Reduce, Reuse, Recycle, Rethink – Continuously Collecting M&V Data to Redefine the Future of Evaluation

Paul Dobrowsky, ERS, New York, NY
Isaac Wainstein, ERS, New York, NY
Patrick Hewlett, ERS, North Andover, MA

ABSTRACT

Over the years, evaluators have collected terabytes of M&V data through program evaluations. Traditionally, the M&V data is used just once—to verify program savings—and is then archived and forgotten. This stems from an outdated M&V approach of collecting data after a program is complete, and only targeting data that will help report on the program at hand.

Advance metering infrastructure (AMI) signals a shift in the availability and potential benefits of customer-level data to evaluation results. But AMI data still has much to prove in its ability to provide granularity in identifying and assessing the savings potential and performance of individual measures. Installing AMI meters is also a typically lengthy undertaking, often taking years to fully integrate within a customer base. For example, New York City's Con Edison has recently begun an AMI deployment plan that will take six years to implement. During that ramp-up time, multiple cycles of a program may have come and gone without realizing the benefits of AMI.

By leveraging archived M&V data with customer insights and continually collected data to achieve statistical precision, we can marry AMI's broad, network-level analytics potential with M&V's deep, measure-level granularity to redefine the speed, effectiveness, and benefits of program evaluation. This targeted framework—termed Energy-focused Deep Granular Evaluation M&V (EDGE M&V)—involves dynamic forecasting models that are flexible enough to continually incorporate new measure-level data and transferrable enough to estimate the savings potential in new territories.

In this paper, we examine the methodology and results of EDGE M&V currently being deployed for one of Con Edison's most aggressive efforts to use efficiency as a grid resource, the Brooklyn Queens Demand Management (BQDM) program. The authors have spearheaded the use of EDGE M&V to not only evaluate but also forecast and optimize grid reductions from 20 MW of program-sponsored lighting measures– with more on the way in 2017 and beyond. Particularly unique to BQDM territory is its network peak of 9:00 p.m. through 10:00 p.m. on summer weeknights; the granularity of EDGE M&V allows program administrators to monitor reduction at this unique peak hour through an interactive dashboard that is continuously updated as new implementation tracking data is input. EDGE M&V can be transferred to other constrained networks by incorporating network information with customer-level billing data and other business characteristics as a first step in determining which measures might alleviate that network's peak.

In short, the objective of EDGE M&V is to provide evaluation-level insights in real time, allowing program administrators to continually optimize program-induced demand reduction per dollar spent. With rapid, in-depth insights into peak savings reductions by measure type and customer type, program administrators are empowered to target high-value customers and evaluate progress as they implement the program.

Introduction

Evaluation is in a state of flux. Two primary contributors are creating pressure for evaluators to produce results more quickly while maintaining high statistical rigor. First, the emergence of AMI data and

Internet of Things (IoT) devices has generated data more quickly, cheaply, and abundantly than ever, creating a renaissance in M&V techniques commonly referred to as M&V 2.0. Second, new demand-based program models that seek to estimate direct impacts on the performance of specific grid networks, as opposed to targeting abstract energy savings goals, make it necessary for utilities to know whether their efforts are working more quickly than the traditional evaluation life cycle. However, not all programs are implemented in areas that have access to AMI or utilize measures that can easily report to the cloud. This paper examines the capability for leveraging preexisting M&V data, collected through a variety of methods including previous program evaluation cycles, to create real-time reporting processes that offer statistical precision in line with classic evaluation. By changing the evaluation perspective from retrospective reporting to forward-looking planning as M&V data is collected, granular performance information can be leveraged long after its initial use to serve the needs of continuously evolving efficiency programs.

Non-Wires Alternatives and the Brooklyn-Queens Demand Management Challenge

Non-wires alternatives (NWAs) efforts represent a new model of energy efficiency and demand response program delivery that focus on specific, demand-based goals instead of more abstract energy-savings benchmarks. NWAs are a growing trend within the industry, and multiple grid-managing authorities such as Pacific Gas and Electric (PG&E),[1] Southern California Edison (SCE),[2] and the Bonneville Power Administration (BPA)[3] are pursuing projects to demonstrate the capability of efficiency, demand response, and renewables to meet grid planning expenses more cost-effectively than traditional solutions. With a performance-based goal of deferring substation or network distribution upgrades, utilities need to know whether the NWAs are making a dent—and more quickly than typical evaluation cycles—while maintaining accuracy that a grid planner can rely on. In addition to the referenced examples, New York's Con Edison is engaged in a widely discussed NWA effort, the Brooklyn Queens Demand Management (BQDM) program, for which ERS has been providing M&V since its inception.

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 specific Brooklyn/Queens network experiencing rapid load growth. The targeted network has a unique load shape, with a peak that extends over 12 hours and reaches maximum load during the evening between 9:00 p.m. and 10:00 p.m. Con Edison prides itself on reliability and therefore needs assurance that the selected solutions will perform as expected during peak periods. A wide range of demand resource technologies are targeted, and reductions that fail to meet reliability and cost standards are not 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 at business segments such as grocery, retail, restaurant, and office. During the first full year of the program's deployment (2015), ERS conducted traditional M&V on SBDI installations with the goal of providing evaluation-grade, measure-level insights and characterizing each business segment's performance for future planning. Results from this study were delivered in early 2016, but as the program had evolved in its first year, it became clear that the traditional M&V approach could not keep up with the dynamic BQDM program. Interdependent stakeholders—including the BQDM program administrators, Con Edison grid planners, and Con Edison's regulators—required real-time results to make informed decisions about the program and its impact on the grid.

[1] http://www.pgecurrents.com/2014/10/30/how-pge-is-using-demand-management-to-make-smarter-use-of-the-grid/
[2] https://www.sce.com/wps/wcm/connect/1ac76183-53c2-4762-8db2-4d52345dfa74/SCE_PRPOverview_1214.pdf?MOD=AJPERES
[3] https://www.bpa.gov/Projects/Projects/I-5/Documents/letter_I-5_decision_final_web.pdf

During this critical juncture, stakeholders and ERS identified some key questions that needed answers: Are current implementation strategies and technologies providing demand relief quickly enough or at the scale expected? How can the program more cost-effectively meet the prescribed goal? How can Con Edison deliver quarterly results to its regulators to report on progress toward meeting BQDM's performance-based goal? These questions created a unique challenge: demand reduction results need to be available quickly to satisfy regulators but also sufficiently accurate to satisfy the utility's grid forecaster that real demand reductions are occurring. Following the initial round of traditional M&V, Con Edison and ERS collaborated to design a method that would address this two-pronged objective.

EDGE M&V: Recycling M&V Data

In addition to providing M&V of the BQDM program in 2015, ERS had been tasked with developing a market characterization of multiple customer segments within the BQDM territory. Through this market study, ERS collected more M&V performance data than would have been necessary in a traditional evaluation; we installed approximately 3,000 metering devices across 250+ small businesses and multifamily units, to profile how Con Edison's customers consume energy at the equipment level throughout all hours of the day.

For the small business population, the data set included over 500 meters targeting lighting operations within 127 businesses. To support the creation of a wide range of equipment load curves, the data collection was highly structured—M&V was organized at a measure level and connected to a set of business-level factors tracked by the program implementer. The outcome of this market characterization was that each installed meter communicated how lighting fixtures operate across several important dimensions:

- ❑ The small business type (e.g., retail, grocery, industrial)
- ❑ The fixture location and type (e.g., exterior fixtures, refrigerated cases, exit signs)
- ❑ The annual operating hours of the business (a classic predictive feature for energy savings)

Importantly, these variables were demographic dimensions that we also expected to be included in measure-level implementation data in the future, as we continued to provide M&V for the program. ERS realized that utilizing the high amount of contextual information contained in this characterization data source would allow us to build a prediction model that could project the hourly operation—reported as a 24-hour coincidence factor (CF) profile on a summer peak-demand day—of each measure within an installed project. Having a granular prediction approach that could adapt to the unique measures at each participating facility is critical because of the BQDM territory's unique 12-hour peak period, as well as its broad base of diverse participants. Site-to-site operation is extremely variable within the small business population at this hour, and high statistical precision is achieved by reducing variability versus our expectation. ERS also expected that this method would provide accurate and granular projections, opening the door to analytics that could spot trends within installations over time and possibly reduce the risk of funding projects that would not achieve significant levels of demand reduction at the hour of greatest need.

The goal was to leverage the 2015 M&V data to create a prediction model for future years that can measure the achieved impact of the BQDM program through analysis of implementer tracking data in real time. We call this targeted and flexible framework of M&V Energy-focused Deep Granular Evaluation M&V (EDGE M&V). ERS proposed this method to Con Edison to meet the unique BQDM challenges, and the idea that the M&V data from 2015 could be recycled for this research was naturally appealing to cost-conscious BQDM program planners. As the model was developed in 2016, ERS would also need to validate its accuracy against traditional M&V standards to create confidence in using this method as a new standard. The creation of this prediction methodology was explained in deeper technical detail within a paper that was presented at the 2017 AESP National Conference.[4] The remainder of this paper explore the tactics and methods that were used to successfully deploy an EDGE M&V strategy and the results achieved.

Methods

Deploying an EDGE M&V strategy requires a paradigm shift from traditional M&V; while traditional M&V is typically retrospective, EDGE M&V looks to the future, leveraging available data to closely predict program progress. An EDGE model relies on implementation tracking data for two pieces of information that are used to provide real-time estimates of program impact:

- **Predicting operation behavior:** Predictive demographic features, such as annual operation hours and business type, as well as project characteristics such as measure location, are collected by the implementers during a project. Such data allows the EDGE model to project operation behavior, represented as a CF, for each installed measure at a project site.
- **Measuring impact:** The predicted nameplate impact of each measure (in kW savings) is used to translate operational forecasts to predicted demand reductions on the grid.

However, the prediction model itself is only one piece of the EDGE puzzle. Regulatory requirements of the BQDM program necessitate quarterly reporting, and grid planners desire high accuracy for continuous forecasting; to successfully design the EDGE M&V strategy, ERS identified three areas of focus that are explored in more detail in the following subsections:

- **Working with the implementation contractor to prepare tracking data:** Close communication with implementers is needed to ensure high availability of granular, structured implementation data for input to the model.
- **Developing and applying the EDGE model**: Deployment of the EDGE system must be robust enough to allow the M&V team to turn around projections in days instead of months.
- **Model validation:** Secondary validation of the analysis method must be completed to prove accuracy of the approach.

Working with the Implementation Contractor to Prepare Tracking Data

This foundational step defines what sort of predictive model can be used for the program; one can only model performance based on relevant inputs available from the implementer's tracking data and previously collected M&V. For the 2016 EDGE model that was used for the BQDM program, the predictive features were founded on data from the 2015 data set of records utilized for evaluation. Since the implementation contractor remained the same between years, there were already processes in place to ensure that the required information was collected for each project.

A significant change for the EDGE system was the frequency with which the data was requested—weekly—and the detail to which it was requested. Previously, implementation data had only been requested when samples were pulled for evaluation study; such frequency is typical of most traditional M&V projects. ERS requested moving to a system that aggregates projects at a measure level, as opposed to a less detailed project summary, for more granular predictive power. Meetings between Con Edison, the implementer, and the ERS team facilitated the development and reporting frequency of this key data

[4] "Using Data Analytics to Bridge the Gap between M&V 1.0 and 2.0," Isaac Wainstein, Patrick Hewlett, and Paul Dobrowsky, AESP National 2017

input to the EDGE model, and the efforts on the part of the implementation team to ensure data quality and availability cannot be understated in accomplishing such real-time and accurate reporting.

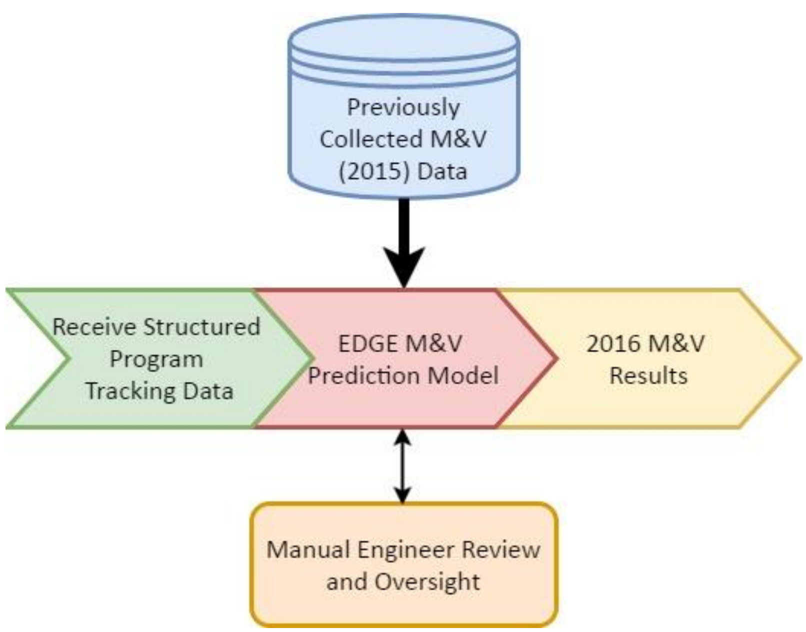
For utilities interested in pursuing an EDGE M&V system, ERS expects that regular coordination with implementers, outside of what may be typical in a traditional program evaluation and as early as possible, will be a key requirement to launching and adjusting the system even after data-sharing practices are in place. One specific example where coordination with implementers would be crucial involves expanding the number of variables used in the predictive model. Expanding the number of input variables creates the opportunity to improve the statistical accuracy of the predictive model by tracking more relevant or applicable information, but it also requires closer collaboration between the M&V team and implementers to collect key variables of interest.

Consider a specific example that is relevant to the BQDM program's EDGE system, wherein the annual operation hours of the target business are used as one of the predictive variables. Annual operation hours intuitively align with calculating energy savings and are routinely collected for SBDI energy efficiency programs, so it was already available within the program tracking data. However, with demand-based programs, annual operation hours do not have as much predictive power. Within the current model this variable implicitly serves as a proxy for open/close time of a business. But we could expect that having access to that information directly would be more directly correlated with lighting operation in each specific hour of the day, increasing the accuracy of our prediction. Adding this information to the tracking data would allow the M&V team to test the efficacy of open/close times compared with annual operation hours. However, implementer costs and capabilities would also have to be considered before new variables are carefully requested.

Developing and Applying the EDGE Model

Deploying the EDGE M&V model requires developing the prediction model itself, creating data pipelines for processing implementation information, and communicating the M&V results with the client. Figure 1 illustrates this process for the 2016 BQDM M&V effort, and the details of the important steps are outlined below.

Figure 1. EDGE Deployment Process



- **Developing the EDGE M&V prediction model:** Creating an EDGE prediction model requires a data set containing M&V performance data of similar measure(s) or program(s) in the past. The available performance data—our dependent variable—must be connected to measure or project features that were historically tracked by implementers and are expected to be included in tracking data for the upcoming program cycle—our independent variables. This data set serves as the training data for the statistical model. Engineers will perform statistical analyses to determine which factors from the implementer data set serve as appropriate variables for prediction of future operation and energy consumption. For example, a program that upgrades window A/Cs could utilize outside air temperature as an independent variable for the EDGE model, although that would not need to be supplied by the implementer.

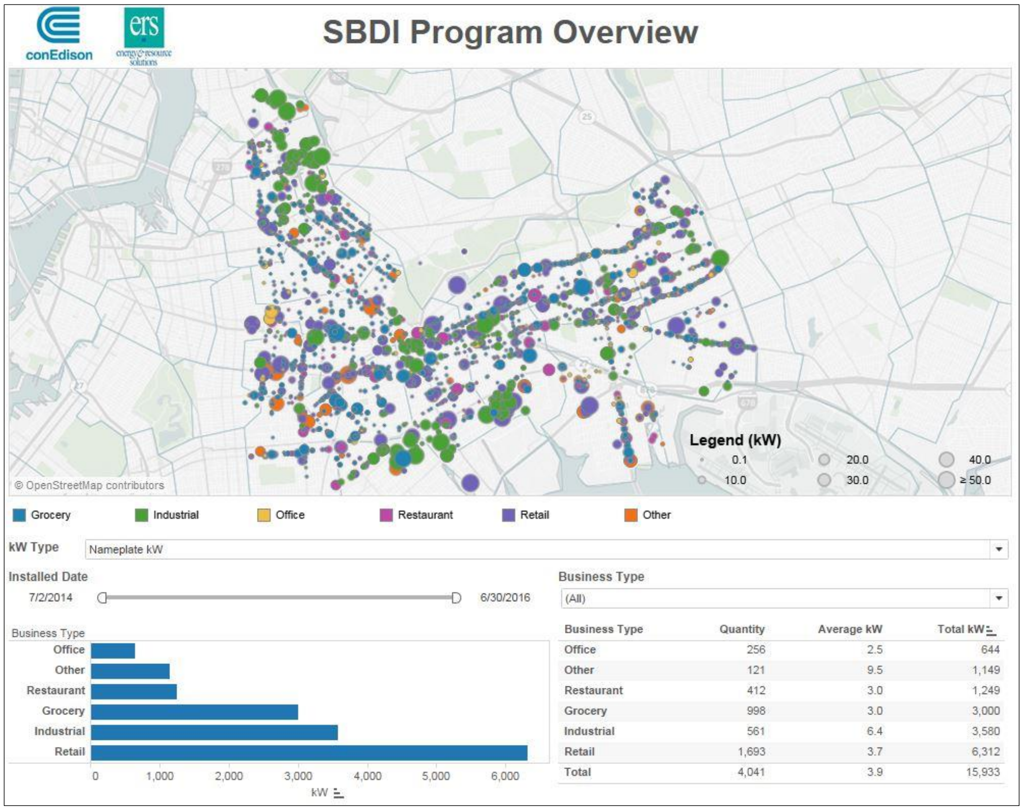
  A complete review of the development of the BQDM EDGE prediction model is outside of the scope of this paper. A deeper dive into the prediction model itself was presented in a paper at the 2017 AESP National Conference.[5] To summarize, the EDGE M&V model was developed prior to the need for projected 2016 M&V results, based on an analysis of the 2015 data. As a result, the features that are used by the model are constrained by information that was available from 2015 M&V data, from which we found that the CF for individual measures within sites was heavily binary, with the majority of CFs less than 5% (off) or greater than 95% (on) in each hour. 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.

- **Processing program tracking data:** The model developed for EDGE analysis needs regularly updated implementation data to make accurate predictions. The implementer provided weekly updates as described in the previous section, and ERS prepared data processing and QC scripts that transform this information into a format that could be interpreted by the predictive model. A consistent data format that was agreed upon with the implementation team facilitated this process and allowed for enhanced confidence in the quality of the output, reemphasizing the value of early buy-in with the implementation contractor.

- **Provide M&V results and analytics insights:** The output from the EDGE model, hourly demand reductions for each measure in the program, was reviewed by engineers to ensure that the projections were within reasonable bounds. This quality control process was critical given the novelty of the effort, ensuring the client that M&V had not transformed into a "black box." Once projections were reviewed, they were aggregated and pushed to a Tableau-based dashboard developed by ERS for program planners and implementers. This dashboard provided an interactive visual platform for program administrators to get up-to-date projections of program impact, but also to explore their program along multiple dimensions including location, business type, project size, and installation date. Reducing the program's barrier to continuously crunching their own granular program data allows program administrators to learn what is working and where there are opportunities to improve the program's efficacy, in near-real time. An example analysis tool that was included in the dashboard is depicted in Figure 2.

[5] "Using Data Analytics to Bridge the Gap between M&V 1.0 and 2.0," Isaac Wainstein, Patrick Hewlett, and Paul Dobrowsky, AESP National 2017

Figure 2. Program Analysis Dashboard



Model Validation

The EDGE model had never been applied prior to BQDM, and while initial validation steps using the training data were optimistic, it was necessary to complete an independent analysis to verify the performance and statistical accuracy of the EDGE M&V projection. Secondarily, independent validation of the technique would also allow the team to diagnose and address any acute issues with the system. To provide an independent validation, ERS completed a tandem M&V effort that mirrored the 2015 data collection, but with a reduced scope of sites. ERS performed site inspections at 48 small businesses receiving lighting upgrades in 2016, installing over 150 additional meters to provide independent M&V data for comparison with the model.

Validation provides multiple benefits to the EDGE M&V strategy. One benefit is that the collected validation data can be folded into the EDGE model in the following cycle, creating a broader training data set, leading to a more complete understanding of the segment, and increasing the power of transferring the model to additional efficiency programs or networks. In addition, this validation phase can be used to test out new predictive features that could increase the power of the model beyond what is possible by simply increasing the amount of data available. For example, within the SBDI sector, the previously described example of collecting facility open/close times would likely be a very effective predictor to experiment within the future.

Findings

The EDGE M&V approach allows evaluators to project the impact of programs in real time by leveraging previously collected M&V data to make predictions about participant operation behavior based on explanatory variables—like location, business type, and operation hours—that are collected by implementers. During application of the EDGE approach in BQDM territory in 2016, there were two primary questions in regards to its efficacy, the answers to which are explored in the sections that follow:

- ❑ Does the EDGE approach provide precise estimates of program impact? How did the validation M&V compare with the model's prediction?
- ❑ Can the EDGE system supply M&V results more quickly than traditional approaches?

Model Precision

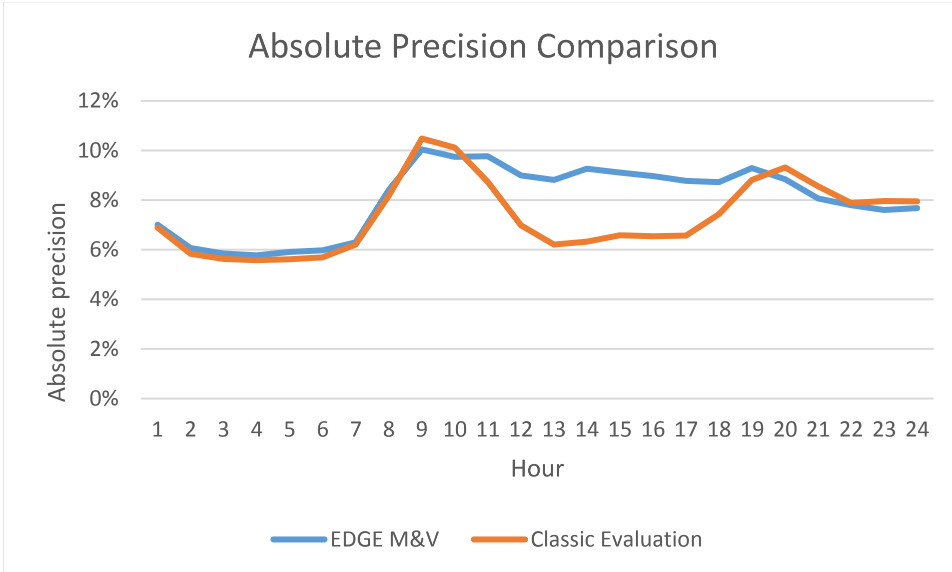
ERS has found that the EDGE model offers results at a precision that matches a traditional M&V approach, based on a comparison of the EDGE model predictions with validation M&V data from 48 sites. The size of the sample was chosen based on the number of sites ERS projected would be required to meet statistical precision requirements (90/10) for normal evaluation efforts.

It's important to note how this comparative analysis was completed. Start by imagining a single site within the sample, which may have 20 or more specific measures—line items within the lighting inventory. On-site, engineers install a selection of metering equipment in a randomly sampled way to represent the 20 line-item measures. It's too onerous to install equipment on every single measure—imagine a parking garage with 1,000 fixtures—and in aggregate, this random sampling approach is sufficiently accurate. Now compare this to how the EDGE model works. When the EDGE model analyzes this project within the sample, it makes an operating prediction for all 20 of the line items within the example site, based on key factors such as business type, annual operation hours, and space type. Both approaches create an estimate of the overall site-level savings profile; however, their answers are based on different sets of information.

To account for how engineers sampled measures within a site, ERS, in addition to providing M&V projections for the overall installed program through predictions at the measure level, ran the model on a more limited subset that only considered measures that were metered. This exercise creates an apples-to-apples comparison between the two methods. This comparative analysis is then rolled up in an aggregate fashion in the same way that classic site-sampling evaluation is handled.

Figure 3, below, compares the M&V results between the EDGE M&V approach and the site-level validation as a function of the most well-known evaluation metric, precision, at each hour of load reduction.

Figure 3. Evaluation Precision Across Methods



There are a couple of key observations. Most critically, in the hours of need—the later afternoon and early evening hours—the precision of the EDGE approach matches that of the traditional M&V approach. The precision, calculated at the 90% confidence interval, also falls within evaluation industry standards of 10% absolute precision within each hour. Precision that matches a site-level metering approach is incredibly exciting, as the predictive analysis now requires no time-intensive site work or additional metering. As we would expect, the site-level M&V is most variable during 8:00 – 9:00 a.m. and in the late afternoon, as those periods are highly variable due to facilities opening and closing, respectively.

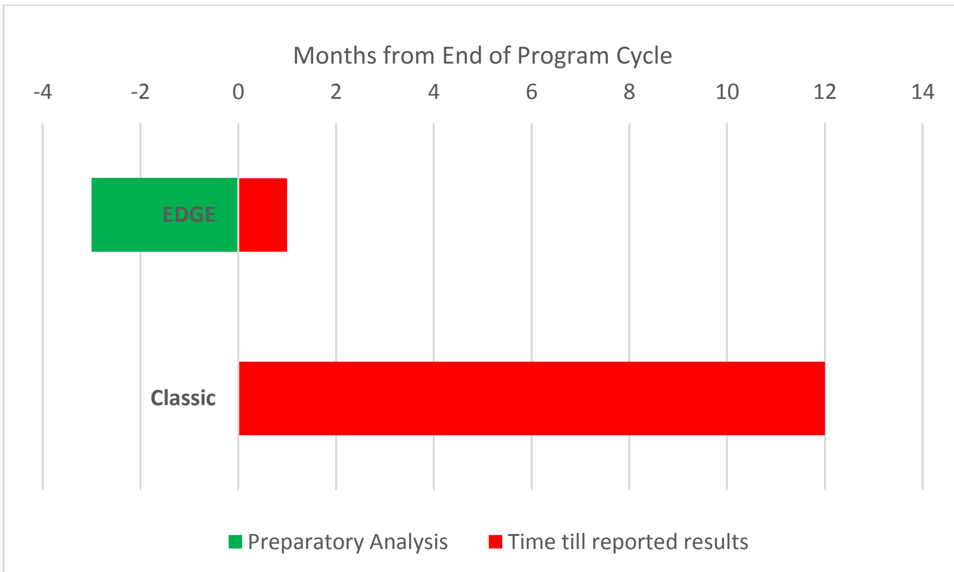
The second primary observation is that during the mid-day period—when most businesses are open—the EDGE model has absolute precision that is roughly 4% worse than the site-level metering. Because we know that operation during this period is actually quite consistent, this suggests that the EDGE model predicts that business operation within typical peak business hours is more variable than it actually is. This weakness of the model will be addressed in future updates to the prediction methodology. For the BQDM program, these early afternoon hours also fall outside of the most critical hours of load constraint (late evening) experienced in the territory.

Real-Time Results

Utilizing the EDGE M&V system, ERS provided M&V results for the program within two weeks of receiving updated implementation data. However, to be conservative and to allow for comprehensive quality control of results, ERS recommends planning up to an extra month for review of the results before reporting. Nonetheless, in contrast to traditional evaluation, EDGE offers a dramatic improvement in the timeliness of results because the M&V results are not being provided only at the end of the program; rather, projections are generated every time new implementation data is made available. Program administrators and regulators can now view accurate snapshots of program performance repeatedly throughout the year.

Figure 4, below, compares the timeline for reporting M&V results between an EDGE M&V approach and traditional evaluation, based on the delivery time of results after the end of a program cycle. Traditional evaluation only pulls a sample following completion of the program, meaning results can take up to a year to be collected and synthesized. In comparison, the EDGE approach provides ongoing and rapid results, but requires more rigorous upfront planning as to how the impact of the program is measured. Preparatory analysis takes place in the months before application of the EDGE system in order to develop the prediction model. By frontloading the analysis in this way, which requires deep communication and coordination with the implementation contractor and other stakeholders, evaluators can deploy the predictive approach in near-real time if planned properly. A traditional M&V approach does include some degree of planning and upfront analysis, for activities such as sample planning, but the rigorous calculation of actual program impacts can only begin once the program cycle has ended and a sample is drawn for inspection and metering.

Figure 4. EDGE vs. Classic M&V Reporting Time



Additional Benefits of the EDGE Model

As highlighted in the discussion of the EDGE model's accuracy, the validation comparison analysis was made based on analyzing only measures for which metering equipment had been deployed. But the results that are reported by the EDGE model are different—they are based on the entire tracking data set. This characteristic introduces another advantage of the EDGE model. Knowing that we have validated the precision of the EDGE prediction on a measure level, we can use the model to have a fuller understanding of the actual performance of the whole program. The EDGE model is not limited by sampling of measures completed by the engineers. One specific example highlights this advantage: exterior fixtures. While ERS had previously recorded data on exterior fixture operation in 2015, exterior fixtures are often not sampled in traditional evaluation because it is typically harder to record performance data for fixtures that are located outside. A traditional aggregate analysis may have shortcomings in characterizing such measures, but the EDGE model can detect exterior fixtures and apply the appropriate operational prediction for every site within the sample after a reasonable amount of upfront performance data collection.

Conclusions

Within the direct context that the EDGE model was launched and validated, Con Edison's BQDM program, EDGE will continue to become more sophisticated as the programs evolve. For example, within the small-business sector, there are still advantages to continuing to perform continuous model validation despite the proven accuracy of the system. Additional site-level validation provides the opportunity to test out new predictive features that can enhance the precision of the model further. Additionally, as more and more small businesses participate in the program, future participants may be less represented by the 2015 foundational data; regular validation will ensure that late participants are properly represented within the model.

More generally, ERS believes that the EDGE approach will expand beyond BQDM and that there are opportunities for many evaluators to develop an EDGE M&V solution for their own clients. The general thesis of EDGE stresses the importance of being prospective about how M&V data will be used in the future. In the BQDM example, we've connected performance data from 2015 to factors such as business type, operation hours, and fixture location to create a prediction model for providing real-time measurements of program impact on the grid. As this paper demonstrates in the validation section, such predictions are comparable to traditional M&V impacts. While the opportunity presented by the BQDM program was extremely unique—a wealth of previously collected highly structured M&V data, and an implementer who was open to adapting to the requirements of this new program—the reporting requirements of this non-wires alternative (NWA) program challenged ERS to provide a cutting-edge approach to evaluation that is not a one-off strategy. Rather, we believe that EDGE M&V is scalable and can become a generalized approach for evaluators that will help them adapt with other disruptions created by modern technological innovations. ERS believes that the benefits passed on to program administrators—enhanced visibility into program performance at earlier dates, the capability to optimize program delivery in real time, and potential cost reductions once models are significantly validated—make expansions of similar approaches throughout other networks and jurisdictions likely.