Will the Measurement Robots Take Our Jobs? An Update on the State of Automated M&V for Energy Efficiency Programs

Jessica Granderson, Samir Touzani, Cody Taylor[1], Samuel Fernandes
[1]US Department of Energy
Lawrence Berkeley National Laboratory

ABSTRACT

Trustworthy savings calculations are critical to convincing regulators of both the cost-effectiveness of energy efficiency program investments and their ability to defer supply-side capital investments. Today's methods for measurement and verification (M&V) of energy savings constitute a significant portion of the total costs of energy efficiency programs. They also require time-consuming data acquisition. A spectrum of savings calculation approaches is used, with some relying more heavily on measured data and others relying more heavily on estimated, modeled, or stipulated data.

The rising availability of "smart" meters and devices that report near-real time data, combined with new analytical approaches to quantifying savings, offers potential to conduct M&V more quickly and at lower cost, with comparable or improved accuracy. Commercial energy management and information systems (EMIS) technologies are beginning to offer M&V capabilities, and program administrators want to understand how they might assist programs in quickly and accurately measuring energy savings. This paper presents the results of recent testing of the ability to use automation to streamline some parts of M&V. In this paper, we detail metrics to assess the performance of these new M&V approaches, and a framework to compute the metrics. We also discuss the accuracy, cost, and time trade-offs between more traditional M&V, and these emerging streamlined methods that use high-resolution energy data and automated computational intelligence. Finally we discuss the potential evolution of M&V and early results of pilots currently underway to incorporate M&V automation into ratepayer-funded programs and professional implementation and evaluation practice.

Introduction

Energy Management and Information Systems (EMIS) span a family of technologies and including energy information systems (EIS), building automation systems, fault detection and diagnostics, and monthly energy analysis tools. These tools have enabled whole-building energy savings of up to 10-20% with rapid paybacks often under three years (Granderson 2011, 2016a) through multiple strategies such as: identification of operational efficiency improvement opportunities, fault and energy anomaly detection, and inducement of behavioral change among occupants and operations personnel.

In addition to *enabling* operational savings, EMIS have begun to automate the *quantification* of whole-building energy savings, relative to a baseline period, using empirical baseline models that relate energy consumption to key influencing parameters, such as ambient weather conditions and building operation schedule (Granderson 2015, 2016b; Kramer 2013a, 2013b; Reddy 1997). Today, the advent of increasingly available interval meter data has enabled the development of more robust baseline models than the monthly models that have traditionally been used to characterize whole-building energy performance (Haves 2014; Katipamula 1998;

Walter 2013). These automated baseline models can be used to streamline the whole-building measurement and verification (M&V) process. This is important because traditional M&V processes using engineering calculations can comprise a significant portion of the total costs of efficiency programs, and require a level of engineering expertise that can challenge scalability.

Although EMIS hold great promise, several questions remain to be answered before energy managers and utility programs can confidently adopt their emerging M&V automation capabilities. Prior work has addressed how baseline models can be objectively evaluated to determine *overall* predictive accuracy, and how public and proprietary models can be tested and compared. However, questions remain as to how these automated approaches can be practically incorporated into practitioner work streams, and the accuracy, cost and time trade-offs with respect to more traditional approaches. In this paper we present the results of recent efforts to use automation to streamline portions of the M&V process. Specifically, we detail the application of automated M&V2.0 to program data, conducted in partnership with a large utility; assessment of the uncertainty of the savings estimation, due to model error; and labor time saving advantages. Finally, we discuss the potential evolution of M&V, and early pilots to incorporate M&V automation into ratepayer funded programs and professional implementation and evaluation practice.

Background: General Baseline Model Performance Assessment

Baseline energy use models characterize building load or consumption according to key explanatory variables such as time-of-day and weather. These baseline models are used for a variety of purposes in EMIS, including near real-time energy anomaly detection, and near future load forecasting, as well as quantification of energy or demand savings. Baseline model accuracy is critical to the accuracy of energy savings that are calculated according to the IPMVP. For both whole-building and measure isolation approaches (IPMVP Options B and C) the baseline model is created during the “pre-measure” period, before an efficiency improvement is made. The baseline model is then projected into the “post-measure” period, and energy savings are calculated based on the difference between the projected baseline and the actual metered use during the post-measure period (EVO 2012). Therefore, the error in reported savings is proportional to the error in the baseline model forecasts.

Prior work established a 4-step statistical procedure that can be used to evaluate the predictive accuracy, of a given baseline model (Granderson 2014, 2015). The test dataset comprises interval meter data and independent variable data, such as outside air temperature, for dozens to hundreds of buildings. These buildings are “untreated” in terms of efficiency interventions. That is, they are not known to have implemented major efficiency measures. The data for each building is divided into hypothetical baseline (i.e., model training) periods and hypothetical post-measure (i.e. model prediction) periods. Meter data from the prediction period is “hidden” from the model. The trained, or fitted model is used to forecast the load throughout the prediction period, and predictions are then compared to the actual meter data that had been hidden. Figure 1 shows an example of actual, and model-predicted data for a 12-month training period and a 12-month prediction period. Performance metrics that quantify the difference between the model prediction and the actual load are calculated and used to characterize accuracy.

Figure 1. Actual and model-predicted energy data, overlaid with outside air temperature, for a 12-month training period and 12-month prediction period.

This testing procedure assesses model performance in general, 'on average' across populations of many buildings. Once it has been determined that a new model of interest performs well *on a population basis*, the practitioner can have reasonable confidence that the model is viable for use in M&V. The model can then be applied to *specific* buildings that have undergone, or will undergo an efficiency intervention. The principles of ASHRAE Guideline 14 (ASHRAE 2014) may be followed to determine whether the model fit is sufficient for each building in the program, and to quantify the uncertainty in the savings that are estimated. Savings and uncertainties for each building can be aggregated for a portfolio level analysis. This approach is described further in the Initial Results section.

Assessment of Open and Proprietary Models, and Initiation of Utility Pilots

Recently, the testing procedure described above was applied to evaluate a set of ten open and proprietary M&V2.0 interval baseline models (Granderson 2016b). The test data set comprised over five hundred buildings, primarily from California, Washington, DC, and Seattle. Overall, the models that were tested were able to predict whole-building energy use with a high degree of accuracy for a large portion of the buildings in the test dataset. For the standard whole-building case of twelve months training followed by twelve months of prediction, the results indicated that the models would meet ASHRAE Guideline 14 fitness requirements for the coefficient of variation of the root mean squared error (CV(RMSE)) for over half the buildings in the data set. The CV(RMSE) is a statistical metric that characterizes the error between the modeled energy use, and the actual metered use. In addition, the models were able to predict a year of energy use with errors that ranged between -1% to 4%, for one quarter of the buildings in the data set.

Given encouraging results from the above analysis and in prior work (Granderson 2014), an effort was initiated to partner with utilities and implementers to pilot these methods on past program data. The desired program and building characteristics included: sites with interval electricity data from commercial buildings with nine-to twelve months of pre- and post-measure data, and measures that generated savings large enough to be detected at the whole building level. In this effort, the authors are currently working with North American utility and implementer partners – one from the northwest, one from the west, and one from the northeast. The initial results that are presented in the following are based on data from one of the partners' program sites.

Initial Results

Savings and Uncertainty

Gross site and portfolio level whole-building savings were computed using an M&V2.0 baseline model for a set of fifty-one commercial buildings. These buildings were retro-commissioned, and in some cases, also received retrofits. Initially a larger set of buildings was considered, however many did not have a sufficient amount of data for the pre- and post-measure period, to confidently apply whole-building level savings estimation. For the fifty-one buildings that were analyzed, between nine and twelve months of interval electric data was available for both the pre- and the post period. The model of interest to the utility partner was the time of week and temperature model that was described and tested alongside nine other models in Granderson 2016b. In this model the predicted load is a sum of two terms: (1) a "time of week effect" that allows each time of the week to have a different predicted load from the others, and (2) a piecewise-continuous effect of temperature. The temperature effect is estimated separately for periods of the day with high and low load, to capture different temperature slopes for occupied and unoccupied building modes.

The following process was used to leverage the benefits of automation, while still adhering to industry best practice as defined in ASHRAE Guideline 14:

1. Automatically fit the model to the data from the measure pre-period.
2. Compute goodness of fit metrics R^2, CV(RMSE) and NMBE from the pre-period meter data, and the fit model[1].
3. Set aside the buildings for which the R^2 was below 0.6 and/or CV(RMSE) surpassed 25% these will require inspection of the data, and engineering expertise to determine whether a better fit can be obtained through adjustments or tailoring.
4. For the remaining buildings automatically compute savings according to IPMVP Option C [EVO 2012]: run the fit model to generate energy use predictions for the measure post period, and compute the energy savings as the difference between the model-predicted energy use and the metered energy use.
    a. For each building, estimate the uncertainty in the calculated savings according the ASHRAE Guideline 14.
5. Aggregate the savings and the uncertainties for each building, to determine portfolio-level results for the cohort.

Of the fifty-one buildings, over thirty-nine 'passed' the suggested threshold for acceptable R^2 and CV(RMSE), and all were fit with an NMBE that met the ASHRAE guidance (NMBE<0.5%). The R^2 value was set empirically, and the CV(RMSE threshold was set conservatively, drawing from the ASHRAE Guideline requirements for projects for which

[1] R^2 is the coefficient of determination, CV(RMSE) is the coefficient of variation, and NMBE is the normalized mean bias error. These metrics are used to characterize different aspects of model error. Formulas to compute these metrics can be found in common statistical references.

uncertainty is *not* computed. For the twelve buildings for which the M&V2.0 model did not meet the guidance for goodness of fit, an inspection of the data was conducted. In five cases it appeared that the documented measure implementation date was incorrect, i.e., a large reduction or increase in energy consumption was observed during the baseline period, and the load profile was consistent with that observed in the measure post period. An example of one of these cases is shown in Figure 2. The majority of the remaining seven buildings for which model fitness was poor revealed seasonal trends that were not captured by the baseline model. In these cases it was not possible to identify a straightforward, justifiable adjustment to obtain a well-fit model – some of these buildings may not be well suited for automated M&V2.0, and would require deeper investigation.

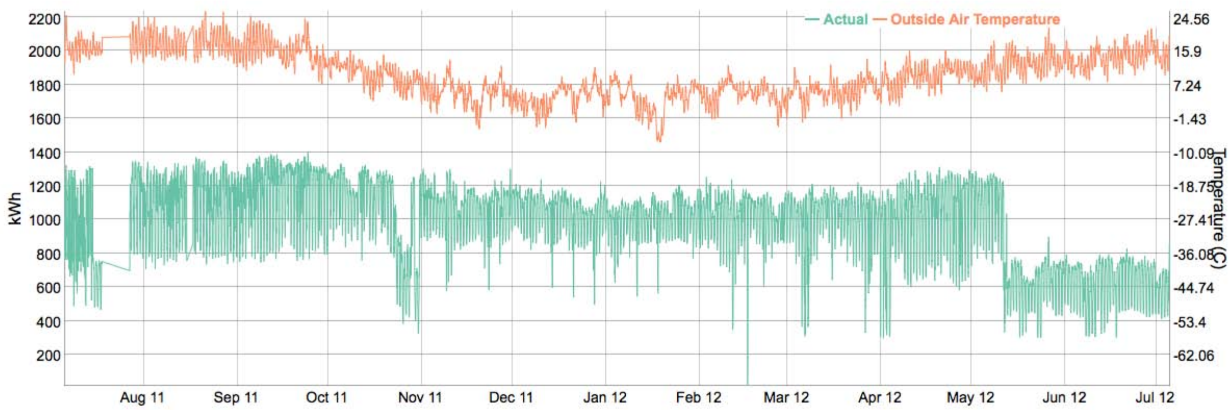

Figure 2. Example of a suspected case of incorrect documentation of measure implementation date; the measure may have been installed in mid-May rather than Mid-July, as reported.

Savings and associated fractional savings uncertainties were computed for each building, for the ninety-five percent confidence level - a level *significantly* higher than the sixty-eight percent advised by the ASHRAE guideline. As for the CV(RMSE) and NMBE assessment, the fractional savings uncertainty was calculated according to the methodology defined in ASHRAE Guideline 14, for models with correlated residuals. Figure 3 shows the results of the savings and the confidence intervals at the ninety-five percent confidence level. Each bar in Figure 3 corresponds to the fractional savings (in percentage) for a site, and each 'whisker' represents the lower (on the left) and the upper (on the right) bound of the confidence interval. The positive and negative values respectively indicate savings and increases in building energy use over the post-measure period.

The results displayed in the Figure 3 show that for thirty buildings savings were estimated with fractional savings ranging from approximately 0.5% to 36.5% and with a median value of 7.3% and a mean of 10%. For the other nine buildings, increases in the energy use were estimated with fractional savings metrics ranging from approximately from -17.15 to -0.1%, and with a median of -1.79% and a mean of -5.85%. The majority of these estimations show acceptable level of uncertainties. ASHRAE Guideline 14 stipulates that the acceptable level of the energy savings uncertainties (i.e. fractional savings uncertainty) is less than 50% of the estimated savings at 68% confidence level, in this study, 32 of thirty nine buildings complied with this rule.

The pooled savings and uncertainties were also calculated, treating the set of thirty-nine buildings as an aggregated portfolio. When the savings were added up across the entire set of 39 buildings, and the uncertainties were propagated, the portfolio-level result was 3.96% savings within a confidence interval of [3.66%; 4.26%] at the 95% confidence level. The fractional saving uncertainties is approximately 7.5% at 95% confidence level and approximately 3.9% at 68% confidence level, which is much lower than the ASHRAE guideline validity level (i.e. 50%).

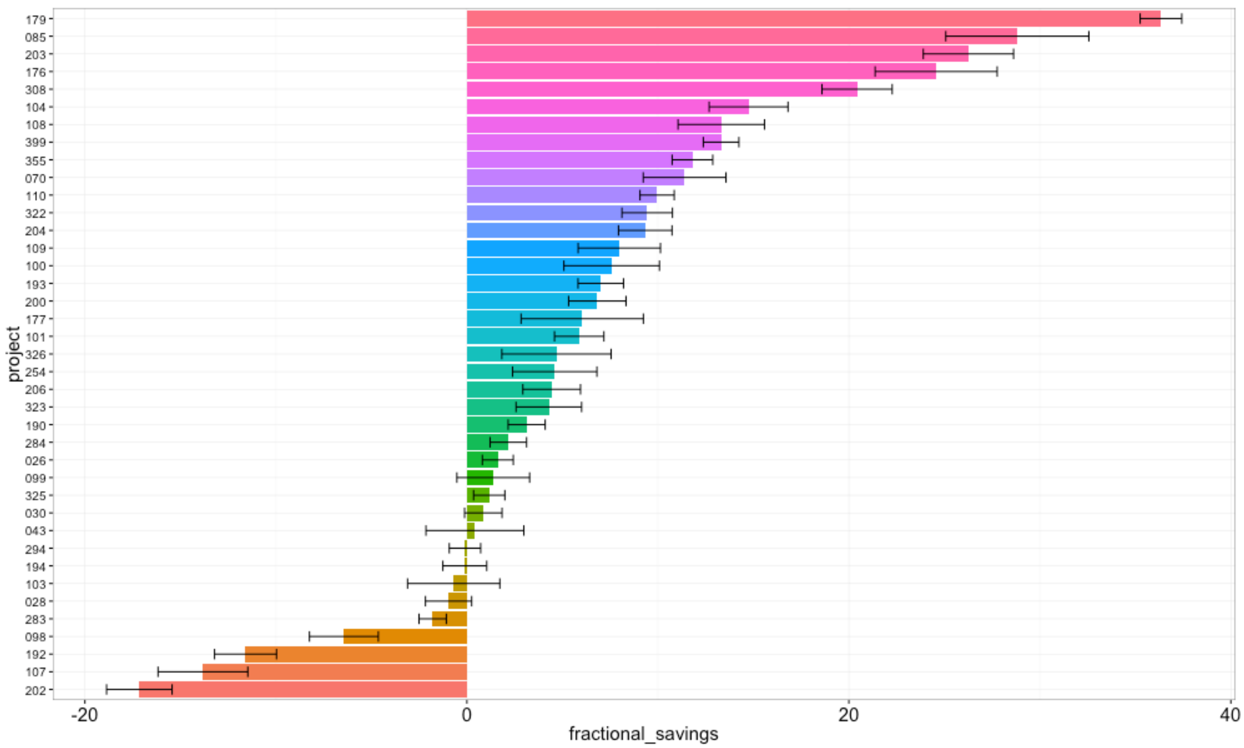

Figure 3. Savings uncertainty ranges for each of 39 buildings, at 95% confidence level

Reductions in Time and Cost

This automated procedure was first worked through and defined by a researcher. To determine the time required to conduct the analysis, it was documented and replicated by a retro-commissioning and energy efficiency program implementer. Based on his prior professional experience, the implementer estimated that traditional whole-building level M&V (using custom fits for each building, as opposed to automated models) would have required 4 days of labor time, while the automated approach required one day. This suggests that the use of automation to streamline M&V, and the associated steps of EM&V process could be a powerful method of reducing costs. Ongoing work with two additional utility and implementer partners will provide further insights as to the potential for labor cost reductions.

Discussion and Conclusions

The initial results of this work indicate that automated M&V2.0 methods can be used to accurately quantify whole-building savings with the advantage of significant time savings relative to more traditional manual approaches. A workflow for practitioners was established to determine which buildings could be treated in this automated manner, and which would need deeper engineering investigations to reliably quantify savings. At the portfolio-level savings of 4% for a set of thirty-nine buildings were quantified within a confidence interval of [3.66%; 4.26%], at the 95% confidence level, which much is higher than the 68% confidence level suggested by ASHRAE as sufficient for transactions based on energy and demand savings in buildings. In the majority of cases, savings and uncertainties for each individual building were also accurate to levels above the ASHRAE Guideline, however when the buildings were pooled and treated as a portfolio, the accuracy of the result improved, i.e., the uncertainties were reduced significantly. It is important to note that these uncertainties are those attributable solely to errors in fitting the model to the baseline data – M&V protocols and guidelines do not address the uncertainty associated with non-routine adjustments that may be made to account for effects such as changes in occupancy or building scheduling. These uncertainties are likely larger than those associated with baseline model error.

While uncertainty is not commonly considered today, it could hold value for evaluating and reducing project and investment risk. For example, ASHRAE's published methods for computing fractional savings uncertainty depend on depth of savings, length of the training and prediction periods, and model CV(RMSE). "Look-up" tables can be used to explore the likelihood that a given model will produce savings estimations that meet uncertainty and confidence requirements, for a specific set of buildings and expected depth of saving. Uncertainty-based approaches are also valuable because they permit an understanding of the risk of under-performance. They align nicely with approaches used by the financial industry, in which investors are accustomed to decision-making under uncertainty and increased use of these concepts could potentially facilitate more effective project financing by private investors.

Since uncertainty analysis has not commonly been used in the industry, we do not know how the accuracy of automated approaches compares to traditional methods. The absence of this type of benchmark highlights the need to engage the regulatory and evaluation community to determine collective acceptance criteria, both in terms of documentation and transparency, as well as quantitative accuracy requirements. With agreed upon acceptance criteria, regulators could require the use of M&V methods that met these criteria, setting a clear bar for evaluation. This could inform further conversation about the demonstrated accuracy and uncertainty of known methods for a given program type. By agreeing to these criteria and methods up front, evaluation risk and ratepayer risk could be diminished. An approach to energy efficiency that includes M&V methods with known uncertainty may also facilitate the effective integration of energy efficiency into state efforts to measurably reduce greenhouse gas emissions by placing efficiency-based GHG reductions in a robust reporting framework similar to that of supply-side resources.

Future Work

We are currently working to support a small group of program administrators in utilizing these M&V 2.0 methods in their programs. Their current purpose is not necessarily to change the way they report savings to their regulators today, but to demonstrate the feasibility of using M&V 2.0 and explore the benefits of doing so. This is likely to be followed by a larger group of program administrators and regulators beginning to pilot M&V 2.0 methods for programs in which existing use baselines are appropriate. This will be facilitated by both program administrators and regulators engaging in discussion of acceptance criteria for M&V. These pilots will provide the opportunity to more thoroughly assess the potential benefits of M&V2.0 in terms of reducing time and cost requirements while maintaining an accurate result; the potential tradeoffs associated with less intensive engineering investigation can also be determined.

Future work will also investigate the development of analytical methods to automatically identify, or 'flag' the potential need for non-routine adjustments – e.g. changes in a building's function, occupancy, or schedule that may need to be accounted for when using avoided energy use to determine measure or project-based savings. If the detection of non-routine adjustments can be automated in ways that are acceptably accurate, then the cost and accuracy of M&V 2.0 can be further improved, and the range of buildings on which they can be used increases.

Finally, we will explore approaches to uncertainty quantification that can be applied to a wider range of models (e.g. machine leaning and highly non-linear techniques), and new ways of measuring high-impact independent variables. For example, if occupancy sensing could be scaled, and made as accessible as weather data for use an independent variable, the performance of M&V 2.0 models could be significantly improved without need for manual adjustment.

Acknowledgement

This work was supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, Building Technologies Office, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The authors would like to thank the members of the project Technical Advisory Committee for their participation and feedback throughout the course of the work. We also acknowledge each of the developers who submitted baseline models for inclusion in this study, and those who contributed to the test dataset.

References

ASHRAE Guideline 14 (2014). ASHRAE Guideline 14-2014 for Measurement of Energy and Demand Savings, American Society of Heating, Refrigeration and Air Conditioning Engineers, Atlanta, GA.

Efficiency Valuation Organization (EVO). International Performance Measurement and Verification Protocol: Concepts and options for determining energy and water savings, Volume I. January 2012. EVO 10000-1:2012.

Granderson, J, Lin, G. Building energy information systems: Synthesis of costs, savings, and best-practice uses. 2016. Energy Efficiency, *In Press*.

Granderson, J, Piette MA, Ghatikar, G. 2011. Building energy information systems: User case studies. Energy Efficiency 4(1): 17-30.

Granderson, J, Price, PN. 2014. Development and application of a statistical methodology to evaluate the predictive accuracy of building energy baseline models. Energy, 66, 981-990.

Granderson, J, Price PN, Jump, D, Addy N, Sohn, MD. 2015. Automated measurement and verification: Performance of public domain, whole-building electric baseline models. Applied Energy 144: 106-113.

Granderson, J, Touzani, S, Custodio, C, Sohn, M.D, Jump, D. and Fernandes, S, 2016. Accuracy of automated measurement and verification (M&V) techniques for energy savings in commercial buildings. Applied Energy 173: 296-308.

Haves, P, Wray, C, Jump, D, Veronica D, Farley, C. Development of diagnostic and measurement and verification tools for commercial buildings**.** Report prepared for California Energy Commission. 2014.

Katipamula, S, Reddy, TA, Claridge, DE. 1998. Multivariate regression modeling. Journal of Solar Energy Engineering, Transactions of the ASME 120(3):177-184.

Kramer, H, Effinger, J, Crowe E. Energy management and information system (EMIS) software technology assessment: Considerations for evaluating baselining and savings estimation functionality. Pacific Gas and Electric, 2013, ET Project Number ET12PGE1311.

Kramer, H, Russell, J, Crowe, E, Effinger, J. Inventory of Commercial Energy Management and Information Systems (EMIS) for M&V Applications, Northwest Energy Efficiency Alliance, 2013, #E13-264.

Reddy, TA, Saman, NF, Claridge, DE, Haberl, JS, Turner WD, Chalifoux AT. 1997. Baselining methodology for facility-level monthly energy use – Part 1: Theoretical aspects. AHSRAE Transactions 103(2):336-347.

Walter T, Price PN, Sohn MD. 2014. Uncertainty estimation improved energy measurement and verification procedures. Applied Energy 130:230-236.