2018 Building Performance Analysis Conference and
SimBuild co-organized by ASHRAE and IBPSA-USA
Chicago, IL
September 26-28, 2018

ANALYSIS OF A LARGE-SCALE DATABASE FOR ENERGY PERFORMANCE MODELING OF EXISTING BUILDINGS

Hye Gi Kim[1], Jae Eun, Sung[1], and Sun Sook, Kim[2]
[1]Department of Architecture, Graduate school, Ajou University, South Korea
[2]Department of Architecture, Ajou University, South Korea

ABSTRACT

As the importance of energy efficiency improvement and management of existing buildings increases, the importance of energy simulation and energy use evaluation of existing buildings is also increasing. However, unlike new building, the level of available information that can be entered into the simulation varies from building to building. Therefore, in various countries were developing and providing a reference model of the existing buildings for evaluating energy performance by using the available information or the standards according to completion year of the building. However, information that can not be gathered is usually input assumed value by the case study or previous studies. The ultimate goal of this study is to replace the assumed values of variables with more accurate and reliable information by calibrating the model output with measurement data. And, it is to provide which not a single reference building but a reference building pakage that can be evaluated to reflect various building characteristics across the country. This study is a basic study for this purpose, to develop a reference building pakage for the national-wide offices according to the existing reference building development method and to present the calibration direction by comparing it with the available information of the existing building.

INTRODUCTION

With the increasing needs for energy efficiency and management of existing buildings [1], the role of energy performance analysis and assessment have been also increasing. Building energy data can be used to track energy performance in buildings and to plan for future energy demand. Also, it can be used in tools and activities that help stakeholders make energy investment decisions and implement energy efficient policies and programs. However, it was difficult to predict and verify energy saving potentials in buildings before energy information was built and standardized. Thus, many countries have collected energy data and spurred several efforts to support data-driven decision-making about the building performance for stakeholders.

Representative analysis model using building energy related data to improve building energy performance can be classified into two types. There a forward model which calculates energy performance by computer simulation based on building features data corresponding to building performance, and on the other hand, an inverse model which to estimate the energy performance of buildings based on actual energy consumption data and weather data. *Table 1* and *Figure 1* presents the characteristics of each model.

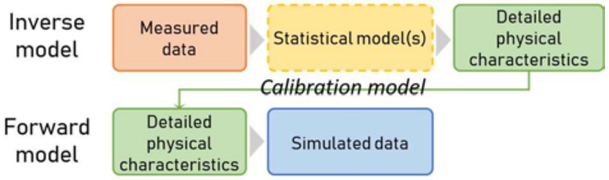


Figure 1.Relationship between forward and inverse model

Forward model (simulation model) may be becoming increasingly important as a tool to assess energy performance in the pre-retrofit phase and to predict the savings when applying energy saving measures (ECMs). However, as can be seen in *Table 1*, it is difficult to reflect uncertainties such as degradation, occupancy and operation of building performance over time because a large of difference level of available information level of existing buildings unlike new buildings. In other words, any single value input due to lack of information on the input variables of the simulation means that it does not accurately reflect the characteristics of the actual energy performance, which means that it is difficult to obtain the expected effect. Therefore, generally arbitrary single value is input through case study, or simulation was performed using reference building which are develop and provided by a national-level.

On the other hand, since the inverse model is intended to determine the mathematical description of the system and to estimate the energy feature based on the measured and known input and output variables, the applicable model is determined according to the information level of the known input and output variables.

Table 1. A characteristic of forward and inverse models

	FORWARD MODELS	INVERSE MODELS
MAIN TARGET	• Generally used for new buildings	• Generally used in existing buildings
ADVANTAGE	• Provide a detailed prediction for building energy performance. • Linked to specific physical building, system and environmental parameters.	• Shortened development time by combining engineering model with statistical models. • Predictor of building performance, given quality prior training data. • Linked to aggregated physical building, system and environmental parameters
LIMIT	• Over-parameterized and under-determined. • Require significant time, effort and expertise for development	• No explicit link between model inputs and physical building simulation input variables – impossible to extrapolate model to compute effect of design or operational changes. • Requires high level of knowledge of both engineering models and statistical models for development.

Inverse model for building energy performance estimation can be broadly classified into three main categories.

(i) Black-box approach: This refers to the use of simple mathematical or statistical models which relate a set of influential input simulation input variables to measured outputs.

(ii) Grey-box/simulation input variable estimation: Grey box approaches differ from black-box approaches in that they use certain key (or aggregated) system simulation input variables identified from a physical system model.

(iii) Detailed model calibration: The final approach uses a fully-descriptive forward model of a building system and tunes the various inputs to match the measured data.This approach provides the most detailed prediction of building performance, given the availability of high-quality input data.

In other words, inverse modeling can overcome the limitations of the forward model in that it can explain the building energy performance based on the energy consumption which reflects degraded building performance by aging and various occupancy characterisctic of existing buildings as shown Fig 1. However, a high level of knowledge is required of both engineering and statistical models for model development. Also, it is difficult to clear identify the connection between model inputs and detailed physical building simulation input variables. Therefore, based on information available on existing buildings, research is in progress in many countries to establish an effective calibration method for the forward model and reflect this in the simulation.

The Korean government also has collected a nation-wide integrated energy consumption database(DB) with about 6.9 million building since 2013 as the importance of building energy data increases. In addition, the Korean government has expanded the scope of the Building Energy Efficiency Certification System [4] from the new building to the existing one based on the recognition of the energy improvement of existing buildings. However, in spite of this awareness of energy efficiency, energy evaluation of existing buildings using simulation is performed by the energy evaluation method for new buildings using standard schedules has been applied.

The ultimate purpose of this study is to develop a reference building pakage(simulation model) that can be used to evaluate and improve the energy performance of existing buildings in Korea using national databases. And this paper is a basic study for the ultimate goal which a new calibration method of the forward model using the inverse model was discussed. Myriad of probabilistic variables were set and compare the measured data with actual office data in Korea to suggest the importance of model calibration and discuss the direction. The process of this study is briefly shown in *Figure 2*. First, we gathered collectible building information from nation-wide database on office buildings. The distribution of the collected building information were analyzed, and the actual distribution was set as a random variable and input to the simulation. The information that can not be gathered were set by assuming myriad random variables based on the completion year of the buildings. Then, energy performance calculation was performed based on various simulation model input variables set. Next, the energy performance calculated by the inputted variables were assumed to be the energy consumption of the existing building, and the information of the input variables were set as the black box. Finally, the energy consumption dabtabase of the actual office building was collected, and the inverse modeling was performed using actual energy consumption and energy performance

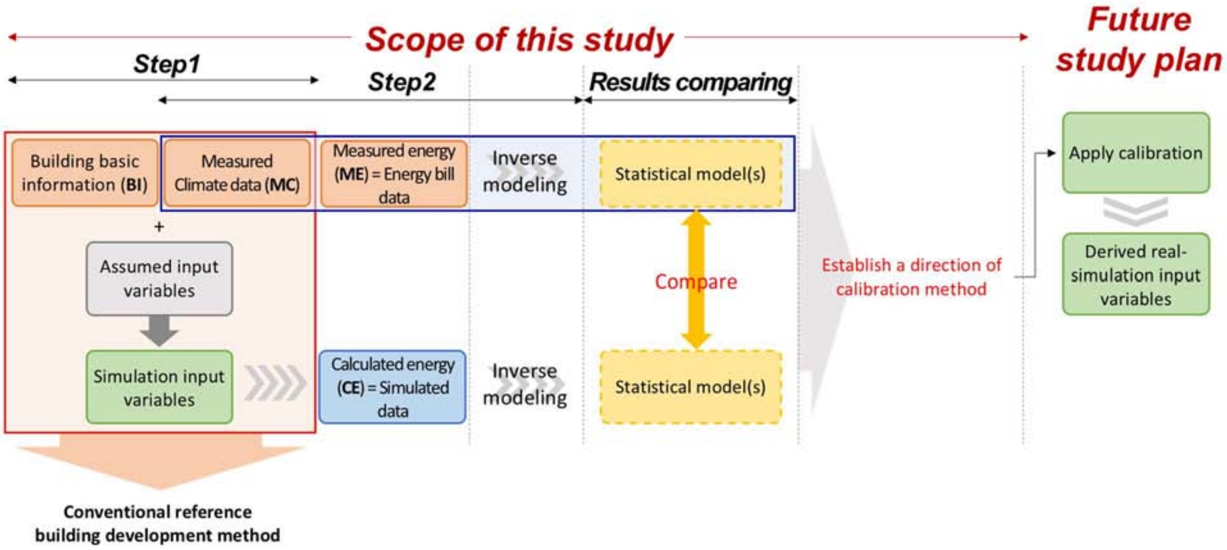


Figure 2. Scope of this study

assumed as energy consumption, lastly the results were compared. And based on the model difference results, we were established a direction to calibration method the assumed random variables.

METHOD

STEP1. Developing a reference building package

(1) Data description
The 4,304 single-use office buildings were selected excluding 2,774 multi-use buildings that were not consistently reflect the characteristics of office buildings among 7,078 buildings registered for use as office buildings. Then, data related to office buildings were collected, as shown in *Table 2*. Next, outlier rejection of energy performance data was conducted using box and whisker plot to exclude data with too high or low annual energy consumption causes distortion in the analysis. Finally, 3,905 building data were selected as subject for analysis from 4,304 single-use office buildings.

(2) Distribution analysis and Input variable assumption
Information in the Korea database is information of total building. Although there is an error due to missing data or error data, it means that it is possible to reflect the most realistic information. Available building information in Korea database are difficult to consider as ECMs. However, since these factors have profound influence on energy performance calculation and as shown in *Figure 3 ~ 8*, do not show distribution characteristics that can be generalized such as uniform or normal distribution. It means that it is necessary to apply the actual probability distribution. Simulation input variables were set to the factors that could affect the energy use of the office buildings, and the probability distribution was divided into the cases where the characteristics of the actual buildings can be reflected or can't be reflected. Because of uncertainty for available inputting of simulation input variable range, various methods of generating random values have been applied to reflect the best reasonable and realistic simulation input variables in the simulation. To set the input range of the items where the building characteristic information is not constructed as data, the existence of the legal standard and the change of the standard were first investigated. In case of occupant density and lighting power density, etc. without legal standard, we set the input value in the simulation as the default value but set the possible input values in various categories.

(3) Establishment of the reference building package
Table 3 presents the basic model characteristics for applying simulation input variables. The basic shape of the model was established according to the characteristics information of Korea. Next, *Table 4* shows simulation input variables and their distribution which use in energy model. Reference building package of office building in Korea were designed by considering the distribution of actual building data which has characteristics of existing office buildings and selecting distribution for generating arbitrary values of parameters which not collected into the integrated database.

STEP2. Inverse modeling

(1) The statistical model using MC & ME (Model I)
3-change point model was used which is one of the ASHRAE Inverse model based on the monthly outdoor

Table 2. Collected office building information data from integrated database in South Korea

	DATA	DEFINITION
BUILDING FEATURES	Location	Sixteen locations by administrative district
	Building Structure	9-structural types including reinforced concrete, brick, etc.
	Roof Type	Four material types slate, concrete, etc.
	Total Floor Area Above Ground (Gross Area)	Total area of each floor of the building {m^2}
	Floor Area Above Ground	Floor area excluding underground area {m^2}
	Number of Floors Above Ground	
	Building Age (or Approval date)	$= \mathit{Date\ of\ analysis} - \mathit{Date\ of\ approval}$ {day or year}
	Building height	Height of a building from ground level {m}
	Aspect Ratio	Collected from Energy Saving Design Standard
ENERGY PERFORMANCE	Electricity energy consumption (E)	-Monthly and Annual data {kWh/m^2} -Collected: 2013 ~ 2016 (four years) -Used: 2013~2015 (three years)
	Gas energy consumption (G)	
	District heating energy consumption (D)	
	Monthly energy use starts and end dates	- Every meter of each building

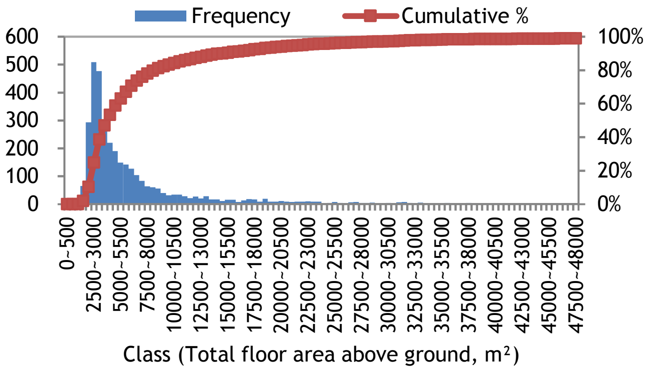


Figure 3. Histogram of total floor area (N=3,491)

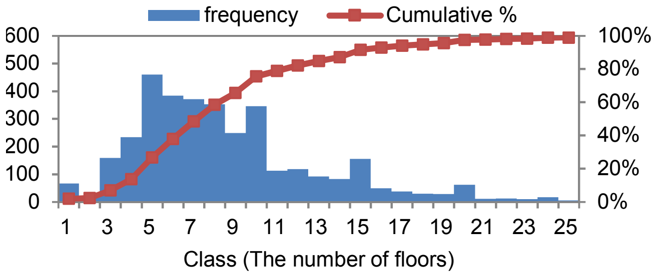


Figure 4. Histogram of the number of floors (N=3,491)

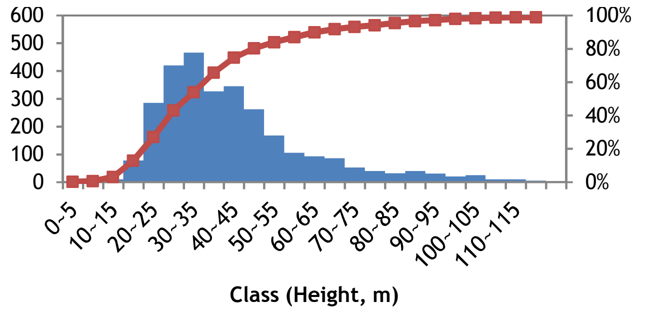


Figure 5. Histogram of building height (N=2,842)

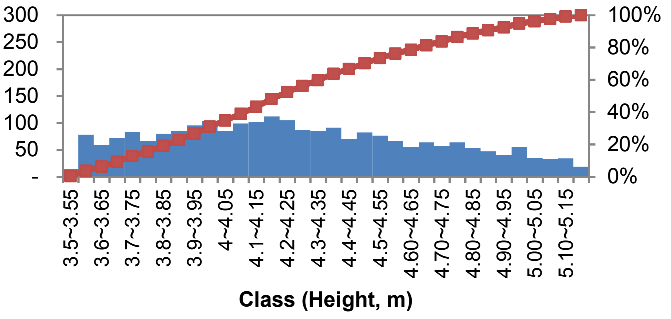


Figure 6. Histogram of the floor height (N=2,352)

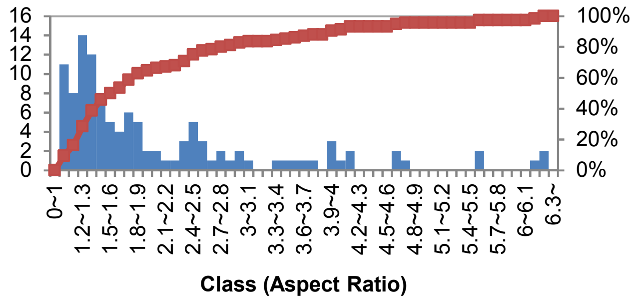


Figure 7. Histogram of the Aspect Ratio (N=116)

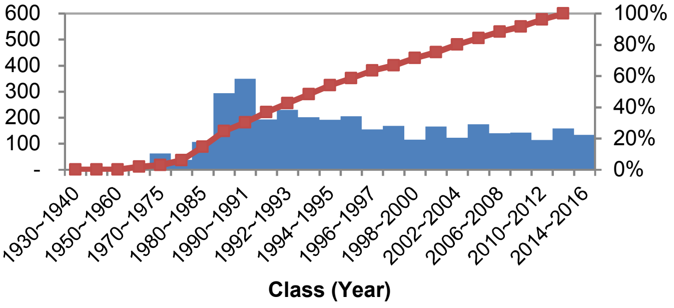
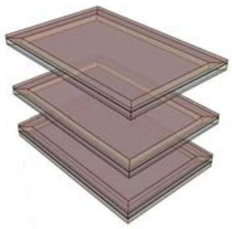


Figure 8. Histogram of Approval year (N=3,461)

Table 3. Basic concept of energy model in Korea office buildings

	**LOCATION**	Korea, Republic of (Using White Box Weather data, ASHRAE)
	FORM	Office building / Rectangle shape
		Stories: Bottom, Mid, Top (Basic) (Apply Multiplier to Mid Story according to the number of floors)
		18 zones (Basic): Core, Perimeter of each side and Plenum (Apply Multiplier to Mid Story according to the number of floors)
		Reinforced concrete structure / Perimeter size: 4.5732 {m}
	HVAC	Packaged MZVAV with plenum zones (Three objects), gas furnace, electric reheat / Economizer per 90.1-2004
	AIR PRIMARY LOOPS	VAV with electric reheat
	SYSTEM EQUIPMENT SIZE	Auto calculation6
	COIL	Cooling: DX: Two-speed / Heating: Gas / Heating: Electricity
	OHERS	Pumps: included / Boilers and Chillers: not included

dry-bulb temperature (Measured climate, MC). The model can derive energy performance and the input variables by where the energy consumption changes based on the relationship between temperature (MC) and energy use (Measured Energy, ME). Using this method, weather-dependent and weather-independent energy use can be derived such as heating, cooling and base-load. *Figure 9* shows the concept of 3-change point model and a following equation is the total energy calculation method by outdoor temperature.

$$E = b_0 + b_1(b_3 - T)^+ + b_2(T - b_4)^+ \quad \cdots\cdots\cdots \quad \mathbf{(i)}$$

Five-parameters which represent the building performing also can be estimated using this method. The horizontal line segment (b_0), represents the base-load such as a combination of lights, plug loads and process loads. The slope coefficient (b_1, b_2) is a function of the building envelope, ventilation/infiltration air, and the efficiency of the heating or cooling system. The break-even temperature (b_3, b_4) is the outdoor mean temperature when cooling or heating is started due to out of the indoor thermal comfort condition. *Figure 10* shows procedure of deriving the simulation parameters using change point modeling. In this case monthly dry-bulb temperature (°C) data were collected from Korea

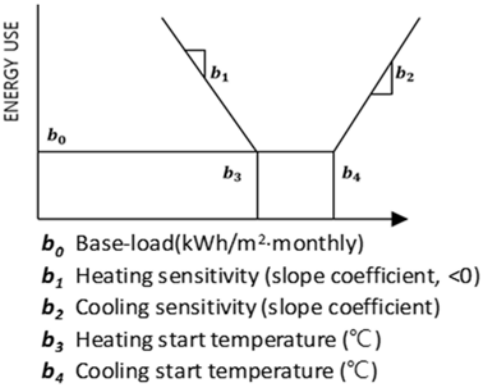


Figure 9. Steady-state single variate models

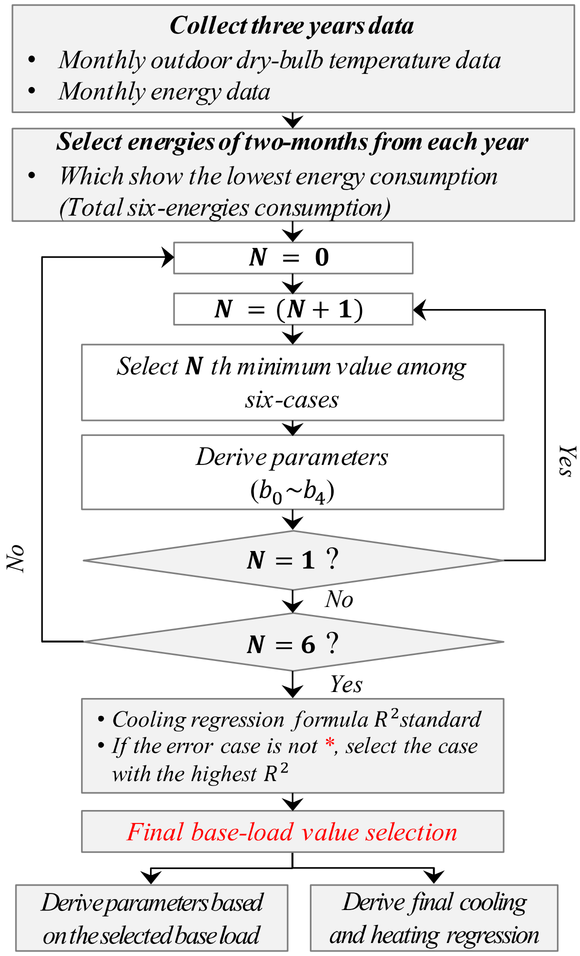


Figure 10. Procedure of deriving the parameters

Table 4. Parametric case for office building in South Korea

	VARIABLE	UNIT	DISTRIBUTION	VALUE OF SIMULATION INPUT VARIABLES	REFERENCE
P1	Orientation	{deg}	Uniform	[0: 45: 359] [min: interval: max]	0 = North Axis
P2	building fabrics: Exterior wall	-	Uniform	Total U-value of an exterior wall {W/m^2-K}: [0.13:0.05: 1.0]	-Energy saving design standard of building
P3	building fabrics: Roof	-	Uniform	Total U-value of a roof {W/m^2-K}: [0.13:0.05: 0.5]	-Energy saving design standard of building
P4	building fabrics: Exterior slab	-	Uniform	Total U-value of an exterior slab {W/m^2-K}: [0.17:0.05: 0.0.47]	-Energy saving design standard of building
P5	Total floor area	{m^2}	Actual	[1,000: 500: 50,000]	-Underground isn't considered (Above ground) - Building integrated database in Korea
P6	Number of floors		Actual	[1:1:60]	
P7	Building height	{m}	Actual	[5:5: 250]	-Building integrated database in Korea
P8	Aspect ratio		Actual	[1.0:0.1:6.3]	- Energy saving design standard
P9	Floor area	{m}	Calculation	$= P5/P6$	-Underground isn't considered (Above ground)
P10	Floor height	{m}	Calculation (constraint)	$= P7/P6$	-P10 should be in the min or max of height distribution of the actual data
P11	Surface vertex (X, Y, Z)	{m}	Calculation	$(1) = \sqrt{((P9/P8))} * P8$ $(2) = \sqrt{((P9/P8))}$ $(3) = P10 - 1.2192$ $(4) = 4.5732$ $(5) = [(1) \text{ or } (2)] - (4)$ $(6) = \mathbf{P10} * floor\ N$	(1): Long-side width, exterior vertex of X (2): Short-side width, exterior vertex of Y (3): Ceiling height (Plenum height = 1.2192m) (4): Perimeter depth, (5): Vertex of the perimeter X, Y (6): From ground level
P12	WWR		Uniform	[0.1:0.1: 0.9]	-Default value: '0.4' or '0.5'
P13	Window height	{m}	Calculation	$P11(3) * P12$	-Width (X, Y) was fixed (same as the wall)
P14	Fenestration vertex (X, Y, Z)	{m}	Calculation	$Z\_Bottom = ((P11(3) + P13)/2) * floor\ N$ $Z\_Top = ((P11(3) + P13)/2) * floor\ N$	-Vertex of width (X, Y) was fixed (same as the wall)
P15	Heating setpoint	{℃}	Uniform	[20:1.2:23.2]	-Setpoint in office area -Consider design supply temperature category -Precedent study and case investigation
P16	Cooling setpoint	{℃}	Uniform	[23.3:1.3:26]	
P17	Fresh air rate	{m^3/s-person}	Uniform	[0.005:0.005:0.02]	-Precedent study -Case investigation -Operational Regulations of the Energy Efficiency Rating of Buildings
P18	Lighting power density	{W/m^2}	Uniform	[4:4:24]	
P19	Office equipment power density	{W/m^2}	Uniform	[10:5:25]	
P20	Occupant density	{m^2/person}	Uniform	[6:3:15]	
P21	Infiltration rate		Uniform	[0.1:0.2:1.0]	
P22	Window U-value	{W/m^2-K}	Normal	**Constraint**: Should be in the min or max of (VT/SHGC) Ratio for the actual data distribution	-IGDB glazing system value
P23	Window SHGC		Uniform		
P24	Window VT		Uniform		

Meteorological Administration and matched integrated database according to the building address information. Also, the error case in Figure 10 means as follows:

(i) $b_3 < b_4$

(ii) $b_1 > 0$ or $b_2 < 0$

(iii) $b_3 >$ max temp. or $b_3 <$ min temp.

(iv) $b_4 >$ max temp. or $b_4 <$ min temp.

Through the scope of Figure 10, 714 error cases were removed, resulting in 3191 final reference building packages among 3905 office buildings.

(2) The statistical model using MC & CE (Model II)

To compare statistical models equally, we performed simulation based on the reference building package established in Step 1. At this time, dry bulb temperature, dew point temperature, humidity, local atmospheric pressure, wind direction and wind speed were replaced for existing EnergyPlus weather file of Korea based on the 2013 ~ 2015 measured climate. Then, Parametric simulation was performed using jEPlus which is a convenient tool for managing large and complex parametric simulations with EnergyPlus engine. The framework of jEPlus is shown in *Figure 11*. Total 1,250 cases were performed.

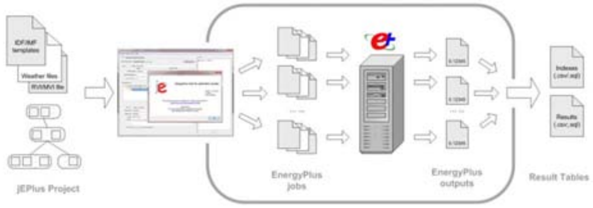


Figure 11. Scope of jEPlus with EenergyPlus

existing office buildings and selecting distribution for generating arbitrary values of simulation input variables which not collected into the integrated database. Finally, the inverse modeling of the calculated energy is also used based on the 3-change point model which is one of the ASHRAE Inverse model based on the monthly outdoor dry-bulb temperature, equally.

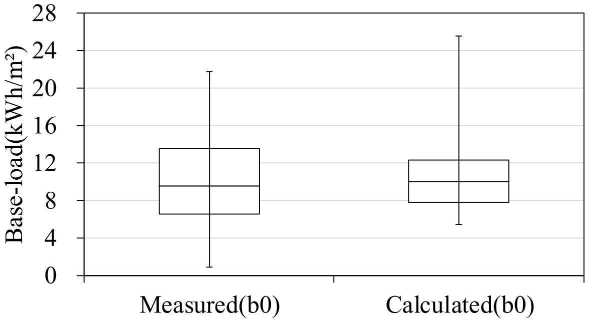


Figure 12. Comparison of base-load between measured and statistical model

DISCUSSION AND RESULT ANALYSIS

Comparison of statistical model

In order to compare the statistical model, we compared the distribution of simulation input variables of all data. *Table 5* and *Figure 12 ~ Figure14* are the results of comparing the statistical model.

Overall, derived statistical model's range based on the CE or ME did not appear to be identical or similar to each other. Although the performance range of the input variables has been largely set in consideration of the existing building, the simulation input variable ranges of the measured energy were found to be wider in most simulation input variables. Particularly, there was a clear difference in cooling sensitivity. However, in the case of the heating sensitivity, the simulation input variable range derived from the calculated energy was wider and the sensitivity was also higher. The reason for this phenomenon is that the Korean government has strengthened the standard around the performance standards related to heating. That is, because the time passed by the strong legal standards, there was no significant performance degradation in the heating sensitivity section, but there was a large difference in the cooling sensitivity because the legal standards related to the cooling performance were not applied strongly.

This suggests that it is necessary to calibration method the uncertainty variable input range of input variables reflecting social characteristics.

Table 5. Comparison table of statistical model

	***b*₀ (kWh/m²)**		***b*₁ (Htg Sensitivity)**		***b*₂ (Clg Sensitivity)**		***b*₃ (Htg Start Temp.)**		***b*₄ (Clg Start Temp.)**	
	Model I	**Model II**	**Model I**	**Model II**	**Model I**	**Model II**	**Model I**	**Model II**	**Model I**	**Model II**
Min	0.86	5.44	-1.58	-0.23	0.18	0.36	9.86	8.52	14.41	6.39
10%	6.56	7.83	-1.85	-0.94	0.40	0.48	13.29	11.28	16.25	12.49
25%	9.59	10.03	-2.20	-1.49	0.71	0.60	14.69	11.92	17.94	13.57
75%	13.59	12.31	-2.56	-1.91	1.45	0.75	16.19	12.96	19.64	14.36
Max	25.78	27.82	-3.10	-7.76	3.11	1.77	20.33	18.40	25.59	17.47

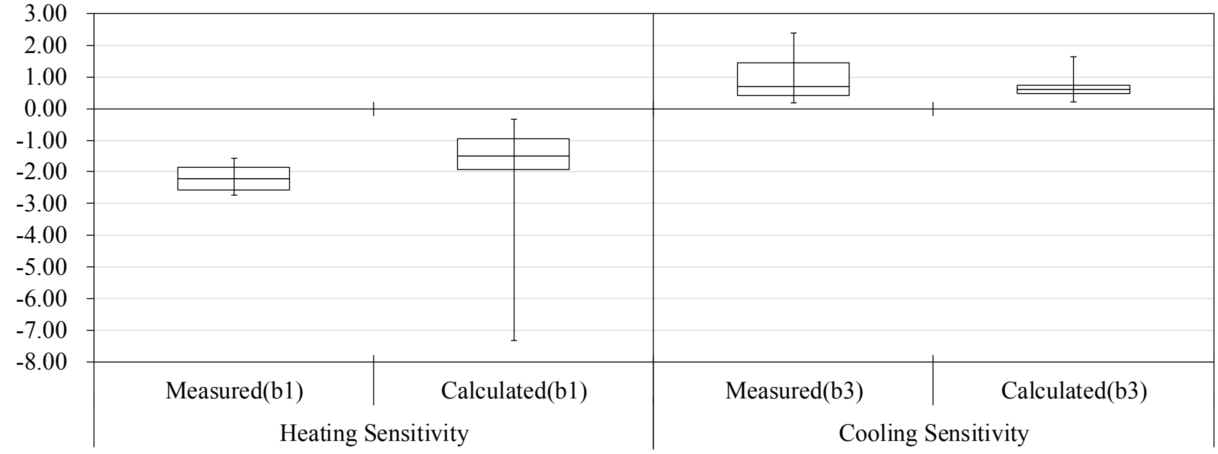


Figure 13. Comparison of heating and cooling sensitivity between measured and statistical model

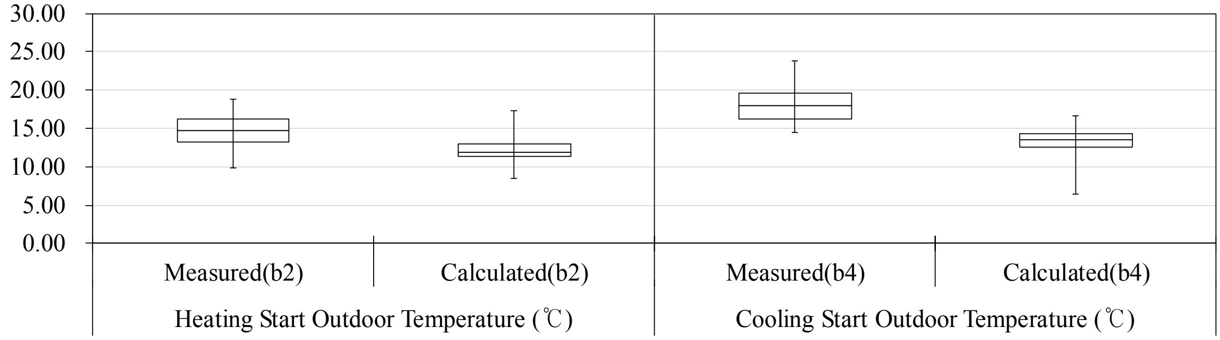


Figure 14. Comparison of heating and cooling sensitivity between measured and statistical model

CONCLUSION

This paper considered the method to provide a structured model for modeling uncertainties in building energy models. A random number generation model reflecting the distribution of actual building information was developed and energy performance was calculated by iterative simulation. Then regression equations with the same input-output variables and simulation input variables were derived so that same factors of calculated and measured energy performance could be compared. Finally, we analyzed the error due to the simulation input variable input of the uncertainty probability distribution by comparing the performance and discussed a method for improving the reliability of the parametric model. The summary of this paper as following:

(1) The actual data distribution analysis and results of comparing the energy performance show that the data do not follow the normal distribution

(2) It means that appropriate energy tracking methods should be provided by energy modeling that reflects the distribution of characteristics of real buildings.

(3) Simulation input variables which were reflected deterioration of buildings can be derived using the change point model.

This research is expected that can help to allow users to select the energy models that better represents actual data distributions by selecting probability distributions based on measured data is presented.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (NRF-2017R1A2B4012057) and the Korea Agency for Infrastructure Technology Advancement(KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant 18AUDP-B079104-0).

REFERENCES

T.Agami Reddy, Springer, Applied Data Analysis and Modeling for Energy Engineers and Scientists, ISBN 978-1-4419-9613-8, (2011)

Kyung et al., Energy performance assessment of existing buidings according to data availability, AIRAH and IBPSA's Australasian Building Simulation Conference, Melbourne, November 15-16, (2017)