Methodologies for Estimating Building Energy Savings Uncertainty: Review and Comparison

Juan-Carlos Baltazar PhD, PE, Yifu Sun EIT, and David Claridge PhD, P.E.

ENERGY SYSTEMS LABORATORY
TEXAS A&M ENGINEERING EXPERIMENT STATION
• Background of building energy savings and savings uncertainty
• Statistics Method and Its Simplification
• Nearest Neighborhood Method
• Bayesian Analysis
• Conclusion
Building energy consumption is illustrated by using time series plot and scattered plot.
Baseline model is developed using the selected period data. Baseline model can be developed by using linear regression technics.

Typical models are summarized below.

<table>
<thead>
<tr>
<th>Mean Model</th>
<th>Simple Linear Regression</th>
<th>3P-CP</th>
<th>4P-CP</th>
<th>5P-CP</th>
</tr>
</thead>
</table>

Proceedings of the 14th International Conference for Enhanced Building Operations, Beijing, China, September 14-17, 2014
Baseline model is employed to estimate the energy use for post retrofit period had no retrofit occurred.

Savings and percentage savings are expressed as

\[
E_{save,m} = \hat{E}_{base,m} - E_{meas,m}
\]

\[
\% E_{save,m} = \frac{\hat{E}_{base,m} - E_{meas,m}}{\hat{E}_{base,m}}
\]

Next question is how to evaluate the risk of the project and the savings uncertainty.
Savings uncertainty represents the absolute reliability of the estimated savings. Statistical matrix expression

\[ \Delta E_{\text{save},m} = t \cdot RMSE \cdot [1' (T'_{\text{post}} (T'_{\text{pre}} T_{\text{pre}})^{-1} T'_{\text{post}} + I) 1]^{0.5} \]

Fractional savings uncertainty represents the relative uncertainty of savings.

\[ \frac{\Delta E_{\text{save},m}}{E_{\text{save},m}} = t \frac{CV(RMSE)}{F \cdot m} [1' (T'_{\text{post}} (T'_{\text{pre}} T_{\text{pre}})^{-1} T'_{\text{post}} + I) 1]^{0.5} \]
**CV(RMSE)** is the coefficient of variation of the root mean square error

\[
CV(RMSE) = \frac{1}{\bar{y}} \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n - p}}
\]

- **T** is the matrix of regressor variables, *pre* and *post* denote the pre-retrofit (baseline period) and post-retrofit.
- **T'** denotes the transpose of **T**.
- **I** is an identity matrix.
Reddy and Claridge derived simplified equation from the matrix expression under two assumptions:

- The temperature distribution for post-retrofit period is the same to the temperature distribution for baseline period.
- The off-diagonal elements in the matrix \((T_{post} (T'_{pre} T_{pre})^{-1} T'_{post} + I)\) are zero.

\[
\frac{\Delta E_{\text{save},m}}{E_{\text{save},m}} \approx 1.26 \cdot t \frac{\text{CVRMSE}}{F \cdot m^{0.5}} [1 + \frac{2}{n}]^{0.5}
\]

\[
\frac{\Delta E'_{\text{save},m}}{E_{\text{save},m}} \approx 1.26t' \frac{\text{CVRMSE}'}{F \cdot m^{0.5}} \left[ \frac{n}{n'} (1 + \frac{2}{n'}) \right]^{0.5}
\]
Sun and Baltazar improved the simplified equation.

- **Monthly Data**

\[
\frac{\Delta E_{\text{save},m}}{E_{\text{save},m}} \approx Y \cdot t \frac{\text{CVRMSE}}{F \cdot m^{0.5}} \left[1 + \frac{2}{n}\right]^{0.5}
\]

\[
Y = -0.00022X^2 + 0.03306X + 0.94054
\]

- **Daily Data**

\[
\frac{\Delta E_{\text{save},m}}{E_{\text{save},m}} \approx Y' \frac{\text{CVRMSE}'}{F \cdot m^{0.5}} \left[\frac{n}{n'} \left(1 + \frac{2}{n'}\right)\right]^{0.5}
\]

\[
Y' = -0.00024X^2 + 0.03535X + 1.00286
\]
The difference between using constant factor simplified equation and matrix equation.

<table>
<thead>
<tr>
<th>Month No.</th>
<th>Building 1</th>
<th>Building 1</th>
<th>Building 1</th>
<th>Building 2</th>
<th>Building 2</th>
<th>Building 2</th>
<th>Building 3</th>
<th>Building 3</th>
<th>Building 4</th>
<th>Building 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ELE</td>
<td>CHW</td>
<td>HHW</td>
<td>ELE</td>
<td>CHW</td>
<td>HHW</td>
<td>ELE</td>
<td>NG</td>
<td>ELE</td>
<td>NG</td>
</tr>
<tr>
<td>3</td>
<td>22%</td>
<td>11%</td>
<td>12%</td>
<td>22%</td>
<td>13%</td>
<td>15%</td>
<td>12%</td>
<td>10%</td>
<td>16%</td>
<td>16%</td>
</tr>
<tr>
<td>6</td>
<td>6%</td>
<td>3%</td>
<td>2%</td>
<td>3%</td>
<td>7%</td>
<td>6%</td>
<td>2%</td>
<td>7%</td>
<td>3%</td>
<td>3%</td>
</tr>
<tr>
<td>9</td>
<td>2%</td>
<td>7%</td>
<td>6%</td>
<td>7%</td>
<td>11%</td>
<td>10%</td>
<td>2%</td>
<td>3%</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>12</td>
<td>5%</td>
<td>14%</td>
<td>14%</td>
<td>5%</td>
<td>11%</td>
<td>10%</td>
<td>4%</td>
<td>4%</td>
<td>4%</td>
<td>4%</td>
</tr>
<tr>
<td>15</td>
<td>10%</td>
<td>19%</td>
<td>18%</td>
<td>10%</td>
<td>16%</td>
<td>15%</td>
<td>12%</td>
<td>12%</td>
<td>10%</td>
<td>10%</td>
</tr>
<tr>
<td>18</td>
<td>14%</td>
<td>23%</td>
<td>22%</td>
<td>17%</td>
<td>22%</td>
<td>21%</td>
<td>17%</td>
<td>15%</td>
<td>17%</td>
<td>17%</td>
</tr>
<tr>
<td>21</td>
<td>18%</td>
<td>27%</td>
<td>26%</td>
<td>20%</td>
<td>25%</td>
<td>24%</td>
<td>18%</td>
<td>18%</td>
<td>19%</td>
<td>19%</td>
</tr>
<tr>
<td>24</td>
<td>22%</td>
<td>31%</td>
<td>30%</td>
<td>21%</td>
<td>27%</td>
<td>26%</td>
<td>22%</td>
<td>22%</td>
<td>22%</td>
<td>22%</td>
</tr>
<tr>
<td>27</td>
<td>25%</td>
<td>33%</td>
<td>33%</td>
<td>25%</td>
<td>30%</td>
<td>29%</td>
<td>26%</td>
<td>25%</td>
<td>25%</td>
<td>25%</td>
</tr>
<tr>
<td>30</td>
<td>27%</td>
<td>35%</td>
<td>35%</td>
<td>28%</td>
<td>33%</td>
<td>32%</td>
<td>28%</td>
<td>28%</td>
<td>28%</td>
<td>28%</td>
</tr>
</tbody>
</table>

Proceedings of the 14th International Conference for Enhanced Building Operations, Beijing, China, September 14-17, 2014
The difference between using improved simplified equation and matrix equation.

<table>
<thead>
<tr>
<th>Month No.</th>
<th>Building 1 ELE</th>
<th>Building 1 CHW</th>
<th>Building 1 HHW</th>
<th>Building 2 ELE</th>
<th>Building 2 CHW</th>
<th>Building 2 HHW</th>
<th>Building 3 ELE</th>
<th>Building 3 NG</th>
<th>Building 4 ELE</th>
<th>Building 4 NG</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0%</td>
<td>4%</td>
<td>4%</td>
<td>0%</td>
<td>0%</td>
<td>1%</td>
<td>8%</td>
<td>10%</td>
<td>4%</td>
<td>4%</td>
</tr>
<tr>
<td>6</td>
<td>5%</td>
<td>1%</td>
<td>1%</td>
<td>13%</td>
<td>11%</td>
<td>10%</td>
<td>9%</td>
<td>4%</td>
<td>13%</td>
<td>13%</td>
</tr>
<tr>
<td>9</td>
<td>1%</td>
<td>3%</td>
<td>3%</td>
<td>10%</td>
<td>8%</td>
<td>7%</td>
<td>1%</td>
<td>0%</td>
<td>4%</td>
<td>4%</td>
</tr>
<tr>
<td>12</td>
<td>2%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>0%</td>
<td>0%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>15</td>
<td>1%</td>
<td>2%</td>
<td>2%</td>
<td>0%</td>
<td>1%</td>
<td>0%</td>
<td>3%</td>
<td>3%</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>18</td>
<td>0%</td>
<td>2%</td>
<td>2%</td>
<td>4%</td>
<td>3%</td>
<td>2%</td>
<td>3%</td>
<td>1%</td>
<td>3%</td>
<td>3%</td>
</tr>
<tr>
<td>21</td>
<td>0%</td>
<td>2%</td>
<td>2%</td>
<td>3%</td>
<td>2%</td>
<td>1%</td>
<td>0%</td>
<td>0%</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>24</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>0%</td>
<td>0%</td>
<td>1%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>27</td>
<td>0%</td>
<td>1%</td>
<td>2%</td>
<td>0%</td>
<td>0%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>30</td>
<td>0%</td>
<td>2%</td>
<td>2%</td>
<td>1%</td>
<td>1%</td>
<td>0%</td>
<td>1%</td>
<td>0%</td>
<td>1%</td>
<td>1%</td>
</tr>
</tbody>
</table>
Lei (2009) and Subbarao et al. (2011) developed the nearest neighborhood method.

For a prediction at a specific temperature, find out an arbitrary given number of closest points to this temperature.

Calculate the model residuals for the all these points.

For a given confidence level, rule out the upper and lower percentiles according to the residuals.

Average the maximum and minimum residuals of the rest points.
Find out an 20 closest points to $60^\circ$F.
Calculate the residuals and sort them from minimum to maximum.

Uncertainty = (|5% percentile| + |95% percentile|) / 2 for 90% Confidence
Compare the uncertainty boundaries for statistics method and neighborhood method.

![Graph comparing energy use estimation and uncertainty boundaries for statistics method and nearest neighborhood method.](graph)

- Baseline
- 90% Confidence Level (Nearest Neighborhood Method)
- 90% Confidence Level (Statistics Method)
- Baseline Period Data

**Outside Dry-bulb Temperature (°F)**

**Daily Energy Use Consumption**

Proceedings of the 14th International Conference for Enhanced Building Operations, Beijing, China, September 14-17, 2014
Comparison of savings uncertainty by using different methods for daily energy use data.
The number of closest points N is 20, the boundary of Nearest Neighborhood method will be more closer to the boundary of statistics method is the N increases.

Nearest Neighborhood method gives a better uncertainty estimation for local (short time) savings.

A large data set is necessary for this method.
Shonder and Im (2012) reported their energy savings and savings uncertainty calculation methodology based on Bayesian inference and Bayesian regression technology.

Bayes inference

\[ p(\theta \mid y) = \frac{p(\theta)p(y \mid \theta)}{p(y)} \quad p(y) = \int p(\theta)p(y \mid \theta)d\theta \]

The integration is for \( \theta \), so \( p(y) \) is a constant.

Bayes inference can also be stated as

\[ p(\theta \mid y) \propto p(\theta)p(y \mid \theta) \]
The matrix expression for linear functions

\[ y = X\beta + \lambda \]

\(\beta\) is the vector of the parameter

\(\lambda\) is a random variable and it is drawn from a normal distribution

\[ \lambda \sim N(0, \sigma^2) \]

Thus Bayesian interference is

\[ p(\beta, \sigma^2 \mid y) \propto p(\beta) p(\sigma^2) p(y \mid \beta, \sigma^2) \]
The prior distribution of all the parameters is assumed as a normal distribution, the mean and variance are zero and $10^8$.

The likelihood is also assumed as a normal distribution, and the function with $y$ given $\beta$, $\sigma^2$ and $X$ is

$$p(y \mid \beta, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma \sqrt{2\pi}} \exp \left[ -\frac{1}{2\sigma^2} (y_i - x_i \beta)^2 \right]$$

It is need to integrate the posterior distribution to characterize the distributions of the parameters.

Usually, the distribution is integrated numerically.
• The two examples results from in Shonder and Im’s paper.
  – monthly data case
  – daily data case

Confidence level for both cases is 95%
Monthly data estimation results are the same for using two methods, because free of correlation for monthly data.

Daily data estimation result using Bayesian method is more accurate, due to with considering the correlation both for savings estimation and savings uncertainty.

Traditional statistics method doesn’t consider the correlation for savings estimation. For uncertainty calculating, it includes a correlation coefficient, which makes the uncertainty bigger.
Matrix equation for savings uncertainty is derived from the statistics fundamental concepts.

Simplified equation is a approximation from the matrix equation, and is the easiest to use.

Improved simplified equation gave a closer result to matrix equation than the simplified equation.

Nearest neighborhood method uses the straightforward concept of percentile. It can be employed to large quantity of data (hourly or daily data), but not applicable to small data group (monthly data).
Bayesian method is more completely consideration, but it is hard to apply. It needs numerical integration and costs longer time and bigger memory.

The savings using traditional statistics method is in the 95% confidence interval of using Bayesian method.

For monthly data, it is not need to use Bayesian method, because it gives the same result as traditional statistics method.

For daily data and highly accuracy requirement, Bayesian analysis is the optimal method.
Thanks !!

JUAN-CARLOS BALTARZ, PhD, PE, BEMP