ESL-IC-14-09-11a

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

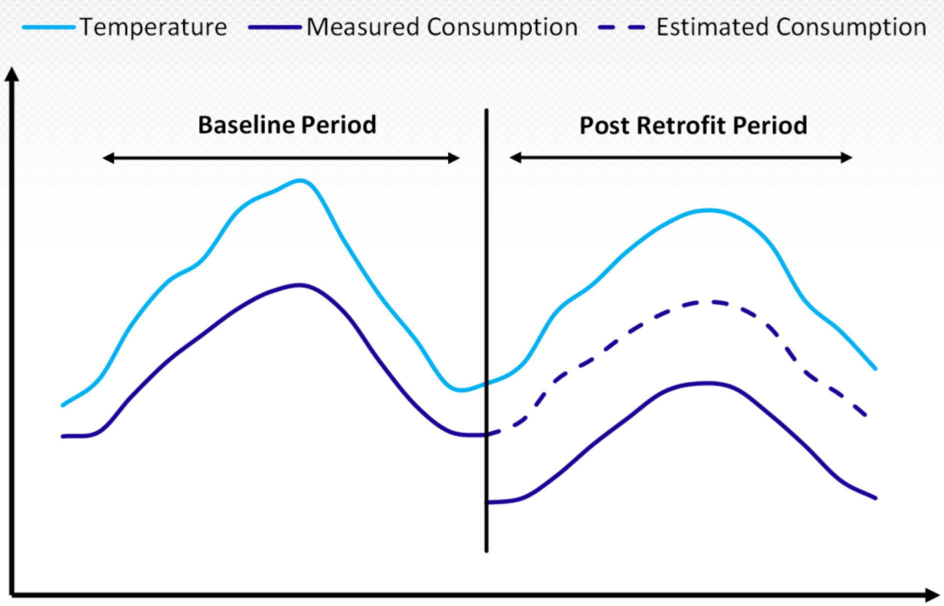
International Conference for Enhanced Building Operations
Tsinghua University – Beijing, China –September 14 -17, 2014

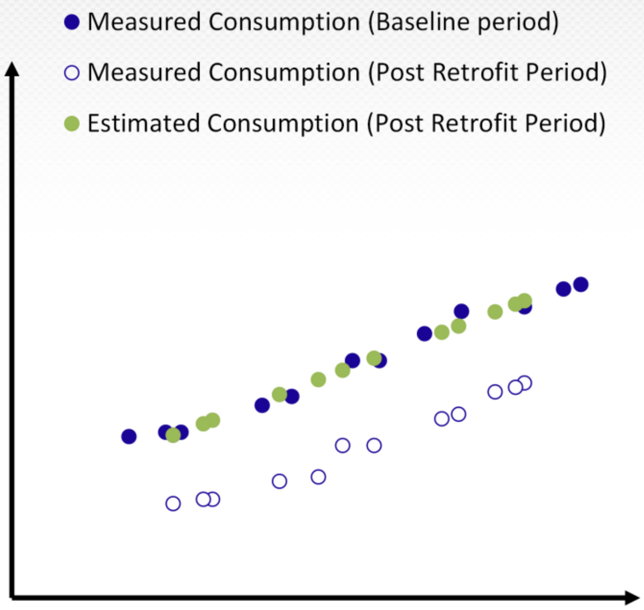
OUTLINE

- Background of building energy savings and savings uncertainty
- Statistics Method and Its Simplification
- Nearest Neighborhood Method
- Bayesian Analysis
- Conclusion

BACKGROUND – BUILDING ENERGY

Building energy consumption is illustrated by using time series plot and scattered plot.



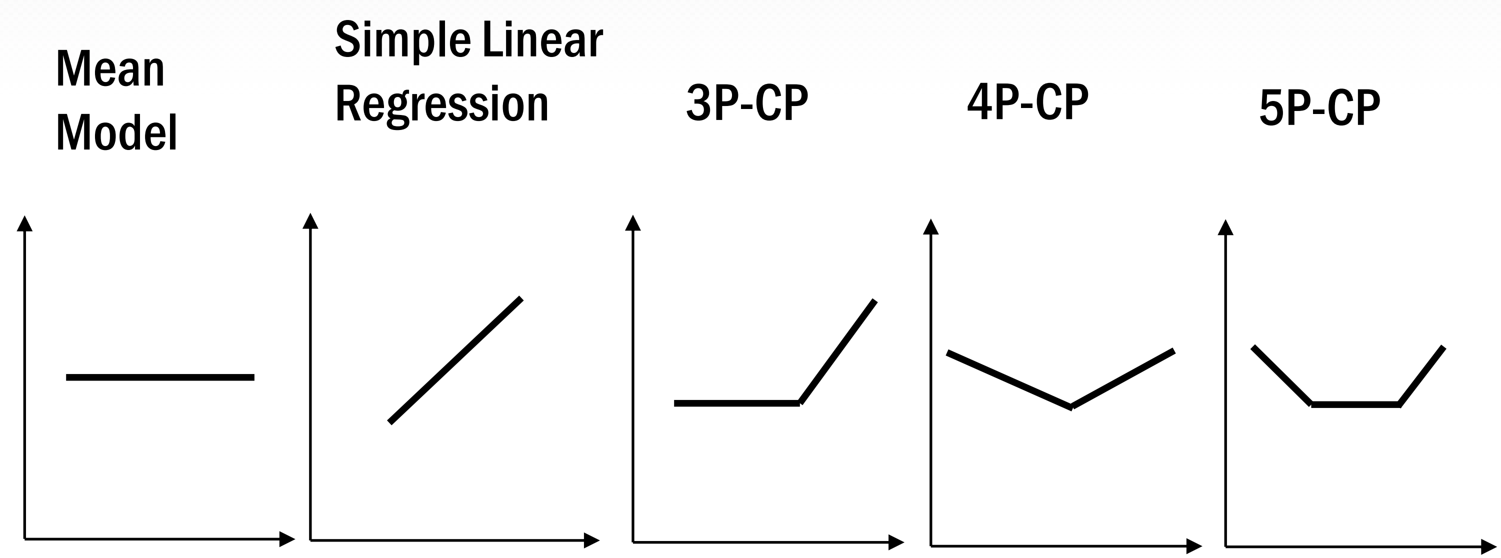


BACKGROUND – BESELINE MODEL

Baseline model is developed using the selected period data.

Baseline model can be developed by using linear regression technics.

Typical models are summarized below.



BACKGROUND – ENERGY SAVINGS

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.

STATISTICS METHOD – METRIX EQUATION

ESL-IC-14-09-11a

Savings uncertainty represents the absolute reliability of the estimated savings. Statistical matrix expression

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

Fractional savings uncertainty represents the relative uncertainty of savings.

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

STATISTICS METHOD – MATRIX EQUATION

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**

STATISTICS METHOD – SIMPLIFICATION

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.**

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

$$\frac{\Delta E_{save,m}}{E_{save,m}} \approx 1.26t' \frac{CVRMSE'}{F \cdot m^{0.5}}[\frac{n}{n'}(1+\frac{2}{n'})]^{0.5}$$

STATISTICS METHOD – SIMPLIFICATION

Sun and Baltazar improved the simplified equation.

- **Monthly Data**

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

$$Y = -0.00022X^2 + 0.03306X + 0.94054$$

- **Daily Data**

$$\frac{\Delta E_{save,m}}{E_{save,m}} \approx Y \cdot t' \frac{CVRMSE'}{F \cdot m^{0.5}}[\frac{n}{n'}(1+\frac{2}{n'})]^{0.5}$$

$$Y = -0.00024X^2 + 0.03535X + 1.00286$$

Proceedings of the 14th International Conference for Enhanced Building Operations, Beijing, China, September 14-17, 2014

STATISTICS METHOD – SIMPLIFICATION

The difference between using constant factor simplified equation and matrix equation.

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

STATISTICS METHOD – SIMPLIFICATION

The difference between using improved simplified equation and matrix equation.

Month No.	Building 1 ELE	Building 1 CHW	Building 1 HHW	Building 2 ELE	Building 2 CHW	Building 2 HHW	Building 3 ELE	Building 3 NG	Building 4 ELE	Building 4 NG
3	0%	4%	4%	0%	0%	1%	8%	10%	4%	4%
6	5%	1%	1%	13%	11%	10%	9%	4%	13%	13%
9	1%	3%	3%	10%	8%	7%	1%	0%	4%	4%
12	2%	1%	1%	1%	1%	0%	0%	1%	1%	1%
15	1%	2%	2%	0%	1%	0%	3%	3%	1%	1%
18	0%	2%	2%	4%	3%	2%	3%	1%	3%	3%
21	0%	2%	2%	3%	2%	1%	0%	0%	1%	1%
24	1%	1%	1%	0%	0%	1%	0%	0%	0%	0%
27	0%	1%	2%	0%	0%	1%	1%	1%	0%	0%
30	0%	2%	2%	1%	1%	0%	1%	0%	1%	1%

NEAREST NEIGHBORHOOD METHOD

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.

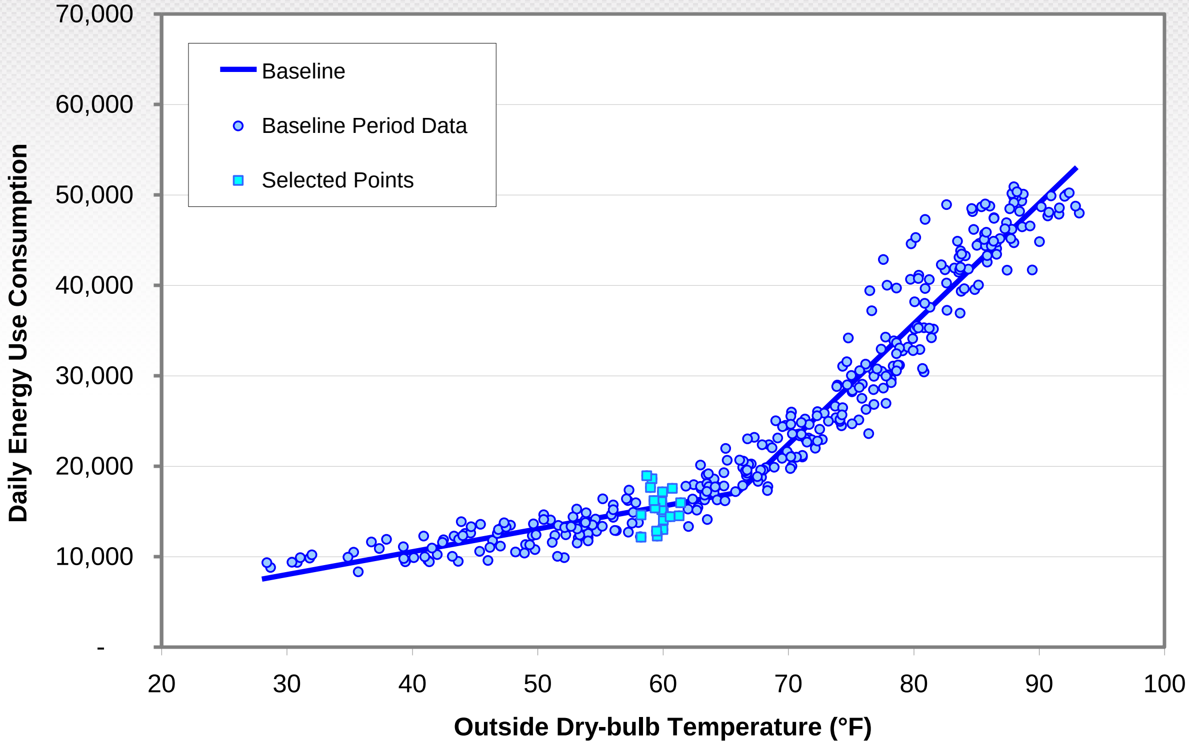
Calculate the model residuals for the all these points.

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.

NEAREST NEIGHBORHOOD METHOD

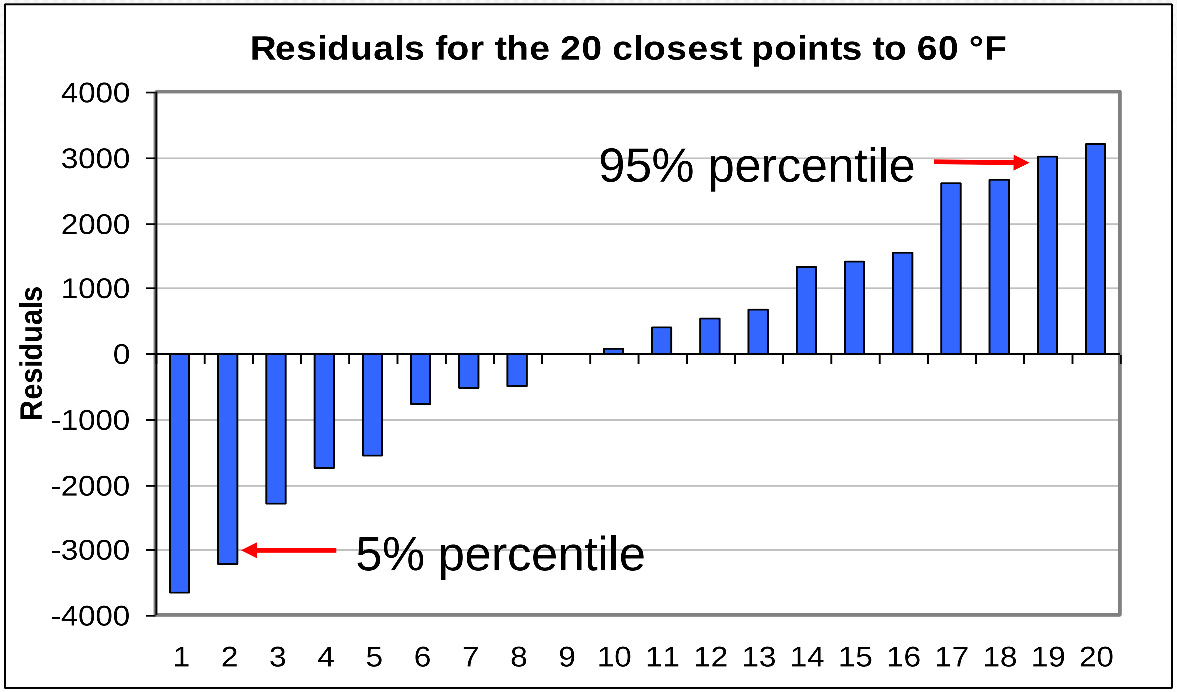
Find out an 20 closest points to 60°F.



NEAREST NEIGHBORHOOD METHOD

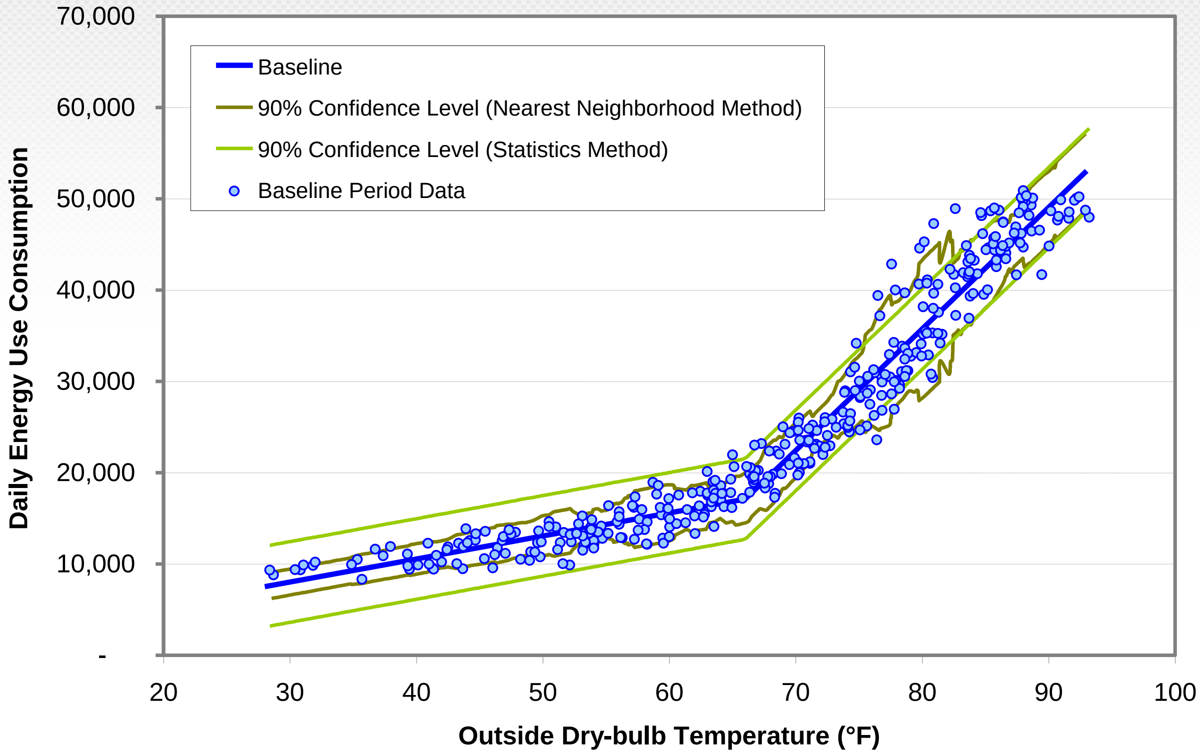
Calculate the residuals and sort them from minimum to maximum.

Uncertainty = (|5%percentile|+|95%percentile|)/2 for 90% Confidence



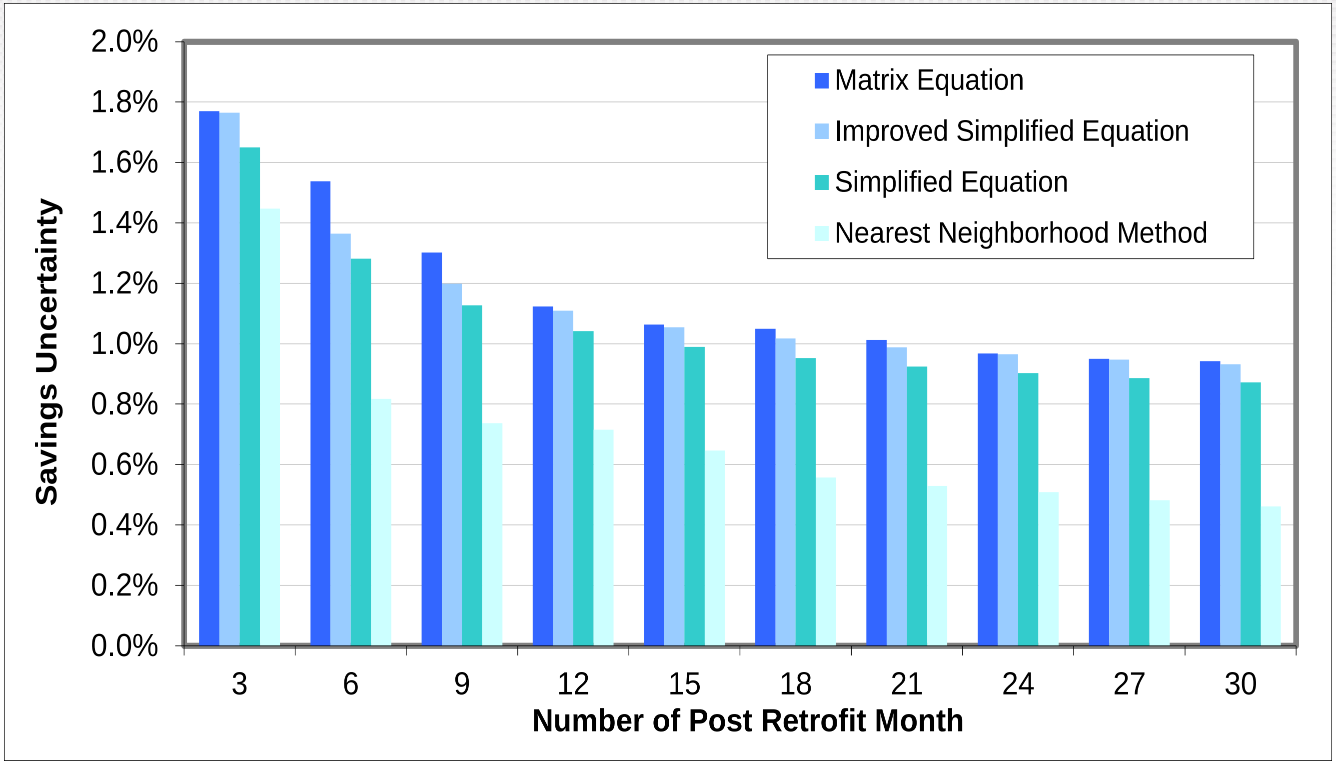
NEAREST NEIGHBORHOOD METHOD

Compare the uncertainty boundaries for statistics method and neighborhood method.



NEAREST NEIGHBORHOOD METHOD

Comparison of savings uncertainty by using different methods for daily energy use data.



NEAREST NEIGHBORHOOD METHOD

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.

BAYESIAN ANALYSIS

Shonder and Im (2012) reported their energy savings and savings uncertainty calculation methodology based on Bayesian inference and Bayesian regression technology.

Bayes inference

$$p(\theta \mid y) = \frac{p(\theta)\,p(y \mid \theta)}{p(y)} \qquad p(y) = \int p(\theta)\,p(y \mid \theta)\,d\theta$$

The integration is for θ, so $p(y)$ is a constant.

Bayes inference can also be stated as

$$p(\theta \mid y) \propto p(\theta)\,p(y \mid \theta)$$

BAYESIAN ANALYSIS

The matrix expression for linear functions

$$y = X\beta + \lambda$$

β is the vector of the parameter

λ is a random variable and the it is drawn from a normal distribution

$$\lambda \sim N(0, \sigma^2)$$

Thus Bayesian interference is

$$p(\beta, \sigma^2 \mid y) \propto p(\beta)\, p(\sigma^2)\, p(y \mid \beta, \sigma^2)$$

BAYESIAN ANALYSIS

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 β, σ^2 and X is

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

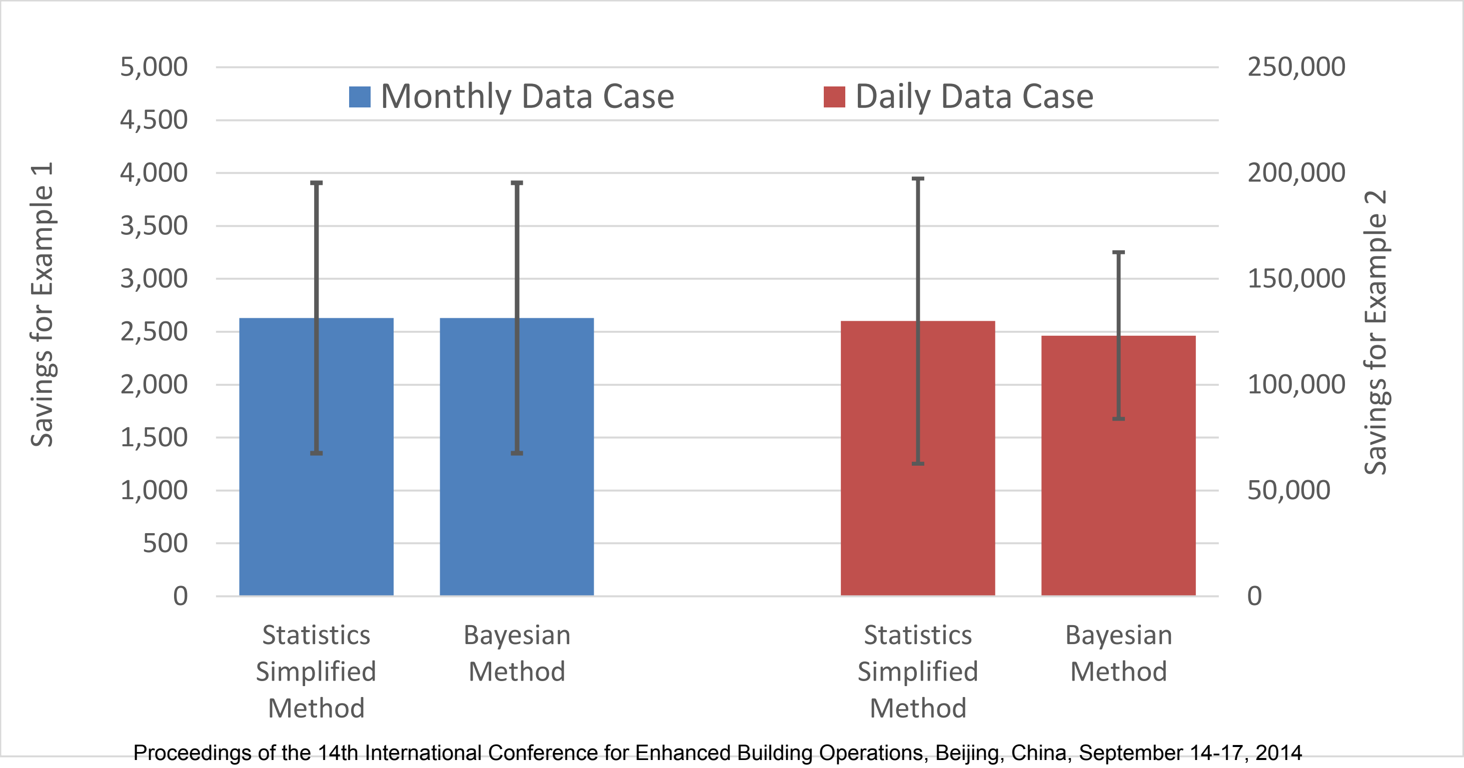
It is need to integrate the posterior distribution to characterize the distributions of the parameters.

Usually, the distribution is integrated numerically.

BAYESIAN ANALYSIS

- **The two examples results from in Shonder and Im's paper.**
  - **monthly data case**
  - **daily data case**

Confidence level for both cases is 95%



Proceedings of the 14th International Conference for Enhanced Building Operations, Beijing, China, September 14-17, 2014

BAYESIAN ANALYSIS

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.

CONCLUSION

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).

CONCLUSION

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 Baltazar, PhD, PE, BEMP