M&V ISSUES AND EXAMPLES
1. Introduction

Measurement errors, defined as the difference between estimates and true values, are an expected outcome of any analytical method. Uncertainty is an assessment of the probability that an estimate is within a specified range from the true value. Errors can be introduced at every stage of the measurement and verification (M&V) process, including sampling, measurement, and adjustment. It is nearly impossible to quantify the effect of every potential source of error. M&V reports often limit uncertainty discussions to random error (especially sampling error and regression error). However, reasonable effort should be made to identify and attempt to minimize every potential source of uncertainty.

The Equation 1 below illustrates the typical approach to determining project savings. This approach involves potential measurement errors at the following stages:

1. estimating baseline energy use,
2. estimating the reporting period energy use, and
3. applying adjustments.

\[
\text{Savings reported for any period} = \text{Baseline Period Energy} - \text{Reporting Period Energy} \pm \text{Adjustments}
\]

1.1. Statistics

Statistics is the science of using data to learn about how a real-life process works. Data are not information unless summarized. Statistics summarizes data through the use of measures of central tendency (mean, mode, and median), measures of dispersion (variance, standard deviation, and standard error), measures of association (correlation, covariance, and regression). Statistics is also about measurement, control and communicating of uncertainty around all these measures. Many decisions cannot be made without statistical techniques.

\[\text{Note: The Joint Committee for Guides in Metrology (JCGM) defines uncertainty as a “parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand. NOTE 1: The parameter may be, for example, a standard deviation (or a given multiple of it, or the half-width of an interval having a stated level of confidence. NOTE 2 Uncertainty of measurement comprises, in general, many components. Some of these components may be evaluated from the statistical distribution of the results of series of measurements and can be characterized by experimental standard deviations. The other components, which also can be characterized by standard deviations, are evaluated from assumed probability distributions based on experience or other information. NOTE 3 It is understood that the result of the measurement is the best estimate of the value of the measurand, and that all components of uncertainty, including those arising from systematic effects, such as components associated with corrections and reference standards, contribute to the dispersion.” Evaluation of measurement data – Guide to the expression of uncertainty in measurement. JCGM 100:2008, September 2008.}\]
1.2. Sources of Uncertainty

Uncertainty is an overall indicator of how well a calculated or measured value represents a true value. Without some assessment of uncertainty, it is impossible to judge an estimate’s value as a basis for decision-making.

Ideally, M&V efforts are designed to determine energy and demand savings with some reasonable accuracy. However, this objective can be affected by either systematic error (i.e., not occurring by chance, including measurement error) or random error (i.e., occurring by chance and often due to using a sample rather than a census to develop the measured value).

1.3. Systematic Error

Many M&V studies do not report any uncertainty measures besides a sampling error-based confidence interval for estimated energy or demand savings values. This is misleading because it suggests potentially incorrect assumptions:

- the confidence interval describes the total of all uncertainty sources (which is incorrect), or
- the other sources of uncertainty are not important relative to sampling error.

Sometimes, however, uncertainty due to measurement and other systematic sources of error can be significant. Sources of systematic errors include the following three areas:

1.3.1. Data Measurement

Data measurement errors are typically caused by meter reading errors, technicians incorrectly recording data, equipment failure, incorrect meter placement, or poor calibration. Measurement errors are best handled by using better metering equipment and improving data collection processes. In some cases, permanent meters and sensors associated with a building automation system (BAS) are used. In these cases, it is important to know the design accuracy of the sensors and whether they have been checked or calibrated. In most M&V applications, data measurement error is ignored. This is appropriate when using utility-grade electricity or natural gas metering equipment, and particularly when metering equipment is of high-caliber.

Human errors need to be tracked through quality control processes, which can then be quantified and included in the overall uncertainty assessment. For mechanical devices – such as meters or recorders – it is sometimes possible to perform tests with multiple meters to assess the measurement variability. However, for most of the data collection devices typically used in M&V studies, it is more practical to use manufacturer or industry study information on the likely amount of error.

1.3.2. Data Collection

Non-coverage errors can occur when some parts of a population are not included in the chosen sample. This may include metering a sample of equipment that is not representative of the population of equipment. The
non-coverage error is reduced by investing in a sampling plan that addresses known coverage issues. The M&V plan must clearly demonstrate that the sample is indeed representative by asking the following questions: Was the sample frame carefully evaluated to determine what portions of the population, if any, were excluded in the sample? If so, what steps were taken to estimate the impact of excluding this portion of the population from the final results? For example, the extrapolation to non-metered equipment must address the needed bias correction.

1.3.3. Data Modeling

Estimates are calculated using models. Some models are fairly simple (e.g., estimating the mean), and others are complicated (e.g., estimating response to temperature through regression models). Regardless, modeling errors may occur for several reasons such as using the wrong model, assuming inappropriate functional forms, including irrelevant information, or excluding relevant information. The M&V plan must illustrate the theoretical rationale for the modeling approach and the inclusion of the relevant variables. When such variables are not available for modeling, their impact on the final estimates must be discussed, and ways of reducing the introduced bias addressed. Also, the plan should address the process used for selecting formulae. Analysts should also consider the following questions: Are the models and adjustments conceptually justified? Has the sensitivity of estimates to key assumptions required by the models been tested or explained? Is the final model sensitive to inclusion or exclusion of explanatory variables? Is it sensitive to individual data points?

1.4. Random Error

Any selected sample is only one of a large number of possible samples of the same size and design that could have been drawn from its population. Random errors are the result of using samples instead of the whole population. This may be the result of metering a sample of lighting fixtures or metering all lighting fixtures for a period of time (i.e., not sampling all the lights all the time).

The fundamental assumption of sampling is that the units not measured will have the same central tendencies as the units that were measured. Due to random chance, however, the units selected for measurement might not be representative of the population as a whole.

The simplest sampling situation is that of randomly selecting \( n \) units from a total population of \( N \) units. In a random sample, each unit has the same probability \( \frac{n}{N} \) of being included in the sample.

In general, the amount of error is inversely proportional to \( \sqrt{n} \). That is, increasing the sample size by a factor \( f \) will reduce the error (improve the precision of the estimate) by a factor of \( \sqrt{f} \).

**Addressing Uncertainty in M&V Plans and Reports**

To summarize, M&V plans need to discuss the expected uncertainty associated with:

1. **Metering equipment.** Discuss the metering equipment chosen. Have the meters been tested? What are the manufacturer’s claims for accuracy?
2. **Modeling approach.** Clearly explain the models to be used. What are the dependent and independent variables? Are there any important variables left out? Why? What is the plan for conducting sensitivity analysis?

3. **Sampling design.** Clearly lay out the sample design. How many end-uses will be metered and for how long? What is the potential impact of a metered period relative to a full year? What is the expected level of confidence and precision?

1.4.1. Definitions of Commonly Used Statistical Terms

1 **Sample Mean (\( \bar{x} \))**

The sample mean is a measure of central tendency and is determined by adding up the individual data points \((x_1, x_2, ..., x_n)\) and dividing by the total number of these data points \((n)\), as follows:

\[
\bar{x} = \frac{\sum(x_1, x_2, ..., x_n)}{n}
\]

2 **Sample Variance (\( s^2 \))**

Sample variance measures the extent to which observed values differ from each other (i.e., variability or dispersion). The greater the variability, the greater the uncertainty in the mean. Sample variance is found by averaging the squares of the individual deviations from the mean. The reason these deviations from the mean are squared is simply to eliminate the negative values (when a value is below the mean), so they do not cancel out the positive values (when a value is above the mean). Sample variance is computed as follows:

\[
s^2 = \frac{\sum(x - \bar{x})^2}{n - 1}
\]

3 **Sample Standard Deviation (\( s \))**

This is the square root of the sample variance described above. The sample standard deviation brings the variability measure back to the units of the data (as the computation above produces an estimate of variation in squared units; e.g., if the data in question are kWh, the variance units are kWh², and the standard deviation units would be kWh).
The standard deviation is a measure of the variability among observations in the data. In normally distributed data, about 68% of observations are within ±1σ of the mean, and 95% are within ±2σ of the mean (note that a large standard deviation indicates greater dispersion of individual observations about the mean.)

**Standard Error (SE) of the Mean of a Sample (s(\bar{x}))**

This measure is used in estimating the precision of the average s(\bar{x}) (estimated through a sample) and is calculated as the sample standard deviation (s) divided by \( \sqrt{n} \).

\[
s(\bar{x}) = \frac{s}{\sqrt{n}}
\]

*SE* is used to estimate the precision of the estimated mean. *SE* is a statistical term that measures the accuracy with which a sample represents a population. In statistics, a sample mean may deviate from the actual mean of a population; this deviation is the *SE* of the population mean (or the true mean) is never observed, yet, the sample mean is. Once we sample and estimate the sample mean and if our sampling was not biased in any way, then this unknown true mean is likely to be within some distance measured in *SE* units from the true mean (we make statements like “we are 90% confident that the true mean is within 2*SE* units of the estimated sample mean”). Assume that, through metering 100 fixtures, we estimate the average hours or use to be 2.5 hours (mean). Also, we estimate the standard deviation to be 1. *SE* = \( \frac{1}{\sqrt{100}} = 0.1 \). We do not know the true hours of use. We only have an estimate of it through the metering of 100 fixtures. However, based on our data, we can say that we are 95% certain that the true value falls between 2.5±2*(.1), or the 95% confidence interval is 2.3 to 2.7 hours.

**Coefficient of Variation (CV)**

The coefficient of variation is simply the standard deviation of a distribution expressed as a percentage of the mean. This is equivalent to the inverse of the signal-to-noise ratio.

\[
CV = \frac{s}{\bar{x}}
\]

**Standard Error of the Regression Estimate**

Also, sometimes referred to as the standard error of the regression, standard error of y or root mean squared error (RMSE). This measure is used in modeling approaches to quantify the amount of variation that exists around the model in question. Similar to the standard deviation, the standard error of the estimate is

\[
s = \sqrt{s^2}
\]
in the same units as the dependent variable, i.e., the variable which is being modeled (which in this context is usually kWh). The standard error of the estimate \( s(\bar{Y}) \) computed as:

\[
s(\bar{Y}) = \frac{\sqrt{\sum(Y_i - \bar{Y})^2}}{n - k - 1}
\]

In this equation, \( \bar{Y} \) is the predicted value of energy (\( Y \)) from the regression model and \( k \) is the number of explanatory variables in the regression equation. \( s(\bar{Y}) \) is used to estimate the precision of the estimated dependent variable produced by a regression model.

1.4.2. **Confidence and Precision**

Statistical methods are available for calculating standard errors for a wide range of estimators (e.g., average energy). Once an estimator’s standard error is known, it is a simple matter to express the estimator’s uncertainty. The most common approach is through a confidence interval – a range of values that is believed with some stated level of confidence – to contain the true population value. The confidence level is the probability that the interval actually contains the population value. Precision provides a convenient shorthand for expressing the interval believed to contain the estimator. For example, if the estimate is average motor use at a facility is 5,300 kWh (obtained by metering a random sample of motors), and the relative precision level (see below) is 14%, then the confidence interval is 5,300 ± 742 kWh. In reporting estimates from a sample, it is essential to provide both the precision and its corresponding confidence level (commonly 90% for M&V applications).

In general, high levels of confidence can be achieved with wider intervals. Narrower (more precise) intervals permit less confidence. In other words, when all else is held constant, there is a trade-off between precision and confidence. Any statement of precision without a corresponding confidence level is incomplete. In the motor example, above, recall the average use from metering a random sample is 5,300 kWh, and the analyst determines this estimate to have 14% relative precision at the 90% confidence level. The analyst may have stated the same estimates with 80% confidence and much lower levels of precision. You can be 90% confident that the true average motor use is 5,300 within ±14% or you can be 80% confident that it is within ±11%. You can get a narrower confidence interval, but you are less confident that the true value will fall within it. You can be nearly 100% confident that the true value will fall within ±200%, for example. Again, reporting one without the other is meaningless.

To calculate the confidence interval for an estimator, first, multiply the estimator’s standard error by a t-value. Then add this product to the estimate itself to obtain the CI upper bound, and subtract the product from the estimate to obtain the lower bound.

---

2 Note the counter-intuitive implication of this standard definition. Low-precision values correspond to narrow intervals and, hence, describe tight estimates. This can lead to confusion when estimates are described as having “low precision.”

3 Some use what is called the z-value here. If the sample size, \( n \), is small, a t-value with \( n-1 \) degrees of freedom is more appropriate than a z-value. The choice of t- versus z-value makes little difference for sample sizes greater than 30. The TINV() function in Microsoft Excel can be used to calculate t-values.
Note that the t-value depends only on the confidence level chosen for reporting results. That is for a given estimate \( \hat{x} \) (e.g., the sample-based estimate of motor use in kWh; while \( x \) is the true unknown motor use), the confidence interval is:

\[
\hat{x} - t \cdot s(\hat{x}) \leq x \leq \hat{x} + t \cdot s(\hat{x})
\]

In this equation, a t-value of 1.645 is used for the 90% confidence level, and a value of 1.960 is used for the 95% confidence level. (These values are tabulated in most textbooks and can be calculated with a spreadsheet.) The absolute and relative precision at the selected confidence level are:

\[
\text{Absolute Precision (} \hat{x} \text{)} = t \cdot s(\hat{x})
\]

\[
\text{Relative Precision (} \hat{x} \text{)} = \frac{s(\hat{x})}{\hat{x}}
\]

The standard error always has the same physical units as the estimator, so absolute precision always has the same physical units as the estimation target. Relative precision, however, is always unit-free and expressed as a percentage.

**Example**

Average use of motors computed from metering a random sample of units is 5,300 kWh. The \( s(\hat{y}) \) is estimated at 450 kWh. Then we have 90% confidence that the true population mean lies within:

\[ 5,300 \text{ kWh} - (1.645 \cdot 450 \text{ kWh}) \text{ and } 5,300 \text{ kWh} + (1.645 \cdot 450 \text{ kWh}) \]

And the precision formulas are:

\[
\text{Absolute Precision (} \hat{y} \text{)} = 1.645 \cdot 450 \text{ kWh} = \pm 740 \text{ kWh}
\]

\[
\text{Relative Precision (} \hat{y} \text{)} = \frac{740 \text{ kWh}}{5,300 \text{ kWh}} = \pm 14\%
\]

In other words, based on the selected sample, the best estimate of the true (unobserved) population mean is the sample mean (5,300 kWh). We are 90% confident that the true value is within 740 kWh or 14% of the sample-based estimate.

If the estimated outcome is large relative to its standard error, the estimator will tend to have a low relative precision value at a given confidence level. (Low precision values are desirable.) However, if the amount of
variability is large relative to the estimated outcome, the precision will be poor. In extreme cases, the precision level may even exceed 100%, in which case the interval for the outcome overlaps zero, and the analyst has no statistically significant evidence that the true population quantity is even positive. For example, if the observed average savings are 5,300 kWh and the associated relative precision (at, say, 90% confidence) is 150%, then we are 90% confident that the true average savings are somewhere between negative 2,650 kWh and 13,250 kWh. In other words, we have no idea what a motor actually uses based on our sample.

Table 1 shows the value of the t-statistic at different confidence levels, which can also be found in statistic tables, books, or online resources.
### Table 1. T-table

Value of t-statistic at different confidence levels

<table>
<thead>
<tr>
<th>Degrees of Freedom (DF)</th>
<th>95%</th>
<th>90%</th>
<th>80%</th>
<th>50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12.71</td>
<td>6.31</td>
<td>3.08</td>
<td>1.00</td>
</tr>
<tr>
<td>2</td>
<td>4.30</td>
<td>2.92</td>
<td>1.89</td>
<td>0.82</td>
</tr>
<tr>
<td>3</td>
<td>3.18</td>
<td>2.35</td>
<td>1.64</td>
<td>0.76</td>
</tr>
<tr>
<td>4</td>
<td>2.78</td>
<td>2.13</td>
<td>1.53</td>
<td>0.74</td>
</tr>
<tr>
<td>5</td>
<td>2.57</td>
<td>2.02</td>
<td>1.48</td>
<td>0.73</td>
</tr>
<tr>
<td>6</td>
<td>2.45</td>
<td>1.94</td>
<td>1.44</td>
<td>0.72</td>
</tr>
<tr>
<td>7</td>
<td>2.36</td>
<td>1.89</td>
<td>1.41</td>
<td>0.71</td>
</tr>
<tr>
<td>8</td>
<td>2.31</td>
<td>1.86</td>
<td>1.40</td>
<td>0.71</td>
</tr>
<tr>
<td>9</td>
<td>2.26</td>
<td>1.83</td>
<td>1.38</td>
<td>0.70</td>
</tr>
<tr>
<td>10</td>
<td>2.23</td>
<td>1.81</td>
<td>1.37</td>
<td>0.70</td>
</tr>
<tr>
<td>11</td>
<td>2.20</td>
<td>1.80</td>
<td>1.37</td>
<td>0.70</td>
</tr>
<tr>
<td>12</td>
<td>2.18</td>
<td>1.78</td>
<td>1.36</td>
<td>0.70</td>
</tr>
<tr>
<td>13</td>
<td>2.16</td>
<td>1.77</td>
<td>1.35</td>
<td>0.69</td>
</tr>
<tr>
<td>14</td>
<td>2.14</td>
<td>1.76</td>
<td>1.35</td>
<td>0.69</td>
</tr>
<tr>
<td>15</td>
<td>2.13</td>
<td>1.75</td>
<td>1.34</td>
<td>0.69</td>
</tr>
<tr>
<td>16</td>
<td>2.12</td>
<td>1.75</td>
<td>1.34</td>
<td>0.69</td>
</tr>
<tr>
<td>17</td>
<td>2.11</td>
<td>1.74</td>
<td>1.33</td>
<td>0.69</td>
</tr>
<tr>
<td>18</td>
<td>2.10</td>
<td>1.73</td>
<td>1.33</td>
<td>0.69</td>
</tr>
<tr>
<td>19</td>
<td>2.09</td>
<td>1.73</td>
<td>1.33</td>
<td>0.69</td>
</tr>
<tr>
<td>20</td>
<td>2.08</td>
<td>1.72</td>
<td>1.32</td>
<td>0.69</td>
</tr>
<tr>
<td>21</td>
<td>2.07</td>
<td>1.71</td>
<td>1.32</td>
<td>0.69</td>
</tr>
<tr>
<td>22</td>
<td>2.06</td>
<td>1.71</td>
<td>1.32</td>
<td>0.69</td>
</tr>
<tr>
<td>23</td>
<td>2.05</td>
<td>1.70</td>
<td>1.31</td>
<td>0.68</td>
</tr>
<tr>
<td>24</td>
<td>2.04</td>
<td>1.70</td>
<td>1.31</td>
<td>0.68</td>
</tr>
<tr>
<td>25</td>
<td>2.03</td>
<td>1.69</td>
<td>1.31</td>
<td>0.68</td>
</tr>
<tr>
<td>26</td>
<td>2.02</td>
<td>1.68</td>
<td>1.30</td>
<td>0.68</td>
</tr>
<tr>
<td>27</td>
<td>2.01</td>
<td>1.68</td>
<td>1.30</td>
<td>0.68</td>
</tr>
<tr>
<td>28</td>
<td>2.00</td>
<td>1.67</td>
<td>1.30</td>
<td>0.68</td>
</tr>
<tr>
<td>29</td>
<td>1.98</td>
<td>1.66</td>
<td>1.29</td>
<td>0.68</td>
</tr>
<tr>
<td>30</td>
<td>1.96</td>
<td>1.64</td>
<td>1.28</td>
<td>0.67</td>
</tr>
</tbody>
</table>

**Note:** Calculate degrees of freedom (DF) using: \( DF = n - 1 \) (for a simple computation like estimating the mean) or \( n - k - 1 \) (for a regression model), where \( n = \text{Sample Size} \) and \( k = \text{Number of Explanatory Variables} \).
1.5. Sample Size Determination

A sample is a subset of a population selected for direct assessment of one or more variables of interest. The sample design describes the exact method by which population members are selected for inclusion in the sample. Every sample design specifies some element of randomness in the sample selection procedure, but the nature of this randomness varies from one design to the next. For example:

» In simple random sampling (SRS), each member of the population has probability \( \frac{n}{N} \) of being selected. If a sample is selected via SRS, then the usual sample mean, and standard error formula will yield valid results.

» In stratified sampling, auxiliary data are used to partition the population into distinct groups, or strata, and then SRS is performed within each group. In this case, stratum weights are needed to extrapolate from strata to the population.

You can minimize sampling error by increasing the fraction of the population that is sampled. Increasing the sample size typically increases cost. As several issues are critical in optimizing sample sizes, the following steps should be used to set the sample size.

1. Select homogeneous populations, technologies, or end uses

In order for sampling to be effective, the measured units should be expected to be the same as the entire population. If there are two different types of units in the population, they should be grouped and sampled separately. For example, when designing a sampling program to measure the operating periods of room lighting controlled by occupancy sensors, rooms occupied more or less continuously (e.g., multiple person offices) should be separately sampled from offices only occasionally occupied (e.g., meeting rooms). Clearly defining the target population being sampled is critical to the validity of the analysis.

2. Determine the desired precision and confidence levels for the estimate

Improved precision requires a larger sample. For example, if you want 90% confidence and ±10% precision, you mean that the range defined for the estimate (±10%) will contain the true value for the whole group (which is not observed) with a probability of 90%. As an example, in estimating the lighting hours at a facility, it was decided to use sampling because it was too expensive to measure the operating hours of all lighting circuits. Metering a sample of circuits provided an estimate of the true operating hours. To meet a 90/10 uncertainty criterion (confidence and precision) the sample size is determined such that, once the operating hours are estimated by sampling, the range of sample estimate (±10%) must have a 90% chance of capturing the true average hours of use. The conventional approach is to design sampling to achieve ±10% precision at the 90% confidence level. However, the M&V Plan needs to consider the limits created by the budget. For example, improving precision from ±20% to ±10% will increase the sample size by four times, while improving it to ±2% will increase the sample size by 100 times.

Note: This is a result of the sample error being inversely proportional to \( \sqrt{n} \). Selecting the appropriate sampling criteria requires balancing accuracy requirements with M&V costs.
3. Decide on the level of disaggregation

Establish whether the confidence and precision level criteria should be applied to the measurement of all components, or to various sub-groups of components. For example, do you wish to have the 90/10 apply to a specific room type or to the building overall?

4. Calculate the initial sample size

Sample size formula is:

\[ n_0 = \left( \frac{Z \cdot CV \%}{e \%} \right)^2 \]

Where:

- \( CV \% \) is the coefficient of variation, the standard deviation divided by the mean
- \( e \% \) is the desired level of relative precision
- \( Z \) is t statistics for the desired confidence level

For example, for 90% confidence, 10% precision, and a CV of 0.5, the initial sample size is:

\[ n_0 = \left( \frac{1.645 \cdot 0.5}{0.10} \right)^2 = 67.7 \]

A CV of 0.5 is often assumed when determining sample sizes when the actual value is not known. We strongly encourage practitioners to use actual values from other similar studies and only use the 0.5 default value when necessary. Also, the 0.5 may be acceptable to homogeneous populations (e.g., hours of use for individual offices), it may understate variability if the measurement is to take place across different use areas.

Since 90/10 confidence/precision is a common target, samples of size 68 are very common.)

One reason that coefficients of variation of 0.5 are often reasonable in M&V work is that the parameters of interest are typically positive for all (or nearly all) projects. If nearly all projects have savings between zero and 200% of the mean savings, and if the savings values are approximately normally distributed, then a CV of 0.5 will apply.

In some cases, (e.g., metering of lighting hours or use), it may be desirable to initially conduct a small sample for the sole purpose of estimating a CV value to assist in planning the main sampling program.
5. Adjust the initial sample size estimate for small populations

The necessary sample size can be reduced if the entire population being sampled is no more than 20 times the size of the sample. For the initial sample size example, above, \( n_0 = 67 \), if the population \( N \) from which it is being sampled is only 200, the population is only three times the size of the sample. Therefore, the “Finite Population Correction Factor” can be applied. This adjustment reduces the sample size \( n_{red} \) as follows:

\[
n_{red} = \frac{n_0 N}{n_0 + N}
\]

Applying the finite population adjustment to the above example reduces the sample size \( n \) required to meet the 90%/±10% criterion to 50.

As the initial sample size \( n_0 \) or the adjusted sample size \( n_{red} \) are determined using an assumed CV, it is critical to remember that the actual CV of the population being sampled may be different. Therefore, a different actual sample size may be needed to meet the precision criterion. If the actual CV turns out to be smaller than the initial assumption in Step 4, the calculated sample size will be larger than required to meet the precision goals. If the actual CV turns out to be larger than assumed, then the precision goal will not be met unless the sample size increases beyond the value computed above.

As sampling continues, the mean and standard deviation of the readings should be computed. The actual CV and required sample sizes should be recomputed. Re-computation may allow early curtailment of the sampling process. It may also lead to a requirement to conduct more sampling than originally planned. To maintain M&V costs within budget, it may be appropriate to establish a maximum sample size. If this maximum is reached after the above recomputations, and the precision goal has not been met, the savings report(s) should note the actual precision achieved by the sampling.

1.6. Determination of Metering Period

The determination of the metering period is often a very important consideration for any M&V project. The key determination is the likely variation throughout the year. Collecting data over the range of independent and dependent variables values is good modeling practice. Any metered load that exhibits weather dependence should ideally be metered over the relevant weather-dependent period. For instance, cooling loads should be metered over the summer and shoulder periods where cooling is required to meet set points. Metering of educational facilities should capture both occupied and unoccupied periods. Industrial metering should cover sufficient variation in production levels to model the probable range of use.

1.7. Modeling

Modeling involves finding a mathematical relationship between a dependent \( Y \) and independent (explanatory) variables \( X \). Models attempt to explain the variations in \( Y \) using the explanatory variables provided by the modeler. For example, if \( y \) is energy use, then weather, size of the facility, orientation,
production levels, occupancy, etc. may all be used to explain the observed energy use. This type of modeling is called multiple regression analysis.

Multiple linear regression equations take the following form:

\[ Y = b_0 + b_1X_1 + b_2X_2 + \cdots + b_kX_k + e \]

The \( b_s \) are the regression coefficients, representing the amount the dependent variable \( Y \) changes when the corresponding independent variable changes by 1 unit. The \( b_0 \) is the constant, which corresponds to the value of \( Y \) that would be estimated if all the independent variables were equal to zero and where dealing with one independent variable, \( b_0 \) can be thought of as the value where the regression line intercepts the \( Y \)-axis. For example, if the \( X_s \) are weather variables, then \( b_0 \) may be the amount of baseload that is not impacted by weather, and the individual \( b \) values explain the amount of energy use that is attributable to weather (e.g., kWh per heating degree days). Associated with multiple regression is \( R^2 \), which is the percentage of variation in the dependent variable explained collectively by all the independent variables.

For example, let \( Y \) be energy use (usually in the form of energy use during a specific time period; one hour, one day, 30 days, etc.). Now let \( X_i \) \((i = 1, 2, 3, \ldots, k)\) represent the \( k \) explanatory variables such as weather, production, occupancy, etc. \( b_i \) \((i = 1, 2, \ldots, k)\) represents the coefficients derived for each independent variable and \( e \) represents the residual errors that remain unexplained after accounting for the impact of the various independent variables. The most common regression analysis finds the set of \( b_i \) values that minimize the sum of squared residual-error terms (these regression models are thus called least-squares models).

An example of the above model for a building’s energy use is:

\[
\text{Monthly Energy Consumption} = 342,000 + (63 \times \text{HDD}) + (103 \times \text{CDD}) + (222 \times \text{Occupancy})
\]

\( \text{HDD} \) and \( \text{CDD} \) are heating and cooling degree days, respectively. \( \text{Occupancy} \) is a measure of percentage occupancy in the building. In this model, 342,000 is an estimate of baseload in kWh, 63 measures the change in consumption in kWh for one additional HDD, 103 measures the change in consumption in kWh for one additional CDD, and 222 measures the change in consumption in kWh per 1% change in occupancy.

1.7.1. Coefficient of Determination (\( R^2 \))

The first step in assessing the accuracy of a model is to examine the coefficient of determination, \( R^2 \), a measure of the extent to which variations in the dependent variable \( Y \) are explained by the regression model. Mathematically, \( R^2 \) is:

\[ R^2 = \frac{\text{Explained Variation in } Y}{\text{Total Variation in } Y} \]
All statistical packages and spreadsheet regression-analysis tools compute the value of $R^2$.

The range of possible values for $R^2$ is 0.0 to 1.0. An $R^2$ of 0.0 means none of the variations is explained by the model, therefore the model provides no guidance in understanding the variations in $Y$ (i.e., the selected independent variable(s) give no explanation of the causes of the observed variations in $Y$). On the other hand, an $R^2$ of 1.0 means the model explains 100% of the variations in $Y$, (i.e., the model predicts $Y$ with total certainty, for any given set of values of the independent variable(s)). Neither of these limiting values of $R^2$ is likely with real data.

In general, the greater the coefficient of determination, the better the model describes the relationship between the independent variables and the dependent variable. Some industry guides suggest 0.75 as a cutoff, but there is no universal standard for a minimum acceptable $R^2$ value, as it is highly dependent on the context. Some portion of a model will always be unexplained, and fit should be assessed with this in mind. For instance, $R^2$ values tend to be higher in time-series analysis, where a single site is analyzed, as a site has less variation from time period to time period (within error) than multiple sites do between each other at any given time period (between error).

The $R^2$ test should only be used as an initial check. Models should not be rejected or accepted solely on the basis of $R^2$. It is important to note that the $R^2$ value is a proportion of the variance of the dependent variable, for example, a low $R^2$ value is less serious when the variance of the dependent variable is small, and correspondingly, a low $R^2$ value becomes increasingly important where the variance of the dependent variable is larger.

Finally, a low $R^2$ may be an indication that some relevant variable(s) are not included, or that the functional form of the model (e.g., linear) is not appropriate. In this situation, it would be logical to consider additional independent variables or a different functional form. However, variables should only be added in the case that they have theoretical merit. Otherwise, they could lead to bias in the estimates.

The $R^2$ value will always increase when you add additional explanatory variables, whether relevant or not. It never decreases. Consequently, a model with more explanatory variables may appear to have a better fit simply because it has more terms. In order to prevent modelers from falling into the trap of maximizing $R^2$ at all costs, it is recommended that Adjusted $R^2$ is used. The Adjusted $R^2$ is a modified version of $R^2$ that has been adjusted for the number of predictors in the model. The Adjusted $R^2$ increases only if the new term improves the model more than would be expected by chance. It decreases when a predictor improves the model by less than expected by chance. The Adjusted $R^2$ can be negative, but it is usually not. The computation of Adjusted $R^2$ uses the following equation where $p$ is the number of explanatory variables used in the regression equation.

\[
\text{Adjusted } R^2 = 1 - \frac{(1 - R^2) \cdot (N - 1)}{N - p - 1}
\]
1.7.2. t-statistic

Since a regression-model coefficient ($b$) is a statistical estimate of the true relationship between an individual $X$ variable and $Y$, it is subject to variation. The accuracy of the estimate is measured by the standard error of the coefficient and the associated value of the t-statistic. A t-statistic is a statistical test to determine whether an estimate has statistical significance.

The standard error of each coefficient is computed by regression software. The following equation applies to the case of one independent variable.

$$s(b) = \sqrt{\frac{\sum(Y_i - \hat{Y}_i)^2 / (n - 2)}{\sum(X_i - \bar{X})^2}}$$

For cases with more than one independent variable, the equation provides reasonable approximation when the independent variables are truly independent (i.e., not correlated with each other). Otherwise, the equation gets very complex, and the M&V analyst is better off using a software package to compute the standard errors of the coefficients. The range within which the true value of the coefficient, $b$, falls is found using (t is the value obtained from t statistical tables):

$$b \pm t \cdot s(b)$$

The standard error of the coefficient, $b$, also leads to the calculation of the t-statistic. This test ultimately determines if the computed coefficient is statistically significant. The t-test is computed by all statistical software using the following equation:

$$t\text{-statistic} = \frac{b}{s(b)}$$

The t-statistic is a value which can be compared to the t-distribution for the appropriate number of degrees of freedom (see Table 1, T-table Value of t-statistic at different confidence levels) in order to decide if the corresponding variable is irrelevant in modeling the dependent variable. This probability is usually referred to as a “p-value.”

All statistical software and spreadsheet provide p-values for the computed regression coefficient. P-value is a number between 0 and 1 and interpreted as:
Small, typically $\leq 0.10$. This indicates strong evidence that the explanatory variable is significant (i.e., does have a significant impact on the dependent variable).

Large, typically $> 0.10$. This indicates weak evidence that the explanatory variable is significant.

Close to 0.10 is considered marginal (could go either way).

Actions to consider in improving the summary statistics of the model:

- select the independent variable(s) with the strongest relationship to energy;
- select the independent variable(s) whose values span the widest possible range (if $X$ does not vary at all in the regression model, $b$ cannot be estimated, and the p-values will be poor);
- gather and use more data points to develop the model; or
- select a different functional form for the model; for example, one which separately determines coefficient(s) for each season in a building that is significantly affected by seasonal weather changes.

1.8. Modeling Errors

When using regression models, as described above, several types of errors may be introduced:

- The mathematical model may not include relevant variables (omitted variable bias).
- The model may include some variables that are irrelevant.
- The model may use inappropriate functional form.
- The model may be based on insufficient or unrepresentative data.

1.8.1. Omission of Relevant Variables

In M&V, regression analysis is the method used to account for changes in energy use. Most complex energy using systems are affected by innumerable explanatory variables. Regression models cannot hope to include all independent variables. Even if it were possible, the model would be too complex to be useful and would require excessive data gathering activities. The practical approach is to include only explanatory variable(s) that are thought to impact energy use significantly.

The omission of a relevant independent variable may be an important error. If a relevant independent variable is missing (e.g., HDD, production, occupancy), then the model will fail to account for a significant portion of the variation in energy. The deficient model will also attribute some of the variations that are due to the missing variable to the variable(s) that are included in the model. The effect will be a less accurate model.

Note that there are two possible consequences to omitted variable bias: less precision and possible bias. Precision is lost in the case where an omitted variable explains changes in energy consumption but is not related to other explanatory variables included in the models. Bias is introduced in the case that an omitted variable is related to both the dependent and independent variable(s).

There are no obvious indications of this problem in the standard statistical tests (except maybe a low $R^2$). Experience and knowledge of the engineering of the system being measured are valuable in addressing this
issue. Simply examining various scatter plots of the parameter of interest plotted against various
independent variables can help in this, as can plotting residuals of preliminary versions of models.

There may be cases where a relationship is known to exist with a variable recorded during the baseline
period. However, the variable is not included in the model due to lack of budget to continue to gather the
data in the reporting period. Such omission of a relevant variable should be noted and justified in the M&V
Plan.

1.8.2. Inclusion of Irrelevant Variables

Sometimes models include irrelevant independent variable(s). If the irrelevant variable has no relationship
(correlation) with the included relevant variables, then it will have minimal impact on the model. However,
if the irrelevant variable is correlated with other relevant variables in the model, it may bias the coefficients
of the relevant variables. Scatter plots of combinations of variables plotted against one another can identify
parameters that are correlated.

Use caution in adding more independent variables into a regression analysis just because they are available.
To judge the relevance of independent variables requires both experience and intuition. However, the
associated p-value is one way of confirming the relevance of particular independent variables included in a
model. Experience in energy analysis for the type of facility involved in any M&V program is necessary to
determine the relevance of independent variables.

1.8.3. Functional Form

It is possible to model a relationship using the incorrect functional form. For example, a linear relationship
might be incorrectly used in modeling an underlying physical relationship that is non-linear. For example,
electricity consumption and ambient temperature tend to have a non-linear (often ‘U’ shaped) relationship
with outdoor temperature over a one-year period in buildings that are both heated and cooled electrically.
(Electricity use is high for both low and high ambient temperatures, while relatively low in mid seasons.)
Modeling this non-linear relationship with a single linear model would introduce unnecessary error. Instead,
it is possible to construct separate linear models for each range of ambient temperatures. Another option is
to build higher order relationships, e.g., \( Y = f(X, X^2, X^3) \). Yet another option is to use a change-point model
(e.g., CDD and HDD rather than temperature).

The modeler needs to assess different functional forms and select the most appropriate among them using
evaluation measures. A useful tool for assessing model functional form is visualization. Often simply
examining scatter plots can make the functional form of the model apparent.

1.8.4. Data Shortage

Errors may also occur from insufficient data either in terms of quantity (i.e., too few data points) or time
(e.g., using summer months in the model and trying to extrapolate to winter months). The data used in
modeling should be representative of the range of operations of the system of interest. The time period
covered by the model needs to include various possible seasons, types of use, etc. This may call for either
extension of the time periods used or increasing sample sizes. For analysis of whole-building data, it is
necessary to obtain both cooling and heating-season data. For an industrial process, it is enough to capture
the range of production levels.
1.8.5. Autocorrelation

Autocorrelation (or “serial correlation”) refers to the relationship between the past and the present values of a time series. As an example, let $Y_t$ represent metered electric load at time $t$. If $Y_t$ is correlated with $Y_{t-1}$ (or $Y_{t-2}$, or $Y_{t-3}$, etc.), then the series is said to be autocorrelated.

Commonly, autocorrelation is discussed in the context of the residuals of a fitted regression model. One key assumption of the least squares regression is that the error terms (or residuals) are independent. When the present value of the dependent variable is correlated with past values, then the amount of error (residual) in time period $t$ is also correlated with the amount of error in time period $t - 1$. In the presence of autocorrelation, regression coefficient estimates remain unbiased, but their standard errors may be biased. Thus, any uncertainty inferences drawn may be incorrect. Option C chapter below provides a detailed description of autocorrelation along with potential remedies.

1.8.6. Prediction Errors

When a model is used to predict an energy value ($Y$) for given independent variable(s), the accuracy of the prediction is measured by the standard error of the estimate ($s(Y)$). As described earlier, this is also known by some other names such as standard error of the regression and root-mean-squared error adjusted for degrees of freedom. In regression modeling, the best single error statistic to look at is the standard error of the regression estimate, which is directly related to the unexplainable variations in the dependent variable ($Y$). This is what your statistics software is trying to minimize when estimating coefficients.

Once the value(s) of the explanatory variable(s) are plugged into the regression model to estimate an energy value ($\hat{Y}$), an approximation of the range of possible values for $\hat{Y}$ can be computed using:

$$\hat{Y} \pm t \cdot s(\hat{Y})$$

Where:

- $\hat{Y}$ is the predicted value of energy ($Y$) from the regression model, at one combination of independent variables
- $t$ is the value obtained from the t-tables (see Table 1)

As described earlier, the standard error of the estimate $s(\hat{Y})$ is computed as:

$$s(\hat{Y}) = \sqrt{\frac{\sum(\hat{Y}_i - Y_i)^2}{n - k - 1}}$$
In this equation, \( k \) is the number of explanatory variables in the regression equation. Again, this often also referred to as the root-mean-squared error (RMSE). The RMSE is a quadratic equation that measures the average magnitude of the error. The differences between the estimated and the observed values are squared and then averaged over the sample. The square root of the average is taken. RMSE gives a relatively high weight to large errors. Dividing the RMSE by the average energy use produces the coefficient of variation of RMSE, or the CV(RMSE), as shown here:

\[
CV(RMSE) = \frac{s(\bar{Y})}{\bar{Y}}
\]

Regression models can also be utilized to directly estimate the change in energy consumption by including an indicator variable equal to 1 in the post-retrofit period and 0 in the pre-retrofit period.

1.8.7. Overfitting

Overfitting occurs when an overly complex model is estimated to the data. This can lead to parameters representing not just the signal but also the noise in the data. This will lead to models that do a poor job of predicting results in other contexts (i.e., outside of the data range in the model).

A good way to protect against overfitting is by using cross-validation techniques. In cross-validation, a portion of the data collected is set aside and not used to estimate the model. Once the model specification is chosen, it is tested against the remaining data. \( R^2 \) and prediction errors can be computed for the test data to give a better estimate of the likely out-of-sample error (error in estimating values left out of the sample, for example, one month of daily energy use data can be used to estimate the model, and another to test the model uncertainty).

1.9. Combining Uncertainty

If the reported savings is the sum or difference of several independently determined components \( (x) \), (i.e., \( Savings \ y = x_1 \pm x_2 \pm \ldots \pm x_p \)), then the uncertainty (or standard error) of the reported savings can be estimated by:

\[
u_c = \sqrt{u_{x_1}^2 + u_{x_2}^2 + u_{x_3}^2 + \ldots}
\]

For example, if savings are computed using the equation:
The uncertainty of the difference (savings) is computed as:

\[
\text{Equation 24}
\]

\[
\begin{align*}
\text{u}_{\text{Savings}} &= \sqrt{\text{u}_{\text{Adjusted Baseline}}^2 + \text{u}_{\text{Reported Period Energy}}^2} \\
\end{align*}
\]

If the reported savings estimate is a product or division of several independently determined components, (i.e., \( \text{Savings} = \frac{x_1 \cdot x_2}{x_3} \)) then the uncertainty of the savings is given approximately by:

\[
\text{Equation 25}
\]

\[
\begin{align*}
\text{u}_c &= \frac{1}{y} \left( \frac{u_{x_1}}{x_1} \right)^2 + \left( \frac{u_{x_2}}{x_2} \right)^2 + \left( \frac{u_{x_3}}{x_3} \right)^2 \\
\end{align*}
\]

A good example of this situation is the determination of lighting savings as:

\[
\text{Equation 26}
\]

\[
\text{Savings} = \Delta \text{Watts} \cdot \text{Hours}
\]

If the M&V Plan requires measurement of hours of use, then “Hours” will be estimated and will have an uncertainty. If the M&V Plan also includes measurement of the change in wattage, then \( \Delta \text{Watts} \) will also be a value with an uncertainty. The relative standard error of savings will be computed using the formula above as follows:

\[
\text{Equation 27}
\]

\[
\text{s}_{\text{Savings}} = \sqrt{\left( \frac{\text{u}_{\Delta \text{Watts}}}{\Delta \text{Watts}} \right)^2 + \left( \frac{\text{u}_{\text{Hours}}}{\text{Hours}} \right)^2}
\]

The components must be independent to use the methods above for combining uncertainties. Independence means that whatever random errors affect one of the components are unrelated to the errors affecting other components.
2. **Option A: Lighting Efficiency**

| Option A | Partially Measured Retrofit Isolation Savings are determined by partial field measurement of the energy use of the system(s) to which an energy conservation measure (ECM) was applied, separate from the energy use of the rest of the facility. Measurements may be either short-term or continuous. The partial measurement means that some but not all parameter(s) may be estimated if the total impact of possible stipulation error(s) is not significant to the resultant savings. A careful review of ECM design and installation will ensure that estimated values fairly represent the probable actual value. Stipulations should be shown in the M&V Plan along with an analysis of the significance of the error they may introduce. |

2.1. **Situation**

Replacing existing lights with higher efficiency models is an extremely common energy efficiency measure. This section compares the measurement errors associated with several methods for metering and calculating the energy savings from a lighting retrofit. It then shows how to calculate uncertainty around those savings. As an example, it uses a vocational school lighting retrofit showing sampling strategies and uncertainty results.

A vocational school replaces existing (baseline) fixtures with more efficient lighting equipment in a year-round training center. The project does not involve any change to existing controls. The contractor is paid based on final estimated savings per terms of an IPMVP-adherent M&V plan.

2.2. **M&V Plan: Run Time Logging - Post Only**

The parties agree to the following terms:

- Baseline and retrofit fixture wattages are based on published tables of wattage (power) for each fixture and are assumed to be known with certainty.
- Lighting hours of use are the key measured parameter\(^4\) and will be estimated based on a metering study of a sample of fixtures. Retrofit hours are assumed to be the same as baseline hours as there will be no change in the lighting controls or building schedule. Metering is conducted during the post-install or performance period.
- The plan will be IPMVP Option A adherent. Three parameters drive the savings calculation; baseline and fixture wattage, quantity of each fixture type, and hours of use. As mentioned above, fixture wattage is determined from lookup tables of published values from independent tests performed to ANSI standards and are known with certainty. Hours are measured. The quantity of each fixture type will be determined through detailed pre- and post-installation inspections.

\(^4\) This example follows standard practice for impact evaluation studies where lighting hours of use is the most commonly measured key parameter for lighting projects enrolled in energy efficiency programs. Evaluators have found hours of use estimates based on schedules and building operator input unreliable, i.e. uncertain. On the other hand, fixture wattages based on published tables and created from ANSI tests have a long track record of matching the power draw of fixtures installed in buildings and are considered accurate. Parties to a performance contract could as plausibly have designated fixture wattage as the key parameter to measure while stipulating the hours of use, and still be Option-A adherent.
The length of the study period will be two months and will be conducted after the project is commissioned and accepted by both parties. The vocational school operates year-round and does not shut down for extended vacation periods. The school closes during national and religious holidays. The demand for lighting is not affected by weather or season. Therefore metering can take place during any two months. Shorter study periods are acceptable provided lighting schedules are near constant as in a commercial office. Baseline and retrofit fixture inventories will be prepared by the contractor and verified by the school. Fixtures will be identified by location and usage group. Usage groups are sub-populations of fixtures with similar operating schedules and lighting hours. Usage groups are an example of stratified sampling, an approach that controls for variance and leads to smaller sample sizes. For estimating the necessary sample size, a coefficient of variation of 0.5 is assumed for all usage groups. The 0.5 is the default based on the record of many lighting hours of use studies. Lower or higher values, corresponding to lower or higher variance, can be used if there is additional available information about the expected variance in the actual lighting hours of use. It is recommended that actual values are used whenever possible. The climate is mild, so the heating and the cooling interactive effects are assumed to be negligible and will not be estimated. Interactive effects are difficult to measure directly. M&V plans can estimate their impacts based on prior studies, which in turn are usually based on computer simulations. Such effects can be significant, especially in climates requiring significant heating and/or cooling. However, actual impacts are highly dependent on HVAC system types, this is why they are often ignored.

2.3. Savings

\[
\text{Savings} = \sum (kW_{\text{baseline}} - kW_{\text{retrofit}})_k \cdot \text{Annual Hours of Use}_k
\]

For each \(k\) usage group.

2.4. Sources of Error

There are two potential types of error for the estimate of annual savings; systematic and random.

The M&V plan stipulates that the fixture wattage is known with certainty and identifies annual lighting hours as the key parameter to be measured. There is a small systematic error associated with the measurement of lighting hours, for example, 0.02% for one logger manufacturer. This error is typically ignored due to its small contribution to total error.

---

5 Length of the meter study period is somewhat dependent on the consistency of the lighting operating schedules. Seasonal schedules may require several studies; one for each season. A stringent alternative is to meter for an entire year. A more relaxed approach would meter for several weeks during a full production or load period and estimate the off-season hours.
2.5. Sample Design

The sample frame is all lighting fixtures involved in the retrofit. The sample unit is a lighting last point of control (LLPC), typically a circuit with a switch that controls a bank of fixtures.

The sample size is determined by the M&V plan requirement that estimated savings be reported at the 90/10 confidence/precision level. A coefficient of variation of 0.5 is assumed and the required sample size is therefore, at 90% confidence (t-statistic=1.645) and ±10% precision, is 68 ($n = 1.645^2 \cdot 0.5^2 / 0.1^2$).

The M&V plan requires oversampling by 10% to allow for some potential loss of data. Therefore the final sample size is 75. The sample is allocated to usage groups based on their share of expected savings. Due to rounding, the final sample size is 77 points (see Table 2).

More sophisticated methods (e.g., Neyman allocation method) allocate the sample based on an expected variance for each usage group. Compared to the simple method used here they would result in fewer meters for usage groups with low expected variance, such as exit signs, and more meters in usage groups with high expected variance.

The sample design and selection are developed using a lighting inventory spreadsheet, usually provided by the lighting contractor. The inventory lists all Lighting Power Circuits (LPC); their location, usage group, annual hours of use, quantity and type for both baseline and efficient fixture. The inventory should contain all the information needed to estimate the project savings.

Sample points are selected by assigning random numbers to each LPC in each usage group, ranking them in descending order, and selecting down the list until the sample size is met. The total sample size is based on the confidence and precision goals in the M&V plan. The sample size is then allocated to each usage group in proportion to its expected savings. Light loggers are installed in one fixture attached to a sampled LPC. Light logger output is a time-stamped data record of when equipment is turned on and off.

Table 2 shows the sample design for the vocational school example.

### Table 2. Sample Design

<table>
<thead>
<tr>
<th>Usage Group</th>
<th>Count of LPC ($N_i$)</th>
<th>Change in connected (kW)</th>
<th>Expected Savings (kWh/Year)</th>
<th>Savings (%)</th>
<th>Sample Size ($n$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labs</td>
<td>97</td>
<td>33</td>
<td>84,872</td>
<td>8%</td>
<td>6</td>
</tr>
<tr>
<td>Offices</td>
<td>182</td>
<td>61</td>
<td>159,136</td>
<td>15%</td>
<td>12</td>
</tr>
<tr>
<td>Classrooms</td>
<td>802</td>
<td>268</td>
<td>700,197</td>
<td>66%</td>
<td>50</td>
</tr>
<tr>
<td>Corridors</td>
<td>134</td>
<td>45</td>
<td>116,699</td>
<td>11%</td>
<td>9</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1,215</strong></td>
<td><strong>407</strong></td>
<td><strong>1,060,904</strong></td>
<td><strong>100%</strong></td>
<td><strong>77</strong></td>
</tr>
</tbody>
</table>

After the two-month metering study has been completed the average annual lighting hours and kWh savings are calculated for each usage group and the building as shown in Table 3.
Table 3. Metered Hours and Estimated Savings

<table>
<thead>
<tr>
<th>Usage Group</th>
<th>Change in connected (kW)</th>
<th>Metered Hours (Annualized)</th>
<th>Verified Savings (kWh/Year)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labs</td>
<td>33</td>
<td>3,781</td>
<td>124,773</td>
</tr>
<tr>
<td>Offices</td>
<td>61</td>
<td>2,905</td>
<td>177,205</td>
</tr>
<tr>
<td>Classrooms</td>
<td>268</td>
<td>2,677</td>
<td>717,436</td>
</tr>
<tr>
<td>Corridors</td>
<td>45</td>
<td>4,233</td>
<td>190,485</td>
</tr>
<tr>
<td>Average or Total</td>
<td>407</td>
<td>2,973</td>
<td>1,209,899</td>
</tr>
</tbody>
</table>

Because the M&V plan stipulates that both the baseline and post-install lighting kW are known with certainty metering error is ignored. Sampling error is the only source of uncertainty for the verified savings. A spreadsheet was used to calculate the confidence interval as shown in Table 4. The key calculations are:

- Standard error for each stratum, \( s(h) = \frac{\text{Standard Deviation}}{\sqrt{n}} \times \text{FPC} \). The FPC is the finite population correction calculated as \( \sqrt{1 - \left( \frac{n}{N} \right)} \), where \( n \) and \( N \) are sample and population stratum sizes.

- Total standard error. \( s(\text{tot}) = \sqrt{\sum (SE_h)^2} \), where \( SE_h \) is the standard error of stratum \( h \).

- Absolute precision. \( s(\text{tot}) \times \text{t-statistic} \). The t-statistic is 1.67, the value for the Sample Size of 77 - 1 or 76 degrees of freedom.

- Relative precision. Absolute precision / total verified energy savings.

The precision formulas for the verified savings are:

- Absolute Precision (kWh\text{save}) = \text{t-statistic} \times s(\text{tot}) = 1.67 \times 62,211 = \pm 103,643 kWh/year

- Relative Precision (kWh\text{save}) = \frac{\text{Absolute Precision}}{\text{kWh\text{save}}} = \frac{103,643}{1,209,899} = \pm 8.57%

The verified savings can be reported as 1,209,899 kWh/year ±62,211, or ±8.57%. We are 90% confident that the true value lies between 1,272,110 and 1,147,688 kWh/year

Table 4. Error Analysis

<table>
<thead>
<tr>
<th>Usage Group</th>
<th>Verified Energy Savings, kWh</th>
<th>Standard Deviation, kWh (s)</th>
<th>Measured CV (s/kWh)</th>
<th>Finite Population Correction (FPC)</th>
<th>Standard Error (s/sqrt(n)) x FPC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labs</td>
<td>124,773</td>
<td>43,362</td>
<td>0.35</td>
<td>0.97</td>
<td>19,392</td>
</tr>
<tr>
<td>Offices</td>
<td>177,205</td>
<td>52,155</td>
<td>0.29</td>
<td>0.98</td>
<td>19,713</td>
</tr>
<tr>
<td>Classrooms</td>
<td>717,436</td>
<td>400,124</td>
<td>0.56</td>
<td>0.97</td>
<td>56,586</td>
</tr>
<tr>
<td>Corridors</td>
<td>190,485</td>
<td>40,590</td>
<td>0.21</td>
<td>0.96</td>
<td>12,238</td>
</tr>
<tr>
<td>Total</td>
<td>1,209,899</td>
<td>62,211</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absolute Precision</td>
<td></td>
<td>103,643</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relative Precision</td>
<td></td>
<td>8.57%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3. Option B: Motor Replacement

Savings are determined by field measurement of the energy use of the systems to which the ECM was applied, separate from the energy use of the rest of the facility. Short-term or continuous measurements are taken throughout the pre- and post-retrofit periods, or in some cases the post period only.

3.1. Situation

Replacing motors with higher efficiency models is a common energy efficiency measure for commercial and industrial facilities. This section compares the measurement errors associated with several methods for metering and calculating the energy savings from motor replacement. It then shows how to calculate uncertainty around those savings. As an example, it uses a set of HVAC fan motors in a hospital showing sampling strategies and uncertainty results. Some simple measurement techniques are used for the purpose of showing the errors they induce, for example the use of run time loggers or recording ammeters.

Motors convert electrical into mechanical energy and are used to serve a variety of loads including pumps, fans, and air conditioning compressors. Motors are sized in the U.S. in horsepower and, in other parts of the world, by kW (HP is equivalent to 0.746 kW). Nameplate sizes indicate the full operational output of the motor, not the power actually consumed by the motor. Motors convert roughly 80% to 96% of electrical to mechanical energy with the remaining energy lost as heat. The low end of the range applies to older, smaller motors and the high end of the range applies to larger and efficient motors.

The energy used by a motor can be estimated by Equation 29:

\[
\text{kWh} = \text{Hours} \cdot \frac{P_{\text{nominal}}}{\text{Efficiency}} \cdot \text{Load Fraction} \cdot \text{Unit Conversion (1 or 0.746)}
\]

Where:

- **Hours** is the length of time the motor runs.
- **\( P_{\text{nominal}} \)** is usually known from motor nameplate. In the US it is indicated as horsepower. Elsewhere it is indicated as kW.
- **Efficiency** is sometimes on the nameplate, may be less well known for older motors, and should be adjusted if the measured load fraction is low.
- **Load Fraction** is the average ratio of the actual load imposed on a motor to the nameplate load for the run time indicated by **hours**. In practice, load fraction is commonly in the range of 0.6 to 1.0. For initially predicting savings prior to the use of M&V, the load fraction is often assumed to be 0.80 or 0.85.
**Unit Conversion** is 1.0 where the motor power is in kW and is 0.746 where the motor power is listed in horsepower.

The energy saved by a motor replacement can be estimated by Equation 30:

\[
\text{Savings} = \left( \text{Hours}_{\text{pre}} \cdot \frac{P_{\text{nominal}}}{\text{Efficiency}_{\text{adjusted}}} \cdot \text{Load Fraction} \cdot (1 \text{ or } 0.746) \right)_{\text{pre}} \\
- \left( \text{Hours}_{\text{post}} \cdot \frac{P_{\text{nominal}}}{\text{Efficiency}_{\text{adjusted}}} \cdot \text{Load Fraction} \cdot (1 \text{ or } 0.746) \right)_{\text{post}}
\]

Where the loads and hours are consistent between the pre- and the post periods, or where the post period hours and loads are most indicative of operating parameters, Equation 30 simplifies to:

\[
\text{Savings} = (1 \text{ or } 0.746) \cdot \text{Hours} \cdot \text{Load Fraction} \cdot P_{\text{nominal}} \left[ \left( \frac{1}{\text{Efficiency}_{\text{adjusted}}} \right)_{\text{pre}} - \left( \frac{1}{\text{Efficiency}_{\text{adjusted}}} \right)_{\text{post}} \right]
\]

Where **Efficiency**\(_{\text{adjusted}}\) is nameplate or nominal efficiency; should be adjusted if measured load fraction is low. The efficiency decreases for motors less than 20 HP at load fractions below 50%.

The equations above are useful for estimating loads and savings. To measure and verify the savings, large uncertainties in the actual operating hours and load usually dictate some form of metering of these parameters.

### 3.2. Measurement Options and Measurement Error

There are several techniques for measuring the savings from replacing motors. They range from the most accurate and resource intensive, where the motor’s power draw is directly measured before and after the installation to the simplest, using motor loggers to estimate runtime. Not all of these techniques are Option B but they are each commonly practiced in evaluating motor systems. The techniques and their associated measurement error are discussed below to illustrate the statistical ramifications of several commonly used practices.

#### 3.2.1. Power Metering, Pre/Post – Stable Loads

Where actual energy consumption of the motor is logged directly during the pre- and post-installation periods uncertainty is limited to annual extrapolation, except for relatively minor meter measurement error. Energy savings can be precisely and accurately calculated; however, this technique is resource intensive compared with less rigorous options. Equation 32 shows how savings are calculated.
\[
\text{Savings (kWh)} = (\text{kWh})_{\text{pre}} - (\text{kWh})_{\text{post}}
\]

Commonly used power meters have measurement errors of less than 0.5%,\(^6\) and the standard error of the savings estimate at a 90% confidence is shown in Equation 33:

\[
\text{Standard Measurement Error}_{\text{pre}} = 0.005 \cdot \frac{\text{kWh}}{1.645} = 0.00304 \text{kWh}_{\text{pre}}
\]

Where the post-installation energy use, for example, is 90% of the pre-installation energy use (equivalent to 10% savings), the standard measurement error for the post case is very similar:

\[
\text{Standard Measurement Error}_{\text{post}} = 0.005 \cdot 0.9 \cdot \frac{\text{kWh}}{1.645} = 0.00274 \text{kWh}_{\text{pre}}
\]

Combining the errors in Equation 33 and Equation 34 yields the measurement error of the savings in Equation 35:

\[
\text{Standard Error of Savings} = \sqrt{0.00304 \text{kWh}^2 + 0.00274 \text{kWh}^2} = 0.0041 \text{kWh}_{\text{pre}}
\]

The standard error of the calculated savings based on the measurement error is therefore small – less than 0.5%. This error becomes proportionally larger where the savings are a small fraction of the baseline at kWh. This is for a case where the energy use for the pre- and post-case is directly measured. The following sections also cover situations where less rigorous measurement techniques are used.

### 3.2.2. Power Metering, Pre/Post – Variable Loads

Where the load varies across the retrofit period, for example varying production of an industrial process, or weather dependent loads, the metered pre- and post-consumption values need to be adjusted. This is necessary for them to be compared under equivalent operating conditions. Where the energy use per product or the weather dependency of energy use is known, the baseline pre-installation or post-installation energy use is adjusted using that dependency. Essentially the metered energy use is converted to what it would have been during the other period for the level of production or the weather pattern that occurred during the post-installation metering period. The choice of whether to adjust the pre- or post-metered energy use will depend on which period was more typical or representative. If the post case is more typical than the base case, energy use must be adjusted. If the pre-case is more typical than the base case, the

\(^6\) Two of the most commonly used portable power meters have accuracies well above 99%. Manufacturer accuracy claims often pertain to a single measurement. In a logging situation, typically many measurements may be taken. With repeated measurements, error may get significantly smaller.
energy use metered during the post case must be similarly adjusted. If neither period is typical, both metered
values would be adjusted to calculate the typical or representative savings.

Consider an example where an industrial production line produced 1,000 units during the metered pre-
installation period and 1,100 units during the metered post-installation period. In Equation 36, the savings
are based on the post case and the metered pre-installation energy use is adjusted as follows (assuming the
pre-installation system had the capacity to produce 1,100 units).

\[
\text{Savings (kWh)} = (\text{kWh})_{\text{pre}} + (100 \text{ units} \cdot \text{Marginal Energy Use}) - (\text{kWh})_{\text{post}}
\]

Where marginal energy use is \( \frac{\Delta \text{kWh}}{\Delta \text{units}} \).

3.2.3. Power Metering, Post-Only

In some cases, it is not practical to measure energy use before an installation. Where the load served by the
motor is relatively unchanging across the retrofit period, metering only after the installation may not
appreciably increase error. An example would be exhaust fans with constant schedules. In these cases,
Equation 32 can be modified into Equation 37 to allow for post only metering:

\[
\text{Savings} = (\text{kWh})_{\text{post}} \cdot \left( \frac{\text{Rated Efficiency}_{\text{post}}}{\text{Rated Efficiency}_{\text{pre}}} \right) - (\text{kWh})_{\text{post}}
\]

The ratio of rated efficiency is typically stable from 50% to 100% of the load for motors above 20HP. For
loads below 50% or below 75% for smaller motors, the ratio of actual efficiency will vary with the ratio of the
nominal efficiencies. The ratio of efficiency should be adjusted where metering identifies low loads. Several
references including a motor guide by DOE can help in adjusting efficiency ratios. As shown in the pre- and
post-metering case, if the meter accuracy is 99%, the combined measurement error is less than 1%. Note
that this assumes that the actual motor efficiencies are exactly equal to the nominally rated efficiencies.

Where the load varies across the retrofit period, for example varying production of an industrial process, or
weather dependent loads, the procedure for converting the metered energy use and associated savings to
typical savings is similar to the pre-post installation example above but is slightly more straightforward.

Consider the same example from above, where an industrial production line produced 1,000 units during the
pre-installation period and 1,100 units during the metered post-installation period. If the post period is
typical, then no change to the equation is necessary. If the pre-installation production of 1,000 units is more
typical, the saving must be reduced as follows in Equation 38:

---

The error induced by not metering the pre-condition can vary widely and is often unknown or unknowable.
With varying loads, the calculations are normalized by weather production or other parameters, but the ultimate accuracy of the normalization is often unknown.

### 3.2.4. Amperage Metering, Post-Only

A less rigorous method than above is to use spot metering to determine the ratio between power and amperage draws of the motor, then monitor the amperage of the motor (see Equation 39). This method is commonly used. Because not all parameters are continuously measured, strictly speaking this is really Option A. The ratio is essentially the product of the voltage and the power factor as shown in Equation 41. Because power factor can change by 10% or more even with moderate variation in motor load, equivalent to a 6\% standard error, spot measuring the ratio instead of metering actual power can increase measurement error appreciably as shown below. In addition, at loads below 50\%,\(^\text{12}\) the power factor drops sharply due to reactive magnetizing current requirements. For example, a motor with a power factor of 0.8 at 80\% load fraction can have a power factor of 0.45 at a 40\% load. Therefore, whenever practical, a power meter should be used and this amperage method should only be used where practical considerations prevent the use of a power meter, such as where one will not fit into a motor disconnect.

\[
\text{Savings} = \left(\text{(kWh)}_{\text{post}} \cdot (100 \text{ units} \cdot \text{Marginal Energy Use})\right) \cdot \left(\frac{\text{Rated Efficiency}_{\text{post}}}{\text{Rated Efficiency}_{\text{pre}}} - 1\right)
\]

\[
\text{kWh Savings} = \left(\text{Hours} \cdot \frac{\text{kW}_{\text{Spot}}}{\text{Amperage}_{\text{Spot}}}_{\text{post}}\right) \cdot \frac{\text{Efficiency}_{\text{post}}}{\text{Efficiency}_{\text{pre}}}
\]

\[
\text{kW} = \frac{\text{Volts} \cdot \text{Amps} \cdot \text{Power Factor}^{\text{12}}}{1,000}
\]

\(^8\) A power factor is the ratio of wattage to the product of volts and amps. It arises from the current and voltage wave forms being out of phase.


\(^10\) Standard error = 10%/1.645 = 6.08\%


\(^12\) For three phase power the equation is V*A*pf*1.73, where V, A, and pf are averaged across the three phases.
The measurement error of the pre-installation kWh from Equation 39 is shown in Equation 42:

\[
\text{Measurement Error} = \sqrt{\left(\frac{s(\text{amps})}{\text{amps}}\right)^2 + \left(\frac{s(\text{hours})}{\text{hours}}\right)^2 + \left(\frac{s(\text{kW/Amp})}{\text{kW/Amp}}\right)^2}
\]

Consider an example where the amperage is 15A with an error of 1%, the hours are 3,200 hours with a standard error of 32 hours, and the ratio of kW/amps is 100 with a standard error of 6%. The 6% is discussed above and arises from errors induced by spot measuring a changing power factor. The resulting error of the estimate of energy usage is then:

\[
\text{Measurement Error} = \sqrt{\left(\frac{0.01 \cdot 15}{15}\right)^2 + \left(\frac{32}{3,200}\right)^2 + \left(\frac{6}{100}\right)^2} = 0.062 \text{ kWh}_{\text{pre}}
\]

Equation 43 calculates the error in the pre-installation energy term of Equation 39. The post-installation energy term will have a similar error, if savings are 10%, then the error post period is about 90% of the pre-period or 0.055. The error of both terms of the equation are combined in Equation 44:

\[
\text{Measurement Error of Savings} = \sqrt{0.062^2 + 0.055^2} = 0.083 \text{ kWh}_{\text{pre}}
\]

This equation essentially shows that the error is dominated by the induced uncertainty by spot metering power and logging amperage. A method that would otherwise produce a measurement error of 0.4% (see equation 35) produces a measurement error of over 20x or 8% because power is not directly measured. The situation could be far worse if the load is low and varying. In these cases, the measurement error induced by amperage only metering could be 20% or more. Measurement errors on the order of 10% or higher could make detecting savings of 10% with any confidence impractical. Conversely, the error in the kW/amps is highly dependent on how much the metered motor load differed from the load when the spot measurement was taken. If the loading is very similar and unvarying, the error will be closer to the error of a simple kW reading and thus less than 1%.

3.2.5. Run Time Logging, Post Only

A less rigorous method that decreases metering cost but that also decreases accuracy and precision is to:
1. collect spot measurements of the motor loads, and;

2. log runtime using external loggers that use the electric field generated by the motor.

This approach eliminates the need to open the electrical panel a second time for meter recovery. Recovery of the motor loggers is simple, where motors are accessible; they are easily collected by non-technical staff. Most would categorize this simple approach as Option A because not all parameters are measured. It is included here for comparison purposes.

\[
\text{kWh Savings} = (\text{Hours} \cdot kW_{\text{spot},\text{post}} \cdot \frac{\text{Efficiency}_{\text{post}}}{\text{Efficiency}_{\text{pre}}}) - (\text{Hours} \cdot kW_{\text{spot},\text{post}})
\]

Errors in this method arise from the difference between the spot measurement of kW taken at the beginning of the metering period and the actual unmeasured kW from the metering period when only hours are logged. This error will be greater than that in Equation 39 because the variation in amperage is added to the variation in voltage and power factor. Variations of a load of 10% and therefore similar errors can be expected unless the operation of the motor is very stable. The combined error in each portion of Equation 46 is then:

\[
\text{Measurement Error} = \sqrt{\left(\frac{s(kW)}{kW}\right)^2 + \left(\frac{s(\text{hours})}{\text{Hours}}\right)^2}
\]

Consider an example where the power is 16 kW, and the standard error in that estimate is 1.6 kW, and the hours are 3,200 hours and the standard error in that estimate is 32 hours. The resulting error of the estimate of energy usage is then:

\[
\text{Measurement Error} = \sqrt{\left(\frac{1.6}{16}\right)^2 + \left(\frac{32}{3,200}\right)^2} = 0.10 \text{ kWh}_{\text{pre}}
\]

Combining the errors of both terms of the equation yields:

\[
\text{Measurement Error of Savings} = \sqrt{0.10^2 + 0.09^2} = 0.135 \text{ kWh}_{\text{pre}}
\]
Table 5 summarizes the four approaches described above and the associated representative measurement errors. Actual errors will vary by project, but the trend will be similar. Power metering has much lower measurement error than amp logging or runtime logging.

<table>
<thead>
<tr>
<th>Measurement Approach</th>
<th>Source of Error</th>
<th>Error Magnitude</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre/post power metering</td>
<td>Meter (minor)</td>
<td>&lt; 0.5%</td>
<td>Equation 33</td>
</tr>
<tr>
<td>Post-only power metering</td>
<td>Meter, pre/post variation(^{13})</td>
<td>&lt; 0.5%</td>
<td>Equation 34</td>
</tr>
<tr>
<td>Post-only amp metering</td>
<td>Meter, pre/post changes, power factor/voltage variation</td>
<td>8%</td>
<td>Equation 44</td>
</tr>
<tr>
<td>Post only runtime logging</td>
<td>Meter, pre/post changes, power factor, amp &amp; voltage variation</td>
<td>13%</td>
<td>Equation 48</td>
</tr>
</tbody>
</table>

Note that these errors do not account for sampling if sampling is required. They assume that the actual motor efficiencies are equal to the rated nominal efficiencies. They also do not account for modeling errors such as when short-term metering must be annualized.

3.3. Example Option B M&V Plan

3.3.1. Situation

A hospital is considering replacing the motors driving fans in their HVAC systems with new, efficient motors. The existing motors are about 20 years old, and the thinking is that rather than replacing the motors as they fail, one-at-a-time, it will save labor and energy and cost to replace them together as a set.

Following are the major points of a verification effort:

- This example shows a plan that is IPMVP Option B-adherent in that it meters all relevant parameters, in this case, total energy consumed in kWh.
- Nameplate efficiency values are used for the base and replacement motors. Where unknown or undocumented, baseline motor efficiency ratings are set to the prevailing baseline at the time. In the U.S., this can be values published in 1992 in the Energy Policy Act (EPACT). In Europe, this could be IE1 values for standard efficiency motors. One difficulty is that average market efficiencies tend to rise prior to a standard release and baseline efficiencies can exceed the minimum standards of the time.
- The length of the study period should cover anticipated variability, for example, the day of week or holiday with different operating hours. In this example, the length of the study will be 2 months prior to the project and 2 months after the project is commissioned. A two-month period helps even out any week-to-week scheduling variations.
- The study goal is to estimate motor savings at the project level with a 90/10 confidence/precision.

\(^{13}\) The measurement (meter) error of metering post only is similar to pre-post metering. There is additional but unknown error introduced by not metering the pre-installation case for all of the post only cases. That error is not measurable or fully knowable. If the post period is considered representative of true operating conditions, then the importance of this uncertainty decreases somewhat.
Table 6. HVAC Supply Fans in a Hospital

<table>
<thead>
<tr>
<th>Motor Function</th>
<th>HP</th>
<th>Estimated Runtime</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supply air fan-1</td>
<td>7.5</td>
<td>5,300</td>
<td>88%</td>
</tr>
<tr>
<td>Supply air fan-2</td>
<td>7.5</td>
<td>6,000</td>
<td>88%</td>
</tr>
<tr>
<td>Supply air fan-3</td>
<td>10</td>
<td>5,300</td>
<td>89%</td>
</tr>
<tr>
<td>Supply air fan-4</td>
<td>10</td>
<td>6,000</td>
<td>89%</td>
</tr>
<tr>
<td>Supply air fan-5</td>
<td>15</td>
<td>6,000</td>
<td>89%</td>
</tr>
<tr>
<td>Supply air fan-6</td>
<td>15</td>
<td>6,000</td>
<td>89%</td>
</tr>
<tr>
<td>Supply air fan-7</td>
<td>15</td>
<td>8,760</td>
<td>90%</td>
</tr>
<tr>
<td>Supply air fan-8</td>
<td>15</td>
<td>8,760</td>
<td>90%</td>
</tr>
<tr>
<td>Supply air fan-9</td>
<td>15</td>
<td>8,760</td>
<td>90%</td>
</tr>
<tr>
<td>Supply air fan-10</td>
<td>15</td>
<td>8,760</td>
<td>90%</td>
</tr>
<tr>
<td>Supply air fan-11</td>
<td>20</td>
<td>4,300</td>
<td>90%</td>
</tr>
<tr>
<td>Supply air fan-12</td>
<td>20</td>
<td>8,760</td>
<td>90%</td>
</tr>
<tr>
<td>Supply air fan-13</td>
<td>20</td>
<td>8,760</td>
<td>90%</td>
</tr>
<tr>
<td>Supply air fan-14</td>
<td>20</td>
<td>6,000</td>
<td>90%</td>
</tr>
<tr>
<td>Supply air fan-15</td>
<td>20</td>
<td>6,000</td>
<td>90%</td>
</tr>
<tr>
<td>Supply air fan-16</td>
<td>20</td>
<td>6,000</td>
<td>90%</td>
</tr>
<tr>
<td>Supply air fan-17</td>
<td>20</td>
<td>6,000</td>
<td>90%</td>
</tr>
<tr>
<td>Supply air fan-18</td>
<td>20</td>
<td>6,000</td>
<td>90%</td>
</tr>
<tr>
<td>Average</td>
<td>16</td>
<td>6,748</td>
<td>89.6%</td>
</tr>
<tr>
<td>St Dev</td>
<td>5</td>
<td>1,525</td>
<td>0.7%</td>
</tr>
<tr>
<td>CV</td>
<td>0.29</td>
<td>0.23</td>
<td>0.8%</td>
</tr>
</tbody>
</table>

3.3.2. Assumptions

» Operating hours are the same for both the baseline and retrofit (post) periods.
» Motor loads are unchanged by the retrofit.
» The base and retrofit efficiencies are accurately stated.
3.3.3. Sample Design

The Sample Frame is all motors involved in the retrofit (18). The sample unit is a motor.

The Sample Size is determined by the M&V plan requirement that estimated savings be reported at the 90/10 confidence/precision level for the population of motors.

Initially, a facility manager considering a motor retrofit will know nameplate horsepower and the nameplate efficiency and can estimate runtime as shown in Table 7.

Using the parameters in Table 6, the facility manager can calculate the approximate consumption of the motors using Equation 49 and an estimated load fraction of 0.85:

\[ \text{kWh} = (\text{Hours}) \cdot \frac{\text{HP}_{\text{nominal}}}{\text{Efficiency}_{\text{nominal}}} \cdot 0.85 \cdot 0.746 \]

Using the baseline and replacement efficiencies, we then calculate approximate savings using:

\[ \text{Savings} = 0.746 \cdot \text{Hours} \cdot 0.85 \cdot \text{HP}_{\text{nominal}} \left[ \left( \frac{1}{\text{Efficiency}_{\text{nominal}}}_{\text{pre}} \right) - \left( \frac{1}{\text{Efficiency}_{\text{nominal}}}_{\text{post}} \right) \right] \]

Table 7 shows the 18 supply fans with estimated energy consumption and planned savings. The savings are normalized by motor HP and in the last column by motor HP and 1,000 hours of expected runtime. It also shows the mean, standard deviation, and coefficient of variation for each parameter.

The CV is 0.35 for estimated savings, but the CV decreased to 0.18 when normalizing for motor capacity.
### Table 7. Supply Fan Saving Estimates

<table>
<thead>
<tr>
<th>Motor Function</th>
<th>HP</th>
<th>Estimated Runtime</th>
<th>Base Eff</th>
<th>Estimated base kWh</th>
<th>Retro Eff</th>
<th>Est. Savings</th>
<th>Save/HP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supply Fan - 1</td>
<td>7.5</td>
<td>5,300</td>
<td>88%</td>
<td>28,643</td>
<td>91.7%</td>
<td>1,156</td>
<td>154</td>
</tr>
<tr>
<td>Supply Fan - 2</td>
<td>7.5</td>
<td>6,000</td>
<td>88%</td>
<td>32,426</td>
<td>91.7%</td>
<td>1,308</td>
<td>174</td>
</tr>
<tr>
<td>Supply Fan - 3</td>
<td>10</td>
<td>5,300</td>
<td>89%</td>
<td>37,761</td>
<td>91.7%</td>
<td>1,112</td>
<td>141</td>
</tr>
<tr>
<td>Supply Fan - 4</td>
<td>10</td>
<td>6,000</td>
<td>89%</td>
<td>42,748</td>
<td>91.7%</td>
<td>1,259</td>
<td>126</td>
</tr>
<tr>
<td>Supply Fan - 5</td>
<td>15</td>
<td>6,000</td>
<td>89%</td>
<td>64,122</td>
<td>92.4%</td>
<td>2,359</td>
<td>157</td>
</tr>
<tr>
<td>Supply Fan - 6</td>
<td>15</td>
<td>6,000</td>
<td>89%</td>
<td>64,122</td>
<td>92.4%</td>
<td>2,359</td>
<td>157</td>
</tr>
<tr>
<td>Supply Fan - 7</td>
<td>15</td>
<td>8,760</td>
<td>90%</td>
<td>92,579</td>
<td>92.4%</td>
<td>2,405</td>
<td>160</td>
</tr>
<tr>
<td>Supply Fan - 8</td>
<td>15</td>
<td>8,760</td>
<td>90%</td>
<td>92,579</td>
<td>92.4%</td>
<td>2,405</td>
<td>160</td>
</tr>
<tr>
<td>Supply Fan - 9</td>
<td>15</td>
<td>8,760</td>
<td>90%</td>
<td>92,579</td>
<td>92.4%</td>
<td>2,405</td>
<td>160</td>
</tr>
<tr>
<td>Supply Fan - 10</td>
<td>15</td>
<td>8,760</td>
<td>90%</td>
<td>92,579</td>
<td>92.4%</td>
<td>2,405</td>
<td>160</td>
</tr>
<tr>
<td>Supply Fan - 11</td>
<td>20</td>
<td>4,300</td>
<td>90%</td>
<td>60,592</td>
<td>93.0%</td>
<td>1,955</td>
<td>98</td>
</tr>
<tr>
<td>Supply Fan - 12</td>
<td>20</td>
<td>8,760</td>
<td>90%</td>
<td>123,438</td>
<td>93.0%</td>
<td>3,982</td>
<td>199</td>
</tr>
<tr>
<td>Supply Fan - 13</td>
<td>20</td>
<td>8,760</td>
<td>90%</td>
<td>123,438</td>
<td>93.0%</td>
<td>3,982</td>
<td>199</td>
</tr>
<tr>
<td>Supply Fan - 14</td>
<td>20</td>
<td>6,000</td>
<td>90%</td>
<td>84,547</td>
<td>93.0%</td>
<td>2,727</td>
<td>136</td>
</tr>
<tr>
<td>Supply Fan - 15</td>
<td>20</td>
<td>6,000</td>
<td>90%</td>
<td>84,547</td>
<td>93.0%</td>
<td>2,727</td>
<td>136</td>
</tr>
<tr>
<td>Supply Fan - 16</td>
<td>20</td>
<td>6,000</td>
<td>90%</td>
<td>84,547</td>
<td>93.0%</td>
<td>2,727</td>
<td>136</td>
</tr>
<tr>
<td>Supply Fan - 17</td>
<td>20</td>
<td>6,000</td>
<td>90%</td>
<td>84,547</td>
<td>93.0%</td>
<td>2,727</td>
<td>136</td>
</tr>
<tr>
<td>Supply Fan - 18</td>
<td>20</td>
<td>6,000</td>
<td>90%</td>
<td>84,547</td>
<td>93.0%</td>
<td>2,727</td>
<td>136</td>
</tr>
<tr>
<td>Average</td>
<td>16</td>
<td>6,748</td>
<td>1</td>
<td>76,130</td>
<td>1</td>
<td>2,374</td>
<td>150</td>
</tr>
<tr>
<td>St Dev</td>
<td>5</td>
<td>1,525</td>
<td>0</td>
<td>27,859</td>
<td>0</td>
<td>820</td>
<td>26</td>
</tr>
<tr>
<td>CV</td>
<td>0.29</td>
<td>0.23</td>
<td>0.8%</td>
<td>0.37</td>
<td>0.6%</td>
<td><strong>0.35</strong></td>
<td>0.18</td>
</tr>
</tbody>
</table>
Using the data in Table 7 and the CV for the savings normalized to motor capacity, we can then calculate the sample size for deriving an estimate of the savings that have a 90% probability of being within 10 percent of the true population mean using the following equation.

\[
\text{Sample Size (n)} = \frac{Z^2 \cdot CV^2}{e^2} = \frac{1.645^2 \cdot 0.18^2}{0.1^2} \approx 9
\]

This sample size needs to be corrected for the small population of 18 motors. By using the following equation, the sample size is corrected for a finite population:

\[
\text{Sample Size (n)} = \frac{n_0 \cdot N}{N + n_0} = \frac{18 \cdot 9}{18 + 9} \approx 6.2
\]

Where \(n_0\) is the calculated sample for infinite population, and \(N\) is the population size. We will round up the corrected sample size to 7 to achieve our desired precision.

Table 8 shows the metered pre and post consumption and calculated and normalized savings for the sample of 7 motors. The average savings is 140 kWh/motor HP; close to but about 7% lower than the estimated 150 kWh/motor HP in Table 7.
### Table 8. Metered Pre and Post Supply Fan Consumption for 7 Sampled Motors

<table>
<thead>
<tr>
<th>Motor Function</th>
<th>HP</th>
<th>Estimated Runtime</th>
<th>Base Eff</th>
<th>Retro Eff</th>
<th>Metered Pre</th>
<th>Metered Pre/HP</th>
<th>Metered Post</th>
<th>Metered Savings</th>
<th>Metered Savings/ HP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supply Fan - 1</td>
<td>7.5</td>
<td>5,300</td>
<td>87.5%</td>
<td>91.7%</td>
<td>27,771</td>
<td>3,703</td>
<td>26,499</td>
<td>1,272</td>
<td>170</td>
</tr>
<tr>
<td>Supply Fan - 2</td>
<td>7.5</td>
<td>6,000</td>
<td>87.5%</td>
<td>91.7%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supply Fan - 3</td>
<td>10</td>
<td>5,300</td>
<td>89.0%</td>
<td>91.7%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supply Fan - 4</td>
<td>10</td>
<td>6,000</td>
<td>89.0%</td>
<td>91.7%</td>
<td>42,731</td>
<td>4,273</td>
<td>41,473</td>
<td>1,258</td>
<td>126</td>
</tr>
<tr>
<td>Supply Fan - 5</td>
<td>15</td>
<td>6,000</td>
<td>89.0%</td>
<td>92.4%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supply Fan - 6</td>
<td>15</td>
<td>6,000</td>
<td>89.0%</td>
<td>92.4%</td>
<td>64,379</td>
<td>4,292</td>
<td>62,010</td>
<td>2,369</td>
<td>158</td>
</tr>
<tr>
<td>Supply Fan - 7</td>
<td>15</td>
<td>8,760</td>
<td>90.0%</td>
<td>92.4%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supply Fan - 8</td>
<td>15</td>
<td>8,760</td>
<td>90.0%</td>
<td>92.4%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supply Fan - 9</td>
<td>15</td>
<td>8,760</td>
<td>90.0%</td>
<td>92.4%</td>
<td>92,856</td>
<td>6,190</td>
<td>90,444</td>
<td>2,412</td>
<td>161</td>
</tr>
<tr>
<td>Supply Fan - 10</td>
<td>15</td>
<td>8,760</td>
<td>90.0%</td>
<td>92.4%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supply Fan - 11</td>
<td>20</td>
<td>4,300</td>
<td>90.0%</td>
<td>93.0%</td>
<td>60,418</td>
<td>3,021</td>
<td>58,469</td>
<td>1,949</td>
<td>97</td>
</tr>
<tr>
<td>Supply Fan - 12</td>
<td>20</td>
<td>8,760</td>
<td>90.0%</td>
<td>93.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supply Fan - 13</td>
<td>20</td>
<td>8,760</td>
<td>90.0%</td>
<td>93.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supply Fan - 14</td>
<td>20</td>
<td>6,000</td>
<td>90.0%</td>
<td>93.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supply Fan - 15</td>
<td>20</td>
<td>6,000</td>
<td>90.0%</td>
<td>93.0%</td>
<td>84,527</td>
<td>4,226</td>
<td>81,800</td>
<td>2,727</td>
<td>136</td>
</tr>
<tr>
<td>Supply Fan - 16</td>
<td>20</td>
<td>6,000</td>
<td>90.0%</td>
<td>93.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supply Fan - 17</td>
<td>20</td>
<td>6,000</td>
<td>90.0%</td>
<td>93.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supply Fan - 18</td>
<td>20</td>
<td>6,000</td>
<td>90.0%</td>
<td>93.0%</td>
<td>82,841</td>
<td>4,142</td>
<td>80,169</td>
<td>2,672</td>
<td>134</td>
</tr>
<tr>
<td>Count</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>285</td>
<td>121,460</td>
<td>16</td>
<td>17</td>
<td>455,523</td>
<td>29,848</td>
<td>440,864</td>
<td>14,659</td>
<td>982</td>
</tr>
<tr>
<td>Average</td>
<td>15.8</td>
<td>6,748</td>
<td>89.5%</td>
<td>92.5%</td>
<td>65,075</td>
<td>4,264</td>
<td>62,818</td>
<td>2,049</td>
<td>140.2</td>
</tr>
<tr>
<td>St Dev</td>
<td>5</td>
<td>1,525</td>
<td>0.8%</td>
<td>0.5%</td>
<td>23,730</td>
<td>965</td>
<td>23,161</td>
<td>620</td>
<td>24.8</td>
</tr>
<tr>
<td>CV</td>
<td>0.29</td>
<td>0.23</td>
<td>0.9%</td>
<td>0.6%</td>
<td>0.36</td>
<td>0.23</td>
<td>0.37</td>
<td>0.30</td>
<td>0.18</td>
</tr>
<tr>
<td>Std Error</td>
<td>8,969</td>
<td>365</td>
<td>8,754</td>
<td>234</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t (90%)</td>
<td>1.94</td>
<td>1.94</td>
<td>1.94</td>
<td>1.94</td>
<td>1.94</td>
<td>1.94</td>
<td>1.94</td>
<td>1.94</td>
<td></td>
</tr>
<tr>
<td>Upper bound</td>
<td>79,095</td>
<td>4,834</td>
<td>76,664</td>
<td>2,460</td>
<td>154.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower bound</td>
<td>51,055</td>
<td>3,694</td>
<td>49,297</td>
<td>1,728</td>
<td>125.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
We can then calculate the upper and lower bounds of the savings estimate as follows, including the t-statistic for 6 degrees of freedom and a finite population correction:

\[
\text{Confidence Interval} = \sqrt{\frac{(N-n)}{N-1}} \cdot t \frac{S}{\sqrt{n}} = \sqrt{\frac{(18-7)}{18-1}} \cdot 1.94 \frac{24.8}{\sqrt{7}} = 14.66 \text{ kWh; RP 10.5%}
\]

The confidence interval and the relative precision, which is 126 to 155 or 89.5% to 110.5% of the 140 kWh/motor HP average, just outside our desired 10% relative precision. In this case increasing the sample size to 8 would increase the precision to our goal of 10%.

3.4. Combining Sampling and Measurement Errors

The confidence interval in this example is based solely on sampling error. In the above calculation, the standard error is:

\[
\text{Standard Error} = \frac{S}{\sqrt{n}} = \frac{24.8}{\sqrt{7}} = 9.37
\]

Typically, the measurement error is ignored, and the sampling error is used to derive confidence intervals. If we combine the sampling error with the previously derived measurement error, the total error is increased, as is the span of the confidence interval. In this example, the power draw of the motors was metered before and after the replacement. The standard measurement error previously derived was 0.004 of the measurement (Equation 35) or using the normalized pre-installation average measurement of 4,264 from Table 8, 0.004 \cdot 4,264 = 11.4. In Equation 55, the sampling error from the example is combined with the measurement error.

\[
\text{Combined Measurement and Sampling Error} = \sqrt{\frac{24.8^2}{7} + \frac{11.4^2}{7}} = 10.31
\]

In this case, the measurement error is low for each measurement but moderate in considering the savings margin. The combined error increases by about 10%. As shown in Equation 56 and the relative precision grows to 11.5%. This can be brought back down to 10% by increasing the sample size. Using the same example and increasing the sample size to 8 decreases the finite population correction and the t-statistic and brings the relative precision down to about 10%.
Confidence Interval = \[ \sqrt{\frac{(N - n)}{N - 1}} \cdot t \cdot \frac{S}{\sqrt{n}} = \sqrt{\frac{(18 - 7)}{18 - 1}} \cdot 1.94 \cdot 10.31 = 16.1 \text{ kW}h; 11.5\% \text{ RP} \]

The situation is quite different in the case where we employ runtime loggers after the installation. As shown in Table 5, the error is 13% of the consumption measurement, or \[ 0.13 \cdot 4,264 = 554 \]. In Equation 57, the sampling error from the example is combined with the measurement error.

Equation 57

\[ \text{Combined Measurement and Sampling Error} = \sqrt{\frac{24.8^2}{7} + \frac{554^2}{7}} = 209.6 \]

In this case, by taking the measurement error into account, the standard error and the width of the confidence interval increases by about 20x. The relative precision increases to over 234% as shown in Equation 58. This essentially illustrates that where savings are small, here they are on the order of 3%, a low accuracy measurement technique will not detect savings.

Equation 58

\[ \text{Confidence Interval} = \sqrt{\frac{(N - n)}{N - 1}} \cdot t \cdot \frac{S}{\sqrt{n}} = \sqrt{\frac{(18 - 7)}{18 - 1}} \cdot 1.94 \cdot 209.6 = 327; 234\% \text{ RP} \]

It is important to note that in these calculations, the errors are assumed to be random with no bias. If, for example, the spot measurements collected prior to installing runtime loggers are not representative of the logging period but rather were collected when the motors were off, the measurements will be biased low, and if this is not discovered or corrected, the measurement errors will be even larger than those calculated in this example.
4. Option C: Whole Facility Regression-Based Savings Estimate

**Option C**

Whole Facility Savings are determined by measuring energy use at the whole facility level.

Option C can be an accurate and cost-effective M&V approach when the energy savings from an upgrade are expected to lead to a significant reduction in total facility energy consumption (typically a minimum of 5-10%). Option C works best when the mathematical relationship between the independent variables explains a large share of the observed variation in consumption. The viability of Option C comes down to the relationship between the “signal” (energy savings) and the “noise” (unexplained variation in consumption).

4.1. Calculating the Precision of a Regression-Based Savings Estimate

The uncertainty (or precision) of a regression-based savings estimate is a standard metric for quantifying the “noise” in the data and is the focus of this section.

4.1.1. Forecast Model

A forecast model of consumption involves creating a statistical model of energy consumption using only data from the baseline period. This mathematical relationship is then projected onto the observed conditions in the reporting period to estimate the adjusted baseline – or what energy consumption would have been in the reporting period absent the changes implemented. Often analysts will want to know up front what the uncertainty will be to determine if Option C is likely to be precise enough for a given project. For example, suppose we predict the uncertainty to be ±50,000 kWh/year and the savings projections for the ECM is 10,000 kWh/year. In this case, the Option C Whole Facility approach would be a questionable choice because a result that states the project saved somewhere between -40,000 kWh and 60,000 kWh is not very useful. An Option A or Option B approach focusing on the affected end-use would be a more viable direction for the project.

Consider the hypothetical establishment shown in Figure 2. This is an intentionally simple example which uses outdoor air temperature without change points. Each point represents a monthly billing period. The values on the y-axis are total monthly kWh consumption, and the values on the x-axis are average outdoor air temperature. The points in the baseline and reporting period are marked by different symbols. Notice that in both the baseline and reporting period, electric consumption is positively correlated with temperature. There is also a general trend of the facility using less energy in the reporting period.
Figure 3 shows the output for a linear regression model constructed using only the observations in the baseline period. The model intercept is 13,609.08 and the model coefficient (or slope) for average outdoor air temperature is 845.6 kWh per degree (°F). The model $R^2$ is quite high and the t-statistic of the slope coefficient is highly significant (23.41). The t-statistic of the model intercept is dependent on how far the intercept is from zero and should not be used when assessing model goodness of fit.
Using the regression coefficients from Figure 3 and the actual temperature values from the reporting period, we can predict what electric consumption would have been in the reporting period absent the improvement. Figure 4 illustrates the calculation. The blue “X” symbols are predicted values for the reporting period using the mathematical relationship between kWh and air temperature determined during the baseline period. For each period, the savings are calculated by comparing the adjusted baseline prediction to the actual consumption reading for the reporting period. The bars at the bottom of the chart represent the kWh savings estimated for the period, by month. Total savings in the reporting period is 54,841 kWh.

That was a simple example where each billing period included the same number of days. Usually, billing periods differ in lengths. If billing periods consist of varying number of days, it is helpful to convert consumption to a daily average so that the number of days can be taken into account. When this is done, a weighted regression must be used: The average daily consumption for a billing period with 35 days must carry more weight in the regression than a billing period with 25 days. If a weighted regression is not used, the model will be biased: after multiplying each period’s prediction of average daily consumption by the number of days in the period, the total consumption predicted will not match the actual total consumption. ASHRAE Guideline 14 Annex B describes an approach for handling billing periods of varying lengths.
Figure 4: Savings Calculation Example

Using the ASHRAE method to estimate uncertainty proceeds as follows:

1. Calculate the mean energy use per period in the baseline period. Call this value $kWh_{base}$.

2. Fit a regression model to the baseline period data. Store the root mean squared error ($RMSE$) – this is a measure of how far predicted consumption typically falls from actually observed consumption.

3. Calculate absolute uncertainty using

$$\text{Absolute Uncertainty} = (1.26) \cdot \left(t_{1 - \frac{\alpha}{2}, n - k}\right) \cdot (RMSE) \cdot \left[1 + \frac{2}{n}\right]^{0.5}$$

In Equation 59, 1.26 is an ASHRAE correction factor\(^{14}\) that approximates some complex matrix algebra, $t_{1 - \frac{\alpha}{2}, n - k}$ is the critical value associated with the desired level of confidence and $n - k$ degrees of freedom, $k$ is the number of parameters being estimated $k$ (= number of independent variables + 1), $n$ is the number

of time periods in the baseline period (12 in this example), and \( m \) is the number of time periods in the reporting period (12 in this example). See table above for \( t \) values.

Using the regression output from Figure 3, the uncertainty for the hypothetical facility at the 90% confidence level is:

\[
\text{Absolute Uncertainty} = (1.26) \cdot (1.81) \cdot (690.78) \cdot \left[ \left(1 + \frac{2}{12}\right) \cdot 12 \right]^{0.5}
\]

\[
\text{Absolute Uncertainty} = \pm 5,895 \text{ kWh}
\]

The relative uncertainty (or fractional uncertainty) in the savings estimate would then be calculated by comparing the absolute uncertainty to the sum of the savings.

\[
\text{Relative Uncertainty} = \frac{5,895}{54,845} = 10.75\%
\]

Advanced practitioners should note that the constant factor 1.26 can be replaced using slightly more complex calculations that still do not require matrix algebra. One approach uses a calculated value “\( Y \)” using the number of measurements in the reporting period (\( m \)) assuming the data are monthly or daily (and thus not significantly autocorrelated). Equations for \( Y \) are shown below.

\[
Y = (-0.00042 \cdot m^2) + (0.03306 \cdot m) + 0.94054
\]

\[
Y = (-0.00024 \cdot m^2) + (0.03535 \cdot m) + 1.00286
\]

Another approach uses an exact algebraic derivation from the matrix algebra. Users should see the footnoted references for further information.

4.1.2 Use of an Indicator Variable for the Reporting Period

A second approach to estimating energy savings from measurements taken by the utility revenue meter is a “pre-post” model that uses data from both the baseline and reporting period, along with an indicator variable, or series of indicator variables, for the reporting period. Consider a data set with 12 months of

---


18 The equations 60 and 61 « Polynomial Correction factor for Monthly Data / for Daily Data » suppose several constraints/limitations in their application. EVO is reviewing the assumptions taken to obtain these formulae. Both formulations, as well as the Correction Factor mentioned in ASHRAE 14 2014 (1.26) in equation 4-8, are primarily intended to avoid the Matrix calculation usually considered as cumbersome. For multivariate regression, since most spreadsheet programs are capable of performing matrix algebra, it is however recommended to calculate the prediction uncertainty using the Matrix calculation formula in ASHRAE 14-2004 B-26.
baseline period consumption data, 12 months of reporting period consumption data, and any relevant explanatory information that could be used as independent variables in the regression specification. The reporting period indicator variable referred to herein as “post,” would be coded as a 0 in the baseline period and as a 1 in the reporting period.

\[ kWH_i = \beta_0 + \beta_1 (X_i) + (Post_i) \cdot [\beta_2 + \beta_3 (X_i)] + \epsilon_i \]

In Equation 60, \( kWH_i \) is metered electric consumption during time period \( i \), \( Post \) is the reporting period indicator variable which equals 1 if time period \( i \) falls in the reporting period, \( X_i \) is the value of the explanatory variable during time period \( i \), and \( \epsilon_i \) is the error term. Note that this model contains an interaction between the post indicator variable and the \( X \) explanatory variable. Also, note that the model could be expanded to include multiple explanatory variables.

When this approach is used, the coefficient of the post-term (\( \beta_2 \)) represents the change in the intercept between the baseline and reporting periods, and the coefficient of the interaction term (\( \beta_3 \)) represents the change in the relationship between \( kWh \) and \( X \) in the reporting period.

Consider the same hypothetical facility explored in Section 4.1.1 where monthly \( kWh \) consumption is the dependent variable, and average outdoor air temperature is the independent variable. Then the interaction (post*air temperature) is computed as a third explanatory variable. Figure 5 shows regression output.

With this approach, the annual savings estimate for the reporting period would be calculated using the ‘post’ coefficient (-837.77 kWh), the “temp_x_post” coefficient (-296.2 kWh per degree), and the average outdoor air temperature from the reporting period (12.601). The calculations are shown below.

\[
\text{Reporting Period Savings} = -12 \cdot \left[ \beta_2 + (\beta_3 \cdot \text{temp}) \right]
\]

\[
\text{Reporting Period Savings} = -12 \cdot [-837.77 + (-296.22 \times 12.601)]
\]

\[
\text{Reporting Period Savings} = 54,845 \text{ kWh}
\]
Figure 5. Regression Output for Post Indicator Variable Example

If normalized savings are the desired outcome, the average level of the explanatory variable in the reporting period could be replaced by a long-run estimate. For this example, where the weather is the independent variable, weather-normalization would be accomplished by replacing the observed outdoor air temperatures with Typical Meteorological Year (TMY) data.

Including interactions between the “post” indicator variable and explanatory variables is an excellent way to develop normalized impact estimates for more complex model specifications. Unfortunately, the estimation of precision is not straightforward because savings are derived via multiple coefficients. The covariance between coefficients must be accounted for when calculating the standard error for a linear or non-linear combination of coefficients. Statistical software packages like SAS, Stata, or R have commands that make this calculation simple to perform, which creates useful options when dealing with daily or hourly data (discussed in Section 4.2.3.4). However, performing the calculation in Excel is a more complex exercise in linear algebra than most analysts will wish to undertake.

There is a wide range of explanatory variables that could be included in a model depending on the nature of the facility and the frequency of the consumption data. For an industrial plant, analysts may find that weather data have little or no explanatory power, but a metric of production (widgets produced, tons of raw material processed, etc.) will be highly significant. If the consumption data being analyzed is daily or hourly, often, indicator variables for the day of the week and hour of the day (or weekday vs. weekend), will allow the model to fit the time-dependent characteristics of the load pattern. Section 4.2 explores special considerations required to produce accurate precision estimates when using daily or hourly data.

4.1.3. Special Considerations for Interval Meter Data

If each record in the dataset describes a different month, then the approaches described above are acceptable for calculating uncertainty. However, if the data are daily or hourly (instead of monthly), extra measures need to be taken. While more granular data allow analysts to fit high-frequency models of energy consumption, such data are also susceptible to autocorrelation. Autocorrelation means that the present value of the series is correlated with past values of the same series. If autocorrelation is present, several of the statistics from the ordinary least squares (OLS) regression output will be biased, and the uncertainty estimates will be incorrect. An important note is that the OLS regression coefficients themselves remain unbiased in the presence of autocorrelation, so the savings estimate derived from OLS does not need any autocorrelation adjustments.

4.2. Managing Autocorrelation

4.2.1. What is Autocorrelation?

Autocorrelation (or serial correlation) refers to the relationship between the past and the present values of a time series. As an example, let $Y_t$ represent metered electric load at time $t$. If $Y_t$ is correlated with $Y_{t-1}$ (or $Y_{t-2}$, or $Y_{t-3}$, etc.), then the series is said to be autocorrelated. That is if the metered load at time $t$ can be estimated by the metered load at time $t-1$ (or any other time), then autocorrelation is present.
Commonly, autocorrelation is discussed in the context of the residuals of a fitted regression model. One key assumption of the OLS approach to regression is that the error terms (or residuals) are independent. When the present value of the dependent variable is correlated with past values, then the amount of error (residual) in period \( t \) is also correlated with the amount of error in period \( t - 1 \). If this is the case, then the key OLS assumption regarding the independence of the error terms is violated. In the presence of autocorrelation, regression coefficient estimates remain unbiased, but their standard errors may be biased. Thus, any uncertainty inferences drawn may be incorrect.

4.2.2. Diagnosing Autocorrelation

This section will provide the reader with three ways to detect autocorrelation:

1. Plot residuals against time
2. Examine an Autocorrelation Function (ACF) plot
3. Perform the Durbin-Watson hypothesis test

To illustrate these methods, consider the simplest pre-post regression model of hourly consumption (kWh) with a reporting period indicator variable that equals 1 in the reporting period. The data to be used in this illustration come from a retail facility from January 2013 to December 2014. Regression output for the simple OLS model is shown in Figure 11. As indicated, by the low \( R^2 \) value, this model lacks explanatory power.

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 17,520</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1032763.48</td>
<td>1</td>
<td>1032763.48</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Residual</td>
<td>69048500.6</td>
<td>17,519</td>
<td>3941.57442</td>
<td>R-squared = 0.0147</td>
</tr>
<tr>
<td>Total</td>
<td>70081264.08</td>
<td>17,519</td>
<td>4000.30048</td>
<td>Adj R-squared = 0.0147</td>
</tr>
</tbody>
</table>

|           | Coef.   | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|-----------|---------|-----------|-------|-----|---------------------|
| post      | -16.35548 | .9486321  | -16.19 | 0.000 | -17.21489 -13.49607 |
| _cons     | 15.7045 | .6707842  | 23.27 | 0.000 | 14.3896 -17.0193    |

Figure 6. Simple Linear Regression Output

One commonly used approach in detecting autocorrelation is a plot of residuals (or standardized residuals\(^{19}\)) against time. Figure 7 shows standardized residuals plotted against time. In the figure, positive residuals tend to follow other positive residuals, negative residuals tend to follow other negative residuals, and there are alternating positive/negative spikes every 12\(^{th}\) hour – all indicative of autocorrelation. The plot suggests that key independent variables (weather) and/or temporal features of the data are not being modeled appropriately.

\(^{19}\) Standardized residuals are calculated as the individual residual divided by the standard deviation of the model residuals.
Figure 7. Standardized Residuals Plotted Against Time

For comparison, Figure 8 shows what a residual plot should look like – no patterns, no cycles, and random spikes rather than spikes every 12th or 24th hour.

Figure 8. Independently Distributed Error

A second plot can be used to diagnose autocorrelation is an ACF plot of residuals. Figure 9 shows an ACF plot of the residuals from the regression model shown in Figure 6.
The large positive spike at Lag = 1 indicates that the residual at time $t$ is positively correlated with the residual at time $t - 1$ (e.g., the first lag). The strength of this correlation decreases as the number of lags approaches 12 but starts increasing again as the number of lags approaches 24 with a peak at Lag = 24. As before, this indicates that the residual at time $t$ is highly correlated with the residual at time $t - 24$. It will be useful to contrast Figure 9 with an ACF plot of the random error series shown in Figure 8. To make the differences between the two ACF plots stand out, Figure 10 is plotted on the same scale as Figure 9.
In comparing the two ACF plots, note the difference in (1) the magnitude of the spikes at each lag (high correlation in Figure 9, hardly any correlation in Figure 10) and (2) the number of spikes that extend beyond the 95% confidence band. In the ACF plot for the residuals, nearly every spike extends beyond the 95% confidence band – this indicates that the residuals do not represent a random error (as OLS assumptions state they should be). In the random error ACF plot, two or three spikes extend beyond the 95% confidence band – this is by chance alone.

Rather than relying on the eyeball test (or perhaps in conjunction with the eyeball test), analysts can perform the Durbin-Watson test to determine if autocorrelation is present. The null hypothesis for the Durbin-Watson test is that the error terms do not exhibit autocorrelation, and the alternative hypothesis is that the residuals do exhibit autocorrelation.

\[ d = \frac{\sum_{t=2}^{T} (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{T} \varepsilon_t^2} \]

In Equation 61, \( \varepsilon_t \) represents the OLS regression residual at time \( t \) and \( T \) represents the total number of time periods. (So, with two years of hourly data, \( T = 17,520 \)). If the test statistic is below the lower critical value or above the upper critical value, then the null hypothesis should be rejected. Tables with these critical values can be found in some Statistics textbooks, but analysts will have better luck searching for such a table online.\(^2\)

\(^2\) Note that the Durbin-Watson test tests for first order serial correlation, meaning the residual at time \( t \) is correlated with the residual at time \( t - 1 \).

\(^2\) For example: https://www3.nd.edu/~wevans1/econ30331/Durbin_Watson_tables.pdf.
4.2.3. Remedial Measures

The presence of autocorrelation does not mean that the model is incapable of estimating savings, but it will not be possible to estimate the uncertainty in the savings estimate using the methods presented in the introduction. If autocorrelation is detected, and uncertainty estimates are required, there are a few ways to proceed:

1. Change the model by (a) adding or deleting variables or (b) trying a functional form other than linear regression. Alternative estimation approaches such as generalized least squares (GLS) regression and feasible generalized least squares (FGLS) regression could also be implemented.

2. Keep the OLS regression estimates and make adjustments to accommodate the serial correlation (estimate robust standard errors, estimate how many of the $n$ observations are truly independent, etc.).

These options may also be approached in tandem. Regarding option 1, adding variables may resolve issues related to omitted variable bias. Including time-related indicator variables (e.g., indicators for each hour of the day, separate regression models for each hour of the day, indicators for each day of the week, an indicator for weekend days, etc.) might combat this. If available, variables describing production schedules or shifts might also help.

Adjusting the model for omitted variables may significantly improve the reporting of the model without fully resolving the serial correlation issue. At this point, it might be easier to make autocorrelation adjustments rather than continue building models in search of an autocorrelation free model. When a forecast model is used to estimate savings and uncertainty, the ASHRAE ‘Adjusted N’ correction is a useful approach to remove autocorrelation related bias from estimates of uncertainty (Section 4.2.3.3). When savings and uncertainty are calculated via a regression model that includes both the baseline and reporting period and the appropriate indicator variable and their interactions with other explanatory variables, Newey-West standard errors can be used to correct the artificially low standard errors introduced by autocorrelation (Section 4.2.3.4).

Note that a one-size-fits-all solution to this problem may not exist, but to ignore autocorrelation entirely would be a grave mistake.

4.2.3.1. Omitted Variable Bias

Conceptually, omitted variable bias is what its name suggests: biased results caused by the omission of an important independent variable. The effect an omitted independent variable has on the dependent variable gets absorbed by the error term, so trends in a residual or ACF plot might indicate that an important variable is being left out.

Often, information regarding exactly what is being left out is hard to come by. However, when the data have a temporal aspect (e.g., time series data), this temporal structure can be modeled directly through time-related indicator variables. Including indicator variables for each hour of the day, each day of the week, and weekends could reduce or remove autocorrelation issues while also substantially improving the model fit.
Consider the simple pre-post regression model considered in Section 4.2.2 and reproduced in Figure 11. In this model, hourly consumption (kWh) is modeled with a reporting period indicator variable that equals 1 in the reporting period.

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs</th>
<th>F(1, 17518)</th>
<th>Prob &gt; F</th>
<th>R-squared</th>
<th>Adj R-squared</th>
<th>Root MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1032763.48</td>
<td>1</td>
<td>1032763.48</td>
<td>17,520</td>
<td>262.02</td>
<td>0.0000</td>
<td>0.0147</td>
<td>0.0147</td>
<td>62.782</td>
</tr>
<tr>
<td>Residual</td>
<td>69048500.6</td>
<td>17,518</td>
<td>3941.57442</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>70081264.1</td>
<td>17,519</td>
<td>4000.30048</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 11. Simple Linear Regression Output**

The regression output shown in Figure 12 illustrates how the model changes when time-related indicator variables are introduced. Note that the savings estimate hardly changed, but its standard error decreased. Also, note that a majority of the hourly indicator variables are highly significant, and the daily indicator variables are statistically significant as well. Finally, statistics describing model fit ($R^2$, $RMSE$) imply this model is far superior to the original regression model. The addition of categorical indicator variables for the hour of day and day of week have significantly improved the explanatory power of the model. That said, model diagnostics (Durbin-Watson test, residual plots) suggest that the model still suffers from autocorrelation, so other autocorrelation corrections are necessary.
Figure 12. Regression with Time-Related Indicator Variables

4.2.3.2. Misspecification

Misspecification can occur in a myriad of ways. One example is when a linear model is used to estimate the relationship between two variables that are not truly linearly related. As an example, consider the plot below.
Variables $X$ and $Y$ are clearly associated with each other, but the association does not appear to be linear through the full range of the data. One can see how implementing a linear model, in this case, would result in positive serial correlation – positive residuals will tend to follow other positives, and negative residuals will tend to follow other negatives. Autocorrelation issues aside, implementing a linear model, in this case, is decidedly wrong, as one of the OLS requirements is that $X$ and $Y$ are linearly related. A change point model that estimates separate linear relationships for different ranges of the $X$ variable is one possible solution. A quadratic model might also provide a better fit to this data.

4.2.3.3. The ASHRAE Correction

The ASHRAE correction for autocorrelation entails estimating the number of independent observations in a total of $n$ observations, denoted $n'$.

$$\text{Equation 62}$$

$$n' = n \cdot \frac{1 - \rho}{1 + \rho}$$

In Equation 62, $\rho$ ("rho") is the autocorrelation coefficient of the residuals. To calculate $\rho$, model the residuals of the OLS model ($\epsilon_t$) as a function of their first lag. The estimate of $\beta_1$ from Equation 63 is the autocorrelation coefficient $\rho$.\textsuperscript{22}

$$\text{Equation 63}$$

$$\epsilon_t = \beta_0 + \beta_1(\epsilon_{t-1}) + \epsilon_t$$

\textsuperscript{22} One could also calculate the correlation coefficient between the residuals and their first lag, or look at a lag plot.
Calculating the autocorrelation coefficient in Excel or any statistical programming software is straightforward. Autocorrelation-corrected estimates of uncertainty can be derived from a forecast model by replacing \(n\) with \(n'\) in the calculation of RMSE, then implementing Equation 59 with the new RMSE. If analysts are using a statistical package that allows for the easy combination of regression coefficients and their standard errors, autocorrelation can be accounted for using Newey-West standard errors.

The ASHRAE correction only corrects for lag1 residuals. Recent research indicates that the actual savings uncertainty exceeds even the corrected estimates. This is likely due to the presence of lag2, lag3, etc. autocorrelation which is not mitigated by the correction factor.

4.2.3.4. Newey-West Standard Errors

In one of the most commonly cited economics papers in the last few decades, Whitney Newey and Kenneth West devised a method that slightly adjusts the OLS calculations to deal with autocorrelation. Note that the Newey-West method does not attempt to adjust the regression coefficients themselves.

In using the Newey-West method, the user must define up to how many lags autocorrelation could persist, denoted \(L\). The Newey-West calculations assume autocorrelation beyond \(L\) can be ignored, so it is advisable to define \(L\) such that it is larger than the cycle of the data. For example, if using hourly electric consumption data, the cycle could be 24, 48 or even 168 hours. Thus, the user should define \(L\) such that \(L \geq 24\) and perhaps try different multiples of 24.

Implementing the Newey-West correction in Excel might prove difficult, as the math involves some matrix algebra and there currently is no Excel package built to handle it. Analysts with access to SAS, R, or Stata can reference user guides on implementing the Newey-West correction.

For illustration, consider a pre-post model that uses three months of hourly summer data in the baseline period and three months of hourly summer data in the reporting period. The form of the pre-post model to be estimated is shown in Equation 64 and model output is shown in Figure 14.

\[
W_{hi} = \beta_0 + \beta_1 (Post_i) + \beta_2 (Temp_i) + \beta_3 (Post_i \cdot Temp_i) + \epsilon_i
\]

23 https://ideas.repec.org/top/top.item.nbcites.html
In this case, the estimate of annual savings is produced via a linear combination of two of the regression coefficients:

\[
\text{Annual Savings} = -8760 \cdot \left[ \beta_2 + (\beta_3 \cdot \text{Average temp in Reporting Period}) \right]
\]

\[
\text{Annual Savings} = -8760 \cdot [-3.68 + (-1.554 \cdot 20.356)]
\]

\[
\text{Annual Savings} = 307,344 \text{ kWh}
\]

But what is the standard error of this estimate? As noted in Section 4.1.2, the estimation of uncertainty is not straightforward when savings are derived from a combination of multiple coefficients. However, common statistical programs have packages that can derive the standard error for a linear (or non-linear) combination of coefficients. Using one such package, an estimate for annual uncertainty (not corrected for autocorrelation) is ±21,439 kWh at the 90% confidence level.

Applying the Newey-West procedure with \( L = 72 \), the standard error of the post-term becomes 21.57 (as opposed to 8.48) and the standard error of the interaction term becomes 1.02 (as opposed to 0.414). With these autocorrelation-corrected standard errors, the uncertainty at 90% confidence increases from 21,439 kWh to 54,417 kWh – nearly three times larger.
5. Option D: New Construction of a Medium Office

Option D involves the use of computer simulation software to predict facility energy use and savings. For options A, B, and C, standard errors and associated uncertainty have what is referred in mathematics as a “closed form” solution. An equation is said to be a closed form if it solves a given problem in terms of functions and mathematical operations. In cases where such functions are not available, the solution can be obtained through simulation modeling.

Option D involves the use of a site-specific simulation model to estimate the savings for a single facility. Option D is traditionally applied to new construction projects and to comprehensive retrofits where there is a desire to account for the interaction of measures. Note that simulations can also be used to estimate savings from a typical measure in a prototypical facility, but that application does not fall under the traditional Option D approach and is not addressed in this Chapter.

Consider the use of a building simulation model to estimate the savings from a retrofit in a building. The model will account for a long list of site-specific factors such as the hours of operation, local weather, envelope construction, HVAC system types, cooling and heating efficiencies, lighting loads, etc. The impacts of the efficiency measures are estimated by changing the values of parameters within the model. The uncertainty in the savings estimate generated by the model is due to a number of factors:

- How accurately the model reflects the operation of the baseline building.
  - For example, if the lighting hours of operation in a building are longer than simulated, the impact of a lighting retrofit will be greater than estimated by the model.
- How accurately the modeling program simulates the impact of each parameter.
  - For example, does the model accurately calculate economizer operation and thus times when the lighting heat is deliberately exhausted from the building?
- How accurately the modeled parameters (such as lighting w/sf) match the actual parameter in both the baseline and post-installation buildings.
  - For example, is the actual change in lighting density greater than modeled because of a change in the specification of the installed fixtures?

The first type of uncertainty is largely addressed through a process of calibrating the model. This chapter does not address the calibration of simulation models. Calibration has traditionally been a key technique by which M&V professionals improved the accuracy of Option D results. Statistical tests can be applied to determine how well the model matches the historical consumption of a facility. The analysis draws on many of the same tests described in the section on regression models. Note, however, that assessing the quality of the calibration does not directly estimate the uncertainty in the resulting savings estimate.

---

24 Watts per square foot.
Uncertainty in the quality of the simulation programs is typically ignored. The major simulation programs are generally accepted to provide realistic calculations. The second error can be significant, however, based on how the modeler chooses to build an individual model. A modeling program may be capable of accurately simulating economizer operation, but the modeling professional may not realize that the economizers in this building are not functioning properly.

In this section, we present three different approaches to conducting simulations to estimate the uncertainty and precision of estimated savings:

- Local One-step-at-a-time (OAT) sampling analysis
- Bootstrap based analysis

5.1. Situation

This example involves the new construction of a medium office building in San Francisco, California. Total electric and natural gas use and savings are to be analyzed for energy efficiency measures that reduce the following parameters in the baseline building by a certain percentage:

- Lighting Power Density (LPD)
- Electric Equipment Power Density (EPD)

5.2. Building Modeling and Analysis Tools

For this example, the OpenStudio (OS) toolchain (free and open source software) will be used. Building Energy Models (BEM) created in the OpenStudio format (.osm) can easily be manipulated by OpenStudio Measures. These model changes can be as simple as changing an existing parameter in the model (e.g., change chiller Coefficient of Performance – COP) or as complicated as changing the entire HVAC system. This concept of creating and manipulating energy models through OpenStudio Measures will allow us to turn building characteristics into variables and determine their sensitivity and savings in a way that is consistent, scalable and easily shared with co-workers, clients or the public as well as allowing for other algorithmic processes such as model calibration and optimization. The approach can be applied with other modeling tools but requires more effort.

The Medium Office from the DOE Commercial Prototype Building Models will serve as our example building, and the energy model will be programmatically generated using the OpenStudio Prototype Buildings Measure. This has the advantage that the reader can easily reproduce the workflow used to create the model and results shown in this section. The sensitivity analyses presented utilize the OpenStudio Parametric Analysis Tool (PAT) for problem definition, and the cloud computing capabilities of OpenStudio Server for

---

25 https://www.OpenStudio.net
26 http://nrel.github.io/OpenStudio-user-documentation/getting_started/about_measures/
27 https://www.energycodes.gov/commercial-prototype-building-models
28 https://github.com/NREL/openstudio-standards
29 http://nrel.github.io/OpenStudio-user-documentation/reference/parametric_analysis_tool_2
30 https://github.com/NREL/OpenStudio-server
algorithm implementation and simulation runs. Both tools are free to use. However, the actual cloud computing time from Amazon must be purchased from a user-specific Amazon cloud computing account (Amazon EC2).

5.3. Energy Model and Variable Characterization

The energy model for this example is created by using the following inputs in the DOE Prototype Building Measure:

- **Building Type:** Medium Office
- **Template:** 90.1-2010
- **Climate Zone:** ASHRAE 169-2006-3C

Where “template” determines the vintage and code requirements for the building. This results in a building that is typical of new construction, a medium office building in San Francisco, CA.

To understand the effect that these parameters have on total building electricity and natural gas usage as well as understanding the sensitivity and uncertainty in the resulting energy savings for these measures, we first need to define valid ranges of the input uncertainty (a minimum and maximum) and a possible distribution type (triangular, uniform, normal, etc.) for each input parameter (LPD and EPD). These input uncertainties can come from many sources such as installation and implementation variations as well as other factors beyond our control. This means that even though we intend to reduce both LPD and EPD by exactly 10% in the building, from a practical standpoint, we will not see a reduction of exactly 10% but more of a distribution of reductions around that 10% target reduction.

To simulate this effect, we need to try and quantify the variation in the input that we will expect to have. For this example, we will consider a 5% variation in the 10% reduction that we wish to apply to the building. This gives us the minimum and maximum values for our distribution and bounds the input parameters at 5% and 15% of the baseline model values of LPD and EPD. We also need to choose a distribution shape to finish quantifying our input uncertainty. We could choose a flat or uniform change from 5% to 15%, a normal or bell curve distribution or one of several others that we think identifies the uncertainty we have about our ability to implement the energy efficiency measures. For this example, we will make the choice of using a triangle distribution for the uncertainty that peaks to its maximum at a 10% reduction and tails off at the bounds as depicted in Figure 15 and Figure 16 with the ranges listed in Table 8. This implies there is a higher probability of achieving the 10% reduction and a decreasing probability to achieve reduction values towards the bounds of 5% and 15%.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Minimum</th>
<th>Mode</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPD (% reduction from baseline)</td>
<td>5 %</td>
<td>10 %</td>
<td>15 %</td>
</tr>
<tr>
<td>EPD (% reduction from baseline)</td>
<td>5 %</td>
<td>10 %</td>
<td>15 %</td>
</tr>
</tbody>
</table>
Figure 15. Lighting Power Density Distribution

Figure 16. Electric Equipment Power Density Distribution

Note: The results of metering studies can be used to determine the range of parameters. Consider the use of a model to estimate the HVAC interaction of a lighting retrofit. A metering study such as is described in the Option A Chapter could be used to determine the likely range of hours of use.
Ranges and distribution types for each variable are typically defined by project constraints or best practices. It should be noted that while the plotted distributions in Figure 15 and Figure 16 do not appear to be triangles, that is just an artifact of plotting them using the histogram function in R.

Next, we explore a method to get an understanding of how the uncertainty in the inputs LPD and EPD translates to uncertainty and precision of estimated energy use and savings.

### 5.4. Local One-step-at-a-time (OAT) Sampling Analysis

We will sample the two variables, one at a time while holding the other constant (at their default values) resulting in an OAT sampling method. This method is also described as a local method since we are holding all the non-sampled variables to their default values. Thus, while we are getting some sense of the effect of the perturbed variable, this perturbation is only with respect to the same locally fixed values for the other variables.

Choosing the “correct” number of samples to use is very problem dependent. For example, a smooth linear problem would require very few samples to understand the variation in the output, while a highly nonlinear problem would require “lots” of points to characterize the nonlinear behavior of the problem. It is typical under these conditions to do a grid refinement study where a small number of points are chosen and used, and then twice as many points are used, and then twice again, this process repeating until the variation or characterization of the outputs becomes “good enough” for us to have confidence that we have chosen the “right” number of points. For illustration purposes, the number of samples in this example for each variable was chosen to be 100. This resulted in a total of 200 simulations to run (100 points x 2 variables = 200 total runs).

Histograms of Total Electric Use and Savings variation due to the variation in the variables are depicted in Figure 17 and Figure 18. In Figure 15, the vertical line at 421,977 kWh is the baseline model Total Electric Use. The green histogram characterizes the Total Electric Use after the change in LPD is applied to the baseline model while the blue represents the change due to the EPD measure. The histograms are plotted as density functions so that they are normalized, with the area under each curve is one. Thus they can be treated as probability distributions. The sensitivity of each variable on Total Electric Use can be determined by comparing the width of each colored distribution to each other. The larger the width of the resulting distribution means a larger sensitivity of that output to the input variable. In our example, the 10% reduction in LPD results in the peak of the green histogram with total electric use around 410,000 kWh with the variation in the results stemming from the variation in the inputs as previously described. The 15% LPD reduction results in the far left of the green distribution which is around 400,000 kWh. Thus, for a 5% additional reduction in LPD, Total Electric Use is reduced by another 10,000 kWh. From Figure 17 we see that EPD results in greater Electric Savings while the uncertainty in that savings is much larger than the LPD case since the width of the distribution is much larger.

The same results can also be seen in Figure 18 by looking at histograms of the energy savings directly. Here, it is easier to see that the energy savings is more sensitive to changes in EPD than in LPD because of the wider distribution. Also, EPD can result in larger energy savings. It should be noted that this analysis does not consider interactions or a combined effect of the two proposed Energy Savings Measures. This is an inherent limitation of OAT sampling.
Figure 17. Total Electric Use (kWh) Variation

Figure 18. Total Electric Savings (kWh) Variation
5.5. Bootstrap Based Analysis

In this section, we attempt to compute the precision of the estimated Total Electric Use and Savings due to both LPD and EPD combined. To create the data necessary, we sample all the Input Variables at the same time. For illustrative purposes, we will use 400 as the sample size. Sampling is conducted using OpenStudio PAT.

The result of sampling 400 values of LPE and EPD changes is 400 values of electric savings. At this point, the savings values are sorted from low to high. The 90% confidence interval is defined by the top and bottom 5 percentiles (the 21st – 381st values in this case). The resulting distribution of savings from simulating changes in both LPD and EPD measures are plotted in Figure 19.

![Figure 19. Range of Electric Savings](image)

The average savings is computed for the range along with the precision (range from 5th to 95th percentile points divided by 2). Average energy savings is estimated at 37,516 kWh and relative precision at ±22%.

One can imagine repeating this analysis for a much larger set of building characteristics. In most buildings, there will be uncertainty in operating schedules, HVAC equipment efficiencies, baseline lighting density, plug load density, etc. Some of these parameters are not changes by the energy efficiency measures but have a significant impact on the savings estimates. The analysis will show the expected range of savings for the expected uncertainty in these inputs.

R code to generate the bootstrap results:

```r
results <- read.csv("ca_2010_lhs_all_5_to_15.csv")
lpd <- results$reducelightingloadsbypercentage.lighting_power_reduction_percent
epd <- results$reduceelecequipmentloadsbypercentage.elecequip_power_reduction_percent
elec <- results$CompareBaseline.total_electric
elec_savings <- results$CompareBaseline.total_electric_savings
```
```r
trim_elec <- sort(elec)[21:381]
trim_elec_savings <- sort(elec_savings)[21:381]
png(paste("trim_elec_savings.png",sep=""), width=8, height=8.0, units="in", pointsize=10, res=200)
hist(trim_elec_savings, breaks=10, freq=F, col=rgb(0,0,1,0.5), main="Total Electric Savings (kWh) from both LPD and EPD reductions", xlab="Total Electric Savings (kWh)")
lines(density(trim_elec_savings),col=rgb(0,1,0,1),lwd=2)
dev.off()
#density.default(trim_elec_savings)
mean_trim <- mean(trim_elec_savings)
distance <- abs(mean_trim - trim_elec_savings[1])
relative_precision <- distance/mean_trim
```