Approximate Bayesian Computation in Parameter Estimation of Building Energy Models

ChuanQi Zhu, Wei Tian, Pieter de Wilde and Baoquan Yin

Abstract Model calibration is a necessary step to create reliable energy models in building retrofit. Bayesian computation in model calibration has attracted more attention because it can make full use of prior knowledge on building parameters. However, the likelihood function is hard to be computed in Bayesian computation due to the complexity of building energy simulation models. Approximate Bayesian computation (ABC) is a likelihood-free method to infer unknown parameters in complicated computational models by approximating the likelihood function with simulation. The ABC method is inherently computationally intensive since a large number of simulation runs are required to find reliable inferred values. This paper proposes a method for combining the ABC technique and the machine-learning method to compute unknown parameters in parameter estimation of building energy models. The results show that this method can provide reliable estimations of unknown parameters when calibrating building energy models.

Keywords Building energy simulation · Approximate bayesian computation · Parameter estimation · Model calibration

C. Zhu · W. Tian (✉)
College of Mechanical Engineering, Tianjin University of Science and Technology, Tianjin 300222, China
e-mail: tjtianjin@126.com

W. Tian
Tianjin International Joint Research and Development Center of Low-Carbon Green Process Equipment, Tianjin 300222, China

P. de Wilde
Chair of Building Performance Analysis, Environmental Building Group, University of Plymouth, Plymouth, Devon PL4 8AA, UK

B. Yin
Tianjin Architecture Design Institute, Tianjin 300074, China

Z. Wang et al. (eds.), *Proceedings of the 11th International Symposium on Heating, Ventilation and Air Conditioning (ISHVAC 2019)*, Environmental Science and Engineering, https://doi.org/10.1007/978-981-13-9528-4_40

1 Introduction

Building energy simulation is the application of physics-based building computer programs to model and predict building energy consumption [1]. Building energy simulation was originally used primarily in the design phase of the building, and now building energy simulations have been applied to all phases of the building life cycle. When performing building energy simulation in existing building retrofit, there is high uncertainty in a number of building variables [2]. This may lead to a large discrepancy between the result of the building energy modeling and the actual building energy consumption.

Building energy model calibration has become a necessary step for building retrofits to effectively evaluate energy savings. Input parameters in a simulation model are tuned to minimize discrepancies between prediction and observed data for building calibration of energy models. Coakley et al. classified building model calibration into two broad categories: manual and automated methods [3]. The main difference between manual and automated calibration is that the specific analytical computation to assist in the calibration process. The manual method mainly depends on the skill and experience by the modeler to adjust the building energy model [4]. In contrast, the automated method employs mathematical and statistical techniques to adjust building energy model. The manual calibration method is a time-consuming process to run simulation engines and calculate the discrepancy between simulation outputs and observed data. Then, input parameters are changed to minimize the error to meet the calibration standard. Manual calibration approaches would introduce the modeler biases into calibration model and not account for uncertainty of input parameters [4]. These shortcomings limit the application of manual approaches in building energy model calibration. Bayesian calibration, one of the automated approaches, can solve these problems by making full use of prior knowledge on uncertainty of input parameters [3]. The Bayesian calibration approach has been used in building energy analysis for calibration of unknown parameters, retrofit analysis, and calibration of sensor errors [5]. However, there are two issues with Bayesian calibration. First, the likelihood function is hard to compute in Bayesian computation due to the complexity of building energy simulation models. Second, the Bayesian calibration method is computationally intensive since a large number of simulation runs are required to find reliable inferred values.

Therefore, this paper presents a novel method that combines the approximate Bayesian computation (ABC) technique with machine-learning to compute unknown parameters in parameter estimation of building energy models. The contributions of this research are two-fold. One is to apply the ABC method in building model calibration to solve the difficulty of computing the likelihood function in Bayesian analysis. The other is to combine machine-learning algorithms with the ABC technique to significantly reduce computational cost of running engineering-based energy models of buildings when applying the ABC in building energy model calibration. Moreover, this study evaluates the suitability of posteriors correction to further improve the accuracy of ABC results in model calibration of building energy.

2 Methodology

2.1 Building Energy Model

Figure 1 shows the rectangular three-story office building studied in this paper. The total floor area is 4500 m^2 with the window-wall ratio of 0.5. The building is assumed to be located in Tianjin, China. The weather data is readily available from the EnergyPlus weather file database (CSWD) [6]. Table 1 shows the main features of the office building. Table 2 shows the unknown input parameter and their possible ranges. The unknown input parameter range was set based on the energy standards of public buildings in China [7]. Specifically, the chosen parameters are equipment power density, occupancy density, infiltration rate, and exterior wall U-value. These parameters have a great influence on building energy consumption, but it is difficult to measure these parameters. Each parameter is set to a uniform distribution with the same probability in their ranges. The thermal properties of the building envelope are based on the energy standards of public buildings in China. Detailed hourly schedules for internal heat gains (occupants, lighting, and equipment) are also derived from this China energy efficiency standard. A fan-coil system is used to provide heating, cooling, and ventilation for this building. In this building, the gas energy consumption is mainly used by the boiler to provide heating for the building. Therefore, the gas data form five months (January, February, March, November, and December) and all twelve months electricity data were used for calibration.

The EnergyPlus V9.0, developed by the US Department of Energy, is used as a simulation engine to compute building energy consumption in this paper. The advantage of EnergyPlus program is that its input data files (IDF files) are ASCII file. This is convenient when modifying the IDF files through the computer language, such as R, MATLAB. For the ABC method, it is necessary to run many building energy models, which requires automation and programming to create thousands of energy models automatically.

Fig. 1 An office building used in building

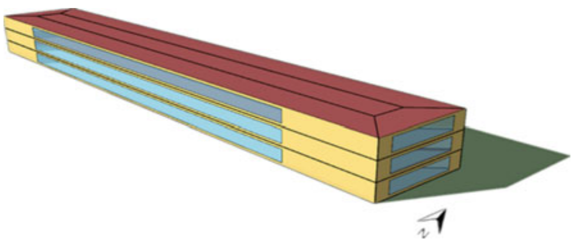

Table 1 Main features of the office building

Component	Parameters	Value	Unit
Envelope	Floor area	4500	m^2
	Floor level	3	–
	Window-wall ratio	0.5	–
	Thermal zone	Core zone with four perimeter zones on each floor	–
	Roof U-value	0.25	W/m^2 K
	Window U-value	2.2	W/m^2 K
	Exterior wall U-vale	See Table 2	W/m^2 K
	SHGC (solar heat gain coefficient)	0.4	–
	Infiltration rate(Air change per hours)	See Table 2	ACH
Internal heat gains	Lighting power density	9	W/m^2 K
	Equipment power density	See Table 2	W/m^2 K
	Occupancy	See Table 2	m^2/person
	Hourly schedules for set-point for heating and cooling, occupants, lights, and equipment	Design standard for energy efficiency of public buildings [8]	–
HVAC systems	Fan-coil system with boiler and chiller		–

Table 2 Unknown input parameter and ranges

Parameters	Range	Unit
Exterior wall U-value	0.3–0.5	W/m^2 K
Infiltration rate	0.3–0.6	ACH
Equipment power density	10–15	W/m^2
Occupancy density	6–10	W/m^2

2.2 *Machine-Learning Models*

In this paper, the EnergyPlus V9.0 is used to compute building energy consumption. However, the use of these physical models to calculate energy consumption is a time-consuming process, especially when there a great number of physical models are explored. In order to solve this problem, there has been increasing interest in applying machine-learning method to construct statistical energy models (also named as meta-models). The machine-learning method utilizes input parameters and energy simulation output to create meta-models that can reduce computation time.

The following five machine-learning methods are used to create meta-models: linear model (LM), support vector machine (SVM), multivariate adaptive regression splines (MARS), bagging multivariate adaptive regression splines (BMARS), and random forests (RF). The R caret package, developed by Max Kuhn, is used to develop these five meta-models [8]. These machine-learning methods are described briefly below. More complete description and theoretical frameworks can be found in [8]. In brief, the LM method uses linear regression to create a linear model between input and output and is the simplest model in these methods. SVM regression is a non-parametric technique because it relies on kernel functions. The MARS is also a non-parametric regression technique and can be seen as an extension of linear models that considers non-linear and interaction terms. The bagging method employs bootstrap aggregation, a general approach to combine a number of models and obtain the averaged predictions for these models. In this analysis, the bagging method is used together with a MARS regression, named as BMARS. The RF is an ensemble learning method for regression that operates by constructing a multitude of decision trees at training time and outputting the mean prediction of the individual trees.

Latin hypercube sampling (LHS) is used in this study to obtain a matrix with 1000 input combinations by sampling the unknown parameter ranges in Table 2 [9]. The R statistical software is used to create 1000 models automatically by using the parameters from the LHS method. The root-mean-square error (RMSE) and the coefficient of determination (R^2) are used as performance measures to choose the meta-models that can achieve the balance between model accuracy and computational cost. The meta-model with the best performance measures is applied to the ABC technique. The NMBE (normalized mean bias error) and CV(RMSE) (coefficient of variation of the root-mean-square error) indicators are used to measure the accuracy of model calibration as per recommendations of ASHRAE Guideline 14-2014 [10].

2.3 Approximate Bayesian Calibration

Bayesian analysis is a statistical method that utilizes the Bayes' algorithm in Eq. (1) to obtain a posterior distribution for the unknown parameter (θ) given the observed data (y) [11]. In this algorithm, $p(\theta)$ is the prior distribution assumed for unknown parameters; $p(y|\theta)$ is a likelihood function; $p(y)$ is the marginal likelihood; $p(\theta|y)$ is the posterior distribution of calibration parameters.

$$p(\theta|y) = \frac{p(y|\theta) \cdot p(\theta)}{p(y)} \propto p(y|\theta) \cdot p(\theta) \tag{1}$$

However, the likelihood function is hard to compute in Bayesian computation due to the complexity of building energy simulation models. Here, we address the issue by using ABC that is well-suited to the complex problems for which the likelihood is either sophisticated or computationally hard to obtain [12]. ABC is a likelihood-free method, widely used for demographic inference in population genetics. In ABC, a set of input variables (θ_i) is sampled from the prior distribution. The input combinations are used to run computer models (such as EnergyPlus models in this case) to obtain the output (y_i). The tolerance error is taken as a value δ. If the discrepancy between target and observed (y) output is less than the tolerance, the input variables are retained. Otherwise, the input variables are discarded. The received variables are considered to have been sampled from the posterior distribution [13]. For this basic ABC algorithm, the accepted variables form the approximate posterior distribution defined by Eq. (2). Compared to the Bayesian algorithm, the likelihood function is replaced by $p(y|\theta) \approx p(||y - y_i|| \leq \delta|\theta)$. This ABC method is called a rejection algorithm.

$$p(\theta|y) \propto p(||y - y_i|| \leq \delta|\theta) \cdot p(\theta) \tag{2}$$

To reduce the computational cost of ABC, two post-simulation approaches (local-linear ridge regression and neural networks) are used for correcting the imperfect match between observed and accepted outputs. The ridge regression assumes a linear function for the purpose of alleviating multicollinearity, while the NN (neural networks) considers a more flexible non-linear correction to reduce the variance of posterior estimations. More comprehensive descriptions and theoretical fundamentals for these methods can be found in Blum and François [14]. The *R* abc package is used to apply three ABC methods in this study [15].

3 Results and Discussion

3.1 Performance of Machine-Learning Models

Figure 2 presents the RMSE and R^2 of the 12 months electricity consumptions and 5 months gas consumption meta-models from internal cross validation. The five-month heating data is selected since most of heating energy occurs in these five months. E01 denotes the electricity use in January and the same description is applied for the electricity in other 11 months. G01 denotes the gas use in January and the same description is applied for the gas use in other four months. Among the five machine-learning methods, the meta-model generated by BMARS is the most accurate model in terms of both RMSE and R^2. The second most accurate model is the MARS model. Table 3 compares the computational time for creating these machine-learning models. The BMARS is the most time-consuming model. In order to maintain a balance between computational cost and model accuracy, the MARS meta-model is chosen for the model calibration with the ABC method.

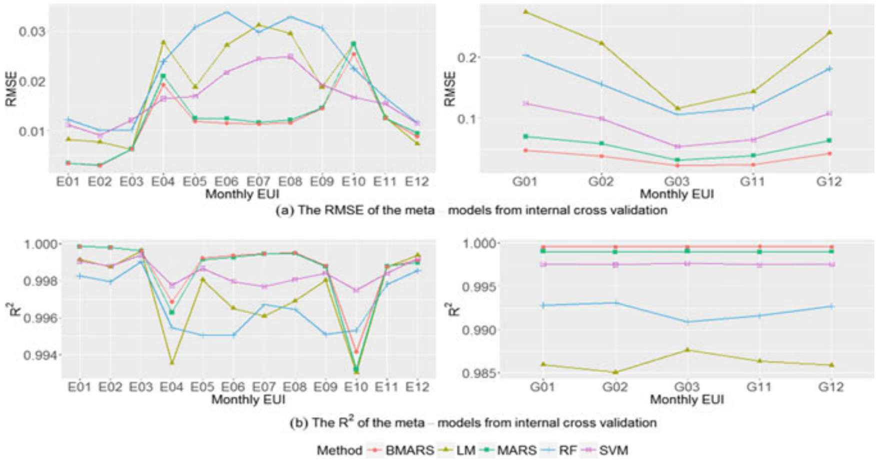

Fig. 2 RMSE and R^2 of the meta-models from cross validation

Table 3 Computational time of constructing five meta-models

Meta-models	LM	RF	MARS	BMARS	SVM
Computation times (second)	17	673	349	5228	460

3.2 *Calibration Results from Approximate Bayesian Calibration*

The posterior distributions of four unknown parameters are presented in Fig. 3. The dotted black lines are the prior uniform distributions, and the black vertical lines indicate the true values of the target building. The solid lines with three different

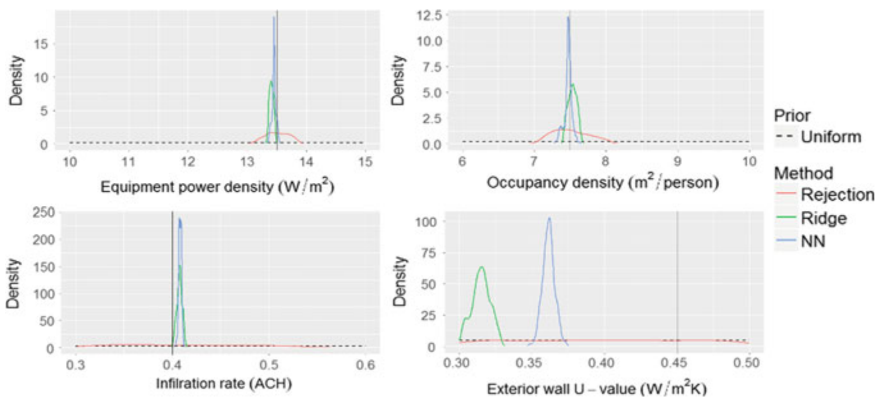

Fig. 3 Posterior distributions for four unknown input parameters from three ABC methods

Table 4 CV(RMSE) and NMBE from three ABC methods

Method	Electricity		Gas	
	CVRMSE [%]	NMBE [%]	CVRMSE [%]	NMBE [%]
Rejection	1.256	0.378	2.366	1.183
Ridge	1.813	0.547	2.596	1.298
NN(neural network)	0.901	0.272	0.950	0.475

colors (red, green, and blue) represent the posterior distributions from three ABC methods. If the posterior distribution of unknown parameters is closer to the corresponding vertical lines, the results of approximate Bayesian calibration are more accurate. The calibrated results from the neural networks method perform the best among these three methods. Neural networks can obviously shrink the range of the unknown parameters better than other two methods. Although infiltration rate and exterior wall U-value were not as important as the other two parameters, neural networks and local-linear ridge regression still can obtain an accurate estimation. The posterior distribution from the rejection method was closer to the prior distribution.

3.3 Evaluation of Accuracy of Model Calibration

ASHRAE Guideline 14-2014 states that the NMBE is less than 5% and the CV(RMSE) is less than 15% for monthly data for a calibrated building energy model. CV(RMSE) can be considered to represent the percent error between the simulation and measured data. NMBE indicates a bias percentage for undershooting (NMBE > 0) or overshooting (NMBE < 0) the actual data during the period of evaluation. Table 4 shows the CV(RMSE) and NMBE of three ABC methods. From Table 4, the CV(RMSES) and NMBE values are much smaller than the requirements of ASHRAE Guideline 14-2014. Hence, the calibration process of using the ABC method in this research can obtain an accurate parameter estimation.

4 Conclusion

This paper implements a method of combining the ABC technique and the machine-learning method to compute unknown parameters in building energy models. The results show that the ABC method can be used for building energy model calibration and these methods can solve the likelihood function problem of using Bayesian calibration. The meta-model of MARS employed in this research provides a good balance between the computational cost and the accuracy of both parameter estimation and energy prediction. From the distributions of unknown parameters, the neural networks can obtain better accurate posterior distribution estimation among three ABC methods. The accuracy of model calibration from three ABC methods can meet the criterion of ASHRAE Guideline 14-2014.

Acknowledgements This research was supported by the National Natural Science Foundation of China (No. 51778416) and the Key Projects of Philosophy and Social Sciences Research, Ministry of Education (China) (contract No. 16JZDH014, approval No. 16JZD014).

References

1. Chong, A., Menberg, K.: Guidelines for the Bayesian calibration of building energy models. Energy Build. **174**, 527–547 (2018)
2. Tian, W., Heo, Y., de Wilde, P.: A review of uncertainty analysis in building energy assessment. Renew. Sustain. Energy Rev. **93**, 285–301 (2018)
3. Coakley, D.: A review of methods to match building energy simulation models to measured data. Renew. Sustain. Energy Rev. **37**, 123–141 (2014)
4. Chaudhary, G., New, J., Sanyal, J., Im, P.: Evaluation of "Autotune" calibration against manual calibration of building energy models. Appl. Energy **182**, 115–134 (2016)
5. Lim, H., Zhai, Z.: Influences of energy data on Bayesian calibration of building energy model. Appl. Energy **231**, 686–698 (2018)
6. Bureau, C.M.: China Standard Weather Data for Analyzing Building Thermal Conditions. China Building Industry Publishing House Beijing, China (2005)
7. MOC.: Design Standard for Energy Efficiency of Public Buildings, China Architecture and Building Press (2015)
8. Kuhn, M., Johnson, K.: Applied Predictive Modeling, Springer (2013)
9. Tian, W.: A review of sensitivity analysis methods in building energy analysis. Renew. Sustain. Energy Rev. **20**, 411–419 (2013)
10. Guideline, A.: Guideline 14-2014, Measurement of Energy, Demand, and Water Savings (2014)
11. Tian, W., et al.: Identifying informative energy data in Bayesian calibration of building energy models. Energy Build. **119**, 363–376 (2016)
12. Sisson, S.A., Fan, Y., Beaumont, M.: Handbook of Approximate Bayesian Computation, Chapman and Hall/CRC (2018)
13. Sunnåker, M., Busetto, A.G., Numminen, E., Corander, J., Foll, M.: Approximate Bayesian computation. PLoS Comput. Biol. **9**(1), e1002803 (2013)
14. Blum, M.G.B., François, O.: Non-linear regression models for approximate Bayesian computation. Stat. Comput. **20**(1), 63–73 (2010)
15. Csilléry, K., François, O., Blum, M.G.: abc: an R package for approximate Bayesian computation (ABC). Methods Ecol. Evol. **3**(3), 475–479 (2012)