. A second approach involves considering whether the lighting fixture of interest was on or off and then using logistic regression to predict the binary state. Both approaches have situations where they are more appropriate, and therefore a standard modeling approach is difficult to identify and apply in all situations.

Validation

Due to the flexibility in the data and modeling approach used, validation is a critical step to ensure that results accurately represent the modeled measure. Classic evaluation metrics, such as error ratio and relative precision, have been useful in communicating the validity of results, yet the enhanced modeling of EDGE M&V will require new metrics and techniques to defend results. See the next section for an example of these validation techniques.

Traditional M&V

The term traditional M&V as used in this paper refers to measuring the performance of individual pieces of equipment using IPMVP Options A or B M&V protocols[2] on a statistically representative sample of sites. The M&V results for the sampled sites are then extrapolated using statistical analysis to represent the entire program.

Why It's Used

The traditional M&V method has been widely used in an evaluation context because its sampling approach allows for a relatively small sample of sites to represent a whole program, for two reasons:

1. **Unbiased sample –** "If the sample projects are selected following an efficient sample design, and if the data collection and site-specific analysis is free of substantial bias, then the statistical analysis can provide an essentially unbiased estimate of each population characteristic of interest and a good measure of the achieved statistical precision" (CA Framework, p. 315).

---

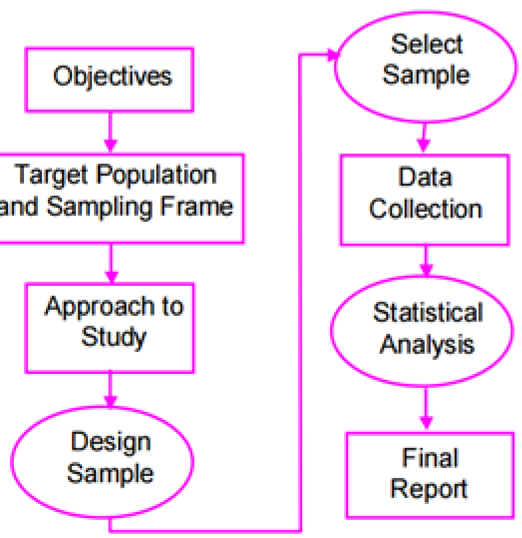
[2] International Performance Measurement and Verification Protocols (IPMVP) Option A calculates savings by "partial field measurement of the energy use of the system(s) to which an ECM was applied" and Option B by "field measurement of the energy use of the systems to which the ECM was applied." The terms are described in full in "International Performance Measurement & Verification Protocol," http://www.nrel.gov/docs/fy02osti/31505.pdf.

2. **More rigorous site analyses –** “By limiting resource-intensive data collection and analysis to a relatively small fraction of all projects, more attention can be devoted to each sample project” (CA Framework, p. 315).

Areas for Growth

The traditional M&V methods described above have provided stability and guidance to the evaluation industry for over a decade. Yet this method has failed to keep up as the industry evolves and as customer-side data becomes more prevalent. Figure 2 presents an example process flow in the 2004 California Evaluation Framework. Note that step two, *target population and sampling frame*, cannot be started until after the program is complete, as designing a representative sample requires a complete population of participating customers. If a sample is drawn early, results are subject to the bias of only representing the beginning stages of the program. Any changes to the program after the sample was drawn will not be captured. Additionally, data collection, statistical analysis, and the final report must also come after the program is complete, leaving traditional M&V timelines far behind that of program implementation.

Figure 2. Traditional M&V Process Flow



With this approach, M&V-based recommendations matriculate well after the studied period ends, possibly limiting their value. Furthermore, traditional M&V results are often not provided in time to decide whether the program should continue at the same funding level, and if so, how future program iterations are designed.

As described in the method overview section, traditional M&V enables relatively small sample sizes; however, these samples reflect a typical accuracy target of 90/10 confidence and precision. Program results with this level of accuracy (90/10) can be described as having a 90% chance of being within 10% of the true value. For example, if the M&V results led to evaluated savings of

2 million kWh at 90/10 confidence and precision, there is a 90% chance the actual program savings are between 1.8 and 2.2 million kWh. But more importantly, there is a 10% chance they are further outside this bound. This precision has been permissible when retrospectively reporting program performance versus goals, but when trying to reliably use energy efficiency as a grid resource, system planners who count on the demand reduction from the program and might think in terms of “6-sigma reliability” (0. 00034%) will deem this an unreliable resource. The precision can be reduced by increasing sample sizes; however, the M&V sample sizes required to achieve grid resiliency standards would not be a cost-effective or sustainable approach, particularly for retrospective results.

In addition to accurately reporting on past program achievements, traditional M&V results can provide insight to forecast the savings of future programs. This is particularly critical when using energy efficiency as a resource, because program efforts and funding should emphasize measures that will alleviate the grid at the specific hours of need. However, due to minimal customer diversity from small sample sizes, results from traditional M&V are often difficult to translate to new populations.

Finally, it is incredibly difficult to accurately pay for grid reduction performance when traditional M&V lags so far behind program implementation. Such an approach would require a payment schedule where the final incentive amount is uncertain until long after the project is installed.

M&V 2.0 (AMI Billing Analysis)

The influx of energy data, or “big data,” from AMI, smart appliances, and scalable data storage has led to many perspectives on what should be considered M&V 2.0. This paper defines M&V 2.0 as an enhanced whole-building-based interval billing data and analysis that compares AMI data in pre- and post-installation cases.

Why It’s Used

Using M&V 2.0 methods, billing analysis is not limited to a narrow sample of sites that undergo rigorous M&V. Instead, a billing analysis approach can be cost-effectively scaled to the entire population. In this model, grid-level consumption is normalized and projected for the pre- and post-installation cases, either at the site level or in aggregate. The difference between the pre- and post-installation cases represents the impact of the program or measure.

With the advent of AMI, M&V 2.0 aspires to improve traditional M&V in multiple ways beyond removing sampling error. With a constant stream of interval billing data, M&V results can be provided in real time shortly after a project is implemented, allowing continuous diagnostics and performance improvements as more post-installation data becomes available. Also, the standardization, transparency, and real-time nature of “smart meter” data are conducive to pay-for-performance contracts. Lastly, since M&V 2.0 can be applied and cost-effectively scaled to a diverse population, results can be more easily extrapolated to predict the performance of current and future programs.

Areas for Growth

The simplicity of AMI analysis also raises serious concerns about the accuracy and bias of the results. It is unclear whether M&V 2.0 accurately models the full building load in order to discern individual measure impacts. To correctly account for accuracy, the model's results must be measured as a percentage of the measure savings and not as a percentage of the building usage. A model that predicts full building load with 90/10 confidence and precision is completely insufficient to accurately model the savings of a measure that accounts for just 10% reduction of building usage.

More concerning, however, are biases within the AMI model. First, characterizing project savings as the difference between pre- and post-installation usage inherently incorporates any other changes (e.g., economic or operational) that happened within the participating building. For example, if in one year a business replaced its lighting yet also increased sales by 10%, both influences on energy use could be attributed to the lighting upgrade measure. Second, a significant portion of participating sites within a program typically do not have clean or comparable billing data that can be used. Hence, these sites are often removed from the analysis altogether, creating bias within the overall program results.

Application of EDGE M&V

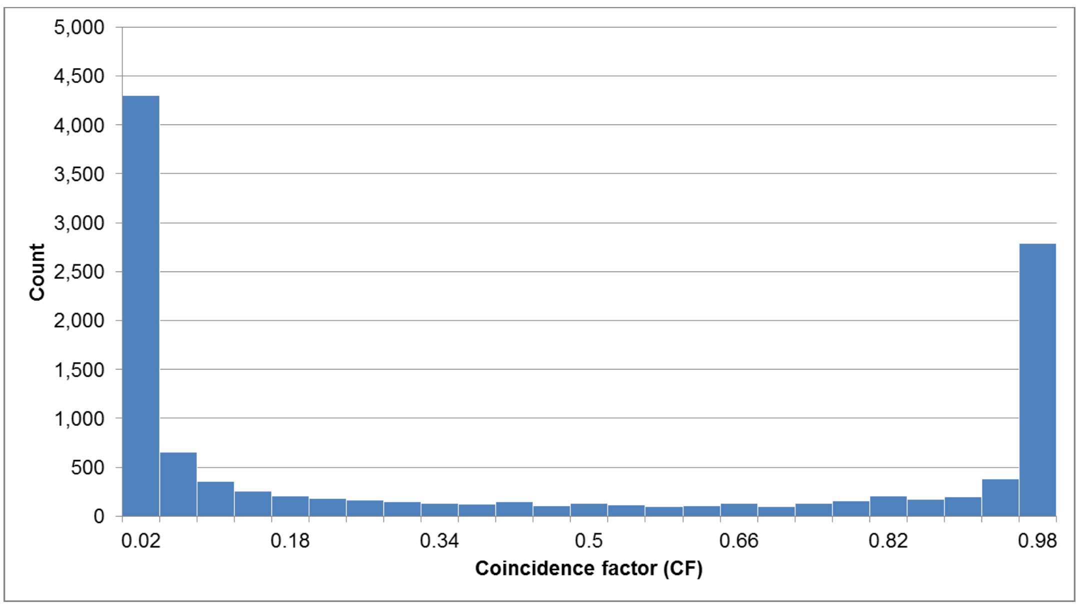
New York's Con Edison has committed to providing more than 52 MW of capacity relief over a 12-hour period using targeted energy efficiency programs to permanently reduce demand in a network experiencing rapid load growth. This program is called the Brooklyn Queens Demand Management (BQDM) program. The utility company prides itself on reliability and wants assurance that the selected resources will perform as expected during peak periods. A wide range of demand resource technologies are targeted, and reductions that fail to meet the reliability and cost standards will not be selected.

A significant portion of the targeted demand reduction for the BQDM program is being achieved by accelerating Small Business Direct Install (SBDI) lighting upgrades over a 4-year time frame. After the first year of implementation, ERS conducted traditional M&V on the program, providing demand reduction results for each hour of a peak-demand day. It was quickly realized that the traditional M&V approach would not work for the BQDM program. Con Edison stakeholders needed real-time results to make informed decisions about this innovative program and its impact on the grid. The BQDM territory's unique 12-hour peak period, with a surprising peak hour of 9:00 p.m. to 10:00 p.m., made it critical to project the program's actual peak demand reduction. Otherwise, the program might be funding projects that were not achieving expected levels of demand reduction at the hour of greatest need. Furthermore, the accuracy of results had to be improved to defend the reductions to grid reliability managers. To achieve this level of accuracy, the sample sizes using traditional M&V were simply untenable.

In response, the EDGE M&V method was applied for the updated SBDI program. The deliverable from this analysis was a 24-hour coincidence factor (CF) profile on a summer peak-demand day. This deliverable was achieved by modeling the CF profile for every measure within the program. The data set used included over 500 meters, spanning 127 businesses. Note that this method featured no additional on-site metering of the targeted program.

The CF for individual spaces and sites was found to be heavily binary, with the majority of CFs less than 5% or greater than 95%, as shown in Figure 3. Hence, a classification approach was found to be an intuitive and accurate way of modeling lighting behavior within small businesses. The problem was split into a three-label classification problem to allow for modeling behaviors that did not fall within the highly binary ends of the distribution.

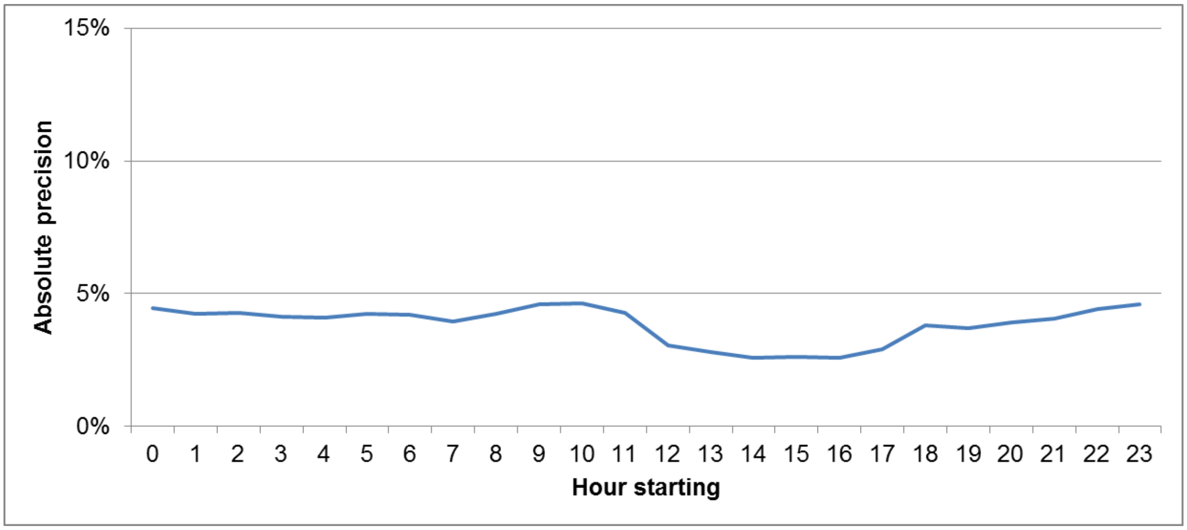
Figure 3. Distribution of Coincidence Factors



A random forest classifier was chosen in the modeling approach for several reasons. The primary reason was the prevalence of non-ordinal categorical variables with multiple labels, such as business type. While the data set was large for M&V standards, encoding the categorical labels into binary dummy variables would have expanded the feature space beyond what was reasonable. Random forest is able to handle multi-label categorical information without additional transformations. In addition, random forest outputs probability expectations of each class – on, off, and intermittent – which were used to calculate an expected value of CF.

Cross-fold validation was used to test the accuracy of the model. The data was subset into separate test data sets to allow for each measure to contribute to the accuracy metrics that would be calculated on the model. The measure-level predictions, based on models that had not observed the test data, were aggregated to the site level and compared to the 90 sites that underwent traditional M&V in the previous year. This allowed the team to contextualize the results in a domain in which they would typically be reported. Figure 4 presents the absolute precision for each hour at the 90% confidence level.

Figure 4. BQDM SBDI EDGE M&V Results



An absolute precision of 5% for a CF of 75% would represent a range of 70%–80%.

Validation, with absolute precision around 5% every hour, indicates that this model is reliable for generalization among the SBDI population. The team deployed the model that reads new tracking data and automatically outputs updated results, allowing results to be presented in an ongoing fashion as the program was implemented. This opens the door to future progress to real-time reporting, dependent upon availability of tracking data. At this stage, new results are available weekly.

Looking Forward

The EDGE M&V method acts as a catalyst for change in more than just the engineering approach. Some program-wide improvements that have been implemented with Con Edison are described below.

- **Dashboards and reporting –** The M&V team is innovating the way results are presented. Results are currently published on a web-based dashboard, viewable by all Con Edison stakeholders and updated on a regular basis. The dashboard also includes a variety of analytical insights to help improve program implementation and forecasting.
- **Model improvement –** This M&V approach is not biased to the data sources or modeling techniques used. The M&V team is constantly testing new modeling

approaches and predictive features from multiple data sources. Better savings prediction leads to lower sample sizes needed for the final savings validation.

- **Data collection –** This paper also highlights a change in perspective of what data M&V collects. This model would not have been possible without Con Edison's initiative to gather comprehensive and representative data about small business customers. This investment is expected to recoup its costs, as it can improve so many other aspects of program operations. And as data is continually collected and organized, we would expect these initial investment costs to continue to drop.
- **Expanding the measure types –** The M&V team is reporting on a variety of measures for the BQDM program and continues to test the validity of this modeling approach beyond small business lighting.

Ultimately, the true test will be considering how the proposed EDGE M&V method contributed to the success of the BQDM program. Did it provide stakeholders with insight to select the right programs and optimal measures, implement them effectively, and defend the results to grid operators? This is the future of EDGE M&V.