Challenging Issues in Stochastic Calibration based on Bayesian Paradigm for Building Energy Model

Young-Jin Kim[1*]

[1*]*Division of Architecture, Architectural Engineering and Civil Engineering, College of Engineering, Sunmoon University, Asan, Chungnam, 336-708, South Korea*
yjkim9943@sunmoon.ac.kr

Abstract

Bayesian calibration has been used to transform the prior distributions of unknown inputs inherited in a building energy simulation model into the trustworthy posterior distributions. It obtains the posterior distributions using a joint distribution composed of a likelihood function and prior distributions of unknown inputs in the Bayesian paradigm. In other words, it provides higher benefits in terms of a stochastic approach than the deterministic calibration and the feasibility of their calibrated results was sufficiently discussed. However, challenging issues in Bayesian calibration still remains as follows: (1) inappropriate selection of prior distributions, (2) truncated sample dataset of the likelihood functions. The aforementioned issues can increase the risks of Bayesian calibration. This paper aims to inform the risks of Bayesian calibration associated with the aforementioned issues through a reference case study. For this study, the Gaussian Process (GP) emulator, which can be regarded as a meta-model of Building Performance Simulation (BPS) tools, was used to reduce the simulation run-time. Bayesian calibration using the GP emulator was implemented with what-if scenarios considering the aforementioned issues. And then the validated models were used for a stochastic retrofit analysis of glazing systems. With the results of the estimated posterior distributions, validation, and stochastic retrofit, this paper presents Bayesian calibration issues regarding the selection of prior distributions and sample dataset of the likelihood functions.

Keywords: Bayesian calibration, Gaussian Process, Emulator, What-if scenario, Building energy model

1. Introduction

Recently, Building Performance Simulation (BPS) tools have been widely used for optimal design, optimal control, and stochastic retrofit analysis. However, the BPS tools decrease accuracy, reproducibility, and reliability of the predicted outputs due to uncertainty sources. To reduce the uncertainty, model calibration techniques that inversely estimate unknown inputs by minimizing difference between measured data and predicted outputs have been used. The model calibration techniques are divided into (1) trial and error method, (2) deterministic calibration, and (3) Bayesian calibration. Among the model calibration techniques, Bayesian calibration can estimate the probability distributions of unknown inputs and produce more accurate and robust predicted outputs than the other calibration techniques even if the computational cost is too heavy a burden under tight budget and schedule time. But it can be resolved using a meta-model or surrogate model of the BPS tools [1-5].

Bayesian calibration obtains posterior distributions using a joint distribution composed of a likelihood function and prior distributions of unknown inputs. In previous studies [6-9], the unknown inputs within the BPS tools were estimated using Bayesian calibration, and their calibrated results were sufficiently validated. And they insisted that Bayesian calibration can be used to deal with various stochastic decision making problems by coupling between the post-processing techniques (*e.g.* optimization algorithm, uncertainty analysis, and sensitivity analysis) and the reliable tuning BPS tools. However, it should be noted that different posterior distributions could be obtained by the inappropriate prior distributions and likelihood functions. And the aforementioned issues have not been presented even now.

In this study, the risks of Bayesian calibration depending on inappropriate selection of prior distributions and truncated sample dataset of the likelihood functions were discussed. For this study, EnergyPlus was chosen as a BPS tool for a given target building, and the prior distributions of unknown inputs were selected by referring to the previous literatures. For the model calibration, the measured data was assumed as the predicted outputs obtained by EnergyPlus in which simulation inputs employed a set of definite values randomly selected within the probability ranges of the unknown inputs. To discuss the risks of Bayesian calibration, this paper implemented Bayesian calibrations in the given what-if scenarios based on the aforementioned issues. And the calibrated models for each what-if scenario were validated using the separate data. The validated models were used for a stochastic retrofit analysis of glazing systems and their results were compared. In other words, this study addresses the issues of Bayesian calibration when it is impossible to accurately obtain the prior distributions and the sample dataset of a likelihood function.

2. Bayesian Calibration and Meta-Model

2.1. Bayesian Calibration

The BPS tool is a mathematical model $\eta(x_D, \theta)$ that approximates the causal relationships of the thermal properties in the real systems. The relationships are realized by numerous deterministic inputs x_m, uncertain inputs θ, and outputs y_m. The deterministic inputs denote definite values (e.g. building location (e.g. latitude, longitude, orientation, and elevation), thermal zoning, system types, control strategy, and simulation run-period) derived from design drawing, specifications, measured data, and literatures, etc. In the contrast, the uncertain inputs denote unknown values having probabilistic characteristics. The uncertain inputs were divided into static and dynamic unknown inputs. The static unknown inputs have non-time-varying natures such as thermal properties of materials (density, conductivity, and specific heat), internal gains of lights and equipment, etc. The dynamic unknown inputs have time-varying natures such as weather data, occupants' behavior patterns, and schedules. Please be noted that the BPS tools need the calibration works for transforming unknown inputs into trustworthy inputs for rational decision making.

In general, the model calibration is an optimization problem wherein unknown inputs are estimated by minimizing the difference between the measured data y_D and the predicted outputs y_m. To improve the abilities of the calibrated model prediction, the model calibration must reflect the unknown inputs as well as two additional uncertain types (discrepancy $\delta(x_D)$ due to model inadequacy and measurement error $\varepsilon(x_D)$ due to sensor noise) as shown in Equations 1-2.

$$y = y_m + \delta = \eta(x_D, \theta) + \delta(x_D) \tag{1}$$

$$y_D = y + \varepsilon(x_D) \tag{2}$$

In Bayesian calibration, the posterior distribution $p(\theta|y)$ can be calculated by multiplying the prior distribution $p(\theta)$ and the observed sample dataset of the likelihood function $p(y|\theta)$ as shown in Equation 3. In other words, the estimation of the posterior distributions is affected by the prior distribution and the sample dataset of the likelihood function.

$$p(\theta|y) \propto p(y|\theta) \times p(\theta) \tag{3}$$

The prior distribution can capture unbiased and non-overfitting probabilistic natures if the measured data are quite plentiful. However, it is not easy to obtain the appropriate measured data regarding all systems in a short building lifecycle. Thus, a few prior distributions are assumed as non-informative prior distributions selected by referring to the existing literatures. The non-informative prior distribution is a flat prior distribution in which the estimation of the posterior distribution is not affected by other data except for the observed sample dataset of the likelihood function. A representative example is a beta distribution having the shape parameters ($a = 1$ and $b = 1$). And it can be easily used when the prior distribution is vague, but it is likely to draw the instability of the posterior distribution and the convergence problem of Bayesian inference. The likelihood function is the probability for obtaining an observed data given the model. It could be affected by the quantity and quality of the measured data. If the measured data are deficient or truncated, the parameter uncertainty increases. And a Markov Chain Monte Carlo (MCMC) method for Bayesian calibration is used to deal with the integration problem of the marginal distribution by generating a sample from the parameter space, but it requires numerous iterative calculations. But, if the number of samples is small, a convergence problem could occur. To solve the problem, there is an approach that the calibration works are performed by only sampling the influential unknown inputs through a screening method (e.g. Morris method). Furthermore, a Gaussian Process (GP) emulator can provide significant probabilistic results having a fast computational speed. With the aforementioned methods, it enables to search a wide parameter space while increasing the number of samples for the MCMC method. In the following section, the GP emulator for Bayesian calibration is described in detail.

2.2. Gaussian Process Emulator

The GP emulator is constructed by (1) training dataset, (2) Gaussian Process, and (3) Bayesian inference [1-5]. The training dataset are generated by the sampling methods (e.g. Simple Random Sampling (SRS), Latin Hypercube Sampling (LHS)). The LHS method is widely used due to its ability to produce significant results with few samples compared to the simple random sampling [10]. With the generated training dataset, the mathematical model and model discrepancy are represented by using the Gaussian Process model that consists of a kernel matrix ($K(x_i, x_j)$) having

zero mean function. In general, the kernel matrix in the Gaussian Process model was assumed as a squared exponential covariance function as shown in Equation 4. The squared exponential covariance function has two hyperparameters (scaling parameter σ_{se}^2 and length scales $l_{1:p}$). And The Gaussian noise (ε_i) is generally expressed as an independent identically distributed (iid) normal distribution with zero mean and variance v_n. The variance in the Gaussian noise is also a hyperparameter. The hyperparameters need to obtain the posterior estimates through the MCMC methods.

$$\sum_{i,j} = k(x_i, x_j) = \sigma_{se}^2 \exp(-\frac{1}{2}\sum_{k=1}^{d}\left|x_{i,k} - x_{j,k}\right|^2 / l_k^2) \tag{4}$$

The MCMC methods can easily estimate the posterior distribution by numerically calculating the multidimensional integration using the Markov chain and Monte Carlo integration [11, 12]. In the MCMC methods, Metropolis-Hastings algorithm, the Gibbs sampling algorithm, and the Hybrid Monte Carlo (HMC) algorithm are used. Among the MCMC sampling methods, the Hybrid Monte Carlo algorithm is a combination of the Metropolis rule and Gibbs sampling, and its sampling and convergence abilities are superior to those of the Metropolis-Hastings algorithm and the Gibbs sampling algorithm [13].

In previous studies [1-5], the aforementioned process was described in detail, and there was sufficiently validated the superiority of Bayesian calibration using GP emulator. However, Bayesian calibration still remains the challenging issues as follows: (1) inappropriate selection of prior distributions, (2) truncated sample dataset of the likelihood functions. In this paper, the issues are discussed.

3. Target Building and Development & Validation of the GP Emulator

3.1. Target Building and What-if Scenarios

A general office building was chosen as the target building and modelled by EnergyPlus as shown in Figure 1. In this study, the model calibration was performed for a virtual building (EnergyPlus model) rather than an existing building. Since, the goal of this study was to examine the risks of Bayesian calibration depending on the inappropriate selection of prior distributions and the truncated sample dataset of the likelihood function, rather than to accurately calibrate an existing building. In the case of the parameter estimation of the prior distribution, it is desirable to use statistical tests that estimate the distribution of the population based on the sample measured in the past (e.g. one-sample Kolmogorov-Smirnov (K-S) test). However, it is not easy to collect the sample for various existing buildings. Thus, most unknown inputs are determined by probabilistic values derived in existing literatures or the judgment and assumption of experts. This leads to different probability distributions of the population.
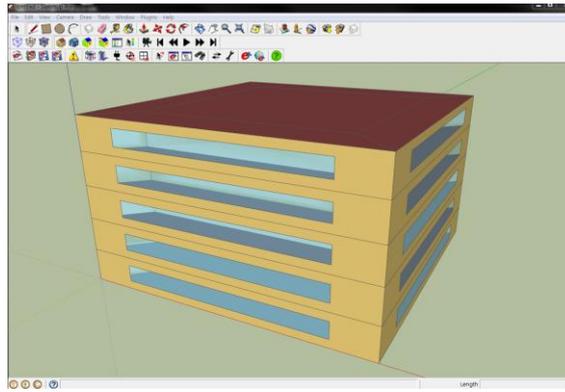
Figure 1. Target Building (Display in OpenStudio)

Riddle and Muehleisen [14] indicated that if the variance ranges using general probability distribution models (e.g., normal distribution, beta distribution, and triangular distribution) can sufficiently explain the uncertainty of unknown inputs, the parameter estimation of the prior distribution does not significantly affect the accuracy of Bayesian calibration. With this mind, a total of 24 normal distributions based on existing literatures [15-19] were used as shown in Table 1.

Table 1. Prior Distribution of the Unknown Inputs

Unknown inputs			Normal distribution	
			Mean	Standard deviation
x1	Gypsum board	Conductivity (W/m-K)	0.43	0.153
x2		Density (Kg/m^3)	1488	501
x3		Specific Heat (J/kg-K)	958	109
x4	Brick	Conductivity (W/m-K)	1.25	0.354
x5		Density (Kg/m^3)	2000	33
x6		Specific Heat (J/kg-K)	840	90
x7	Heavyweight concrete	Conductivity (W/m-K)	1.68	0.54
x8		Density (Kg/m^3)	2310	225
x9		Specific Heat (J/kg-K)	840	90
x10	Insulation	Conductivity (W/m-K)	0.039	0.014
x11		Density (Kg/m^3)	38	27
x12		Specific Heat (J/kg-K)	1072	298
x13	Tile	Conductivity (W/m-K)	0.932	0.406
x14		Density (Kg/m^3)	1610	436
x15		Specific Heat (J/kg-K)	818	89
x16	Glass	U-factor (W/m^2-K)	2.74	0.274
x17		SHGC (dimensionless)	0.76	0.076
x18	Infiltration rate	Air Change per Hour (1/hour)	0.5	0.17
x19	Indoor set-point temperature	Heating	21.5	0.43
x20		Cooling	27	0.54
x21	People	Person per area	0.215	0.1075
x22		Activity level (W/m^2)	125	12.5
x23	Light	Power density (W/m^2)	18.22	3.644
x24	Equipment	Power density (W/m^2)	16.15	3.23

To deal with the inappropriate variance ranges of the prior distribution, the variances of the unknown inputs were changed. For the sample dataset of the likelihood function, monthly total energy consumptions during two years (2009 and 2010, in Seoul) were calculated by randomly generating the definite inputs extracted

from the variance ranges of the unknown inputs. Table 2 shows the what-if scenarios regarding the selected prior distributions and sample dataset of the likelihood function.

Table 2. What-if Scenarios Depending on the Priors and the Likelihood Function

| Scenarios		Descriptions
1	Prior distribution	Mean: refer to Table 1, Standard deviation: refer to Table 1
	Likelihood function	Monthly energy consumption (2009 and 2010)
2	Prior distribution	Mean: refer to Table 1, Standard deviation: ±5% of the mean
	Likelihood function	Monthly energy consumption (2009 and 2010)
3	Prior distribution	Mean: refer to Table 1, Standard deviation: ±10% of the mean
	Likelihood function	Monthly energy consumption (2009 and 2010)
4	Prior distribution	Mean: refer to Table 1, Standard deviation: refer to Table 1
	Likelihood function	Monthly energy consumption (2010)
5	Prior distribution	Mean: refer to Table 1, Standard deviation: ±5% of the mean
	Likelihood function	Monthly energy consumption (2010)
6	Prior distribution	Mean: refer to Table 1, Standard deviation: ±10% of the mean
	Likelihood function	Monthly energy consumption (2010)
7	Prior distribution	Mean: Table 1, Standard deviation: Table 1
	Likelihood function	Monthly energy consumption (random sampling of 12-month data)
8	Prior distribution	Mean: refer to Table 1, Standard deviation: ±5% of the mean
	Likelihood function	Monthly energy consumption (random sampling of 12-month data)
9	Prior distribution	Mean: refer to Table 1, Standard deviation: ±10% of the mean
	Likelihood function	Monthly energy consumption (random sampling of 12-month data)

3.2. Development and Validation of the GP Emulator

For the training dataset, a total of 200 samples were generated using the LHS method. 150 samples and 50 samples were used to construct and validate the GP emulator, respectively. In this study, the GP emulators of what-if scenarios # 4-9 replaced those of what-if scenarios # 1-3 having the same prior distributions and wider likelihood functions. Figures 2-3 show results comparing between the GP emulator and EnergyPlus using a boxplot and the two-sample K-S test. The two-sample Kolmogorov-Smirnov (K-S) test was used to find whether the populations of the two samples have same probability distribution or not. The null hypothesis is that the two populations have the same distribution, and the alternative hypothesis is the opposite of the null hypothesis. If p-value (0.0-1.0) is smaller than 0.05, the null hypothesis is rejected. As shown in Figures 2-3, the probabilistic results of the GP emulator are similar to those of EnergyPlus and the null hypothesis was accepted. In other words, the GP emulator can provide significant probabilistic results having
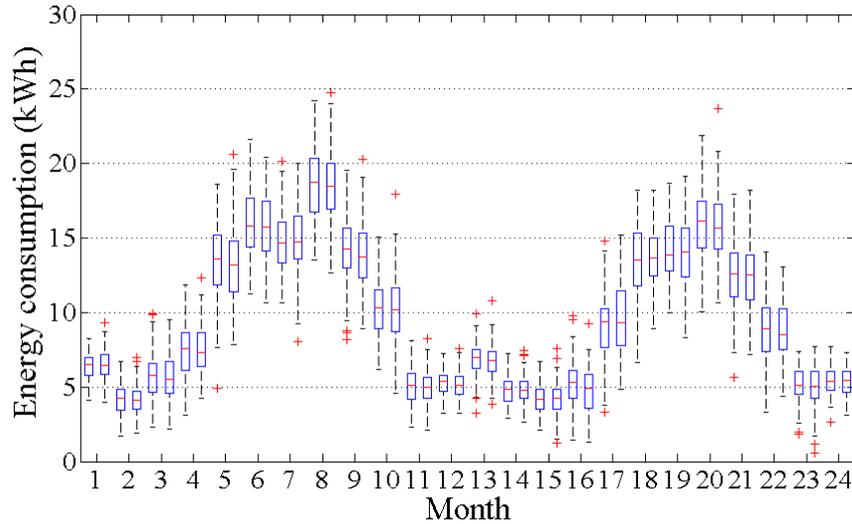
fast computational speed. In this study, the GP emulator was used for Bayesian calibration.
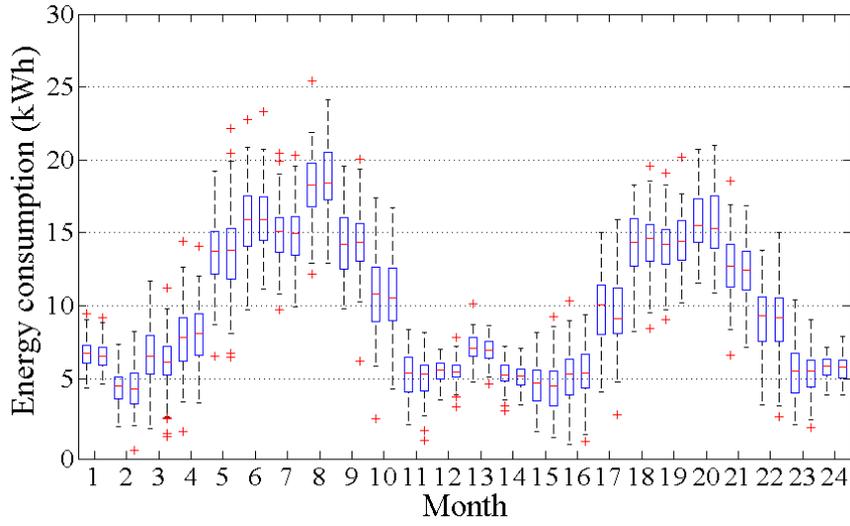
(a) Scenario #1
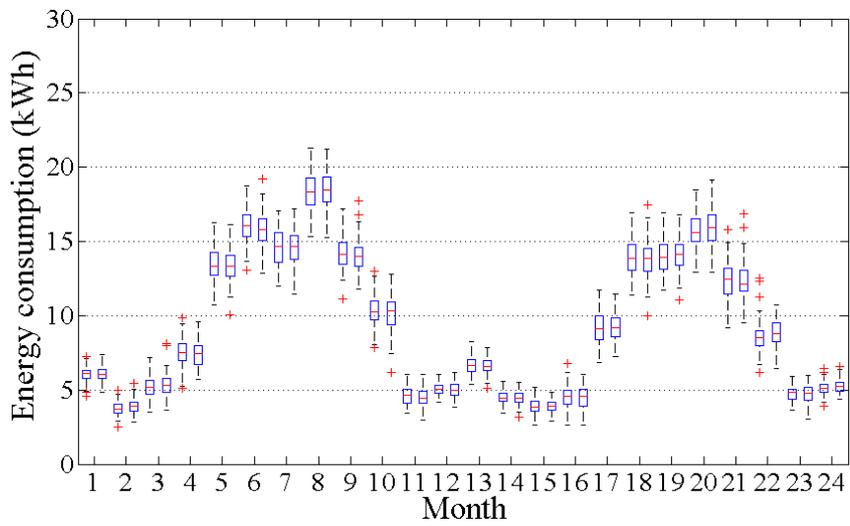
(b) Scenario #2

(c) Scenario #3

Figure 2. Comparison between EnergyPlus (left) and the GP Emulator (right)
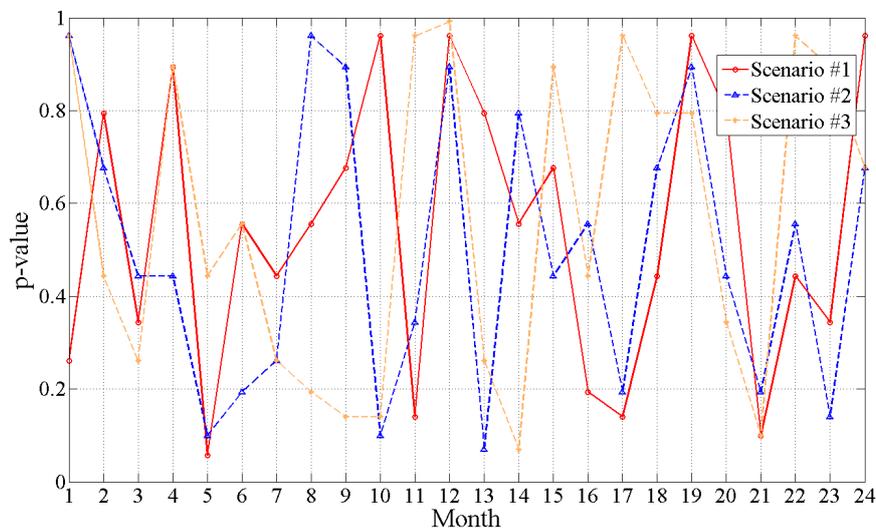
Figure 3. P-values Using the Two-sample K-S Test

4. Calibration Results

4.1. Priors vs. Posteriors

According to what-if scenarios #1-9, Bayesian calibration using the Hybrid Monte Carlo algorithm was implemented. The number of sampling and burn-in period were set to 20,000 and 2,000, respectively. The burn-in period was used to avoid the effect of the initial value. Tables 3-5 show the results of the posterior distributions. In the results, the coefficient of variations of the posterior distributions was smaller than those of the prior distributions. It can be inferred that Bayesian calibration can reduce the uncertainty. However, it should be noted that

the posterior distributions according to what-if scenarios #1-9 were different. In case of the what-if scenarios #1, 4, and 7, the definite values were in ranges of the 95% confidence interval (from 2.5% to 97.5%) even if they had the different sample dataset of the likelihood function. In the contrast, the definite values of the others were out of range of the 95% confidence interval. For an example, the posterior distributions for the density of the gypsum board (X2 in Tables 3-5) in the what-if scenarios # 2, 3, 5, 6, 8, and 9 were shifted to the left comparing the definite values. In other words, Bayesian calibration was very sensitive to the inappropriate selection of the prior distribution, rather than the truncated sample dataset of the likelihood function. Figure 4 shows the definite value, prior distributions, and posterior distributions of the density of gypsum board.

Table 3. Calibration Results for What-if Scenarios #1-3
(Std. dev.: Standard Deviation)

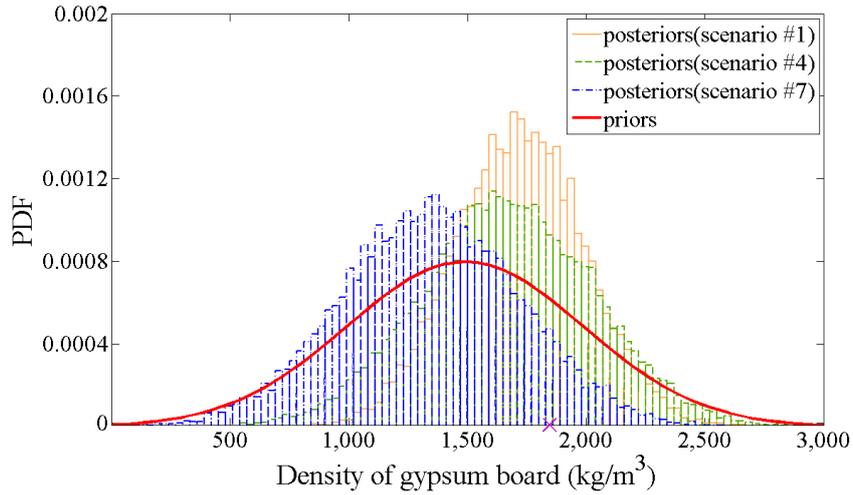|	Definite value	Posterior distribution												
		What-if scenario #1				What-if scenario #2				What-if scenario #3			
		Mean	Std. dev.	Confidence interval		Mean	Std. dev.	Confidence interval		Mean	Std. dev.	Confidence interval	
				2.5%	97.5%			2.5%	97.5%			2.5%	97.5%
X1	0.46	0.42	0.11	0.19	0.64	0.44	0.02	0.40	0.48	0.42	0.03	0.35	0.49
X2	1853.09	1722.83	271.94	1189.83	2255.82	1503.93	49.29	1407.33	1600.53	1458.51	120.78	1221.78	1695.24
X3	932.04	890.60	100.74	693.15	1088.05	909.49	42.98	825.25	993.74	917.47	65.92	788.27	1046.66
X4	1.01	1.21	0.29	0.64	1.78	1.27	0.05	1.18	1.37	1.21	0.09	1.04	1.38
X5	2010.49	2011.87	27.46	1958.05	2065.70	2009.03	58.93	1893.52	2124.53	2078.26	119.57	1843.91	2312.61
X6	880.31	896.02	100.58	698.88	1093.15	823.19	30.66	763.11	883.28	852.29	54.05	746.35	958.22
X7	1.52	1.57	0.35	0.89	2.25	1.65	0.06	1.52	1.77	1.70	0.11	1.47	1.92
X8	2452.39	2416.13	163.19	2096.28	2735.97	2278.02	65.13	2150.37	2405.68	2019.27	109.74	1804.17	2234.37
X9	696.66	815.34	61.71	694.38	936.29	851.00	26.15	799.76	902.25	825.53	38.77	749.54	901.52
X10	0.04	0.04	0.01	0.03	0.05	0.04	0.001	0.035	0.041	0.04	0.002	0.036	0.043
X11	35.00	31.86	19.44	-6.24	69.96	38.23	1.18	35.92	40.54	37.20	2.82	31.66	42.73
X12	1200.44	1037.06	239.66	567.33	1506.79	1090.00	41.42	1008.82	1171.18	1094.50	75.37	946.77	1242.23
X13	0.77	0.84	0.29	0.27	1.41	0.93	0.04	0.86	1.01	0.99	0.09	0.82	1.17
X14	1936.54	1850.34	344.71	1174.72	2525.97	1696.82	74.21	1551.37	1842.27	1603.56	94.43	1418.47	1788.65
X15	881.35	812.06	79.16	656.90	967.22	831.02	30.40	771.44	890.60	772.80	47.89	678.92	866.67
X16	2.39	2.60	0.17	2.26	2.93	2.39	0.08	2.22	2.55	2.26	0.12	2.03	2.48
X17	0.66	0.67	0.02	0.62	0.71	0.68	0.01	0.66	0.69	0.65	0.01	0.62	0.68
X18	0.39	0.38	0.01	0.36	0.41	0.40	0.01	0.38	0.41	0.42	0.01	0.40	0.43
X19	20.94	21.01	0.16	20.70	21.31	20.78	0.18	20.43	21.13	21.17	0.19	20.81	21.54
X20	27.26	27.03	0.25	26.55	27.52	26.92	0.24	26.45	27.39	27.23	0.32	26.61	27.85
X21	0.27	0.28	0.04	0.21	0.36	0.22	0.01	0.21	0.24	0.24	0.01	0.22	0.25
X22	121.48	118.55	10.99	97.01	140.10	130.74	4.46	121.99	139.49	136.17	7.84	120.80	151.54
X23	17.34	14.55	1.56	11.50	17.60	17.92	0.54	16.86	18.97	19.17	0.84	17.53	20.81
X24	15.91	16.61	1.34	13.99	19.23	15.95	0.63	14.72	17.18	15.59	1.07	13.49	17.69

Table 4 Calibration Results for What-if Scenarios #4-6
(Std. dev.: Standard Deviation)

	Definite value	What-if scenario #4				What-if scenario #5				What-if scenario #6			
		Mean	Std. dev.	Confidence interval		Mean	Std. dev.	Confidence interval		Mean	Std. dev.	Confidence interval	
				2.5%	97.5%			2.5%	97.5%			2.5%	97.5%
X1	0.46	0.44	0.14	0.16	0.71	0.43	0.02	0.39	0.47	0.43	0.03	0.36	0.50
X2	1853.09	1669.57	365.49	953.20	2385.94	1482.70	68.19	1349.05	1616.35	1439.45	131.45	1181.81	1697.08
X3	932.04	938.47	105.68	731.33	1145.61	937.32	41.92	855.15	1019.49	979.34	74.00	834.30	1124.38
X4	1.01	1.26	0.28	0.71	1.81	1.26	0.06	1.15	1.38	1.22	0.11	1.00	1.43
X5	2010.49	2004.22	29.11	1947.16	2061.27	2026.39	74.50	1880.36	2172.42	2074.81	114.29	1850.80	2298.82
X6	880.31	850.23	87.96	677.83	1022.64	846.52	37.41	773.20	919.84	830.34	66.33	700.34	960.35
X7	1.52	1.58	0.38	0.82	2.33	1.67	0.08	1.52	1.82	1.71	0.14	1.43	1.98

X8	2452.39	2336.44	185.03	1973.79	2699.10	2258.04	102.35	2057.45	2458.64	2277.01	136.31	2009.84	2544.18
X9	696.66	793.44	72.82	650.71	963.16	827.18	37.99	752.71	901.64	808.36	66.02	678.95	937.76
X10	0.04	0.03	0.01	0.02	0.05	0.04	0.002	0.036	0.042	0.04	0.002	0.032	0.041
X11	35.00	44.38	20.95	3.33	85.44	37.77	1.67	34.49	41.04	38.70	3.57	31.69	45.71
X12	1200.44	1083.75	266.39	561.62	1605.88	1087.05	53.98	981.26	1192.84	1084.24	99.69	888.85	1279.62
X13	0.77	0.87	0.33	0.22	1.52	0.95	0.04	0.86	1.03	0.93	0.08	0.77	1.09
X14	1936.54	1667.22	319.10	1041.79	2292.66	1588.27	76.73	1437.88	1738.66	1633.56	126.98	1384.68	1882.45
X15	881.35	782.96	78.30	629.50	936.42	811.34	35.57	741.61	881.07	796.94	61.74	675.93	917.94
X16	2.39	2.82	0.19	2.45	3.20	2.57	0.11	2.35	2.79	2.45	0.17	2.11	2.78
X17	0.66	0.72	0.05	0.63	0.81	0.75	0.03	0.70	0.80	0.71	0.04	0.63	0.78
X18	0.39	0.39	0.01	0.36	0.41	0.41	0.01	0.39	0.42	0.41	0.01	0.39	0.43
X19	20.94	21.22	0.22	20.78	21.66	21.03	0.23	20.58	21.47	21.04	0.28	20.49	21.59
X20	27.26	27.66	0.35	26.98	28.34	27.58	0.33	26.94	28.22	27.56	0.50	26.58	28.55
X21	0.27	0.31	0.05	0.22	0.40	0.22	0.01	0.20	0.24	0.22	0.01	0.19	0.25
X22	121.48	125.53	9.88	106.20	144.92	125.89	6.11	113.92	137.86	120.59	10.00	100.98	140.20
X23	17.34	14.83	2.27	10.39	19.28	18.46	0.76	16.89	19.94	19.07	1.27	16.59	21.56
X24	15.91	16.38	1.66	13.12	19.64	16.42	0.69	15.08	17.77	16.74	1.23	14.32	19.16

Table 5. Calibration Results What-if Scenarios #7-9
(Std. dev.: Standard Deviation)

	Definite value	What-if scenario #7				What-if scenario #8				What-if scenario #9			
		Mean	Std. dev.	Confidence interval		Mean	Std. dev.	Confidence interval		Mean	Std. dev.	Confidence interval	
				2.5%	97.5%			2.5%	97.5%			2.5%	97.5%
X1	0.46	0.45	0.15	0.16	0.74	0.43	0.02	0.39	0.48	0.43	0.04	0.35	0.50
X2	1853.09	1340.95	381.21	593.78	2088.12	1463.14	71.34	1323.31	1602.97	1410.24	135.81	1144.04	1676.43
X3	932.04	935.18	115.17	709.46	1160.91	945.46	46.09	855.12	1035.81	921.27	77.67	769.04	1073.51
X4	1.01	1.35	0.32	0.73	1.98	1.26	0.06	1.15	1.38	1.23	0.11	1.02	1.45
X5	2010.49	1993.54	33.95	1927.00	2060.08	2040.38	74.99	1893.40	2187.36	2074.79	120.78	1838.05	2311.52
X6	880.31	850.83	79.20	695.60	1006.06	843.04	36.56	771.39	914.69	843.10	71.65	702.66	983.54
X7	1.52	1.91	0.40	1.12	2.71	1.69	0.08	1.54	1.85	1.66	0.14	1.38	1.93
X8	2452.39	2235.20	185.99	1870.66	2599.74	2182.21	103.57	1979.21	2385.20	2015.88	138.14	1745.12	2286.64
X9	696.66	796.60	66.23	666.78	926.42	791.27	38.93	714.97	867.57	802.61	62.81	679.51	925.71
X10	0.04	0.04	0.01	0.01	0.06	0.04	0.002	0.035	0.042	0.04	0.003	0.030	0.042
X11	35.00	42.28	20.50	2.09	82.47	37.91	1.77	34.45	41.38	38.25	3.48	31.43	45.07
X12	1200.44	1123.52	287.70	559.62	1687.42	1081.75	48.61	986.48	1177.02	1062.07	102.51	861.14	1262.99
X13	0.77	0.98	0.33	0.33	1.63	0.94	0.05	0.85	1.03	0.92	0.09	0.75	1.09
X14	1936.54	1632.59	387.69	872.72	2392.45	1596.30	74.65	1449.99	1742.60	1604.31	148.75	1312.76	1895.87
X15	881.35	835.82	74.06	690.66	980.97	819.50	36.95	747.07	891.92	795.27	63.36	671.08	919.47
X16	2.39	2.64	0.21	2.23	3.06	2.61	0.11	2.39	2.82	2.52	0.18	2.15	2.88
X17	0.66	0.70	0.03	0.64	0.76	0.73	0.03	0.68	0.78	0.70	0.04	0.63	0.78
X18	0.39	0.41	0.02	0.37	0.46	0.42	0.01	0.39	0.44	0.43	0.01	0.41	0.46
X19	20.94	21.27	0.21	20.85	21.68	20.90	0.25	20.42	21.39	21.35	0.27	20.81	21.89
X20	27.26	27.11	0.27	26.59	27.63	27.45	0.33	26.80	28.09	27.87	0.45	27.00	28.74
X21	0.27	0.29	0.04	0.21	0.38	0.22	0.01	0.20	0.24	0.23	0.02	0.20	0.26
X22	121.48	130.96	11.27	108.87	153.05	126.81	6.41	114.25	139.37	132.49	9.35	114.17	150.82
X23	17.34	13.17	2.32	8.62	17.72	18.66	0.78	17.13	20.19	20.13	1.38	17.42	22.83
X24	15.91	16.79	1.82	13.22	20.37	16.41	0.73	14.99	17.84	16.65	1.43	13.84	19.46

(a) What-if scenarios #1, 4, and 7
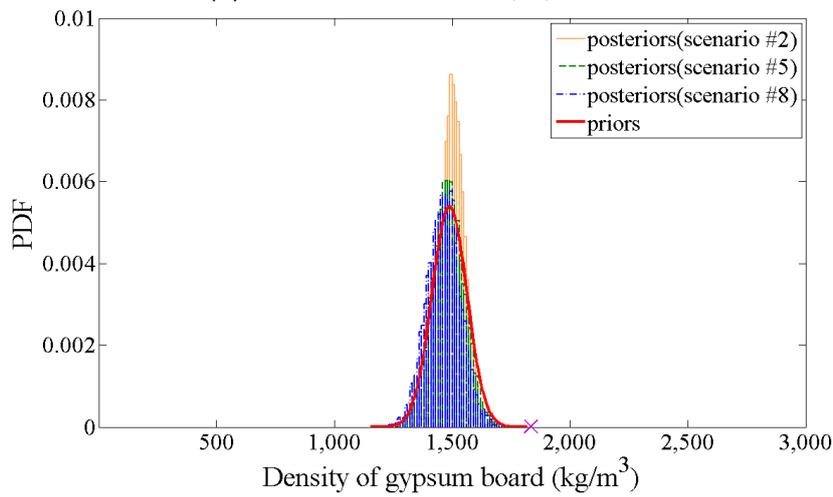
(b) What-if scenarios #2, 5, and 8
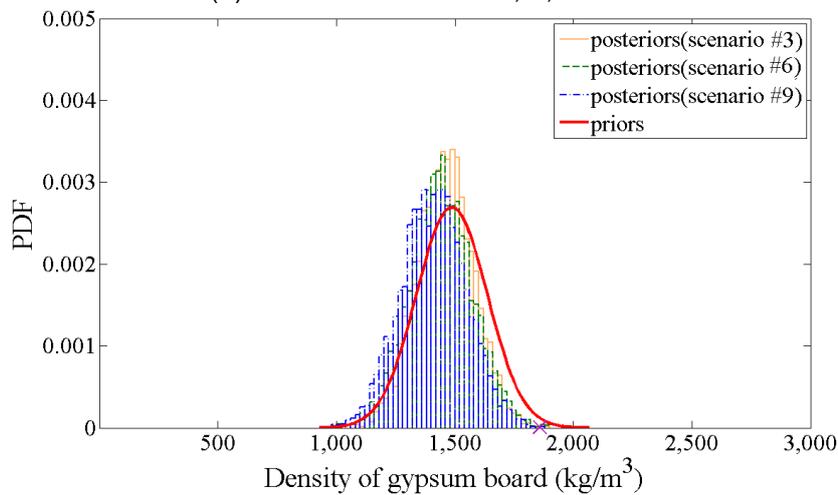
(c) What-if scenarios #3, 6, and 9

Figure 4 Comparison between Definite Value (x mark) vs. Prior Distribution (Normal Distribution), and Posterior Distribution (Histogram) for Density of the Gypsum Board

As shown in Figure 4, both the prior and posterior distributions of the what-if scenarios # 2, 5, and 8 were biased toward the left side comparing with the definite values due to the prior distribution with a small variance. In other words, the appropriate selection of the prior distribution was very important for the estimation of the posterior distribution. Due to the aforementioned issue, a few experts may select a non-informative prior distribution having a wide variance range. However, it should be noted that the selection of the prior distribution with a wide variance range could lead to increase the uncertainty of the posterior distribution due to the convergence problem of Bayesian inference. As shown in Figure 4(a), when the variance of the prior distribution is wide, the width of decrease of the variance of the posterior distribution is small even if the definite value were in ranges of the 95% confidence interval. To obtain the reliable prior distribution, the measured dataset of various buildings must be secured and filtered through the stochastic techniques. And then the filtered dataset can be used to estimate population of the prior distribution. In the following section, the validation results comparing between the measured data using EnergyPlus with definite values and predicted outputs using the calibrated model during the separate weather data (2011 year in Seoul) are described in detail.

4.2. Model Validation

Table 6 shows the validation results using the Coefficient of Variance of the Root Mean Square Error (CVRMSE). In the results, the calibrated models can provide more accurate predicted outputs than those of the uncalibrated model. It should be noted that the validation results can be improved even if the definite values were out of range of the 95% confidence interval. In other words, the calibration works only reduces the difference between the predicted outputs and measured data regardless of whether it is unbiased posterior distributions or not. In other words, it is desirable to additionally validate if the posterior distributions of the unknown inputs were properly estimated. In the case of the CVRMSE results depending on the different sample dataset of the likelihood function, the more the sample dataset of the likelihood function were truncated, the more CVRMSE values increase (what-if scenario #1 vs 4 and 7, what-if scenario #2 vs 5 and 8, and what-if scenario #3 vs 6 and 9). It can be inferred that the calibration works were affected by the truncated sample dataset of the likelihood function. Figure 5 shows the results comparing between the predicted probabilistic energy consumptions and the measured energy consumption.

### Table 6. CVRMSE Results Using Separate Data

What-if scenarios	CVRMSE (%)	
	Prior distribution	Posterior distribution
1	6.89	0.39
2	5.16	0.27
3	4.55	0.49
4	6.89	0.73
5	5.16	0.99
6	4.55	0.59
7	6.89	0.99
8	5.16	0.97
9	4.55	0.74
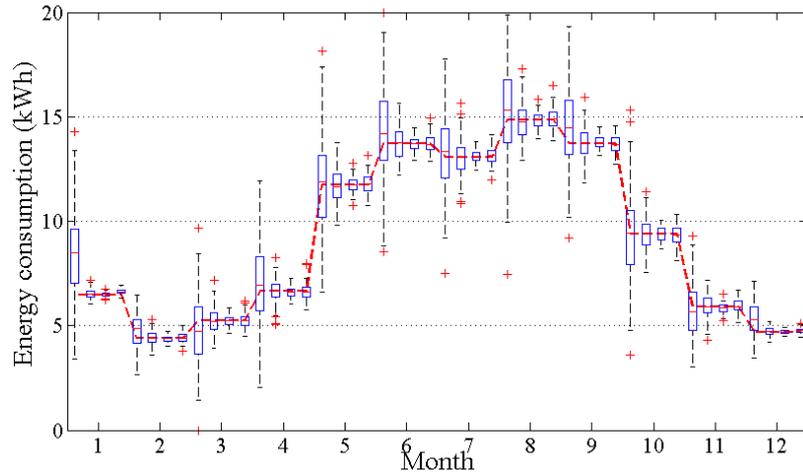
(a) Comparison between measured data (solid line) vs. uncarlibrated model
(boxplot) vs. calibrated model of what if scenarios #1, 4, and 7 (boxplot)
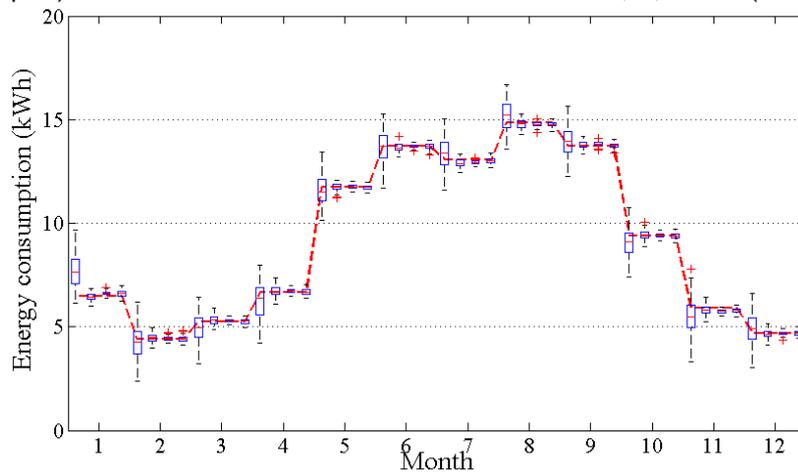
(b) Comparison between measured data (solid line) vs. uncarlibrated model
(boxplot) vs. calibrated model of what if scenarios #2, 5, and 8 (boxplot)
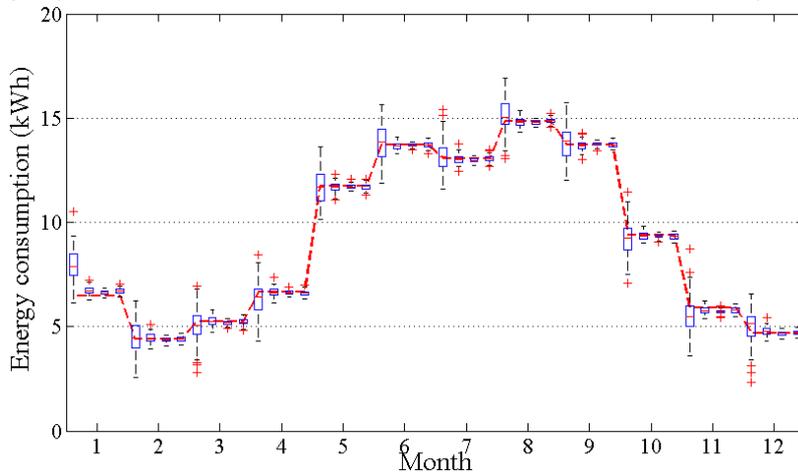
(c) Comparison between measured data (solid line) vs. uncarlibrated model
(boxplot) vs. calibrated model of what if scenarios #3, 6, and 9 (boxplot)

**Figure 5. Validation Results of the Calibrated Models Using Separate
Data**

5. Case Study

In this study, a stochastic retrofit analysis of glazing systems for the design alternatives (Table 7) was implemented. For comparing with probabilistic results of each design alternative, thermal properties of the glazing system such as U-factor and Solar Heat Gain Coefficient (SHGC) are assumed to follow a normal distribution with 10% standard deviation of definite values. To generate the sample, the LHS method was used and the number of sample is set to 200.

Table 7. Design Alternatives for Stochastic Retrofit Analysis of Glazing Systems

Design alternatives		U-factor (W/m^2-K)	SHGC (dimensionless)
1	Clear 3 mm	3.160	0.763
	Air 6 mm		
	Clear 3 mm		
2	Loe clear 6 mm	1.770	0.568
	Air 13 mm		
	Clear 6 mm		
3	Green 6 mm	3.096	0.510
	Air 6 mm		
	Clear 6 mm		

Table 8 shows the probabilistic results of the total annual energy consumption using the calibrated model. In the results, alternative #3 has a lower mean and standard deviation than the others in all the what-if scenarios. It means that the identical decision making result can be obtained if the probabilistic predicted outputs are used only for the relative comparison. However, the different results depending on the what-if scenarios may lead to a mistaken optimal solution in terms of a decision making problem requiring accurate probabilistic predicted outputs such as stochastic optimal design and control.

Table 8. Stochastic Retrofit Analysis Using the Calibrated Model (kWh)

What-if scenarios	Alternative #1		Alternative #2		Alternative #3	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
1	114.04	8.50	107.61	7.39	**99.16**	**6.75**
2	112.56	5.34	106.48	3.85	**97.87**	**3.34**
3	113.37	5.50	107.27	5.02	**98.27**	**4.54**
4	111.40	10.01	104.49	9.38	**95.90**	**8.58**
5	108.29	5.49	100.77	4.73	**92.49**	**3.97**
6	110.31	6.81	105.82	5.79	**99.22**	**5.44**
7	112.60	9.69	105.89	8.71	**97.97**	**8.04**
8	109.62	5.63	102.03	4.81	**93.67**	**4.13**
9	111.02	7.10	106.16	6.07	**99.02**	**5.64**

5. Conclusions

This paper presented the challenging issues regarding the inappropriate selection of prior distributions and truncated sample dataset of the likelihood functions in terms of Bayesian calibration. To deal with the issues, this paper implemented

Bayesian calibration based on the what-if scenarios in view of the different prior distribution and likelihood function in a given reference case study. The calibration results indicated that the inappropriate selection of prior distributions had a greater effect on the accuracy of the posterior distributions than the truncated sample dataset of the likelihood function. To solve the problem, a few experts may assume a wide probability range for a vague prior distribution. However, the estimated posterior distributions using wide variance ranges of the prior distributions could lead to increase the uncertainty. In other words, the appropriate selection of the prior distribution is very important, and it is thought that building simulation societies or research institutions need to perform studies on the detailed parameter estimation for the prior distributions. In respect of validation results of the calibrated model using separate data, the more the sample dataset of the likelihood function were truncated, the more model uncertainty increased. In other words, it is important to acquire the undamaged sample dataset. In a stochastic retrofit analysis of glazing systems, the results showed that the calibrated model could yield the same solution about each what-if scenario, but different probabilistic results derived from risks of Bayesian calibration will be inappropriate for the stochastic optimal design and control. In other words, the appropriate selection of the prior distribution and a likelihood function is important to acquire meaningful information and knowledge from robust Bayesian calibration.

Acknowledgements

References

[1] M.C. Kennedy, A. O﹐ Hagan, "Bayesian calibration of computer models", Journal of the Royal Statistical Society, Series B, vol. 63, no. 3, **(2001)**.

[2] J. Oakley, A. O﹐ Hagan, "Probabilistic sensitivity analysis of complex models: a bayesian approach", Journal of the Royal Statistical Society, Series B, vol. 63, no. 3, **(2004)**.

[3] M. Goldstein, J.C. Rougier, "Bayes linear calibrated prediction for complex systems. Journal of the American Statistical Association, Journal of the American Statistical Association, vol. 101, no. 475, **(2006)**.

[4] C.E. Rasmussen, C.K.I. Williams, Gaussian Processes for Machine Learning, the MIT Press **(2006)**.

[5] Y.J. Kim, K.U. Ahn, C.S. Park, I.H. Kim, "Gaussian emulator for stochastic optimal design of a double glazing system", Proceedings of the 13th IBPSA Conference, **(2013)**; August 25-28, Chambery, France.

[6] Y.S. Heo, "Bayesian calibration of building energy models for energy retrofit decision-making under uncertainty", Ph.D. thesis, Georgia Institute of Technology, **(2011)**.

[7] A.T. Booth, R. Choudhary, "Decision making under uncertainty in the retrofit analysis of the UK housing stock: Implications for the Green Deal", Energy and Buildings, vol.64, **(2013)**, pp. 292-308.

[8] Y.S. Heo, D. Graziano, L. Guzowski, R. Muehleisen, "Evaluation of Calibration Efficacy under Different Levels of Uncertainty", Proceedings of the 13th IBPSA Conference, **(2013)**; Chambery, France.

[9] Y.J. Kim, K.C. Kim, C.S. Park, I.H. Kim, "Deterministic vs. Stochastic Calibration of Energy Simulation Model for an Existing Building", The 2nd Asia conference of International Building Performance Simulation Association, **(2014)**; Nagoya, Japan.

[10] S. de Wit, G. Augenbroe, "Analysis of uncertainty in building design evaluations and its implications", Energy and Buildings, vol. 34, **(2002),** pp. 951-958.

[11] W.R. Gilks, S. Richardson, D. Spiegelhalter, "Markov Chain Monte Carlo in Practice", Chapman and Hall, **(1996)**.

[12] D. Gamerman, H.F. Lopes, "Markov Chain Monte Carlo-Stochastic Simulation for Bayesian Inference", Chapman and Hall, **(2006)**.

[13] M. Alfaki, "Improving efficiency in parameter estimation using the Hamiltonian Monte Carlo algorithm", Master thesis, University of Bergen, **(2008)**.

[14] M. Riddle, R.T. Muehleisen, "A Guide to Bayesian Calibration of Building Energy Models, ASHRAE/IBPSA-USA Building Simulation Conference", **(2014);** Atlanta, GA, USA.

[15] I.A. Macdonald, P. Strachan, "Practical application of uncertainty analysis", Energy and Buildings, **(2001)**; Atlanta, GA, USA.

[16] I.A. Macdonald, "Quantifying the effects of uncertainty in building simulation", Ph.D. thesis, University of Strathclyde, **(2002)**.

[17] SAREK, "Equipment Engineering Handbook", The Society of Air-conditioning and Refrigerating Engineers of Korea, **(2004)**.

[18] C.J. Hopfe, "Uncertainty and sensitivity analysis in building performance simulation for decision support and design optimization". PhD thesis, Technische Universiteit Eindhoven, **(2009)**.

[19] DOE, "EnergyPlus 8.0 Input/Output Reference: The Encyclopedic Reference to EnergyPlus Input and Output", US Department Of Energy, **(2013)**.

Authors

Young-Jin Kim, received the B.S., M.S. and Ph.D. degrees from Sungkyunkwan University. Young-Jin Kim now works at architectural engineering, Sunmoon University. His research interests include building energy simulation, BIM to BEM interface, optimal design & control, Monte-Carlo simulation, and inverse model.