Applied Energy 231 (2018) 686–698

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Influences of energy data on Bayesian calibration of building energy model

Hyunwoo Lim, Zhiqiang (John) Zhai*

Department of Civil, Environmental, and Architectural Engineering, University of Colorado Boulder, Boulder, CO 80309-0428, USA

HIGHLIGHTS

- Statistical classification methods can categorize building energy data.
- Energy data from different groups provide more information than those from the same group.
- Bayesian calculation time and accuracy are distinct for different selections of the data.
- Using informative data keeps similar accuracy reducing 44% in computing time compared to the use of all data.

ARTICLE INFO

Keywords:
Building energy model
Bayesian calibration
Utility data
Informative energy data
Sensitivity analysis

ABSTRACT

Every building has different (and fuzzy) characteristics and contains complex sub-systems that affect each other. Therefore, significant uncertainties exist when modeling an entire building as a system. Calibration is necessary and able to reduce many sources of these uncertainties. Bayesian calibration is one of the automatic calibration methods that has been utilized in various applications. However, few researches were found that investigated the influences of quality and quantity of measured data used for the calibration. Moreover, Bayesian calibration requires considerable computing cost due to the inherent iteration attribute. This paper proposes the use of informative data to produce more accurate Bayesian calibration with reduced computing time. The measured energy data are classified by statistical classification methods. Using different energy measurement data, the study compares and analyzes the calibration outcomes with three criteria: input parameter estimation accuracy, energy use prediction accuracy, and overall computing time. The results show that the calculation time and the accuracy of the calibration are distinct for different selections of the data for calibration. Proper data should be used in comprehensive consideration of purpose, computing time and accuracy of calibration. Using informative data for calibration is able to keep similar accuracy but with 44% reduction in computing time compared to the use of all data.

1. Introduction

Building energy simulation tools have been widely used for energy use estimation, optimal design and control, and energy conservation analysis of buildings. Differences between simulation results and the actual measured data of a building are inevitably due to the complexity and uncertainty associated with the simulation tools. Uncertainty is caused by insufficient building information, simplification of modeling processes, and the behavior of building occupants. Model calibration techniques have been used to reduce differences between simulated and measured data. Calibration tunes input parameters of the simulation model to minimize discrepancies. Among various calibration techniques, Bayesian calibration has received increasing attention in the building energy model domain due to its expandability and accuracy. Bayesian calibration estimates the probability distribution of unknown input variables and provides posterior distributions based on the observed data.

Lim and Zhai [1] compared various Bayesian calibration techniques used in building energy simulation. The Bayesian calibration framework proposed by Kennedy and O'Hagan has been used for various purposes in the field of building energy modeling: calibration of unknown input variables [2–7], retrofit analysis [2,3,8,9], comparison with traditional calibration methods [10,11], use of simplified models [10,12], influence of input data on uncertainty [13], sensor calibration [14,15], estimation of system parameters [16] energy meter calibration [17] and prediction of building stock energy use [18–23].

Despite the benefits of Bayesian calibration, such as accounting for uncertainty and improving model accuracy, Bayesian calibration is time-consuming. Simplified energy models and meta-models can be adopted to reduce the calculation time of Bayesian calibration. These

* Corresponding author.
E-mail address: john.zhai@colorado.edu (Z.J. Zhai).

https://doi.org/10.1016/j.apenergy.2018.09.156
Received 8 June 2018; Received in revised form 12 September 2018; Accepted 17 September 2018
0306-2619/

Nomenclature

Acronyms, abbreviations

ACH	air changes per hour
AE	annual electricity energy use
AG	annual gas energy use
AT	annual total energy use
CVRMSE	coefficient of variation of root-mean-square error
EG	electricity and gas energy use
E	electricity energy use
G	gas energy use
MCMC	Markov chain Monte Carlo
MLR	multiple linear regression
RMSE	root-mean-square error
SVI	sensitivity value index
SRC	standardized regression coefficient
T	total energy use (sum of electricity and gas energy use)

Subscripts

i	parameters
j	sensitivity methods

Symbols

R^2	coefficient of determination
θ	calibration parameter
μ	mean
$\bar{y}$	mean of observed value
n	number of observations
k	number of sensitivity methods
m	number of target outputs
y	observed value
r	Pearson correlation coefficient
$\hat{y}$	predicted variable value
σ	standard deviation
l	target output
V	value from the sensitivity analysis

simplified models, such as resistance-conduction models, utilize relatively simple equations to express and represent buildings and associated sub-systems. Meta-models are based on input and output data from the original model or actual measurements. Lim and Zhai [24] compared five meta-models to determine the impact of meta-model accuracy on Bayesian calibration. They found that all five meta-models significantly reduced computation time as compared to the original model. Nagpal et al. [25] demonstrated that a meta-model is faster than traditional approaches while providing a sufficiently accurate estimation.

Furthermore, Bayesian calibration requires appropriate measured energy data to calibrate the original model. Previous studies have used as much energy data as possible for Bayesian calibration, as shown in Table 1. However, in the practical application of calibration, it is often impossible to obtain all of the data that is needed due to privacy issues, sensor errors, missing data, and related factors. In addition, even if detailed data is accessible, large amounts of data can take a significant amount of time to collect and analyze.

There has been a lack of discussion regarding the quantity and quality of building energy data needed for Bayesian calibration. Research is needed to define the amount of data required for accurate Bayesian calibration, and how to distinguish usable and useless data. It is also necessary to examine how the quantity and quality of the energy data impact the accuracy of Bayesian calibration.

Tian et al. [26] indicated this issue and identified informative data in Bayesian calibration using correlation analysis and hierarchical clustering methods. The research described in this paper was inspired by the work of Tian and argues that building energy data should be selectively used to increase the accuracy of Bayesian calibration and to reduce computational time. This paper aims at determining the informative energy data using statistical analysis. Furthermore, it analyzes the effect of informative and uninformative energy data on Bayesian calibration through case studies. The case studies were compared using three criteria: simulation time, the coefficient of variation of root-mean-square error (CVRMSE) to true input parameter values, and CVRMSE to observed outputs (monthly and annual energy use intensity).

2. Methodology

The overall methodology and calibration process are shown in Fig. 1. The first step is to build energy models based on known building information, such as building geometry, envelope, and HVAC systems. The second step is parameter screening using sensitivity analysis. The uncertainty of the building model is defined by choosing unknown parameters and associated probabilities. A set of input variables is sampled from a defined distribution of unknown parameters. The input combinations are fed into the building energy simulation program to

Table 1
Building energy data used in the Bayesian calibration.

Author	Year	Calibration method	Calibration target	Resolution	Data type
Heo [8,27]	2011, 2012	Bayesian	Individual building	12 months	Gas (heating) energy
Tian and Choudhary [18]	2012	Inverse problem/ Bayesian	Building stock	Annual	Gas (heating) energy
Booth et al. [20]	2012	Bayesian	Building stock	61 days	Electricity energy
Zhao et al. [28,29]	2012, 2016	Inverse problem	Building stock	Annual	Total energy
Kim et al. [12]	2013	Bayesian	Individual building	Annual	Heating and cooling energy
Manfren et al. [9]	2013	Bayesian	Individual building	12 months	Electricity and gas energy
Heo et al. [13]	2015	Bayesian	Individual building	12 months	Total energy
Kang and Krarti [6]	2016	Bayesian	Individual building	12 months	Electricity and gas energy
Lim and Zhai [24]	2017	Bayesian	Individual building	12 months for electricity and 5 months for gas	Electricity and gas energy
Kristensen et al. [30]	2017	Bayesian	Individual building	Six-hourly, daily, weekly, and monthly for 1 year	District heating energy
Sokol et al. [31]	2017	Bayesian	Building stock	12 months	Electricity and gas energy
Nagpal et al. [25]	2018	Bayesian	Individual building	12 months	Electricity, steam, and chilled water energy