

Measurement and Verification System (M&V) For a DSM Bidding Program

David Sumi, Hagler Bailly Services, Madison, WI
Doug O'Brien, Wisconsin Electric Power Company, Milwaukee, WI
Waiman Mok, Hagler Bailly Services, Madison, WI

ABSTRACT

Measurement and Verification System (M&V) for a DSM Bidding Program

Introduction

This paper presents a description of a real-time monitoring and verification feedback system (M&V) for Wisconsin Electric Company's (WE) Demand-Side Management (DSM) Bidding Program. The M&V system incorporates a set of performance measures and implementation procedures developed for this program in 1996, and then continued over the duration of WE's 1997-1998 program. The system is now being implemented again for the 1999 version of the program. In addition to verifying installation, measuring hours of operation, and measuring customer satisfaction, this M&V system is unique in that it also captures free-ridership and persistence information.

The purpose of this paper is to describe the key issues that need to be addressed for each type of performance measure and describe the research that has addressed these issues over the past three years. In particular, the paper examines some changes in Energy service companies' (ESCO) delivery of products and services that are attributable to results produced by the M&V system. These changes have contributed to greater market competitiveness for these Energy service companies.

Summary of the DSM Bidding Program

Wisconsin Electric's DSM Bidding Program is targeted at residential, farm, and small commercial customers giving them access to both electric and gas energy efficiency products and services. From WE's perspective, the program achieves two objectives: (1) it meets net energy savings goals set by the Public Service Commission of Wisconsin (PSCW); and (2) it encourages transformation toward a self-sustaining market for energy conservation products and services through encouraging third party involvement (the Energy service companies).

Wisconsin Electric contracts on an annual basis with multiple Energy service companies to implement the program. These contractors deliver a turnkey implementation that includes program design, marketing, service/product delivery, and administration. Since the inception of the program in 1996, these contractors have delivered energy efficient products and services to over 100,000 customers throughout the service territory of WE. These customers range from single-family homes, apartment complexes, farms, grocery stores, restaurants, hotels and motels, retail stores, schools and colleges, hospitals, nursing homes to warehouses. There are over 120 energy efficiency measures included in the program covering a broad range of end uses (building shells/envelope insulation, lighting, refrigeration, space heating and cooling, clothes washers, etc.). So far, WE has invested over \$30 million in the program.

Summary of the M&V System

Since 1996 the M&V system has been used to document contractor performance in delivering project-specific kWh and therm savings, and has been the basis for performance-based compensation for WE's contractors. The M&V system has four goals:

- To ensure the net savings targets for the DSM Bidding Program are being achieved
- To correct the net savings claimed by the Energy service companies
- To gain acceptance for the DSM Bidding Program with regulators
- To help move Wisconsin Electric into a competitive environment of market-driven products and services, and to gain experience with customer-based performance measures potentially applicable to other product and service offerings

The purpose of M&V for this program is to (1) develop performance measures for ESCO-delivered energy efficiency products/services, and then (2) conduct data collection and analysis necessary for producing estimates of performance (corrected net energy savings) used to adjust ESCO payments. As stated above, the following performance measures are estimated: measure installation verification; customer satisfaction; free-ridership; hours of operation; and persistence. Energy efficiency measure (EEM-specific) assumptions for gross savings and the performance measures have been developed primarily from the Wisconsin Statewide Database with modifications agreed to by WE, the evaluator, the PSCW, and the Energy service companies. These assumptions, and their measurement and verification, are agreed to in the Energy service companies' contracts with WE.

In periodic meetings with each of the Energy service companies (either in-person or telephone), Hagler Bailly Services staff gather information concerning each Energy service companies' planned products/services (by fuel and customer type) and delivery mechanisms. The distinctions between these "program" types are directly related to the types of survey questions that must be designed to conduct the M&V across the performance measures. Over the past three years, Hagler Bailly Services's performance measures have addressed ESCO program efforts that can be grouped into six general types. These types reflect separate products/services and/or delivery mechanisms, summarized as follows:

- *Comprehensive Assessment Direct Install*
- *Comprehensive Assessment Non-Direct Install*
- *Follow-Up Visit Direct Install*
- *Follow-Up Visit Non Direct-Install*
- *Mail-Order*
- *Retail Sales*
- *Other Direct Install*

Summary of M&V System Implementation

Two data collection efforts are conducted concurrently to estimate the five performance measures outlined earlier. First, telephone surveys are conducted annually with approximately 4,000

to 5,000 customers participating in one (or more) of the programs across all of the Energy service companies involved in the program. The total sample size can vary according to the volume of participating customers, as reflected in the claimed energy savings (invoices) submitted to WE by the Energy service companies.

The telephone surveys are an ongoing effort to collect time-sensitive and actionable feedback on customer reactions to the participation process and their decisions regarding measure installation. The surveys are implemented on a monthly basis with a specified percentage of program participants. The overall sample design incorporates stratification variables including: contractor, measure type, customer sector, date of installation, and size of expected savings. Monthly sampling plans are developed that include not only these stratification variables but also data collection for measure types requiring additional attention (i.e., whenever results appear to be deviating from the prescribed assumptions for the performance measure).

Second, telephone survey data collection is augmented with strategically sampled on-site visits to verify measure installation, location and usage, and retention. As telephone survey samples are completed, these results are examined to develop other samples of on-sites, which are then used to further examine measure location and usage issues for specific ESCO products/services.

The remainder of this paper describes the following aspects of the M&V System:

- The assumptions and key issues for specific performance measures as they apply to different energy efficiency measures and energy end-use categories (from the Wisconsin Statewide Database)
- The sampling implications for the estimation of the performance measure
- The specification of the analysis methods for ongoing estimation of performance measures, including the procedures for assessing whether performance results are significantly at variance with prescribed assumptions (standards).

The paper concludes by focusing on how the M&V results are used. A critical capability of the M&V system is that Energy service companies be provided results in as close to a real-time mode as possible (e.g., within 70 days for claimed energy savings). This has enabled Energy service companies to make performance-based changes to program delivery or target markets – and then receive prompt M&V feedback regarding impacts of program changes. Finally, examples of these M&V-stimulated program changes are provided in the paper.

The Concept of Performance Measurement Control Bands

For each month of the DSM Bidding Program, each implementation contractor (ESCO) submits a report to WE and Hagler Bailly Services - Madison that details the energy conservation measures (ECM's) installed and the associated kWh or therm savings claimed. An *Electronic Achievement Reporting* guideline has been provided by Hagler Bailly Services and WE that prescribes the layout of this information for all contractors (this guideline is designed to serve WE's invoicing requirements as well as the M&V System). These savings are expressed as "net" savings; that is, they combine assumed gross savings from the ECM with assumed factors for free-ridership and persistence.

Assumptions for ECM-specific gross savings, free-ridership, and persistence are obtained primarily from the Wisconsin Statewide Database.

The M&V performance measurement system then tests these ECM-specific free rider and persistence assumptions with customer samples drawn from the participants in each contractor's programs. In addition to free-ridership and persistence, two other specific performance measures also directly influence net savings: ECM installation verification and hours of operation. Monthly estimates for these four performance variables *together* are reported to WE using analysis procedures that recommend "corrected net savings." The corrected net savings integrate the effects of performance on the four variables relative to the standards and assumptions, and are used by WE to determine any adjustments to be made to a contractor's payments.

This section of the paper describes in general the methods for assessing whether performance falls within "control bands" relevant to each variable. This involves estimating the width of the control band using the following assumptions and parameters:

- Wisconsin Electric's assumptions for specific performance variables (e.g., a requirement of 100 percent performance on installation verification for all ECM's), as well as their assumptions for individual ECM's (e.g., 95 percent persistence after one year for "hard-wired" ECM's; or 25 percent free-ridership on compact fluorescents).
- Use of either an 80 percent confidence interval (CI) for two-tailed statistical tests of a performance variable when the value for the variable can vary both above and below WE's assumption (as with free-ridership, persistence, and hours of operation), or a 90 percent CI for one-tailed tests of a performance variable when the value for the variable can vary in only one direction from WE's assumption (as with measure installation verification). The CI calculations incorporate a correction factor for finite populations.
- The monthly sample sizes for the ECM. Since the monthly sample sizes tend to yield rather small n's at the contractor/ECM level, a succession of monthly samples is cumulated for purposes of calculating ECM-specific confidence intervals (Bayesian statistical procedures are used to cumulate the monthly sample distributions for specific performance variable survey responses). Also, for the initial months of data collection in a given program year, ECM-specific sample sizes per contractor are aggregated (by technology, customer segment, or -- if necessary -- by contractor, i.e., all customers participating in a contractor's offerings) to ensure sufficient sample size to yield reasonable confidence intervals across the performance variables.

If the performance assumption, or standard, for any of the four variables falls outside its control band, the percentage variance is calculated (weighted) in terms of kWh or therm impacts. Weighting by kWh/therm impacts yields a more appropriate estimate of performance because it takes into account the relative savings attributable to a given ECM.

Based on these impacts, an adjustment to net savings is made for that month for that ECM (or aggregation of ECM's). The adjustment reflects the sum of all ECM-specific, percentage variances that are outside of the relevant control bands for each of the four performance variables. For each month, the net savings adjustments are summed across all ECM's.

This monthly total net savings adjustment is the basis for determining any adjustments to contractor payments. The difference between net savings claimed by a contractor and the corrected net savings (i.e., the net savings adjustment) are appropriately allocated to either kWh or therms. All payment adjustments are made by WE.

For the verification of installation and free-ridership measurement, contractors are provided M&V results within 70 days for claimed energy savings reported as prescribed in the *Electronic Achievement Reporting* guidelines. A purpose of the 70-day turn-around for M&V results is to enable contractors to examine performance variable results that are outside of control bands *and make appropriate changes in a timely fashion*. For the persistence and hours of operation measurement, M&V results are available approximately 8 months after receipt of names from the contractors.

Examples of the performance assumptions and applicable statistical tests are discussed below for all five performance measures. For each of the four performance measures that can affect net savings, the analysis procedures that produce recommended corrected net savings are discussed.

Measure Installation Verification

This particular performance measure assesses an essential responsibility of each contractor: to deliver/install the ECM's and accurately report which ones were delivered/installed. Measure installation verification directly influences estimates of the effective lifetime of an ECM and thus affects adjustments to claimed net savings.

Key Issues and Assumptions. Wisconsin Electric assumes that 100 percent of the measures delivered/installed by a contractor (i.e., claimed toward net savings payments) are in fact installed. Verifying 100 percent of the installations can be problematic for programs where ECM's are not directly installed by the contractor (or a subcontractor delivering program services for the contractor). For example, ECM's obtained by customers through the mail or ECM's delivered to commercial/farm customers may warrant more careful installation verification.

Data Elements. Survey questions are designed to determine what percent of the ECM's reported by contractors were actually installed (any additional adjustments for ECM's installed but subsequently removed will be calculated as part of the *persistence* performance measurement described below). Questions are asked about each ECM received through the program to determine the variance, if any, from the assumed 100% verification. The exact wording of the survey questions vary depending upon the contractor's product/service and/or delivery mechanism.

Generally, the survey questions verify that the same quantity of ECM's were actually received. Customers are then asked how many of the ECM's were ever installed. For some ECM's (e.g., programmable thermostats), additional information is collected to verify not only installation, but *appropriate* installation (for the programmable thermostat example, the thermostat should have actually been programmed to automatically adjust the temperature settings). For the ECM verification measurement, customers receive a verification value of 100% if the ECM entered in the *Electronic Achievement Reporting* was installed *regardless of whether it is still installed*. Each ECM never installed in WE's service territory receives a value of 0%.

Sampling. A key issue for this performance measure is to sample and complete telephone surveys for a sufficient number of *customer-installed* ECM's. The on-sites are used as a diagnostic means to further investigate the ECM and customer (segment) types that seem most vulnerable to non-installation.

Analysis Method and Control Band Assessment. Installation verification is estimated at the individual ECM level. Since WE assumes 100 percent installation for all ECM's reported by contractors, the control band is evaluated using a one-tailed test and a 90 percent confidence interval (CI). For any given ECM, the "width" of the control band is a function of (1) the appropriate CI test and (2) the monthly sample sizes for the ECM (cumulated across successive months to provide greater sample precision). Typically, verification rates are about 85 percent with a confidence interval of 78 and 92 percents.

Customer Satisfaction

Our goal in developing the customer satisfaction measures for the telephone survey is to design a set of measures that monitor satisfaction and provide diagnostic information on the sources of dissatisfaction. Specifically, this section of the survey determines customers' opinions of contractors' contacts, contractors' knowledge and responsiveness, quality (including clean-up) of installations, and ECM performance. As with all reporting of M&V performance, contractor-specific results on customer satisfaction measures are provided only to that contractor and WE.

Key Issues and Assumptions. Unlike the other performance measures, the objective of the satisfaction measurement is to provide WE and each of the contractors with feedback and diagnostic information on sources of customer dissatisfaction. These data are not used to adjust claimed net savings. Rather, they are used to specifically highlight areas where each contractor is providing satisfactory (or more than satisfactory) performance or areas where they need improvement.

Measures of satisfaction are tracked over time to determine if improvements have been made where needed, or if any new problem areas are occurring. To provide additional feedback, the telephone survey asks customers who provide low satisfaction ratings why they are dissatisfied. This produces a list of problems reported by customers that can be used to identify potential areas for correction.

Two types of satisfaction measurements are relevant for this program. The first type focuses on customer satisfaction with various aspects of the program and service delivery: How satisfied are customers with their phone contact? The responsiveness of ESCO staff? The amount of the incentive or product discount?

The second type centers around customer satisfaction with the performance of the program ECM's: How satisfied are customers with the performance of each ECM? How is their satisfaction affecting persistence?

A minimum level of satisfaction performance needs to be established in order to provide meaningful feedback. This target goal must take into account the fact that across the contractors there are a number of different delivery mechanisms, all of which could impact satisfaction. For example, based on our past research, we would expect satisfaction to be very high for a direct-install program

that involves less time, effort, and cost on the part of the participant. On the other hand, there may be lower levels of satisfaction with a mail order (retail sales) program where the customer is expected to request the product, pay for a portion of it, wait for the product to be delivered, then figure out how to install and use the product on their own.

Based on our past experience with different types of programs and delivery mechanisms, and our review of the 1996 through 1998 DSM Bidding Program contractors' performance, a reasonable target minimum is that 85 percent of customers must be satisfied with a contractor's performance. For scale items, satisfied is defined as '7' or more on a 10-point scale where 0 means "not at all satisfied" and 10 means "very satisfied."

Sampling Implications. Since satisfaction is measured at both the customer level and the ECM level, two different types of samples are necessary. The ECM-specific satisfaction estimate is obtained from the same samples used for the other performance measures. Those samples are representative of ECM's, not of customers. Therefore, a random sample of customers is necessary to obtain estimates of customer satisfaction that are representative of the population of all customers.

Free-ridership

One of the DSM Bidding Program's most critical performance measures, free-ridership, requires continuous monitoring. Our goal in developing free rider performance measures is to construct a set of direct questions that ask about actions customers would have taken in the absence of a contractor's installation (or delivery) of a energy conservation measure.

Key Issues and Assumptions. Most of the issues pertain to applying free rider-type questions to the different contractor delivery mechanisms (i.e., energy audit, direct mail). The rebate-style programs are not a difficult problem. For these programs, we ask whether the respondent would have (1) installed same quantity of ECM at the same time; (2) installed same quantity ECM at a later date; and (3) purchased ECM(s) of a similar efficiency level. These timing, quantity, and efficiency dimensions of free-ridership are appropriate for this type of program delivery.

However, questions measuring customer intentions on these dimensions do not apply as well to the other program types. For example, questions about efficiency do not work well for compact fluorescent ECM's. Here, the customer either adopts the ECM or not—there are no efficiency choices. Similarly, questions that remind or inform program participants of the incentive amount they received from a contractor for each ECM are only applicable to the rebate/coupon program type. Thus, different free rider questions are used to reflect the separate products/services and/or delivery mechanisms of the different program types.

Sampling. Within the residential sector programs and energy conservation measures, certain ECM's have differing free rider percentage assumptions depending on whether the ECM is installed in a single-family (SF) or a multifamily (MF) dwelling. Though this distinction does not modify our sample designs (the sample sizes for the M&V System are too small to use another stratification variable), we still need to break out separate free rider analysis results for SF vs. MF installations. If the sample sizes are too small for this break-out, a weighted free rider estimate is developed to account for different free rider percentage assumptions across ECM installations.

Persistence

Like the ECM-specific free-ridership estimates, the persistence of ECM's is a critical factor for potential adjustments to net impact calculations. The design of the telephone surveys and the on-site data collection incorporate measurements for several key dimensions of savings persistence.

Key Issues and Assumptions. Whether conducted on site or by telephone, the performance measurement for persistence captures data on the current status of ECM's installed at the customers' location. The key is to use separate question modules that are appropriate for the various technologies across the different program types (delivery mechanisms) used by the contractors.

Data Elements. The estimation of measure persistence does not occur until customers have had the ECM delivered/installed for at least six months, and no longer than 12 months. Where applicable, this allows sufficient time for customers to install the ECM and to have acquired experience with it.

In the telephone survey, the persistence questions determine:

- Whether the ECM was ever installed/removed.
- For all ECM's previously installed, the survey determines if the ECM is currently in place.
- If removed, reason(s) for removal or disablement.
- Whether removals are likely to persist into the future, or if they were caused by factors that are unlikely to recur.
- Other factors affecting potential future loss (or extension) of ECM life, such as customer dissatisfaction and/or unmet expectations regarding the operation of the ECM and its energy savings.

Analysis Method and Control Band Assessment. The Wisconsin Statewide Database assumptions applying to the persistence performance variable specify that all hard-wired ECM's will have annual, target persistence rates of 95 percent. An example of a hard-wired ECM is hot water pipe wrap. Soft-wired ECM's are assumed to have annual persistence rates of 85 percent. These include low flow showerheads, faucet aerators, and hot water tank temperature set-back. Some ECM's are considered hard-wired in some applications (e.g., compact fluorescents installed in multifamily dwellings) but soft-wired in others (e.g., compact fluorescents in single family dwellings).

Persistence is estimated at the ECM level using the applicable hard-wired/ soft-wired assumption, which means that information on the installation setting must also be gathered (e.g., multifamily vs. single family dwellings). Since the persistence assumptions have two values that vary by ECM, the control band is evaluated using a two-tailed test and a 80 percent CI.

Hours of Operation

Like free-ridership and measure installation verification, contractor payments are subject to adjustment if the operating-hours estimates assumed in the net savings impact calculations vary from observed operating hours. The performance measurement challenge for this variable is to reliably and cost-effectively collect the relevant data on observed operating hours.

Key Issues and Assumptions. Measurements of average annual operating hours are usually defined as the equivalent number of hours that the equipment would be running if it were operating at full load. Precise estimates of operating hours require run-time metering, such as lighting loggers. However, telephone surveys can be used to reliably obtain equipment operating hours by asking respondents to report what percentage of a particular type of equipment (e.g., rebated fluorescent lighting) is on during each hour in a typical week. While these hourly data can then be used in equations for determining annual operating hours, collecting this detailed information tends to dominate the entire telephone survey interview. Very little other information can be elicited in a typical 15- to 20- minute survey.

Thus, for the operating hours variable, we conduct a less data-intensive approach for the telephone survey.

- First, the hours of operation performance variable is estimated using only lighting ECM's (however, WE reserves the option to include additional ECM's in the hours of operation performance variable, e.g., motors).
- Second, the telephone survey questions focus on assessing whether average use meets *two-thirds of the hours of operation assumption from the Statewide Database*. Since this assumption represents *annual* hours of use needed to yield the assumed kWh savings, we incorporated a seasonal (monthly) correction factor. Our assessment of whether lighting use meets the two-thirds threshold (or exceeds 133% of the standard) adjusts for two seasonal variables: possible hours of sunlight and percentage of possible sunshine. Using monthly climatological data for Wisconsin, we have prepared an adjustment factor that is applied to the monthly survey responses to the hours of use questions. This factor corrects for the differences across the months in the percentage of darkness, as follows:
 - (Assumed hours) * (%dark index for month surveyed) = adjusted standard
- Third, if the telephone survey response indicates that average use *exceeds* the hours of operation assumption from the Statewide Database by *more than one-third*, there will be an opportunity for an adjusted increase to net savings.

M&V System Reporting

The basic objective of the reporting task is clear: the M&V system must provide *ongoing* and *timely* feedback to WE, each contractor, and to the Wisconsin Public Service Commission. This means conducting monthly implementation and reporting activities, rather than producing a single report at

the end of the project (although a final compilation report that combines all monthly reports is submitted to WE).

The ability of WE and its contractors to make necessary adjustments and improvements to the DSM Bidding Program is heavily dependent upon timely and accurate reporting of monthly and cumulative results. This is particularly so since the contractors are guaranteed 70-day M&V turnaround on claimed savings.

The monthly reporting consists of a series of spreadsheets – one for each contractor and a cumulative spreadsheet aggregating across all contractor activity. The spreadsheet has the following features:

- By fuel type, and subtotaled by customer segment, the following columns:
 - measure category code
 - measure category description
 - payment savings claimed
 - *verification* savings adjustment
 - *free-ridership* savings adjustment
 - *persistence* savings adjustment
 - *hours of operation* savings adjustment
 - total savings adjustment (corrected net savings)
 - total savings adjustment percentage
- And, the above columns in a spreadsheet that combines electricity and gas to MMBtu
- A summary spreadsheet that denominates the savings (claimed and adjusted) in dollars, using the per kWh and per therm payment amounts specified in the agreements between WE and the contractors (enabling WE to consider contractor payment adjustments that correspond exactly to the corrected net savings)

Examples of M&V-stimulated Changes to ESCO Programs

A critical capability of the M&V system is that energy service companies be provided results in as close to a real-time mode as possible (e.g., within 70 days for claimed energy savings). This has enabled energy service companies to make performance-based changes to program delivery or target markets – and then receive prompt M&V feedback regarding impacts of program changes. Some examples of these M&V-stimulated program changes are provided in this section.

The performance measure most frequently associated with program changes by Energy service companies has been free-ridership. Fairly dramatic M&V results were obtained in 1996 for a residential appliance turn-in program (refrigerators/freezers) offered by one contractor. Though the agreed-upon standard for free-ridership was 20 percent, the M&V results consistently indicated free-ridership at approximately 90 percent. The program was subsequently dropped by the ESCO since the negative adjustments to payments were significant.

In contrast, a contractor offered a residential water kit that was delivered through the mail. Though a 45 percent free-ridership rate had been agreed upon, M&V results indicated 20 percent free-ridership – a result that prompted the contractor to significantly increase their efforts with this program because of the positive adjustments to the contractor's payments.

A similar outcome occurred for a commercial CFL adapter program, a direct install offering from an ESCO. Where the free-ridership standard was 39 percent, M&V results were approximately 17 percent which again produced advantageous adjustments to invoice payments.