Calibrating Dynamic Building Energy Models using Regression Model and Bayesian Analysis in Building Retrofit Projects

Wei Tian[1], Qiang Wang[1], Jitian Song[1], and Shen Wei[2]
[1] College of Mechanical Engineering, Tianjin University of Science and Technology, Tianjin 300222, China
[2]Building Performance Analysis Group, University of Plymouth, Plymouth, Devon PL4 8AA, United Kingdom

Abstract

In building retrofit analysis, the whole-building calibrated simulation approach can be used to calculate potential energy savings by developing the physical simulation model of a building. The most difficult part in this process is to calibrate simulation outputs of an energy model to measured energy data. A new methodology of combining statistical regression models and Bayesian analysis is introduced in this paper to infer the distributions of unknown inputs based on measured energy data and other available information on input data. This method consists of three steps: (1) create and run EnergyPlus models to construct an original data set; (2) construct full linear models (or other variations) to approximate the relationships between model inputs and simulated energy consumption; (3) perform Bayesian analysis to estimate unknown inputs based on regression models, available input information, and measured energy data. Two case studies indicate this method can provide fast and reliable inputs.

1 Introduction

The building sector accounts for about 35% of global final energy consumption in 2010 (IEA 2013). Therefore, the building sector plays a significant role in substantially reducing energy consumption and associated carbon emissions world widely. Another interesting point is that half of existing buildings will be still standing in 2050 (IEA 2013). Hence, more attentions should be focused on retrofitting existing building stock.

The calibrated simulation approach is an effective method to analyse the potential effects of various building energy efficiency measures for existing buildings (ASHRAE 2013; Fumo 2014). This method is to tune the inputs in a building energy model in order to achieve a close match between measured and modeled energy data. This procedure is often called calibration. Then this calibrated model can be used to estimate the energy savings for various retrofit measures. The calibration approaches can be either deterministic or probabilistic methods, depending on the method used in the analysis. In the field of building energy analysis, most existing studies are based on the deterministic approach (O'Neill et al. 2011; Pan et al. 2007), which is to obtain input parameters by using the manual trial-and-error method. Therefore, this deterministic method is very time-consuming and the results are more dependent on the analyst. In building energy models, many combinations of input values may result in the same energy use (ASHRAE 2013). However, this deterministic method can only find a limited number of possible solutions, so they may not represent the actual operation conditions in a building. Thus, the energy savings estimations based on these limited solutions may be unreliable. In contrast, the probabilistic approach can deal with this problem in an efficient and effective way by treating the unknown inputs as random variables with probability density functions (PDFs). These PDFs can quantify the input uncertainties in actual buildings. This probabilistic calibration method (also called Bayesian analysis) has great potentials in analysing energy consumption for building retrofit projects (Heo and Zavala 2012; Tian and Choudhary 2012).

Therefore, this study is to develop a method of combining the regression models and Bayesian analysis to estimate the distributions of unknown inputs in energy models for building retrofit projects. An office building and a retail building are used as case study buildings to demonstrate how to implement this method. The advantage of this method is that the approach is transparent and flexible. The regression method can not only have full linear model, but also have several variations, including transformation of inputs and outputs, which can be suitable for more complicated relationships between inputs and energy use in different buildings. The software packages used in this analysis are available in the public domain, which means the methods can be easily implemented in other building retrofit projects.

This paper is structured as follows. Section 2 introduces the methodology used in this study, including the creation of the building energy models, the implementation of the regression analysis, and the use of the Bayesian analysis. Then the results from regression analysis and Bayesian analysis are presented and discussed in Section 3. The conclusions from this study are described at the end of the paper (Section 4).

2 Methodology

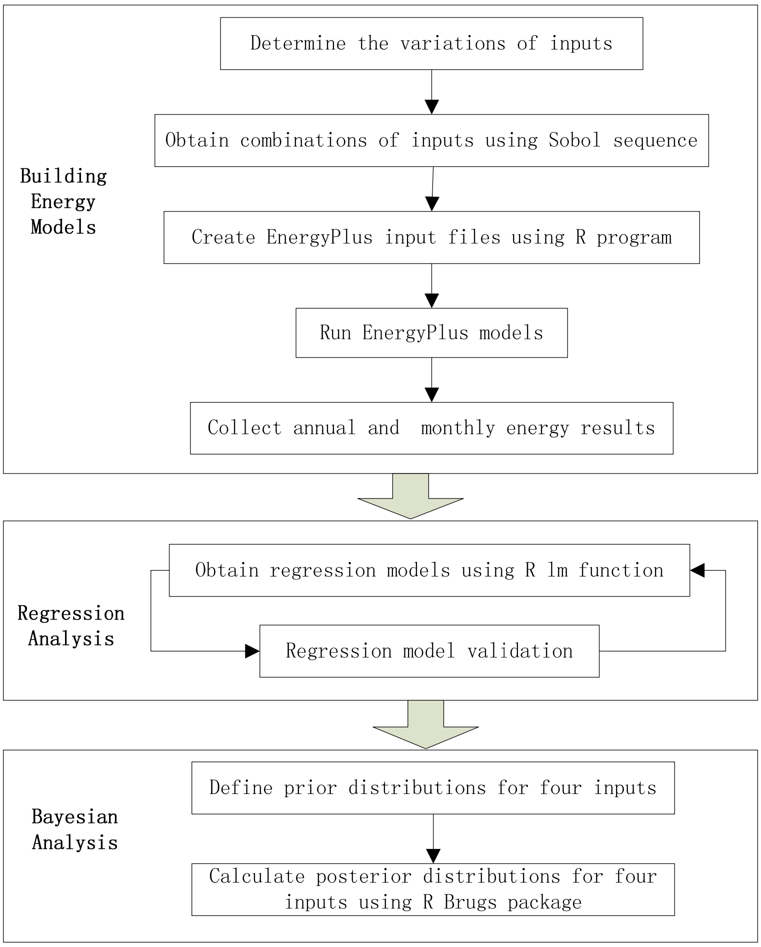


Figure 1: The flow chart for input parameter estimation using regression models and Bayesian analysis

This section is to describe the methodology (as shown in Figure 1) for estimating the input parameters for dynamic building energy models in building retrofit projects. The three main steps are: (1) creating building energy models, (2) applying regression models, (3) using Bayesian analysis.

Building energy models

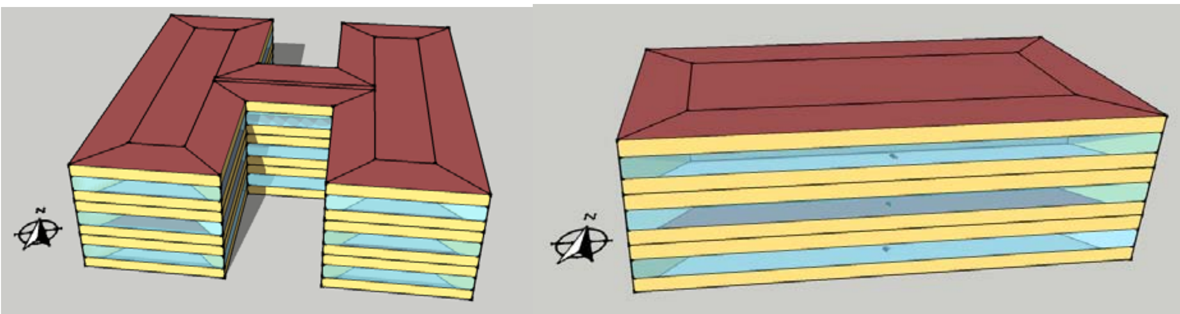
Figure 2 shows an office building and a retail building used in this analysis. Table 1 summarizes the main characteristics of these two buildings. Table 2 lists the parameter values of inputs, which can be changed in energy models and regarded as the possible ranges in these two buildings. The office building is a three-storey H-shape building with a total floor area of 3900 m^2 and the retail building is a three-storey rectangular building with a 2400 m^2 floor area. The parameters related to thermal performance in these two buildings are mainly derived from Chinese design standard for non-residential buildings (MOC 2005). These two buildings are assumed to be located in Tianjin, China and the corresponding weather data is downloaded from the EnergyPlus website (DOE 2013).

Table 1: Main characteristics of an office building and a retail building

Component	Item	Office	Retail
Envelope	Floor area (m2)	3900	2400
	Floor levels	3	3
	Window-wall ratio	0.40	0.40
	Zone number	45	15
	Wall U-value (W/m^2K)	0.36	0.36
	Roof U-value (W/m^2K)	0.28	0.28
	Window U-value and SHGC (solar heat gain coefficient)	See Table 2	
	Infiltration rate (air changes per hour)	0.4	0.4
Internal heat gains	Lighting power density (W/m2)	11	16
	Daylighting	no	Lights will be off when daylighting is above threshold 400 lux
	Equipment gain, occupant density	See Table 2	
	Hourly schedules for occupants, lights, and equipment	Office and retail buildings in China standard (MOC 2005)	
HVAC		Fan coil system with hot water and chilled water coils, heating provided by district heating system, cooling from centrifugal chiller and water-cooled cooling tower, ventilation requirements: $30m^3$/(h person) for office, $20m^3$/(h person) for retail	

The simulation is carried out using EnergyPlus program, developed by the Department of Energy, USA (DOE 2013). EnergyPlus has been widely used in the field of building energy simulation and has also been tested extensively. The advantage of using EnergyPlus is that the input data file (IDF) for EnergyPlus models is text format, which can be easily managed to edit or change for parametric analysis, since many EnergyPlus models are required for regression analysis in this research. The EnergyPlus models are automatically created using the R statistical program, so the building energy simulation is directly linked to the statistical analysis needed in this study. R program is a free software environment for statistical computing with many advanced, latest statistical functions (R Development Core Team 2013).

The combinations of inputs (as defined Table 2) are constructed using Sobol sequence, one of quasi-random low-discrepancy sequences (LDS). LDS (also called quasi-random numbers) are designed to create uninform sample points (Kucherenko et al. 2011). The Sobol sequence can meet three requirements: (1) best uniformity as the number of sampling iterations approaches infinity; (2) good distributions for small initial sets; (3) fast computation (Kucherenko et al. 2011). Therefore, this sampling method is used in this work to obtain the combinations of four parameters with the uniform distributions (Table 2). The sampling number is chosen as 40 based on the recommendation from Levy and Steinberg (2010). The extra 40 simulation runs are performed to validate the regression energy models created from the first 40 runs. In total, there are 80 EnergyPlus simulation models used in this study.



(a) office building (b) retail building

Figure 2: The 3-D EnergyPlus models for an office building and a retail building

Table 2: Variables used in regression analysis

Variable	Short names	Unit	Office	Retail
Window U-value	WinU	W/m^2K	1.5-3.5	1.5-3.5
Window SHGC (solar heat gain coefficient)	SHGC	-	0.3-0.7	0.3-0.7
Occupant density	Occupant	m^2/person	4-8	3-5
Equipment peak heat gain	Equipment	W/m^2	13-20	11-15

Regression model

After the annual and monthly energy simulation results are collected, the various types of regression models can be obtained, depending on the input number and model accuracy. Regression models can be divided into parametric and non-parametric models (Faraway 2006). The parametric models include linear and non-linear predetermined regression forms, whereas the non-parametric regression method is more flexible without a formulaic way of defining the relationship between inputs and outputs (Faraway 2006).

The analyst usually should choose full linear models in the field of building energy analysis for calibration problems in the first place. There are at least three reasons for this choice. First, the simple linear models are usually more robust and have less overfitting compared to non-parametric models (Qiao et al. 2009). Second, the study on calibration problems for building energy models is often focused on a limited number of input parameters, maybe less than ten or even five. Therefore, it is unlikely to have strong non-linear relationships between these input factors and energy use in a specific building. If there are more than ten parameters, the analyst should firstly run sensitivity analysis to choose the first four or five important factors influencing building energy use for calibration study. Actually, if it is assumed that there are large variations for over ten inputs in building energy analysis, it may not be a good choice to select the whole-building dynamic energy simulation for building retrofit analysis. Third, the linear models have many variations, including the transformation of both inputs and outputs. For example, the interaction between inputs and second-order terms may be added to increase the predictive capability if the full linear model has high prediction errors. Many transformations may be also used, such as log, inverse, square functions (Faraway 2005).

The accuracy of regression models is evaluated using two metrics: R2 and RMSE (root mean square error) in this research. R2 is the coefficient of determination to indicate the proportion of the total variations of response (building energy in this case) explained by the regression models. The higher R2, the more accurate the regression model is. RMSE is the root mean square error to indicate the average errors of regression models. This value has the same unit as the building energy consumption in this study. A lower RMSE indicates a better model. It should be emphasized that the R2 or RMSE from the training data set cannot be used to accurately assess the predictive performance of regression models for the future data set. The training errors consistently decrease with an increase in model complexity, possibly due to overfitting. However, the test error may be increase at some point. Therefore, the extra 40 simulation runs are carried out to test the predictive performance of regression models in this study.

Bayesian analysis

In this study, the four input parameters (listed in Table 2) will be estimated using annual and monthly energy use for heating and electricity. The model inputs will be calculated two times based on two data sets. One is annual heating and annual electricity data (i.e. two energy data only). The other is to include three monthly heating data and three monthly electricity energy data (January, February, and March). The purpose of choosing two data sets is to assess the sensitivity of estimated inputs with respect to the number of available energy data in building retrofit analysis.

To assess the accuracy of estimation method, it is necessary to know the true values of input parameters. In actual buildings, there are usually very high uncertainties on the values of these input parameters. Hence, this research is to evaluate the reliability of the estimation method by using the specified inputs and the energy consumption simulated from EnergyPlus models. Therefore, we assume that the energy data is available and the input parameter will be inferred using Bayesian computation. As a result, these inferred inputs can be compared to the true values in order to accurately assess the validity of the estimation method. Note that the 40 data sets used for creating regression models described in the last section cannot be used in this Bayesian analysis. This is because we will evaluate the accuracy of both the regression models and the Bayesian computation. In this research, we randomly choose one of the 40 extra data set that have been used to validate the regression models in the last section. This random data set is used for Bayesian computation.

In Bayesian statistical analysis, all the uncertainty should be regarded as probabilities (Christensen et al. 2011). In this case, the input parameters in building energy models can be modeled using probabilities, called prior and posterior distributions. The prior distributions of building input parameters can be obtained from previous documents or expert knowledge. Then the new posterior distributions will be computed based on the Bayes Theorem to combine the data and prior distributions (Cowles 2013),

$$p(\theta|y) = p(y|\theta)\, p(\theta) / p(y) \quad (1)$$

where $p(\theta|y)$ is the posterior distribution containing the updated knowledge, $p(y|\theta)$ is the likelihood, $p(\theta)$ is the prior distribution, $p(y)$ is the marginal likelihood. Therefore, the posterior probability is proportional to the prior probability times the likelihood. If there is a large data sample, then the influence of prior becomes smaller. In contrast, for a small data sample, the prior distribution tends to have a predominant effect on the posterior distribution. In the case of calibrating the building energy models, this means the calculated building inputs would more depend on the prior information (i.e. how much we know the building investigated in the project) when the number of available energy data is small due to missing data or other issues.

Modern Bayesian analysis is usually performed to obtain the posterior distributions using Markov chain Monte Carlo (MCMC) method. This is because the integrals cannot be evaluated using calculus in most of actual problems and the numerical approximations become necessary, especially MCMC simulation method. The MCMC method has become the main computational approach in Bayesian analysis because this method can draw samples from high-dimensional posterior densities (Congdon 2007). The shortcoming of this method, however, is that the samples may not be independent. Then it is crucial to conduct convergence diagnostics for posterior distributions in order to make appropriate inference. Note that there is no best approach to assessing the convergence of MCMC posterior distributions (Congdon 2007). Therefore, two methods are used in this study: (1) the two chains whether to converge to the same distributions; the Brooks–Gelman–Rubin (BGR) statistic (Cowles 2013). The two chains with diverse starting values are run for 20,000 iterations with a burn-in of 5,000 iterations in this research. Two posterior distributions from two chains are compared to ensure the convergence of the distributions. The BGR statistics is to test a ratio of parameter interval lengths of multivariate, which should be 1 for the convergence of distributions (Cowles 2013).

The Bayesian model in this research can be expressed as,

$$E_i \sim N(\mu_i, \sigma^2) \quad (2)$$

$$\mu_i = \beta0_i + \beta1_i WinU + \beta2_i SHGC + \beta3_i Occupant + \beta4_i Equipment \quad (3)$$

where E_i is the annual or monthly energy data, i is the nth month for heating or electricity, ~ means the stochastic node (or called random variable), N is the normal distribution, μ_i is the mean of energy use in the n^{th} month, σ^2 is the variance of distributions, β0 to β4 denote the regression coefficients from the regression models described in the last section "Regression model". WinU, SHGC, Occupant, and Equipment are four parameters that need to be estimated using Bayesian computation (for full names of these four input parameters, please see Table 2). The four prior distributions are defined as follows,

$$WinU \sim unif(1.5, 3.5) \quad (4)$$

$$SHGC \sim unif(0.3, 0.7) \quad (5)$$

$$Occupant \sim unif(4,8) \text{ for office building, } unif(3,5) \text{ for office building} \quad (6)$$

$$\text{Equip} \sim \text{unif}(13,20) \text{ for office building, } \text{unif}(11,15) \text{ for office building} \quad (7)$$

$$\sigma \sim \text{dunif}(0, 20) \quad (8)$$

where unif represents the continuous uniform distributions and the remaining terms are as explained above.

R BRugs package is used for Bayesian analysis in this research (Thomas et al. 2006). This package is an interface to the OpenBUGS program for Bayesian analysis using MCMC sampling. OpenBUGS software is an open source Bayesian analysis environment. The advantage of using BRugs package is to make full use of advanced statistical functions in R environment. For example, the non-standard semi-parametric regression can be implemented using this package (Marley and Wand 2010). Another important feature is to directly use R coda package for convergence diagnostic (Plummer et al. 2006).

3 Results and discussion

This section is to describe the results from both the office and retail buildings. Firstly, regression results are described for two buildings and, then, the inputs are estimated using Bayesian analysis.

Regression results

The regression results are summarized in Table 3 and Table 4. As can be seen from these two tables, the R2 for testing sets is close to 1, which indicates that the regression models are very accurate. The RMSE (root mean square error) for two test sets is very low, which also suggests the regression models are reliable to predict the energy consumption in these two buildings.

The regression performance for the office building is shown in Table 3. As might be expected, the test errors from both R2 and RMSE are slightly higher than those from the training set. However, the differences are small and, therefore, there is no significant overfitting for the regression equations. The R2 values from both training and test data sets are very close to 1. This means that these eight regression equations can be used to accurately estimate the energy consumption in the conditions, which are different from regression analysis. The RMSE values are very low, which also suggests the reliability of regression models.

The performance for the eight regression equations in the retail building is summarized in Table 4. All the R2 from eight test cases is above 0.95, which indicates that the models are accurate enough for the prediction of annual or monthly energy (heating and electricity). This conclusion can be also obtained from RMSE in this table.

As also can be seen from Table 3 and Table 4, the accuracy of regression models for electricity is higher than that for heating energy in the office building based on the R2 from testing data sets, whereas the opposite is true for the retail building. This difference may be due to the daylighting control (Table 1), which makes the electricity consumption in the retail building harder to accurately predict.

As discussed above, the sixteen regression equations obtained from the full linear models can accurately predict the energy consumption in these two buildings. Hence, the transformation of inputs and outputs (such as inverse, log, square, interaction terms) will not be implemented in this study. Note that the transformations or other variations based on linear models may be necessary in other cases to improve the predictive performance of regression models. If there is no significant improvement after simple transformation, more complicated non-parametric methods may be needed to model complex relationships between building inputs and energy consumption.

Table 3: Performance of the regression models for the office building

Energy	Time period	Training		Test	
		R2	RMSE	R2	RMSE
Electricity	Annual	0.998	0.549	0.996	0.922
	January	0.996	0.053	0.993	0.072
	February	0.999	0.025	0.998	0.041
	March	0.992	0.091	0.986	0.144
Heating	Annual	0.972	3.621	0.966	4.175
	January	0.973	0.893	0.967	1.020
	February	0.973	0.764	0.968	0.875
	March	0.972	0.415	0.964	0.498

Table 4: Performance of the regression models for the retail building

Energy	Time period	Training		Test	
		R2	RMSE	R2	RMSE
Electricity	Annual	0.985	0.991	0.978	1.245
	January	0.966	0.092	0.970	0.092
	February	0.951	0.098	0.950	0.105
	March	0.979	0.090	0.974	0.103
Heating	Annual	0.985	3.133	0.984	3.213
	January	0.985	0.753	0.984	0.773
	February	0.985	0.651	0.984	0.672
	March	0.984	0.385	0.984	0.391

Calibration results

As described in the "Methodology" section, the model inputs will be estimated using Bayesian computation in two data sets. One is that the annual heating and electricity are available, while the other is that three monthly electricity and three monthly heating energy data (January, February, and March) are available.

Figure 3 shows the estimated model inputs using annual heating and electricity data in the office building. Figure 4 illustrates the calculated inputs using six monthly energy data, including electricity and heating energy data in January, February, and March. The red vertical lines in Figures 3 and 4 denote the actual parameter values. The posterior means for occupant density and equipment heat gains are close to actual values, while the estimated window U-value and SHGC are very different from actual parameters as shown in Figure 3. After the six monthly energy data are used, the accuracy for four parameters has increase a lot and the posterior means are similar to actual values as shown in Figure 4. The variations of four estimated inputs are also decreased significantly due to the use of six energy data. Since the estimation in this study did not use all the monthly data, this means the methodology proposed here is also suitable for the situation when there is missing data. It is common that the analyst may not have all the energy data due to measurement errors or building maintenance. Hence, this flexibility is very useful for accurately estimating the inputs from incomplete energy data in buildings.

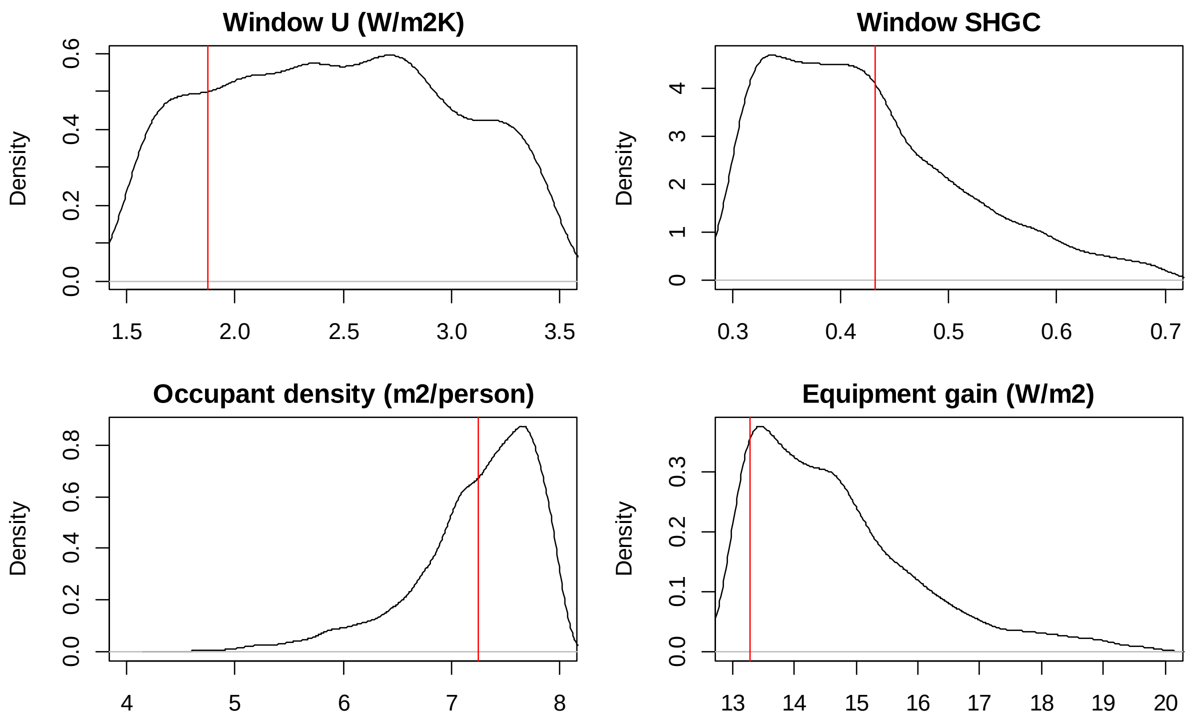


Figure 3: Posterior distributions of four parameters based on annual heating and electricity energy use in the office building

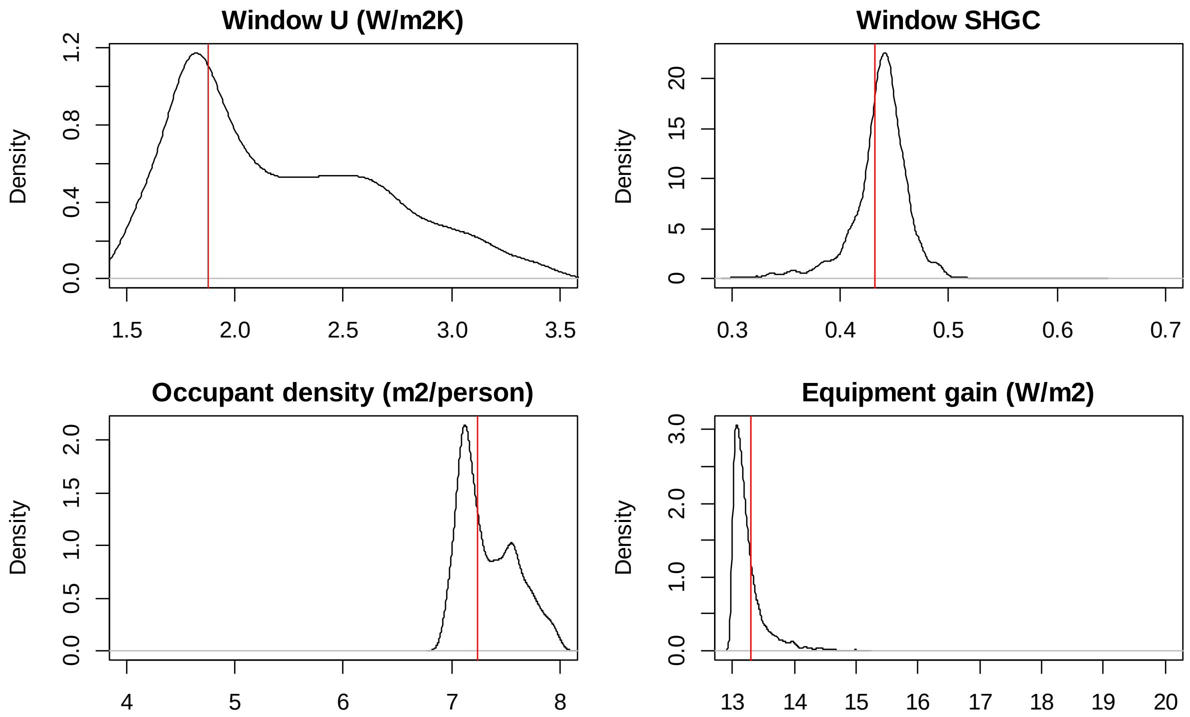


Figure 4: Posterior distributions of four parameters based on monthly heating and electricity energy use for January, February, and March in the office building

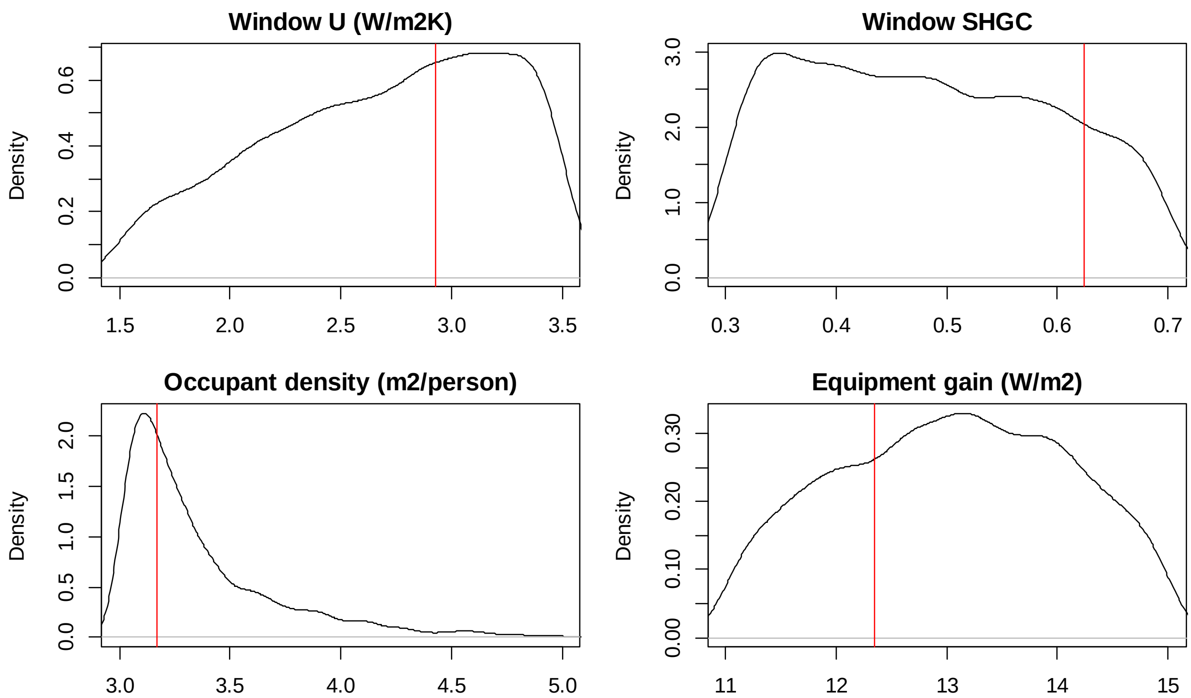


Figure 5: Posterior distributions of four parameters based on annual heating and electricity energy use in the retail building

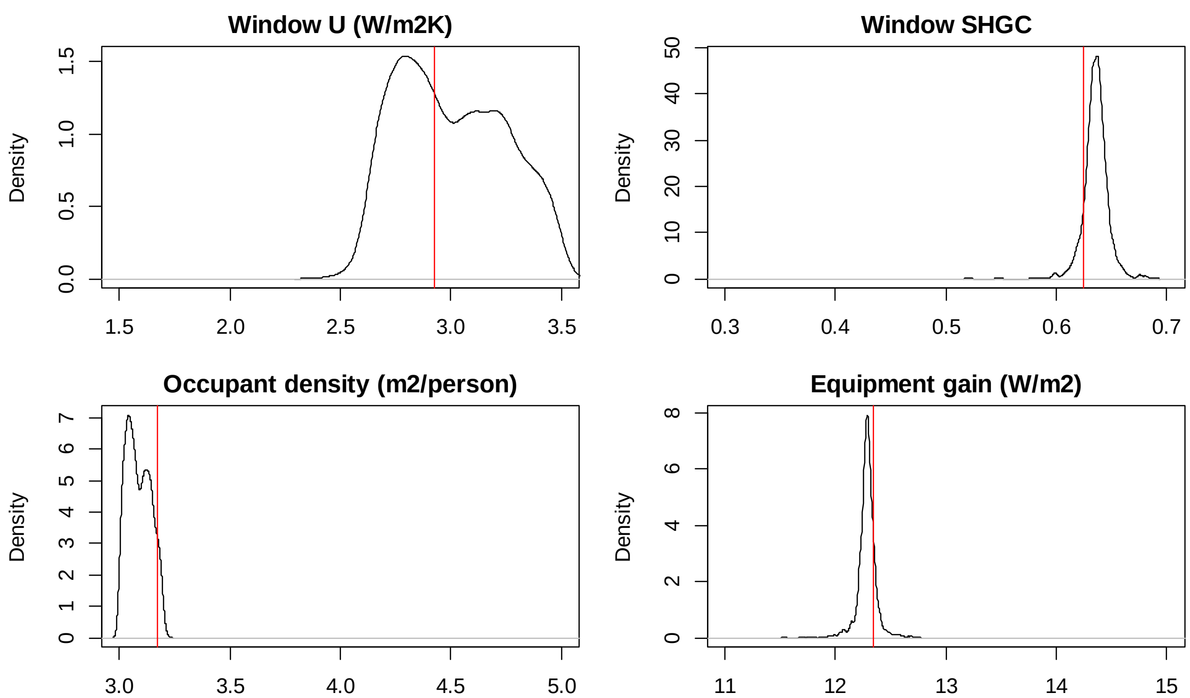


Figure 6: Posterior distributions of four parameters based on monthly heating and electricity energy use for January, February, and March in the retail building

Figure 5 and 6 show the posterior distributions of four parameters for the retail building from two annual and six monthly energy data, respectively. The red vertical lines in these two figures denote the true values. The results show that the estimated values from Bayesian analysis using six energy data are more reliable compared to those from the two annual values. The variations of the posterior distributions are also small as shown in Figure 6.

Convergence diagnostics

As discussed in the last section, the results from the six monthly energy data are more accurate than those from the two annual energy data. Therefore, this section is to apply convergence diagnostics to the results from the six energy data in both the office and retail buildings. Note that the convergence diagnostics of posteriors should be performed before Bayesian inference. In this study, however, the purpose of discussing the posterior inference in the first place is to emphasis the differences of estimated inputs between the two annual and the six monthly energy data.

Table 5: Summary of building parameter estimation from the two chains in the office building

Parameter	Chain number	Quantile 2.5%	25%	50%	75%	97.5%	Actual
Window U-value (W/m^2K)	1st	1.53	1.73	2.03	2.43	3.29	1.87
	2nd	1.55	1.89	2.32	2.78	3.35	
Window SHGC	1st	0.36	0.42	0.44	0.45	0.47	0.43
	2nd	0.37	0.43	0.44	0.46	0.48	
Occupant density (m^2/person)	1st	6.96	7.08	7.24	7.47	7.94	7.24
	2nd	6.96	7.17	7.40	7.66	7.96	
Equipment gain (W/m^2)	1st	13.01	13.08	13.17	13.33	14.11	13.28
	2nd	13.01	13.08	13.17	13.35	14.11	

Table 6: Summary of building parameter estimation from the two chains in the retail building

Parameter	Chain number	Quantile 2.5%	25%	50%	75%	97.5%	Actual
Window U-value (W/m^2K)	1st	2.63	2.87	3.12	3.31	3.48	2.93
	2nd	2.62	2.78	2.91	3.13	3.44	
Window SHGC	1st	0.61	0.63	0.64	0.64	0.66	0.62
	2nd	0.61	0.63	0.63	0.64	0.66	
Occupant density (m^2/person)	1st	3.01	3.06	3.11	3.15	3.19	3.17
	2nd	3.01	3.04	3.07	3.12	3.18	
Equipment gain (W/m^2)	1st	12.12	12.26	12.30	12.33	12.51	12.34
	2nd	12.10	12.25	12.29	12.32	12.50	

Table 5 shows the summary of posterior distributions for four parameters from two chains in the office building. The 50% quantiles (i.e. median) from two chains are very similar, which are both close to the true values listed in the last column in Table 5. The window

U-value is slightly different from actual value, while the other three estimated parameters are very accurate and the corresponding variations are also small. These differences may be due to the fact that that the window U-value is the least sensitive parameter influencing energy consumption in this case study. As a result, the estimated mean and the variations cannot be accurately calculated. Since the window U-value is the least important factor among four parameters affecting energy consumption, this variation do not have significant influences on the predicted energy use for building retrofit project. The Brooks–Gelman–Rubin (BGR) statistic is also below 1.1 for these four parameters, which indicates there is no significant non-convergence.

Table 6 lists the quantiles and the true values for four parameters in the retail building. The results are similar to Table 5 for the office building. Window U-value is the least accurate estimate because this factor is also the least important variable influencing energy consumption in the retail building. The accuracy of the remaining three parameters is very high as shown in Table 6. The relative errors for all these parameters are below 5%. The results from the two chains are very close, which suggests that the posterior distributions may converge. This can be confirmed from the Brooks–Gelman–Rubin (BGR) statistic of below 1.1 for two chains.

4 Conclusions

This paper describes how to apply the combination of regression models and Bayesian analysis for calibrating dynamic building energy models in building retrofit projects. Two case studies (an office building and a retail building) are used to demonstrate the methodology. The following conclusions can be drawn from this study.

(1) The combination of regression models and Bayesian analysis can have advantages in providing both complete and fast results compared to manually trail-and-error method in building retrofit analysis.

(2) The regression models can include both a full linear model and its variations (i.e. transformation of both inputs and outputs). If the predictive performance is still not acceptable for a specific building after transforming the linear models, non-parametric models can be used to improve the performance of regression models.

(3) The method proposed in this study can be used when there are missing energy data in buildings. The results can be still very accurate even only using six energy values in this study, containing three monthly electricity data and three monthly heating data.

Future research will explore the effects of three factors (prior, regression forms, and daily energy data) on the accuracy of posterior distributions of building parameters. The prior distributions may be specified more accurately (such as normal or triangle distributions) if the analyst can obtain more detailed information in the building investigated in the project. The variations of linear models can be used to check whether the results can be further improved using different regression models. If it is assumed that daily energy data are available, it will be very interesting to see whether the estimated inputs would be more accurate and how the calculation time would change accordingly.

5 References

ASHRAE. 2013. Handbook of Fundamentals, Atlanta: American Society of Heating, Air-Conditioning and Refrigeration Engineers: Inc.

Christensen, R., W. O. Johnson, A. J. Branscum, and T. E. Hanson. 2011. *Bayesian ideas and data analysis: An introduction for scientists and statisticians*: CRC Press.

Congdon, P. 2007. *Bayesian statistical modelling*. Vol. 704: Wiley.

Cowles, M. K. 2013. *Applied Bayesian Statistics: With R and OpenBUGS Examples*: Springer.
DOE. 2013. EnergyPlus V8.1, October 2013, Department of Energy, USA.
Faraway, J. J. 2005. *Linear models with R*. London: CRC Press.
Faraway, J. J. 2006. *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*. Vol. 66: Chapman & Hall.
Fumo, N. 2014. A review on the basics of building energy estimation. *Renewable and Sustainable Energy Reviews* 31 (0):53-60.
Heo, Y., and V. M. Zavala. 2012. Gaussian process modeling for measurement and verification of building energy savings. *Energy and Buildings* 53 (0):7-18.
IEA. 2013. Transition to Sustainable Buildings -- Strategies and Opportunities to 2050, International Energy Agency (IEA).
Kucherenko, S., D. Albrecht, and A. Saltelli. 2011. Comparison of Latin Hypercube and Quasi Monte Carlo Sampling Techniques. In *Eighth IMACS Seminar on Monte Carlo Methods August 29 – September 2, 2011, Borovets, Bulgaria*.
Levy, S., and D. M. Steinberg. 2010. Computer experiments: A review. *AStA Advances in Statistical Analysis* 94 (4):311-324.
Marley, J. K., and M. P. Wand. 2010. Non-standard semiparametric regression via BRugs. *Journal of Statistical Software* 37 (5):1-30.
MOC. 2005. *GB50189-2005. Energy Conservation Design Regulation for Public Buildings. Ministry of Construction (MOC) of P.R.China China Planning Press (2005) (in Chinese).*
O'Neill, Z., B. Eisenhower, S. Yuan, T. Bailey, S. Narayanan, and V. Fonoberov. 2011. Modeling and Calibration of Energy Models for a DoD Building. *ASHRAE Transactions* 117 (2).
Pan, Y., Z. Huang, and G. Wu. 2007. Calibrated building energy simulation and its application in a high-rise commercial building in Shanghai. *Energy and Buildings* 39 (6):651-657.
Plummer, M., N. Best, K. Cowles, and K. Vines. 2006. CODA: Convergence diagnosis and output analysis for MCMC. *R news* 6 (1):7-11.
Qiao, Z., L. Zhou, and J. Z. Huang. 2009. Sparse linear discriminant analysis with applications to high dimensional low sample size data. *International Journal of Applied Mathematics* 39 (1):48-60.
R Development Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. , URL http://www.R-project.org/.
Thomas, A., B. O'Hara, U. Ligges, and S. Sturtz. 2006. Making BUGS open. *R news* 6 (1):12-17.
Tian, W., and R. Choudhary. 2012. A probabilistic energy model for non-domestic building sectors applied to analysis of school buildings in greater London. *Energy and Buildings* 54:1-11.