Renewable and Sustainable Energy Reviews 143 (2021) 110930

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: http://www.elsevier.com/locate/rser

Review on building energy model calibration by Bayesian inference

D. Hou [a], I.G. Hassan [b], L. Wang [a,*]

[a] Centre for Zero Energy Building Studies, Department of Building, Civil and Environmental Engineering, Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, H3G 1M8, Canada

[b] Mechanical Engineering Program, Texas A&M University at Qatar, Engineering Building, Education City, Al Rayyan, P.O. Box 23874, Doha, Qatar

ARTICLE INFO

Keywords:
Bayesian
Uncertainty
Calibration
Stochastic modeling
Building energy
Markov Chain Monte Carlo

ABSTRACT

A building energy model (BEM) is essential for understanding building energy consumption, evaluating energy-saving measures, and developing associated codes, standards, and policies. The calibration of BEM helps to ensure the accuracy of the model, whereas it remains a challenge. Conventional manual or automated methods are mostly deterministic and neglect the inherent uncertainties of BEM. In comparison, the recent development of the stochastic BEM calibration based on Bayesian inference has gained attention, whereas many are baffled by its underlying theory, strengths, limitations, and implementations. There are also various mathematical models and tools in the literature, making it hard for selection. This paper aims to unravel the myths about the Bayesian inference and critically review various implementation options with a series of model selections suggested so that a user would be able to employ the Bayesian inference calibration at the end of the paper. We also hope that the review contributes to facilitating a broader implementation of the method for BEM calibrations. First, an overview is summarized for the current status and development of Bayesian inference calibration in building energy modeling. Second, the theory and methodology of model calibration, Bayesian statistics, and Markov Chain Monte Carlo are illustrated. Third, the implementation of Bayesian inference is described, including several practical issues such as BEM determination, unknown calibration parameters number, their ranges and distributions, Meta-model selections, and programming languages based on the statistical package R. The review ends with conclusions and future work identified.

1. Introduction

In a recent report of the International Energy Outlook by the U.S. Energy Information Administration [1], the gross domestic product (GDP) between 2018 and 2050 is expected to grow 1.5%/year in the countries of the OECD and 3.8%/year in non-OECD countries. Meanwhile, the world energy consumption is anticipated to rise by nearly 50% by 2050, and worldwide energy-related CO_2 emissions grow at an annual average rate of 0.6% for the same period. Although the industrial sector is still the largest energy consumer, the building sector energy consumption has increased drastically over the past decades as a result of rapid population growth and urbanization process, higher requirements for indoor air quality and comfort, more indoor time, more diversified building functions, and global climate change. The combined

Abbreviations: ASHRAE, American Society of Heating, Refrigerating and Air-Conditioning Engineers; ANN, Artificial Neural Network; BEM, Building Energy Model; BMARS, Bagging Multivariate Adaptive Regression Splines; CDEM, Community Domestic Energy Model; CRPS, Continuous Rank Probability Score; CVRMSE, Coefficient of Variance of Root Mean Squared Error; DOE, Department of Energy; ECM, Energy Conservation Measure; EPC, Energy Performance Coefficient Calculator; EPSCT, Energy Performance Standard Calculation Toolkit; ESCO, Energy Service Company; EST, Energy Savings Trust; GA, Genetic Algorithm; GP, Gaussian process; GSHPS, Ground Source Heat Pump System; HMC, Hamiltonian Monte Carlo; HPS, Heat Pump System; IES-VE, Integrated Environmental Solutions Virtual Environment; INLA, Integrated Nested Laplace Approximation; IPMVP, International Performance Measurement and Verification Protocol; LM, Linear Model; MAP, Maximum A Posteriori; MARS, Multivariate Adaptive Regression Splines; MBE, Mean Bias Error; MC, Monte Carlo; MCMC, Markov Chain Monte Carlo; MLR, Multiple Linear Regression; M&V, Measurement and Verification; NMEB, Normalized Mean Bias Error; NN, Neural Network; NST, Nash-Sutcliffe Efficiency Coefficient; NUTS, No-U-Turn Sampler; OECD, Organization of Economic Cooperation and Development; PAM, Posterior Approximation Method; PDF, Probability Density Function; RF, Random Forest; RFVI, Random Forest Variable Importance; RMSE, Root Mean Square Error; SRC, Standardized Regression Coefficient; SUSDEM, Stochastic Urban-Scale Domestic Energy Model; SVI, Sensitivity value index; SVM, Support Vector Machine; UBEM, Urban Building Energy Modeling; VIC, Virtual In-Situ Sensor Calibration.

* Corresponding author. Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, H3G 1M8, Canada.
E-mail addresses: leon.wang@concordia.ca, lzwang@gmail.com (L. Wang).

https://doi.org/10.1016/j.rser.2021.110930
Received 15 April 2020; Received in revised form 16 February 2021; Accepted 1 March 2021
Available online 12 March 2021

energy consumptions from the residential and commercial sections worldwide will increase from around 20% in 2018 to 22% in 2050 of the world total delivered energy, corresponding to an increase from 91 quadrillions to 139 quadrillions British thermal units (Btu) for the same period with an average annual growth of 1.3%. For the world's largest economic entities, building energy consumption increases more significantly than the world average: for the U.S., the largest building energy consumer in the world, the end-use energy consumption by the residential sector and commercial sector was about 21 quadrillion Btu, equal to 28% of total U.S. end-use energy consumption in 2019 [2]; for the second-largest consumer, China, in 2016, around 20% of total energy use was consumed by commercial building sector [3]. In the European Union, buildings are accounted for about 41% of the final energy consumption in 2016 [4].

To slow down the increasing building energy consumptions, different stakeholders and organizations have undertaken various measures and actions. Governments adopted a variety of policies to promote more utilization of renewables such as wind, solar power, and biomass energy. Professional associations such as ASHRAE tailored their standards and codes for more energy-efficient designs and operations of high-performance buildings. Among research communities, most recent developments on smart buildings [5], smart cities [6,7], smart grids [8], Internet of Things (IoT) [9], and various advanced data-driven control strategies have started to contribute to optimizing building energy usages. During this process, computer simulations using BEMs play a crucial role: a successful BEM can provide many insights into the complicated building physics and evaluate different energy-saving measures. On the other hand, the performance of a building energy model is subject to many uncertainties from the model itself (e.g., model-form uncertainty) and the inputs (i.e., parameter uncertainty) [10]. The uncertainties are often inevitable due to the complexities of a building and its system, and many model parameters. The model-form uncertainties originate from numerical approximations, quality of computer programming and coding, and underlying assumptions of building models. For example, it remains a challenge to model the dynamic correlations and interactions among multiple physics components, including building envelopes, facility responses, interior impacts (e.g., occupants and appliances), and exterior impacts (e.g., weather conditions and impacts from neighboring buildings such as microclimates and shading). For the parameter uncertainties, hundreds to thousands of inputs/parameters are often required to create a building model. It is estimated that for a typical building energy model developed in EnergyPlus [11], about 3000 input parameters need to be specified [12]. The parameter uncertainties can thus be introduced through the input data from 1). the outdoors, including the long-term and short-term macroclimate and microclimate information around the building, 2). the building itself, including building material discrepancies from the design to the built process, property transformations and function changes during operations, and 3). the indoor building parameters from the randomness of occupancy behaviors and equipment deteriorations during the service life.

To reduce the uncertainties and align simulation results with measured data, model calibration is an integral step for developing a reliable and accurate BEM, which can then be applied to building optimization, retrofit analysis, fault detections, and diagnoses, and advanced model-based controls. From a mathematical perspective, model calibration is a searching process for the highly-parameterized model in an undetermined search space with a large number of independent and interdependent parameters. Model calibration often includes manual and automated calibrations [13]. The former heavily relies on a user's expertise in building science and simulation and his/her knowledge about the target building. So a few key parameters are manually selected and tuned to obtain the simulation results close to the measured information from audited and monitored energy usage data. Manual calibration is, therefore, a very time-consuming, labor-intensive, and cost-expensive process, and the manually calibrated model is often questionable due to the limited expertise of the user and the complexity of the calibrated problem.

Automated calibration is a non-user-driven and mathematically-based process in which an objective function or penalty function is defined for matching simulation results with measured data [14]. Although the input variables under search and the actual physical properties may not match each other well, they should have physical significance and meaning in reality. With the mathematical/statistical methods coded in a computer program, the calibration activity can be iterated automatically for a large batch of simulations with many combinations of parameters. The automated search process is considered complete when the calibration error rate (the difference between simulation and measurement) is less than a threshold criterion, or the calibration activity runs long enough and should be stopped by a given time. In this case, the group of input parameters of the specific simulation with the lowest error rate is selected as the calibration result. When both criteria are often applied, there is a trade-off because automated calibration is often an iterative total permutation process with a heavy computing burden. With the recent development in computing power and advanced mathematical and statistical methods, the calibration process is speeded up. So automated calibration is always more efficient and faster than manual calibration. Nowadays, combined with online metering, continuously automated calibration becomes possible in buildings [15].

Conventional calibrations are often deterministic, leading to estimated parameters far off from their original values when the training data is qualitatively/quantitatively insufficient. Sometimes, the deterministic calibration approaches are inadequate and even risky. For example, for a building retrofit project, different ECMs can be evaluated by BEMs following the IPMVP [16]. However, because of the deterministic process, the ECMs are compared by their absolute performance without uncertainties quantified. As a result, there exist unknown risks of the underperformance of a suggested ECM, which may not be expected from the performance contract. In practice, to avoid the risk, ESCOs generally provide building owners with one fixed minimum guaranteed savings for each selected ECM based on the rules of the thumbs of the experts. It was reported that the experts' subjective judgment is estimated to be between 60% and 70% of the deterministic energy-saving prediction [17].

Besides, with more big data available from the advancement of sensor technologies, and more research on larger scale BEMs, e.g., urban-scale analysis, the calibrations of BEMs face unprecedented challenges: (a) how to extract high-quality information from big data and use them as much as possible for model calibrations while maintaining acceptable computing costs; (b) how to adequately consider the impact of measurement errors on calibrations; (c) how to reduce uncertainties in assumed data and model parameters for a large scale problem; (d) how to understand the causes of errors and improve the accuracy of BEMs. In some previous work, the average error of around 69% of UBEM has been considered the acceptable level [18].

As a scientific way to interpret and quantify these uncertainties and risks, Bayesian inference has gained interest recently. Bayesian inference is "the process of fitting a probability model to a set of data and summarizing the results by a probability distribution on the parameters of the model and unobserved quantities such as predictions for new observations" [19]. Based on the literature review, the history with the central mark stones of Bayesian statistics and inference is shown in Fig. 1. It was first proposed by Reverend Thomas Bayes from England in his 1763 doctoral dissertation [20], then developed by Pierre Simon Laplace to form the Bayes theorem in France and then spread to other parts of Europe in the nineteenth century. But because of limited computing resources, its development did not gain momentum until the twentieth century thanks to the development of MCMC [21] and modern computers. In 2001, Kennedy and O'Hagan systematically illustrated the Bayesian calibration of computer models [22], signifying the boom of Bayesian calibration and inference. Since then, it has been utilized in a

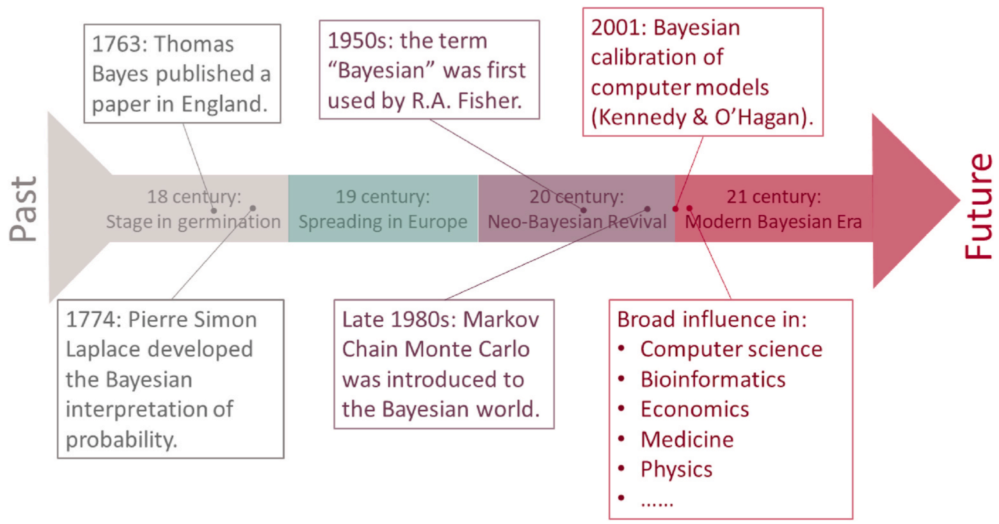


Fig. 1. Bayesian statistics development.

variety of topics, including environment [23–26], hydrology [27–29], transportation [30], and medicine research [31], etc. By propagating parameters using probabilistic analysis, Bayesian inference incorporates uncertainties into the approximations of real systems by computer models. Combining multiple sources of information at different scales and with different reliabilities, the inadequacy of a model, which is revealed by the discrepancy between the predictions and observed data, can be corrected [13].

For building energy modeling, one of the first applications of Bayesian inference may be presented by Heo [32]. A Bayesian inference begins with modelers' prior knowledge and expertise beyond the impact of observations in the forms of the prior distributions of model parameters. Then these prior distributions are incorporated into the building simulation and mapped into a probability distribution of model outputs. The prior distributions may then be updated and improved based on field observations and output distributions via the Bayes' rule. In the end, the posterior distribution of selected parameters can be obtained and employed to forecast building energy demand and consumption or retrofit benefits more accurately and reliably.

Since the posterior distribution is intractable analytically, approximation methods are needed. Variational Inference is an optimization-based method from machine learning. It fits a variational distribution to the posterior under an objective function to improve computational efficiency and more suitable for the high-dimensional problem. Researchers already started to employ it on occupancy behavior, building energy/thermal performance, forecast of wind speed, and solar irradiation. Sadeghi et al. [33] trained a visual preference model using Variational Inference with 565 observations of 75 participants. Different models based on different combinations of variables were developed. It is found that the best model's prediction accuracy is 0.69, which is satisfactory compared to the acceptable value of 0.33. Lee et al. [34] used Automatic Differentiation Variational Inference to train a high-dimensional thermal preference model with 5454 latent variables. Then the trained model was used as a prior for personalized thermal preference models to learn new occupants. Results show that the proposed method can better predict performance, even with limited data. Garstens et al. [35] demonstrated the Variational Inference's potential to building energy measurements and savings. The method was applied to the example from IPMVP to estimate the mean value of 12 monthly measurements. Simulations show that, with 100,000 draws, the case can be done, and the result is stable and converges on the posterior distribution in 10.76s on a middle-range laptop computer. Besides, Liu et al. [36] employed Variational Inference to approximate the posterior parameter distribution of a spatial-temporal neural network to forecast wind speed with probability. It is concluded that the proposed method is better than other point forecast models by performing 14.1% lower in terms of RMSE value. The authors also leveraged Variational Inference on solar irradiation prediction [37]. An ensemble spatio-temporal deep learning model was proposed for solving the problem with the collaboration of Variation Inference to quantify the uncertainty. The results demonstrate that the proposed model outperforms the other four models in terms of RMSE, MAE, and NSE in all months. These machine learning-based models show great potentials for inference studies. On the other hand, they are relatively new, and the publications are still limited compared to other methods, especially for building energy model calibration.

Another method is MCMC that has been the dominant paradigm to approximate posterior distributions for decades. It's a sampling-based approach and often unbiased, guaranteed to converge to the true posteriors [41], which has been widely studied, extended, and applied [38]. In the field of building energy, MCMC is more commonly used. Details about its application is presented in Section 2.

When compared to conventional calibration approaches, the benefits of Bayesian inference-based calibrations have been previously revealed. Kim et al. [42] compared the deterministic calibration based on a constrained optimization method and the stochastic Bayesian calibration. It is shown that the Bayesian calibration performed better and reduced the variance of unknown inputs. By considering the sensor error of −3%, the MBE can reduce from 4.53% to 0.25% using the Bayesian calibration instead of the deterministic method. Pavlak et al. [43] compared the traditional least-square approaches and the Bayesian inference for a gray box model. For the cases with uniform priors and noises neglected, the two methods had similar performances, whereas, for the noisy situations, especially when the noise level is over 25%, the Bayesian inference had prior knowledge of the system and outperformed the traditional method. For the load calibration, the values of MBE performed by the conventional method and Bayesian method are 347 and -24, respectively. And for the cumulative error, the Bayesian method can reduce it from 0.869% to −0.059%. Similar conclusions were also reached by Kim and Park [44]. For the calibration of heat extraction of an AHU, the MBE calculated based on the deterministic and Bayesian method are −22.46 and 9.44, respectively. In Zhang et al.'s study [45], the calibration performance of the Bayesian inference method and GA are compared. It is found that the GA method performs worse accuracy. This may be because for the GA method, it only minimizes the explicitly defined objective functions. If a specific error metric is not included in the objective function, its calibration ability will be limited and can lead to the error metrics' poor performance. And how to define a suitable objective function is difficult. In contrast, Bayesian inference calibration calculating the conditional distribution of calibration parameters given the observations instead of an explicit optimization objective function makes a more stable performance no matter what error metrics are employed. Rouchier et al. [46] compared the forecast performance of a BEM calibrated using Bayesian inference and Kalman Filter. Results show that the Bayesian-based stochastic model is more robust and offers more reliable forecasting since parameter uncertainties are considered. These studies all highlighted the unique feature of the Bayesian inference of the inclusion of uncertainties and associated possibilities into calibrations, making it a highly promising calibration method for building energy analyses.

There existed some previous but brief introduction of Bayesian calibrations applied to the building sectors [47,48]. It was also mentioned in a review on the uncertainty analysis in building energy assessment as an inverse uncertainty quantification method [10]. In Lim and Zhai's review of stochastic modeling for building stock energy predictions, the basic theory of Bayesian inference and its application in UBEM were presented [49]. However, owing to fewer implementations at that time, the authors did not thoroughly introduce Bayesian inference applications in individual BEMs. The review may be complete but a bit general, and they did not provide details on the methods, tools, and settings of key Bayesian parameters for a regular user, so he/she still may not know how to proceed with implementation and realization. Coakley et al. [13] summarized the characteristics of the Bayesian calibration: natural incorporation of uncertainties in the calibration process, correction of model inadequacy, and the combination of multiple sources of information. On the other hand, currently, no comprehensive and informative review study focuses on the Bayesian application specifically to building energy modeling with current development status, detailed Bayesian principles unraveled, and actual implementations through practical software tools illustrated. Adopting the Bayesian-inference-based BEM calibration is mostly hindered by the theoretical myths and the various but non-systematical options of implementation tools.

As a result, this literature review is developed under the enlightenment and belief that Bayesian inference, as an innovative calibration approach, will carry forward in the following decades in building energy simulations and other similar topics. In this review, the framework and calibration process of Bayesian calibration in building energy modeling are presented. Current research status, limitations, challenges, and future work are also discussed. Besides, for those who are interested in Bayesian statistics but without a strong mathematical/statistical background, this paper clarifies the theory and application for them. The paper is organized as follows: following the introduction, the previous studies on Bayesian calibrations of building energy fields are summarized in Section 2; then theoretical fundamentals are demonstrated schematically in Section 3; the implementation of Bayesian inference including its programming, especially by using the R language, is presented in Section 4; conclusions and identified future research work are at the end of the paper. As a critical review, many figures and tables are original contributions by default after synthesizing the information collected.

2. Applications

Bayesian inference is developed quickly after Kennedy and O'Hagan's publication in 2001, which illustrated a Bayesian calibration procedure of computer models [22]. One of the first applications of Bayesian inference to building modeling was presented by Heo [32]. Since then, it has gained researchers' interest in the building energy field and has been applied from building energy models to the related research field, such as occupancy behavior calibration, thermal property estimations, and sensor calibrations. This section summarized the previous studies focused on building energy calibration to the end of Sep. 2020.

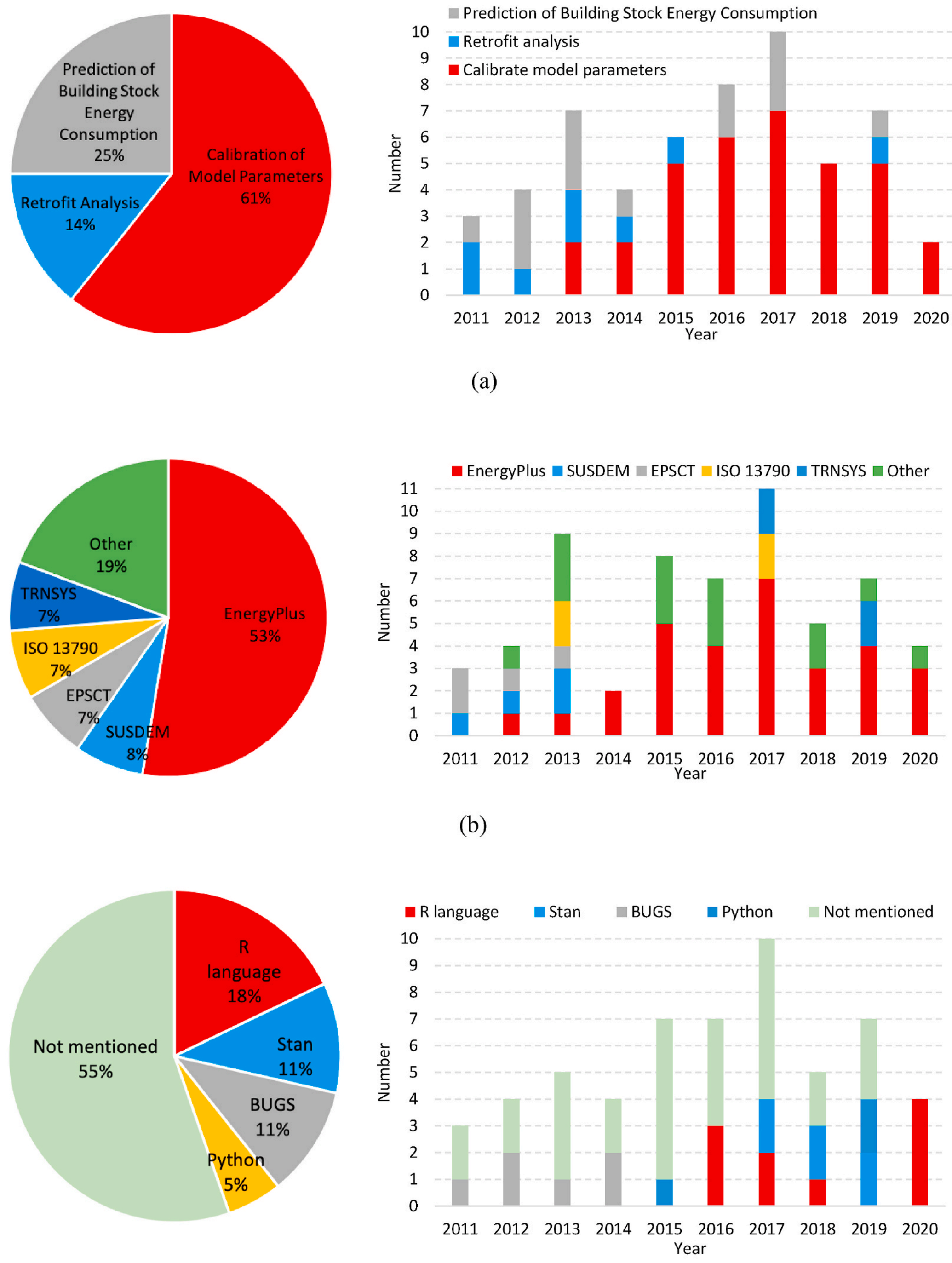
During the past decade, more than 50 papers, including doctoral dissertations, journal/conference papers on building energy modeling, were found to be closely related to Bayesian inference, as summarized in Fig. 2 and Table 1. The Bayesian inference was first applied to retrofit analysis and prediction of building stock energy consumption for the application domains. Later, more research focuses on increasing the calibration accuracy using individual buildings as an example, including the improvement of the MCMC algorithm and the impact of the prior distribution, Meta-model, and informative data. Dr. Choudhary and her colleagues, Dr. Heo and Dr. Tian, from the University of Cambridge, might be the most active researchers in this field. Their studies based on Bayesian inference began in 2011 and continue until now. During this period, several other researchers employed Bayesian inference through cooperation with Dr. Choudhary. Meanwhile, research teams within the University of Colorado Boulder have been focusing on this topic as well. From Fig. 2, it is shown that EnergyPlus and R language are the most often used BEM and program environments for Bayesian inference calibration.

The most widespread application of Bayesian inference in BEM is the calibration of unknown model parameters, either continuous or discrete. Due to the concerns over computing resources and time, a sensitivity analysis is often necessary to identify the key parameters with the most significant impacts before the calibration is deployed. The key parameters will be ranked and then fed into the calibration process. Although for each calibrated parameter, the Bayesian-based calibration result is a PDF in which each value's possibility can be regarded as a degree of certainty, the mean value or mode value of the PDF is potentially improperly considered as calibrated results by some researchers and used to calculate the calibration tolerance criteria such as CVRMSE and conduct further study. This utilization approach ignored the unique feature of Bayesian inference as a probabilistic method. From a statistic perspective, the Bayesian parameter estimation is precisely the entire posterior distribution, not a single number. Instead, it maps each unique parameter value onto a plausibility value [50].

For the application of Bayesian-based stochastic building energy models, there are two main fields: predictions of building stock energy consumption and retrofit analysis. There are two primary approaches to applying Bayesian inference to the prediction of building stock energy consumption. For the first method [51], individual reference building models are selected, calibrated using Bayesian inference, and then aggregated for the energy prediction of the building stock. Two aggregation methods are available: in the 1st method, every posterior distribution of the representative building from the iterated Bayesian inference is combined to represent one building stock type; in the 2nd method, each distribution is used to describe a different building type. For multiple building classifications, the result for every building type is combined to represent the whole building stock. Booth et al. [52] applied this method to 35 flats in the UK by using the average daily values of measured energy consumptions as observations but not each energy consumption of the 35 units. The results showed that there were minimal discrepancies between the Bayesian calibrated model outputs and the observation data.

For the second method, combining regression analysis and Bayesian inference, the measured macro-level data (e.g., at the district, urban, national level) are used to calculate the micro-level models (e.g., at an individual building level). Booth et al. [53] leveraged the proposed method to the area of Salford in Greater Manchester, UK, containing approximately 86,400 households. Tian and Choudhary [54] applied this method to the school buildings in London, UK. In the study by Yamaguchi et al. [55] to calculate the supermarkets' energy consumption at the urban scale, it was revealed that the building insulation performance might impact significantly on the seasonal and weekly energy consumption. In contrast, its effect on annual energy use might be modest.

As the other application of stochastic models, retrofit analysis has been done in several studies. One of Heo's early studies in his Ph.D.



Notes: the "Stan" includes interfaces using Stan engine, such as RStan and PyStan.

(c)

Fig. 2. Summary of annual publications focus on Bayesian inference calibration of BEM based on different perspectives (a) application domain; (b) building simulation tool; (c) Bayesian program environment.

dissertation in 2011 [32] focused on building retrofit analysis. Later, similar studies were conducted to include retrofit modeling and risk analysis for individual buildings and building stock. The energy-saving performances of different ECMs are presented with the degree of belief, based on which the risk analysis can be further conducted. This method protects the interests of ESCOs, which cannot be obtained when they are provided with a fixed minimum guaranteed saving. Booth and Choudhary [56] applied a Bayesian-calibrated stochastic model to predict the energy savings of retrofit measures of approximately 15,000 houses in the UK, considering not only the installation costs and future energy prices but also the lifetime carbon savings and increased thermal comfort. The posterior distribution was distorted when using the Meta-models instead of the original BEMs in Bayesian inference to save computing time. For the case of MLR, the errors were within 0.71%. It was confirmed from the conclusion of Lim's study [51] that the distorted posterior distribution can be used to evaluate the effects of ECMs.

Table 1
Summary of publications about BEM calibration based on Bayesian inference.

Author	Year	Study scope	Building Scale	Building Simulation Tool	Bayesian Program Environment	Emulator/ Surrogate	Sensitivity Analysis	Posterior Distribution Estimation	Data Resolution	Data Type	Ref.
Booth and Choudhary	2011	Calibrating a bottom-up engineering-based housing stock model	Stock	SUSDEM	BUGS: WinBUGS	MLR	N/A	Gibbs	Annual	Electricity and gas consumption	[57]
Booth et al.	2012	Prediction of stock energy consumption	Stock	SUSDEM	–	N/A	Morris	–	Daily	Electricity Consumption	[52]
Booth and Choudhary	2013	Retrofit analysis	Stock	SUSDEM	–	N/A	Morris	–	Annual	Total energy consumption	[56]
Booth et al.	2013	Prediction of stock energy consumption	Stock	CDEM; EST; SUSDEM	BUGS: WinBUGS	N/A	N/A	Gibbs	Annual	Total energy consumption	[53]
Heo	2011	Retrofit analysis	Individual	EPSCT	–	GP	Morris	Metropolis-Hastings	Monthly	Gas consumption	[32]
Heo et al.	2011	Retrofit analysis	Individual	EPSCT	–	GP	Morris	Metropolis-Hastings	Monthly	Gas consumption	[58]
Heo et al.	2012	Retrofit analysis	Individual	EPSCT	–	GP	Morris	Metropolis–Hastings	Monthly	Gas consumption	[59]
Heo et al.	2013	Retrofit analysis	Individual	EPSCT	–	GP	Morris	Metropolis–Hastings	Monthly	Gas consumption	[60]
Heo et al.	2015	Calibration efficacy under different uncertainty level	Individual	EnergyPlus	–	GP	Morris	Metropolis–Hastings	Monthly	Electricity consumption	[61]
Heo et al.	2015	Retrofit analysis	Individual	Normative model	–	GP	Morris	Metropolis–Hastings	Monthly	Total energy consumption	[62]
Tian and Choudhary	2012	Building stock energy modeling; Retrofit analysis	Stock	EnergyPlus	BUGS: OpenBUGS	N/A	SRC; MARS	Gibbs	Annual	Gas consumption	[54]
Tian et al.	2016	Identification of informative energy data	Individual	EnergyPlus	R: BRugs	MLR	SRC; RFVI	Gibbs	Monthly	Electricity consumption; Gas consumption	[63]
Choudhary	2012	Prediction of stock energy consumption	Stock	Statistic model	BUGS: WinBUGS	N/A	N/A	Gibbs	Annual	Total energy consumption	[64]
Choudhary and Tian	2014	Influence of district feature on energy consumption	Stock	Statistic model	BUGS	N/A	N/A	Gibbs	Annual	Total energy consumption	[65]
Kim et al.	2013	Performance comparison of different energy models	Individual	ISO 13790; EnergyPlus	–	GP	Morris	Metropolis-Hastings	Monthly	Electricity consumption	[66]
Kim et al.	2014	Comparison to conventional calibration method	Individual	EnergyPlus	–	GP	Morris	MAP	Daily	Gas consumption	[42]
Kim et al.	2014	Decision making of HVAC system	Individual	EnergyPlus	BUGS: WinBUGS	MLR	Morris	Gibbs	Daily	initial construction cost and total energy consumption	[67]
Kim	2015	Calibrating unknown parameters	Individual	EnergyPlus	–	GP	N/A	Hybrid Monte Carlo	Monthly	Total energy consumption	[68]
Kim and Park	2016	Comparison to conventional calibration method	Individual	EnergyPlus	–	GP	Morris	Metropolis–Hastings	Monthly	Heat extraction; Gas consumption	[44]
Manfren et al.	2013	Calibrating unknown parameters	Individual			GP	N/A			Electricity and gas demand	
Yamaguchi et al.	2013	Prediction of stock energy consumption	Stock	Statistic model	–	N/A	N/A	–	Weekly; Annual	Energy consumption	[55]
Pavlak et al.	2014	Comparison to conventional calibration method	Individual	RC model	–	N/A	N/A	Metropolis	Hourly	Sensible zone loads and corresponding temperatures	[43]
Li et al.	2015	Calibrating unknown parameters	Individual	EnergyPlus; EPC	–	MLR	Lasso	Metropolis	Daily; Monthly	Chilled water consumption (daily), peak demand of chilled water (monthly)	[69]
Li et al.	2015	Calibrating unknown parameters	Individual	EnergyPlus	–	MLR	Lasso	Metropolis	Daily; Monthly		[70]

(*continued on next page*)

Table 1 (*continued*)

Author	Year	Study scope	Building Scale	Building Simulation Tool	Bayesian Program Environment	Emulator/ Surrogate	Sensitivity Analysis	Posterior Distribution Estimation	Data Resolution	Data Type	Ref.
										Chilled water consumption (daily), peak demand of chilled water (monthly)	
Li et al.	2016	Comparison of different meta-models	Individual	EnergyPlus	–	MLR; GP	Lasso	Metropolis	Daily; Monthly	Chilled water consumption (daily), peak demand of chilled water (monthly)	[71]
Chong and Lam	2015	Calibrating unknown parameters	Individual	EnergyPlus	Python: PyMC	GP	None	Metropolis-Hastings	Hourly	Electricity consumption; Gas consumption	[72]
Chong and Lam	2017	Comparison of different MCMC methods	Individual	EnergyPlus	–	GP	Morris	Metropolis; Gibbs; NUTS	Hourly	Cooling energy consumption	[73]
Chong et al.	2017	Selection of representative subset of the entire dataset and its performance	Individual	TRNSYS; EnergyPlus	R: RStan	GP	Morris	NUTS	Hourly	Electricity consumption	[38]
Chong and Menberg	2018	Introduction of Bayesian inference	Individual	EnergyPlus	Stan	GP	Morris	NUTS	Monthly	Electricity consumption	[74]
Chong et al.	2019	Continuous-time model calibration	Individual	EnergyPlus	Stan	GP	Morris	HMC	Monthly	Electricity consumption	[75]
Henze et al.	2015	Distinguish normal and abnormal energy usage profile	Individual	Reduced-order BEM	–	N/A	None	Metropolis	Hourly	Energy end-use	[76]
Braulio-Gonzalo et al.	2016	Prediction of stock energy consumption	Stock	EnergyPlus	R	INLA	None	MCMC	Annual	Heating and cooling demand; discomfort heating and cooling hours	[77]
Kang and Krarti	2016	Calibrating unknown parameters	Individual	eQUEST	–	GP	Local sensitivity analysis	Metropolis Hastings	Monthly	Total energy consumption	[78]
Muehleisen	2016	Introduction of Bayesian inference	Individual	OpenStudio	–	–	Morris	–	Monthly	Total energy consumption	[47]
Zhao et al.	2016	Efficient energy model development at a city scale	Stock	Normative model	R	MLR	Absolute t statistic values	MCMC (coordinate directions algorithm (CDA))	Annual	Total energy consumption	[79]
Sokol et al.	2017	Prediction of stock energy consumption	Stock	EnergyPlus	–	N/A	N/A	Defined by authors	Monthly; Annual	Electricity consumption; Gas consumption	[80]
Kristensen et al.	2017	Performance comparison of measurements under different temporal resolution	Individual	ISO 13790:2008	–	GP	Sobol	Metropolis-Hastings	Six-hourly; daily; weekly; Monthly	District heating energy	[81]
Kristensen et al.	2017	Building clusters and building stock energy modeling	Stock	ISO 13790:2008	–	GP	Sobol	Metropolis-Hastings	Annual	District heating energy	[82]
Lim	2017	Prediction of urban-scale energy consumption; ECM analysis	Stock	EnergyPlus	R	MLR	SVI	Metropolis-Hastings	Annual	Total energy consumption	[83]
Lim and Zhai	2017	Performance comparison of different meta-models	Individual	EnergyPlus	R	MLR; NN; SVM; MARS; GP	SVI	Metropolis-Hastings	Monthly	Electricity consumption; Gas consumption	[84]
Lim and Zhai	2018	Identification of informative energy data	Individual	EnergyPlus	R	MLR	SVI	Metropolis-Hastings	Monthly	Electricity consumption; Gas consumption	[85]
	2017			TRNSYS	Stan	GP	Morris	NUTS	15 min		[86]

(*continued on next page*)

Table 1 (*continued*)

Author	Year	Study scope	Building Scale	Building Simulation Tool	Bayesian Program Environment	Emulator/ Surrogate	Sensitivity Analysis	Posterior Distribution Estimation	Data Resolution	Data Type	Ref.
Menberg et al.		Calibrating unknown parameters	Individual (GSHPS)							Inlet and outlet temperature of the heat pump of load side	
Menberg et al.	2019	Influence of error terms in Bayesian calibration	Individual (HPS)	TRNSYS	Stan	GP	Morris	NUTS	15 min	Inlet and outlet temperatures on both load and source sides	[87]
Yuan et al.	2017	Performance of proposed posterior distribution estimation method	Individual	EnergyPlus	–	GP	Pre-defined	Gibbs; PAM	Monthly	Electricity consumption	[88]
Yuan et al.	2017	A simultaneous calibration and parameter ranking method	Individual	EnergyPlus	–	GP	GP based method	Gibbs	Monthly	Electricity consumption	[89]
Yuan et al.	2019	Retrofit analysis	Individual	EnergyPlus	–	GP	N/A	Gibbs	Annual	Total energy consumption	[90]
Raillon and Ghiaus	2018	Calibrating unknown parameters	Individual	RC model	–	GP	Predefined	Metropolis-Hastings	10 min	Indoor temperature	[91]
Rouchier et al.	2018	Comparison to deterministic calibration method	Individual	RC model	–	N/A	Predefined	Metropolis-Hastings	Hourly	Indoor temperature and heating power	[46]
Zhang et al.	2018	Calibrating unknown parameters	Individual	EnergyPlus	R: RStan	MLR	Morris	NUTS	5 min	Calculated heating energy consumption based on the measured radiant system inlet/outlet water temperature and water mass flow rate.	[92]
Zhang et al.	2019	HVAC control optimizing	Individual	EnergyPlus	–	GP	Morris	NUTS	Hourly	Average indoor air temperature and heating demand	[45]
Chen et al.	2019	Prediction of stock energy consumption	Stock	IES-VE	–	GP	None	–	Monthly	Heat demand	[93]
Rysanek et al.	2019	Calibrating unknown parameters	Individual	TRNSYS	Python: PyMC	N/A	–	Metropolis-Hastings	Hourly	Electricity consumption	[94]
Yi et al.	2019	Calibrating unknown parameters	Individual	EnergyPlus	Python: PyMC	ANN	None	Adaptive metropolis algorithm	Annual	Gas and electricity consumption	[95]
Ahmadi et al.	2020	Calibrating unknown parameters	Stock	Statistic model	R	N/A	N/A	–	Annual	Energy demand	[96]
Zhu et al.	2020	Calibrating unknown parameters	Individual	EnergyPlus	R	LM, SVM, MARS, BMARS, RF	Sobol	Approximate Bayesian computation	Monthly	Heating and electricity consumption	[97, 98]

Notes: "–" means the information was not mentioned.

3. Methodology

3.1. Model calibration

From a statistical perspective, a model calibration process can be expressed as [22]:

$$y(x) = \eta(x, t) + \delta(x) + \varepsilon_m \tag{1}$$

where y and η are the field observation and simulation output, respectively. x represents the model input, and t represents the model parameter. $\delta(x)$ is the model error due to the model input x while ε_m is the random observation error, which is often assumed to follow a Gaussian distribution, i.e. $\varepsilon_m \sim N(0, \sigma_m^2)$. With the same model input x, the model parameter t can significantly affect the simulation output accuracy. The process of model calibration is about adjusting the model parameters and forcing within the margins of the uncertainties. Its objective is to obtain a model that can represent the process of interest within acceptable criteria. Note here for simplicity, we use the singular form of the parameter. In the case of many parameters, the singular form can represent the vector form of multiple parameters.

3.2. Bayesian inference

As the footstone of all Bayesian statistics, Bayes' theorem was first proposed by Reverend Thomas Bayes in his doctoral dissertation [20] and can be described as:

$$Posterior = \frac{Probability\ of\ the\ data\ \times Prior}{Average\ probability\ of\ the\ data} \tag{2}$$

The probability of an event is inferred based on the prior knowledge of conditions that might be related to the event. Bayesian inference is one application of Bayes' theorem and can be written as:

$$p(\theta|y) = \frac{p(y|\theta)\cdot p(\theta)}{p(y)} \propto p(y|\theta)\cdot p(\theta) \tag{3}$$

where $p(\theta|y)$ is the posterior distribution of the unknown parameter θ based on known observation y. $p(y|\theta)$ is the likelihood function of observation conditional on the unknown parameter. $p(\theta)$ is the prior distribution of the unknown parameter which is the marginal probability that means it is irrespective of the outcome of another variable, and $p(y)$ is the probability of the observation that is marginal as well to normalize $p(y|\theta)$. Therefore, the posterior probability is proportional to the product of the prior probability and the likelihood.

3.3. MCMC algorithm

In reality, not all problems can be solved analytically using Bayesian inference since the integrals of the likelihood can be computationally costly or are sometimes impossible to be calculated. Compared to Variational Inference, which is an optimization-based approximation method, MCMC is a sampling-based approach widely applied to building energy field to solve the parameter estimation problem with two components. One is the well-known Monte Carlo method. It is a computational algorithm to solve statistically challenging problems relying on repeated random samplings and approximate the target value (e.g., mean value) using the independent samples' results. The other is the Markov Chain method for solving a sequence of possible events in which the probability of each event depends only on the state attained in the previous event. By integrating MCMC and Bayesian inference, posterior distribution can be estimated efficiently.

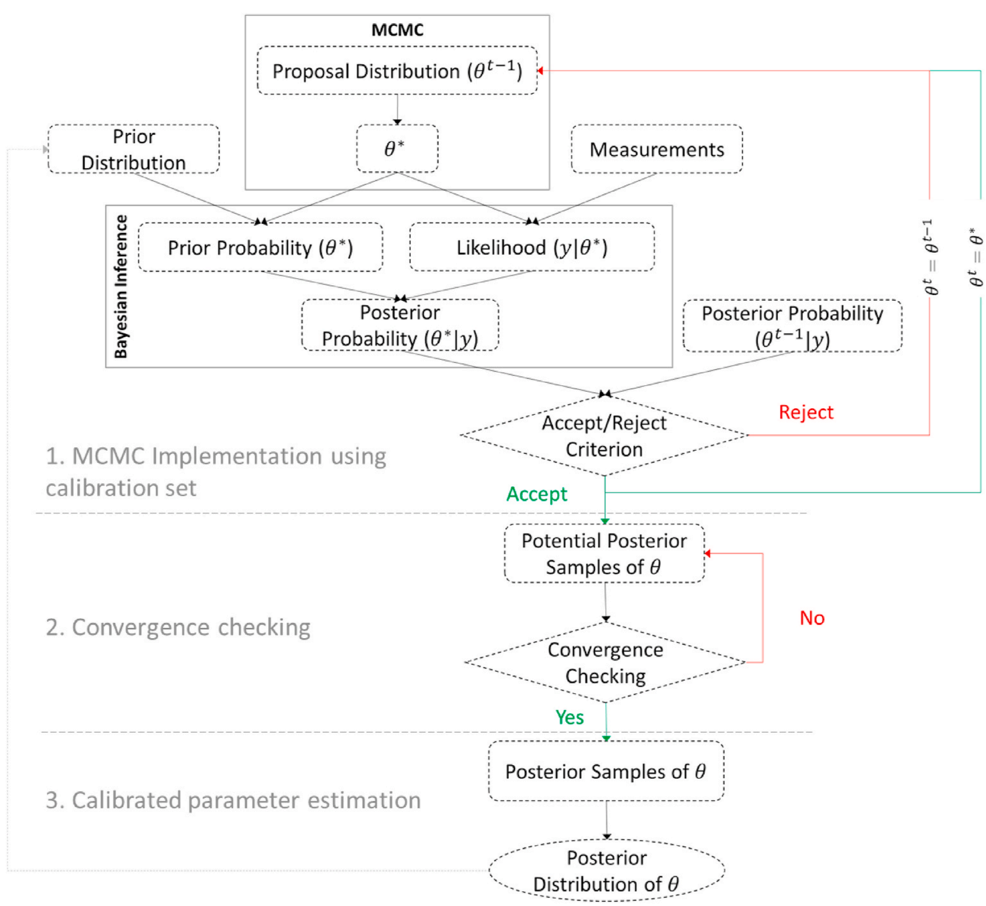


Fig. 3. Schematic of Bayesian inference calibration using MCMC.

Here, we developed Fig. 3 to illustrate the Bayesian inference calibration using MCMC, which is the most important step of the Bayesian inference BEM calibration (in the next section). In the first step, a value within the proposal distribution is set arbitrarily to represent the unknown model parameter θ^1. Then combining with predefined prior distribution and Bayesian inference, the posterior probability of the unknown parameter θ^1 conditional on observations y is obtained. Then it proposes a second value θ^* random sample from the proposal distribution based on the characteristic parameter of θ^1 and repeats the procedure in the next time step. Hereto, θ^1 and θ^*, and their based $p(\theta^1|y)$ and $p(\theta^*|y)$ are obtained. An acceptance-rejection criterion is applied to determine which one moves in the right direction to approach the posterior distribution. The satisfactory sample will be regarded as θ^2, which is used in the next time step meanwhile regarded as a member of the potential posterior sampling trunk. Iterations are conducted, and the chain's convergence is checked. Finally, only randomly generated values after the convergence point apply to the statistics of the posterior distribution.

For the acceptance-rejection criterion, different MCMC algorithms adopt different criteria. Generally, it is classified into either a "random walking" group or a gradient-based group. Metropolis algorithm is the origin of several different algorithms for unknown posterior distributions [99]. This algorithm assumes that the sampling proposal distribution should be symmetric. Acceptance probability is defined as:

$$r = min\left\{\frac{p(\theta^*|y)}{p(\theta^{t-1}|y)}, 1\right\} \tag{4}$$

where θ^* is the proposal unknown model parameter.

The generated value at the t time step is determined as:

$$\theta^t = \begin{cases} \theta^*, \ p(\theta^*|y) > p(\theta^{t-1}|y) \\ \theta^{t-1}, \ others \end{cases} \tag{5}$$

To avoid automatically rejecting θ^*, which could be because the acceptance probability is less than one, and to allow asymmetric proposals, an advanced version of Metropolis, which is called Metropolis-Hastings [100], was developed. After the calculation of Eq. (4), a random value u is drawn from a Uniform distribution (0, 1) and compared to the acceptance probability r. Then Eq. (5) is revised as:

$$\theta^t = \begin{cases} \theta^*, \ u < r \\ \theta^{t-1}, \ others \end{cases} \tag{6}$$

When θ^t represents a vector, $\theta^t = [\theta_1^t, \theta_2^t, \dots, \theta_n^t]$, and the sample of θ_j^t is updated according to the distribution specified by $p(\theta_j^t|\theta_1^t, \dots, \theta_{j-1}^t, \theta_{j+1}^{t-1}, \dots, \theta_n^{t-1})$, it features the Gibbs [100] algorithm to better estimate posterior with fewer samples.

For the gradient-based algorithm, HMC is a typical representation. It avoids the random walk behavior inherent by using the first-order gradient information to determine how it moves through the target distribution [101]. The properties of HMC allow it to converge to the target distribution more quickly for a complicated high-dimensional problem. However, HMC requires users to provide values of two hyperparameters: a step size ε and the number of steps L, making it difficult and time consuming to tune. To mitigate the challenges of tuning, the NUTS was developed [101]. NUTS uses a recursive algorithm to automatically tune the HMC algorithm without requiring user intervention or the time-consuming tuning runs. Studies about Bayesian inference in building energy modeling have shown that NUTS is one of the most practical and efficient sampling methods [73].

To diagnose the convergence to the posterior distribution, trace plot, trace rank plot, and Gelman-Rubin statistics are always applied, which will be illustrated in the next section.

4. Realization

The framework of Bayesian inference applied to the BEM calibration is detailed and shown in Fig. 4. Note that Fig. 4 was based on Tian et al. [63] with more details and steps: e.g., the measurement preparation was added as the first step, and the tool for each step by the R language was included for the implementation and realization of Bayesian inference. The first step is to analyze and understand the measurement data. Then the second step is to develop the energy model for the target building. To reduce the number of calibrated model parameters and the computing cost, the third step is the parameter screening to select the most important inputs and model parameters. Prior distributions of the unknown model parameters are defined to represent parameter uncertainty. During this step, parametric simulation is conducted to create an input-output dataset used for sensitivity analysis. The fourth step of the informative data selection is optional to reduce the computing cost for better Bayesian performance. Based on the input-output dataset, a Meta-model is developed in the fifth step to replace the original BEM to save the computing time for the next step. The sixth step is the Bayesian inference of the posterior distribution based on MCMC. In the end, the seventh step is the validation and analysis of the calibrated model and parameters. The computational burden by the Bayesian inference calibration framework is heavily intensive for parameter screening, Meta-model generation, posterior estimation, and further simulation for energy prediction. Although it can be conducted in various ways, how the whole process is implemented will profoundly affect the estimation accuracy and efficiency. In the remainder of this section, each step will be illustrated in detail for the corresponding implementation based on a combination of base functions and packages of the R language [102, 103]. The R language is one of the widely used statistical tools. This review shares the experiences and knowledge of using the R language as one single programming environment to fulfill the whole Bayesian calibration procedure.

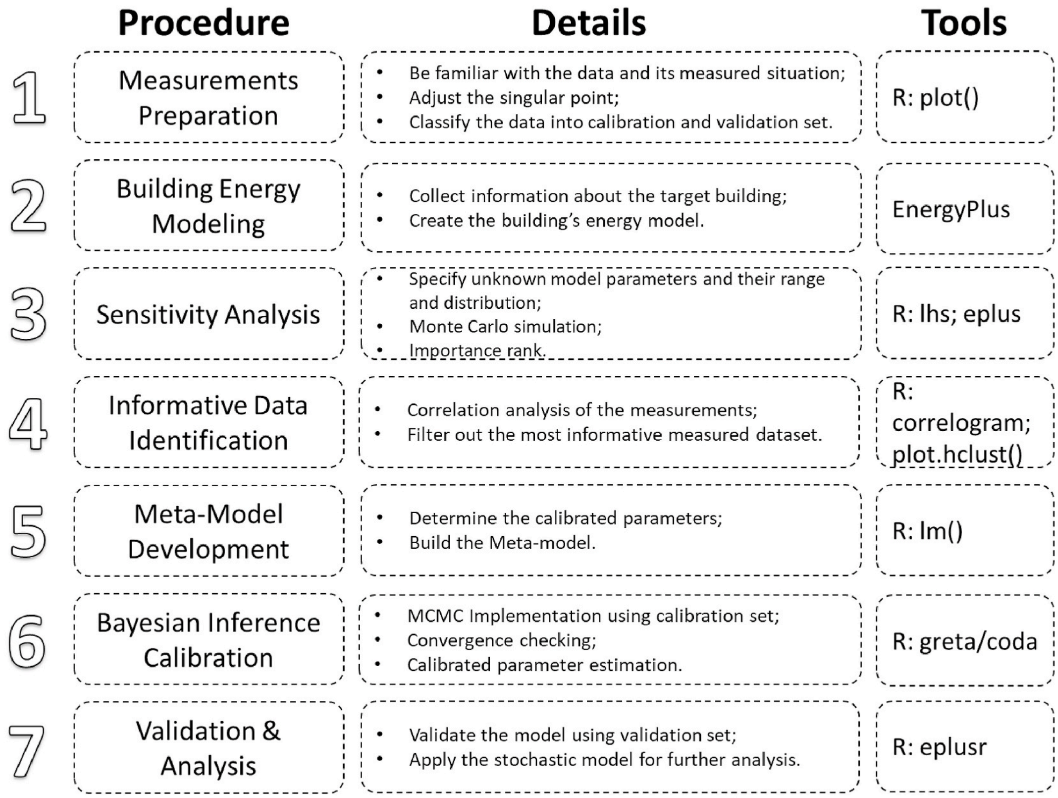
4.1. Step 1 – measurements preparation

Measurement data are the target of calibration. The quality and quantity of measurements will strongly affect calibration accuracy. To be familiar with the measured data, such as the type, resolution, primary trend, and measuring conditions, should be the first step for a Bayesian modeler. Due to sensor/metering errors, unexpected events, and conditions, there often exist outlier data. Low-quality data will undoubtedly reduce the calibration performance and accuracy. Using the R base function "plot()" to visualize the measurements, these singular points can be identified and removed or adjusted accordingly.

4.2. Step 2 - building energy modeling

The second step is to build the energy model based on reasonable assumptions of unknown parameters and collected building information from audits, site-visits, surveys, and design documents. A range of simulation tools include but are not limited to DOE-2 [104], EnergyPlus [11], TRNSYS [105], ESP-r [106], or user-developed models. Under the same conditions, different BEMs should attain consistent calibration results. However, the selection of a BEM should consider the model development feasibility, the calibration problem, and its further application. It is also important to select a tool suitable for parametric studies because many simulations will be conducted during the sensitivity analysis and Bayesian inference steps later. Fig. 2 indicates that EnergyPlus is one of the most commonly used tools since its input data file can be modified as a text file for easier editing and integration with other programming environments. In addition, within the R language, there is a package "eplusr" which was designed for EnergyPlus. Therefore, EnergyPlus is introduced here.

To evaluate the impact of different energy models on the performance of Bayesian calibration for a target office building, Kim et al. [66]



Notes: for the "Tool" column, words with brackets means the R base functions while others represent the R packages.

Fig. 4. Procedure of Bayesian inference calibration for building energy models.

developed two models by using a simplified calculation method (ISO 13790) and dynamic simulation tool (EnergyPlus). It is concluded that the simplified approach can perform comparably to the dynamic model. Li et al. [69] conducted a similar study using a dynamic model (EnergyPlus) and a reduced-order model (EPC). It is demonstrated that the calculation results for EPC are better than EnergyPlus. But this may be caused by fewer calibration parameters in EPC than in EnergyPlus.

4.3. Step 3 - sensitivity analysis

Ideally, with sufficient measurements and computer resources, all the uncertain parameters should be included in the calibration parameter set. The posterior distributions may still be uncertain when the data is insufficient in quantity/quality. It indicates the information provided by data is limited rather than a failure of calibration activity. Many parameters and inputs could also manifest different levels of uncertainties and significances on simulation outputs. The identification of dominant parameters cannot merely rely on arbitrary parameter selections from modelers' knowledge but should be based on a scientific process, i.e., a sensitivity analysis. Tian [107] summarized various sensitivity analysis methods in the Bayesian inference framework and their corresponding R packages. But, how to do the parametric simulation to generate the input-output dataset used for sensitivity analysis, especially based on R, was not mentioned.

For the parametric simulation, prior distributions and ranges of selected unknown parameters should be determined first. Then MC simulation is employed to propagate simulations whose model parameters' values are randomly selected from the predefined ranges using a specific sampling method to perform simulation runs iteratively. Here, the Latin Hypercube Sampling (LHS) method [71] is recommended, which can be realized using the R "lhs" package since it provides good convergence of parameter space with relatively fewer samples. To perform the MC simulation and collect input-output dataset automatically, an R package named "eplusr" developed by Jia [108] is suggested to use EnergyPlus directly in R. "eplusr" enables programmatic navigation, modification of EnergyPlus, parametric simulations, and retrievals of outputs. The obtained input-output dataset is then employed to identify the dominant model parameters that strongly affect the outputs.

The importance ranking results may vary with different combinations of sensitivity methods and outputs depending on the variety of fundamental algorithms and conditions of each sensitivity analysis method [109]. To avoid the potential inconsistency, Lim and Zhai proposed a new sensitivity analysis method, SVI, to account for the differences in sensitivity analysis methods and target outputs [84]. Eq. (7) shows how SVI is applied to recognizing and comparing the importance rankings from different sensitivity analysis methods through the normalization and aggregation process.

$$\sum_{l=1}^{m}\frac{\frac{\sum_{j=1}^{k}\left(\frac{v_{i,j}}{\sum_{i=1}^{n}|v_{i,j}|}\right)}{k}}{m\cdot k}\times 100 = Sensitivity\ Value\ Index\ (SVI)\ (\%) \tag{7}$$

where V is the value of a sensitivity analysis method, i is a parameter, n is the total number of the parameters, j is a sensitivity method, k is the total number of sensitivity methods, l is the target output, and m is the total number of target outputs.

The most important parameter is ranked as 1, and only parameters with smaller rankings will be selected for the subsequent Bayesian inference. How many parameters and what parameters to be selected is "balancing art." It would become computationally costly to select many parameters, whereas a calibration with few parameters may not be adequate to identify the uncertainties and disclose the hidden information. It was suggested that a maximum of ten parameters be selected based on the importance ranking for acceptable performance of Bayesian inference [110].

Chong and Menberg [74] investigated the influence of the number of calibration parameters on the posterior distribution by increasing the parameter number from 2 to 6 with one resolution. They found that over-parameterization occurs, indicated by the increase of posterior uncertainty. Especially when the observations are insufficient, this phenomenon becomes obvious in Kang and Krarti's study [78], which concluded that both the CPU time and the posterior errors increased gradually as the calibrated parameter number increased.

The informative levels of prior distributions also affect the calibrated posterior results. Chong and Menberg [74] studied Bayesian inference with three levels of prior distributions (non-informative: a uniform prior; weakly informative: a normal prior with a large standard deviation; specifically informative: a normal prior with a small standard deviation). From the calibration results, it was observed that for the non-informative prior, the posterior distribution is also relatively uninformative and primarily driven by the measured data. In contrast, the result from a specific prior distribution is highly constrained, suggesting that the posterior is driven primarily by the prior, and the influence of measured data is limited. Nevertheless, the weakly informative prior shows the best performance by balancing the flat and highly informative priors.

Lim and Zhai [84] investigated the impacts of a range of prior distributions on the posterior distributions and predictions. The study was conducted using a 30% extension of the original range of uniform distributions. The results showed that the prior distribution with a narrower range performed better. By extending the range, the prior distribution becomes less informative, and its capability to generate an accurate posterior distribution is weakened.

4.4. Step 4 - informative data identification

Many observation data, e.g., monthly/annual building energy usage data and hourly/weekly indoor thermal conditions, can be collected through site visits and sub-metering. It may be beneficial for more accurate calibration results to include many measurements, which may result in high computing costs in the meantime [81,85]. Therefore, it is essential to analyze and understand the collected information deeper through correlation analysis and hierarchical clustering methods by using the R "correlogram" package and "plot.hclust()" base function. It is shown that by the classification of the collected information, the informative data from different groups led to reliable results at low computing cost [63,85].

4.5. Step 5 - Meta-model development

In building energy modeling, many models are developed, including white-box, black-box, and gray-box models. Many of these models are quite complicated and could make computing expensive when applied to Bayesian calibrations [84]. For example, it was estimated that if a DOE reference medium office building developed using EnergyPlus used in the MCMC process that 100,000 iteration number was used, the computing time might exceed 70 days [84]. One solution is to use a Meta-model, or a surrogate model, which is a simplified representation or approximation of the computer model but with lower computing costs and acceptable accuracy. The Meta-model is often developed as the correlation based on the input-output dataset. Six Meta-models are often used in building energy and system analysis: MLR [70,79,80,111], NN [112–114], SVM [115], MARS [116], GP [48,79,111,116,117], and PR [80]. Their implementation in R is summarized in Ref. [84]. Table 1 shows that the GP emulator and MLR model are the two commonly used Meta-models. The GP model is with superior accuracy, while the computing cost could increase much with the augmentation of sample size and calibration parameters [71]. The MLR emulator is a relatively simpler and faster model with a lower overfitting risk when many parameter variations are involved [110].

The following statistical criteria can be used to define the performance of a Meta-model:

Coefficient of determination (R2):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{8}$$

Where $\hat{y}_i$ is a predicted variable value for period i, y_i is an observed value for period i, $\bar{y}$ is the mean of the observed value, and n is the sample size.

RMSE:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{n}} \tag{9}$$

CVRMSE:

$$CVRMSE = \frac{RMSE}{\bar{y}} \tag{10}$$

In Kang and Krarti's study [78], by utilizing a Meta-model instead of the actual building energy model to estimate two unknown input parameter values, the CPU time reduced from 40 h to just 5 min at a marginal loss of accuracy.

To analyze the Meta-model's influence on Bayesian calibration, Lim and Zhai [84] researched a total of five types of Meta-model: MLR, NN, SVM, MARS, GPE using 100 training samples. For the first four Meta-models, their developing time ranges from 0.05 s to 4.03 s, among which the MLR model is the fastest one. While generating a GP emulator, it took almost 20 min. The accuracy rankings of five meta-models differ for each monthly energy model. Using averages of absolute errors for 100 testing data as the index shows that the GP is the most accurate while the MLR is the least accurate. The non-linear model performs better than the linear model (MLR) due to its inherent limitation to represent the non-linear complex building energy model. For the MCMC with 100,000 chains, MLR was the fastest one (2.2 min), while GP took about 48.2 h. For the posterior distribution accuracy, the GP is the best one, while MLR is the worst one, but it still is accepted.

Li et al. [71] developed three MLR models as the emulator to optimize prediction performance. For a linear-main (LM) emulator, only the significant main effects are included. For a linear-interaction (LI) emulator, both significant main effects and interaction effects are considered. Based on the LI emulator, quadratic effects are considered in the linear-quadratic (LQ) emulator as well. Results show that the performance of LQ is quite close to GP, which could provide the most informative posterior distribution, while the posterior distribution based on LM is less informative. But for the normal MBE estimation, the LM performance is as good as a GP, and LI is the best selection based on the CVRMSE.

4.6. Step 6 - Bayesian calibration

This step is to apply the actual Bayesian inference to calibrate the unknown key parameters based on the Meta-model developed in the previous step. This is implemented through MCMC [118] for saving computing time, and several common MCMC algorithms and the fundamental theory were introduced in an earlier section. To study the effectiveness of three different MCMC algorithms: Metropolis, Gibbs sampling, and NUTS, Chong and Lam [73] diagnosed their convergence performance. It was found that NUTS can achieve adequate convergence to the posterior distribution the fastest, with a significantly reduced number of iterations. An improved Metropolis-Hastings algorithm was proposed in Ref. [91] by using gradient and second-order Metropolis-Hastings to improve the algorithm's tuning. In two calibration cases of 18 and 19 unknown parameters, the proposed algorithm was more robust than the conventional Metropolis-Hastings.

To diagnose the convergence to the posterior distribution, trace plot, trace rank plot, and Gelman-Rubin statistics are always applied. A trace plot merely plots the samples in sequential order, joined by a line. The trace plot of each parameter is often the first task for an analyst to diagnose common problems. A healthy chain typically has three features: stationarity, good mixing, and convergence. For stationarity, it is defined when the mean value of the chain is quite stable from beginning to end. Good mixing means that the chain rapidly zig-zags around to explore the full region. Convergence means that when more than one chains are used, multiple independent chains stick around the same region of high probability. An example of an effective trace plot is shown in Fig. 5a.

When many chains are employed in Bayesian inference, it is hard to read trace plots since chain traces are overlapped, so some pathologies in some chains are hidden. In this situation, the trace rank plot or trank plot is a better way to visualize the chains by plotting the ranked samples' distribution. The lowest sample gets rank "1". It is stacked histograms of ranked samples. In a "healthy" chain, these histograms should be reasonably uniform, without significant chain spiking above or below the others, like Fig. 5b.

Moreover, Gelman-Rubin statistics ($\widehat{R}$) is often applied in diagnosing the convergence of the Markov Chain. $\widehat{R}$ is the ratio of between-chain variance to within-chain variance, which is based on the concept that, if multiple chains have converged, there should be little variability between and within the chains. For convergence, $\widehat{R}$ should be approximately 1 ± 0.1 [73].

Several software MCMC packages have been developed since 1997, including the initial release of WinBUGS [119], one version of BUGS, a software package for performing Bayesian Inference Using Gibbs Sampling, before which new users had to create everything from scratch. WinBUGS is stable and still available but will not be further developed. OpenBUGS [120] is another version of BUGS, which will be developed further. Another early and preferred programming environment is Stan [121], featured by probability models, inference algorithms for model fitting and predictions, and posterior analysis. However, these three packages do not use general programming languages such as Python, so users must first learn their specialized programming language. In R, an abundance of packages was developed for the MCMC estimation. Some of them are the interfaces to specific software tools like RStan [122] and BRugs [123], while others are independent R packages without the need to understand any other language like Stan and BUGS. But some packages are designed with limited capability, e.g., only applicable for GP. Two elegant R packages are recommended here. The first one is "greta" [124], which uses Google TensorFlow directly in R. Simple examples and explanations are shown on its website to provide a straightforward way for beginners to build their own MCMC models. A plot function is provided by "greta" to visualize and check the relationship between the unknown parameters in forming an MCMC model to guarantee its corrections. The package is fast, even for the massive dataset, and runs on CPU clusters and GPUs. The other suggested package is "coda" [125]. It provides functions for summarizing and plotting the output from MCMC simulations, as well as diagnostic tests of convergence to the equilibrium distribution of the Markov Chain, which is more comprehensive and suitable for intermediate users.

4.7. Step 7 - validation and additional analysis

The final step of the framework of Bayesian inference in BEM is to estimate and validate the calibrated model and conduct further analysis based on the calibrated model. For the validation, there are two types of criteria: point estimation and probabilistic estimation. For point estimation, it is the same as the conventional calibration and validation method. As usual, the mean value or a mode value of the posterior distribution is selected as the representative point to do the estimation. Criteria like NMBE and CVRMSE are frequently applied, and their tolerance is shown in Table 2.

MBE is defined by the average of the differences of the simulated energy consumptions and the measured data for all the intervals over a given period.

$$MBE(\%) = \frac{\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)}{n} \times 100 \tag{11}$$

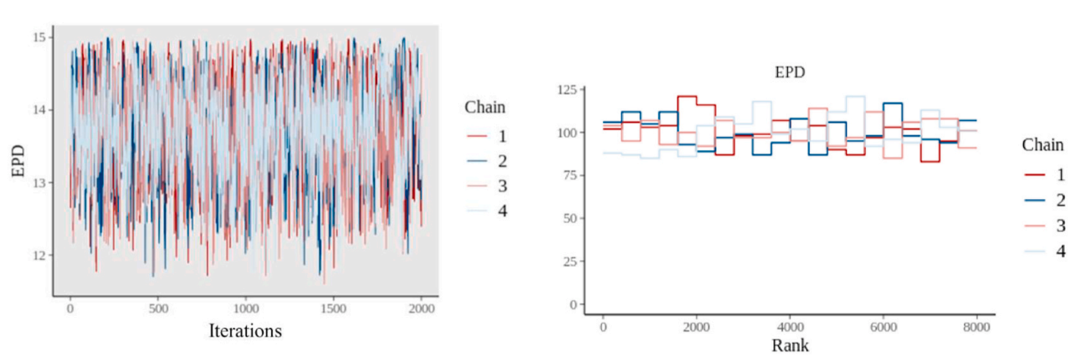
where M is the measured kilowatt-hours or fuel consumption during the time interval, S is the simulated kilowatt-hours or fuel consumption during the same time interval.

NMBE is a normalization of the MBE index. It quantifies the MBE by

Table 2
Acceptable calibration tolerances.

Standard/Guideline	Acceptable Value[a]			
	Monthly		Hourly	
	NMBE	CVRMSE	NMBE	CVRMSE
ASHRAE Guideline 14 [126]	±5%	15%	±10%	30%
IPMVP [127]	20%	–	5%	20%
FEMP [128]	±5%	15%	±10%	30%

[a] Lower values indicate better calibration.



(a)   (b)

Fig. 5. Examples of (a) trace plot; (b) trace rank plot.

dividing it by the mean of measured values ($\overline{y}$).

$$NMBE(\%) = \frac{1}{\overline{y}} \frac{\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)}{n} \times 100 \tag{12}$$

For probabilistic estimation, CRPS is an index to evaluate the performance of the whole distribution. It estimates how close the predictive distributions and corresponding observations are, and has been widely used in forecast verification [129], and have also been applied to building performance predictions [130]. When the predictive distribution is obtained from the MC simulation, the score can be calculated as:

$$CRPS = E_F|Y - y| - \frac{1}{2}E_F|Y - Y^{'}| \tag{13}$$

where F is the predictive distribution of random variable Y represented by the sample set, y is a single observation, E_F is the expectation over F, $Y^{'}$ is an independent random variable with identical distribution as Y. This identical distribution can be obtained by random permutations of the sample set F. A larger CRPS value indicates more discrepancy between the predictive and observed distributions. More details can be found in Ref. [129].

5. Conclusions and future work

This paper has reviewed the status and development of applying Bayesian inference in calibrating BEMs. Compared with the conventional deterministic calibration methods, the advantages of Bayesian inference calibration are:

1) When the calibration measurements are qualitatively/quantitatively insufficient, for traditional methods, the estimated model parameters can be far off from their original values; however, for Bayesian calibration, since the uncertainties are considered, the calibration results are more stable and reasonable;
2) For the traditional calibration method, the results are often deterministic. While for the Bayesian inference calibration method, the results are derived from quantitative stochastic analysis and with possibilities that can be regarded as a degree of belief. The Bayesian-based calibrated model is more comprehensive and reliable in its analysis.

Since Bayesian inference is a new calibration technique, which is fundamentally different from conventional approaches, and both Bayesian inference and MCMC algorithms involve many statistics and various options during the implementation, it can be quite challenging for new users to understand its underlying theory, methodology, and implementation. The learning curve is perhaps the most critical factor limiting the adoption of Bayesian calibration. This paper helps to enlighten beginners with explanations and details to ease their learning curves and facilitate the migration from the deterministic calibration to the stochastic one. A generic procedure of Bayesian inference calibration of BEMs is summarized. The corresponding implementation for each step of the process based on the R language is detailed as well. Here is a list of the conclusions, contributions, and future work in summary:

a. Depending on different levels of reliabilities, singular measurement points exist. This kind of measurement should be removed or adjusted to ensure the proper informativity of the measurement data.
b. Under the same conditions, different BEMs should attain consistent calibration results. However, the BEM selection should consider the model development feasibility, the calibration problem, and its further application. Since many simulations are needed for sensitivity analysis and Meta-model development, the feasibility of automatic parametric simulations and the attainment of input-output datasets should be considered.
c. Prior distributions of calibrated parameters can impact the calibration results, especially when the measurement data are insufficient. Informative distributions are suggested since they can provide more information. A pilot study is suggested for a Bayesian experimental design or Bayesian model update to reduce the data collection effort and the computational load.
d. Sensitivity analysis is a crucial step for Bayesian inference calibration since the determined calibrated model parameters will be selected based on the importance of ranking. The selection of the calibration parameter number is a balancing art. Although it is suggested that the number should be less than 10, more studies should be explored owing to its insufficient application cases.
e. When the measurements are redundant, an informative data identification process can be conducted to filter out the most informative combination of different data to reduce the computing time while maintaining the calibration performance.
f. Meta-model development can be critical, and a Meta-model could reduce the computing time dramatically. It shows that the GP model is relatively more accurate but needs longer computing time. In comparison, the MLR model is simpler and more computing-efficient, and its accuracy depends on the specific calibration scenario, so how to improve MLR accuracy still needs more future work.
g. For the MCMC algorithm, currently, NUTS seems one of the preferred methods. More efficient algorithms should be explored to increase sampling efficiency further and to ensure the chains' convergence in shorter steps. Convergence checking of the chains should be done to make sure the samplers come from the posterior distribution. In addition, the parameters of the MCMC model should be appropriately set up to aid the convergence.
h. It is still an issue whether the calibrated building parameter distributions can accurately represent the real distribution. Zhao et al. [79] argued that the calibrated building parameter distributions should be regarded as the "best guess" of the real world, which is agreed by McElreath [50]. Future research should be conducted on the relationship between numerically estimated distributions and actual real-world ones. Although it may be possible to develop more accurate building energy models to reflect reality with the advancements of science and computer technology, there is a trade-off between the effort to develop the model and the added value from the increased accuracy. A balance should be maintained depending on applications and real needs. Again, this also shows the necessity of more high-quality measurement data and more Bayesian calibrated BEMs applications in the near future.

Currently, the application of Bayesian inference to BEM calibration is still limited. There are no sufficient studies to support specific conclusions from previous studies, such as the relationship between the calibrated parameter number and the phenomenon of "over-fitting"/"over-parameterization," which means fewer/more model parameters than necessary. In addition, how to apply Bayesian inference to a BEM for a particular application can be a challenge. So more applications and demonstrations are needed to provide more examples for new users to master this technique. Also, it seems that no studies have yet compared the calibration performance using different types of BEMs (e.g., white model, gray model, and black model), especially at the urban scale. More studies are necessary not only on the improvement of Bayesian inference and MCMC algorithms from a statistical perspective but also on the strengths and weaknesses of Bayesian inference that should be explored more in the field of building energy modeling and other building-related topics.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This paper was made possible by an NPRP award [#NPRP11S-1208-170073] from the Qatar National Research Fund (a member of The Qatar Foundation) and the Advancing Climate Change Science in Canada Program [#ACCPJ 535986-18] from the Natural Sciences and Engineering Research Council of Canada. The statements made herein are solely the responsibility of the authors.

References

[1] U.S. Energy Information Administration. International energy Outlook. Outlook 2019;484:70–99. https://www.eia.gov/outlooks/ieo/pdf/ieo2019.pdf.
[2] U.S. Energy Information Administration. How much energy is consumed in U.S. buildings?. 2019. https://www.eia.gov/tools/faqs/faq.php?id=86&t=1.
[3] University BERC of T. China building energy use 2018. 2018.
[4] Rousselot M. Energy efficiency trends in buildings. 2018.
[5] Buckman AH, Mayfield M, Beck SBM. What is a smart building? Smart Sustain Built Environ 2014;3:92–109. https://doi.org/10.1108/SASBE-01-2014-0003.
[6] Batty M, Axhausen KW, Giannotti F, Pozdnoukhov A, Bazzani A, Wachowicz M, et al. Smart cities of the future. Eur Phys J Spec Top 2012;214:481–518. https://doi.org/10.1140/epjst/e2012-01703-3.
[7] Albino V, Berardi U, Dangelico RM. Smart cities: definitions, dimensions, performance, and initiatives. J Urban Technol 2015;22:3–21. https://doi.org/10.1080/10630732.2014.942092.
[8] Stuart Borlase. Smart grids: infrastructure, technology, and solutions. Taylor & Francis; 2013. https://doi.org/10.1007/978-1-4615-0981-3_11.
[9] Gubbi J, Buyya R, Marusic S, Palaniswami M. Internet of Things (IoT): a vision, architectural elements, and future directions. Future Generat Comput Syst 2013; 29:1645–60. https://doi.org/10.1016/J.FUTURE.2013.01.010.
[10] Tian W, Heo Y, de Wilde P, Li Z, Yan D, Park CS, et al. A review of uncertainty analysis in building energy assessment. Renew Sustain Energy Rev 2018;93: 285–301. https://doi.org/10.1016/j.rser.2018.05.029.
[11] Crawley DB, Lawrie LK, Winkelmann FC, Buhl WF, Huang YJ, Pedersen CO, et al. EnergyPlus: creating a new-generation building energy simulation program. Energy Build 2001. https://doi.org/10.1016/S0378-7788(00)00114-6.
[12] Chaudhary G, New J, Sanyal J, Im P, O'Neill Z, Garg V. Evaluation of "Autotune" calibration against manual calibration of building energy models. Appl Energy 2016;182:115–34. https://doi.org/10.1016/j.apenergy.2016.08.073.
[13] Coakley D, Raftery P, Keane M. A review of methods to match building energy simulation models to measured data. Renew Sustain Energy Rev 2014;37:123–41. https://doi.org/10.1016/j.rser.2014.05.007.
[14] Chaudhary G, New J, Sanyal J, Im P, O'Neill Z, Garg V. Evaluation of "Autotune" calibration against manual calibration of building energy models. Appl Energy 2016. https://doi.org/10.1016/j.apenergy.2016.08.073.
[15] Chong A, Xu W, Chao S, Ngo NT. Continuous-time Bayesian calibration of energy models using BIM and energy data. Energy Build 2019;194:177–90. https://doi.org/10.1016/j.enbuild.2019.04.017.
[16] Valuation Organization Efficiency. International Performance Measurement & Verification Protocol: concepts and options for determining energy and water savings, vol. 1; 2010.
[17] Hansen SJ, Brown JW. Investment Grade Energy Audit: making smart energy choices. Lilburn, GA: Fairmont Press, Inc; 2004.
[18] Quan SJ, Li Q, Augenbroe G, Brown J, Yang PP-J. Urban data and building energy modeling: a GIS-based urban building energy modeling system using the urban-EPC engine. Plan. Support Syst. Smart Cities 2015:447–69.
[19] Gelman A, Carlin J, Stern H, Dunson D, Vehtari A, Rubin D. Bayesian data analysis. Chapman and Hall/CRC; 2013.
[20] Bayes TR. An essay towards solving a problem in the doctrine of chances. Phil Trans Roy Soc Lond 1763;53:370–418.
[21] Gamerman Dani, Lopes Hedibert F. Markov chain Monte Carlo: stochastic simulation for Bayesian inference. Chapman and Hall/CRC; 2006.
[22] Kennedy MC, O'Hagan A. Bayesian calibration of computer models. J R Stat Soc Ser B (Statistical Methodol 2001;63:425–64. https://doi.org/10.1111/1467-9868.00294.
[23] Van Oijen M, Rougier J, Smith R. Bayesian calibration of process-based forest models: bridging the gap between models and data. Tree Physiol 2005;25: 915–27. https://doi.org/10.1093/treephys/25.7.915.
[24] Steel D. Bayesian statistics in radiocarbon calibration. Philos Sci 2001;68: S153–64. https://doi.org/10.1086/392905.
[25] Lehuger S, Gabrielle B, Oijen M van, Makowski D, Germon JC, Morvan T, et al. Bayesian calibration of the nitrous oxide emission module of an agro-ecosystem model. Agric Ecosyst Environ 2009. https://doi.org/10.1016/j.agee.2009.04.022.
[26] Arhonditsis GB, Papantou D, Zhang W, Perhar G, Massos E, Shi M. Bayesian calibration of mechanistic aquatic biogeochemical models and benefits for environmental management. J Mar Syst 2008. https://doi.org/10.1016/j.jmarsys.2007.07.004.
[27] Kuczera G, Kavetski D, Renard B, Thyer M. A limited-memory acceleration strategy for MCMC sampling in hierarchical Bayesian calibration of hydrological models. Water Resour Res 2010;46:1–6. https://doi.org/10.1029/2009WR008985.
[28] Hall JW, Manning LJ, Hankin RKS. Bayesian calibration of a flood inundation model using spatial data. Water Resour Res 2011;47:1–14. https://doi.org/10.1029/2009WR008541.
[29] Tianfang Xu, Valocchi Albert J. A Bayesian approach to improved calibration and prediction of groundwater models with structural error, vol. 51; 2015. p. 9290–311. https://doi.org/10.1002/2015WR017200.A.
[30] Van Hinsbergen CPIJ, Van Lint HWC, Hoogendoorn SP, Van Zuylen HJ. Bayesian calibration of car-following models. IFAC Proc 2009. https://doi.org/10.3182/20090902-3-US-2007.0049.
[31] Whyte S, Walsh C, Chilcott J. Bayesian calibration of a natural history model with application to a population model for colorectal cancer. Med Decis Making 2011; 31:625–41. https://doi.org/10.1177/0272989X10384738.
[32] Heo Y. Bayesian calibration of building energy models for energy retrofit decision-making under uncertainty. 2011. p. 1–116.
[33] Sadeghi SA, Lee S, Karava P, Bilionis I, Tzempelikos A. Bayesian classification and inference of occupant visual preferences in daylit perimeter private offices. Energy Build 2018;166:505–24. https://doi.org/10.1016/j.enbuild.2018.02.010.
[34] Lee S, Karava P, Tzempelikos A, Bilionis I. Inference of thermal preference profiles for personalized thermal environments with actual building occupants. Build Environ 2019;148:714–29. https://doi.org/10.1016/j.buildenv.2018.10.027.
[35] Carstens H, Xia X, Yadavalli S. Bayesian energy measurement and verification analysis. Energies 2018;11:1–20. https://doi.org/10.3390/en11020380.
[36] Liu Y, Qin H, Zhang Z, Pei S, Jiang Z, Feng Z, et al. Probabilistic spatiotemporal wind speed forecasting based on a variational Bayesian deep learning model. Appl Energy 2020;260:114259. https://doi.org/10.1016/j.apenergy.2019.114259.
[37] Liu Y, Qin H, Zhang Z, Pei S, Wang C, Yu X, et al. Ensemble spatiotemporal forecasting of solar irradiation using variational Bayesian convolutional gate recurrent unit network. Appl Energy 2019;253:113596. https://doi.org/10.1016/j.apenergy.2019.113596.
[38] Chong A, Lam KP, Pozzi M, Yang J. Bayesian calibration of building energy models with large datasets. Energy Build 2017;154:343–55. https://doi.org/10.1016/j.enbuild.2017.08.069.
[41] Zhang C, Butepage J, Kjellstrom H, Mandt S. Advances in variational inference. IEEE Trans Pattern Anal Mach Intell 2019;41:2008–26. https://doi.org/10.1109/TPAMI.2018.2889774.
[42] Kim YJ, Kim KC, Park CS, Kim IH. Deterministic vs. Stochastic calibration of energy simulation model for an existing building. Proceeding ASim 2014;2014.
[43] Pavlak GS, Florita AR, Henze GP, Rajagopalan B. Comparison of traditional and bayesian calibration techniques for gray-box modeling. J Architect Eng 2014;20. https://doi.org/10.1061/(ASCE)AE.1943-5568.0000145.
[44] Kim YJ, Park CS. Stepwise deterministic and stochastic calibration of an energy simulation model for an existing building. Energy Build 2016. https://doi.org/10.1016/j.enbuild.2016.10.009.
[45] Zhang Z, Chong A, Pan Y, Zhang C, Lam KP. Whole building energy model for HVAC optimal control: a practical framework based on deep reinforcement learning. Energy Build 2019;199:472–90. https://doi.org/10.1016/j.enbuild.2019.07.029.
[46] Rouchier S, Rabouille M, Oberlé P. Calibration of simplified building energy models for parameter estimation and forecasting: stochastic versus deterministic modelling. Build Environ 2018;134:181–90. https://doi.org/10.1016/j.buildenv.2018.02.043.
[47] Muehleisen RT, Bergerson J. Bayesian calibration-what, why and how. In: 4th int high perform build conf; 2016. p. 167.
[48] Riddle M, Muehleisen RT. A guide to Bayesian calibration of building energy models. In: 2014 ASHRAE/IBPSA-USA build simul conf; 2014. p. 276–83. https://doi.org/10.13140/2.1.1674.9127.
[49] Lim H, Zhai ZJ. Review on stochastic modeling methods for building stock energy prediction. Build Simul 2017;10:607–24. https://doi.org/10.1007/s12273-017-0383-y.
[50] McElreath R. Statistical rethinking: a Bayesian course with examples in R and Stan. Chapman and Hall/CRC; 2018.
[51] Lim H. Prediction of urban-scale building energy performance with a stochastic-deterministic-coupled approach. University of Colorado at Boulder; 2017.
[52] Booth AT, Choudhary R, Spiegelhalter DJ. Handling uncertainty in housing stock models. Build Environ 2012;48:35–47. https://doi.org/10.1016/j.buildenv.2011.08.016.
[53] Booth AT, Choudhary R, Spiegelhalter DJ. A hierarchical bayesian framework for calibrating micro-level models with macro-level data. J Build Perform Simul 2013;6:293–318. https://doi.org/10.1080/19401493.2012.723750.
[54] Tian W, Choudhary R. A probabilistic energy model for non-domestic building sectors applied to analysis of school buildings in greater London. Energy Build 2012;54:1–11. https://doi.org/10.1016/j.enbuild.2012.06.031.
[55] Yamaguchi Y, Suzuki Y, Choudhary R, Booth A. Urban-scale energy modeling of food supermarket considering uncertainty. In: Proc. BS; 2013.
[56] Booth AT, Choudhary R. Decision making under uncertainty in the retrofit analysis of the UK housing stock: implications for the Green Deal. Energy Build 2013;64:292–308. https://doi.org/10.1016/j.enbuild.2013.05.014.
[57] Booth A, Choudhary R. Calibrating micro-level models with macro-level data using bayesian regression analysis. In: Proc build simul 2011 12th conf int build perform simul assoc; 2011. p. 641–8.
[58] Heo Y, Augenbroe G, Choudhary R. Risk analysis of energy-efficiency projects based on bayesian calibration of building energy models. In: 12th conf. Int. Build. Perform. Simul. Assoc.; 2011. p. 2579–86.

[59] Heo Y, Choudhary R, Augenbroe GA. Calibration of building energy models for retrofit analysis under uncertainty. Energy Build 2012;47:550–60. https://doi.org/10.1016/j.enbuild.2011.12.029.

[60] Heo Y, Augenbroe G, Choudhary R. Quantitative risk management for energy retrofit projects. J Build Perform Simul 2013;6:257–68. https://doi.org/10.1080/19401493.2012.706388.

[61] Heo Y, Graziano DJ, Guzowski L, Muehleisen RT. Evaluation of calibration efficacy under different levels of uncertainty. J Build Perform Simul 2015;8:135–44. https://doi.org/10.1080/19401493.2014.896947.

[62] Heo Y, Augenbroe G, Graziano D, Muehleisen RT, Guzowski L. Scalable methodology for large scale building energy improvement: relevance of calibration in model-based retrofit analysis. Build Environ 2015;87:342–50. https://doi.org/10.1016/j.buildenv.2014.12.016.

[63] Tian W, Yang S, Li Z, Wei S, Pan W, Liu Y. Identifying informative energy data in Bayesian calibration of building energy models. Energy Build 2016;119:363–76. https://doi.org/10.1016/j.enbuild.2016.03.042.

[64] Choudhary R. Energy analysis of the non-domestic building stock of Greater London. Build Environ 2012;51:243–54. https://doi.org/10.1016/j.buildenv.2011.10.006.

[65] Choudhary R, Tian W. Influence of district features on energy consumption in non-domestic buildings. Build Res Inf 2014;42:32–46. https://doi.org/10.1080/09613218.2014.832559.

[66] Kim YJ, Yoon SH, Park CS. Stochastic comparison between simplified energy calculation and dynamic simulation. Energy Build 2013;64:332–42. https://doi.org/10.1016/j.enbuild.2013.05.026.

[67] Kim YJ, Ahn KU, Park CS. Decision making of HVAC system using Bayesian Markov chain Monte Carlo method. Energy Build 2014;72:112–21. https://doi.org/10.1016/j.enbuild.2013.12.039.

[68] Kim YJ. Challenging issues in stochastic calibration based on bayesian paradigm for building energy model. Int J Smart Home 2015;9:127–42. https://doi.org/10.14257/ijsh.2015.9.5.13.

[69] Li Q, Gu L, Augenbroe G, Jeff Wu CF, Brown J. A Generic approach to calibrate building energy models under uncertainty using Bayesian inference. In: 14th int conf IBPSA - build simul 2015, BS 2015, conf proc; 2015. p. 2913–21.

[70] Li Q, Gu L, Augenbroe G, Jeff Wu CF, Brown J. Calibration of dynamic building energy models with multiple responses using Bayesian inference and linear regression models. Energy Procedia 2015;78:979–84. https://doi.org/10.1016/j.egypro.2015.11.037.

[71] Li Q, Augenbroe G, Brown J. Assessment of linear emulators in lightweight Bayesian calibration of dynamic building energy models for parameter estimation and performance prediction. Energy Build 2016;124:194–202. https://doi.org/10.1016/j.enbuild.2016.04.025.

[72] Chong A, Lam KP. Uncertainty analysis and parameter estimation of HVAC systems in building energy models. In: 14th int conf IBPSA - build simul 2015, BS 2015, conf proc; 2015. p. 2788–95.

[73] Chong A, Poh Lam K. A comparison of MCMC algorithms for the bayesian calibration of building energy models. In: 15th IBPSA conf; 2017. p. 494–503.

[74] Chong A, Menberg K. Guidelines for the Bayesian calibration of building energy models. Energy Build 2018;174:527–47. https://doi.org/10.1016/j.enbuild.2018.06.028.

[75] Chong A, Xu W, Chao S, Ngo NT. Continuous-time Bayesian calibration of energy models using BIM and energy data. Energy Build 2019;194:177–90. https://doi.org/10.1016/j.enbuild.2019.04.017.

[76] Henze GP, Pavlak GS, Florita AR, Dodier RH, Hirsch AI. An energy signal tool for decision support in building energy systems. Appl Energy 2015;138:51–70. https://doi.org/10.1016/j.apenergy.2014.10.029.

[77] Braulio-Gonzalo M, Juan P, Bovea MD, Ruá MJ. Modelling energy efficiency performance of residential building stocks based on Bayesian statistical inference. Environ Model Software 2016;83:198–211. https://doi.org/10.1016/j.envsoft.2016.05.018.

[78] Kang Y, Krarti M. Bayesian-Emulator based parameter identification for calibrating energy models for existing buildings. Build Simul 2016;9:411–28. https://doi.org/10.1007/s12273-016-0291-6.

[79] Zhao F, Lee SH, Augenbroe G. Reconstructing building stock to replicate energy consumption data. Energy Build 2016;117:301–12. https://doi.org/10.1016/j.enbuild.2015.10.001.

[80] Sokol J, Cerezo Davila C, Reinhart CF. Validation of a Bayesian-based method for defining residential archetypes in urban building energy models. Energy Build 2017;134:11–24. https://doi.org/10.1016/j.enbuild.2016.10.050.

[81] Kristensen MH, Choudhary R, Petersen S. Bayesian calibration of building energy models: comparison of predictive accuracy using metered utility data of different temporal resolution. Energy Procedia 2017;122:277–82. https://doi.org/10.1016/j.egypro.2017.07.322.

[82] Kristensen MH, Choudhary R, Pedersen RH, Petersen S. Bayesian calibration of residential building clusters using A single geometric building representation. Proc 15th IBPSA Conf 2017:1294–303.

[83] Lim H. Prediction of urban-scale building energy performance with a stochastic-deterministic-coupled approach. 2017.

[84] Lim H, Zhai ZJ. Comprehensive evaluation of the influence of meta-models on Bayesian calibration. Energy Build 2017;155:66–75. https://doi.org/10.1016/j.enbuild.2017.09.009.

[85] Lim H, Zhai Z (John). Influences of energy data on Bayesian calibration of building energy model. Appl Energy 2018;231:686–98. https://doi.org/10.1016/j.apenergy.2018.09.156.

[86] Menberg K, Heo Y, Choudhary R. Efficiency and reliability of Bayesian calibration of energy supply system models university of Cambridge , department of engineering , Cambridge , UK. Cambridge , UK: University of Cambridge , Department of Architecture; 2017. p. 1594–603.

[87] Menberg K, Heo Y, Choudhary R. Influence of error terms in Bayesian calibration of energy system models. J Build Perform Simul 2019;12:82–96. https://doi.org/10.1080/19401493.2018.1475506.

[88] Yuan J, Nian V, Su B. A meta model based bayesian approach for building energy models calibration. Energy Procedia 2017;143:161–6. https://doi.org/10.1016/j.egypro.2017.12.665.

[89] Yuan J, Nian V, Su B, Meng Q. A simultaneous calibration and parameter ranking method for building energy models. Appl Energy 2017;206:657–66. https://doi.org/10.1016/j.apenergy.2017.08.220.

[90] Yuan J, Nian V, Su B. Evaluation of cost-effective building retrofit strategies through soft-linking a metamodel-based Bayesian method and a life cycle cost assessment method. Appl Energy 2019;253:113573. https://doi.org/10.1016/j.apenergy.2019.113573.

[91] Raillon L, Ghiaus C. An efficient Bayesian experimental calibration of dynamic thermal models. Energy 2018;152:818–33. https://doi.org/10.1016/j.energy.2018.03.168.

[92] Zhang Z, Chong A, Poh Lam K, Pan Y, Zhang C, Lu S. A deep reinforcement learning approach to using whole building energy model for HVAC optimal control ASHRAE multidisciplinary task group on occupant behavior in buildings view project international energy agency energy in buildings and communities program. 2018.

[93] Chen S, Friedrich D, Yu Z, Yu J. District heating network demand prediction using a physics-based energy model with a bayesian approach for parameter calibration. Energies 2019;12:3408.

[94] Rysanek AM, Fonseca JA, Schlueter A. Bayesian calibration of a building energy model by stochastic optimisation of root-mean square error. Prepr Submitt to Appl Energy 2019. https://doi.org/10.3929/ethz-b-000349836.

[95] Yi DH, Kim DW, Park CS. Parameter identifiability in Bayesian inference for building energy models. Energy Build 2019;198:318–28. https://doi.org/10.1016/j.enbuild.2019.06.012.

[96] Ahmadi S, hossien Fakehi A, vakili A, Haddadi M, Iranmanesh SH. A hybrid stochastic model based Bayesian approach for long term energy demand managements. Energy Strateg Rev 2020;28:100462. https://doi.org/10.1016/j.esr.2020.100462.

[97] Zhu C, Tian W, Yin B, Li Z, Shi J. Uncertainty calibration of building energy models by combining approximate Bayesian computation and machine learning algorithms. Appl Energy 2020;268:115025. https://doi.org/10.1016/j.apenergy.2020.115025.

[98] Zhu C, Tian W, Wilde P De, Yin B. Approximate Bayesian computation in parameter estimation of building energy models. Springer Singapore; 2020. https://doi.org/10.1007/978-981-13-9528-4.

[99] Metropolis N, Rosenbluth WA, Rosenbluth NM, Teller HA, Teller E. Equation of state calculations by fast computing machines. Jounal Chem Phys 1953;21:1087–92.

[100] Geman S, Geman D. Stochastic relaxation, Gibbs distributions, and the bayesian restoration of images. IEEE Trans Pattern Anal Mach Intell 1984. PAMI-6.

[101] Hoffman DM, Gelman A. The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. J Mach Learn Res 2014;15:1593–623.

[102] Ihaka R, Gentleman RR. A language for data analysis and graphics. J Comput Graph Stat 1996;5:299–314.

[103] Team RCR. A language and environment for statistical computing. 2013.

[104] Winkelmann FC, Birdsall BE, Buhl WF, Ellington KL, Erdem AE, Hirsch JJ, et al. DOE-2 supplement. 1993. United States, Version 2.1E.

[105] Klein SA. TRNSYS-A transient system simulation program. Univ Wisconsin-Madison, Eng Exp Stn Rep 1988. 38–12.

[106] Strachan P. ESP-r: summary of validation studies. Analysis 2000. 0–8.

[107] Wei T. A review of sensitivity analysis methods in building energy analysis. Renew Sustain Energy Rev 2013;20:411–9. https://doi.org/10.1016/j.rser.2012.12.014.

[108] Jia H. eplusr n.d. https://hongyuanjia.github.io/eplusr/.

[109] Menberg K, Heo Y, Choudhary R. Sensitivity analysis methods for building energy models: comparing computational costs and extractable information. Energy Build 2016. https://doi.org/10.1016/j.enbuild.2016.10.005.

[110] Tian W, Wang Q, Song J, Wei S. Calibrating dynamic building energy models using regression model and Bayesian analysis in building retrofit projects. ESim 2014; 2014.

[111] Manfren M, Aste N, Moshksar R. Calibration and uncertainty analysis for computer models - a meta-model based approach for integrated building energy simulation. Appl Energy 2013;103:627–41. https://doi.org/10.1016/j.apenergy.2012.10.031.

[112] Yang J, Rivard H, Zmeureanu R. Building energy prediction with adaptive artificial neural networks. In: IBPSA 2005 - int build perform simul assoc 2005; 2005. p. 1401–8.

[113] Kalogirou SA, Bojic M. Artificial neural networks for the prediction of the energy consumption of a passive solar building. Energy 2000. https://doi.org/10.1016/S0360-5442(99)00086-9.

[114] Aydinalp M, Ugursal VI, Fung AS. Modeling of the space and domestic hot-water heating energy-consumption in the residential sector using neural networks. Appl Energy 2004. https://doi.org/10.1016/j.apenergy.2003.12.006.

[115] Eisenhower B, O'Neill Z, Narayanan S, Fonoberov VA, Mezić I. A methodology for meta-model based optimization in building energy models. Energy Build 2012. https://doi.org/10.1016/j.enbuild.2011.12.001.

[116] Tian W, Song J, Li Z, de Wilde P. Bootstrap techniques for sensitivity analysis and model selection in building thermal performance analysis. Appl Energy 2014. https://doi.org/10.1016/j.apenergy.2014.08.110.
[117] Kim YJN, Aim KU, Park CS, Kim IH. Gaussian emulator for stochastic optimal design of a double glazing system. In: Proc BS 2013 13th conf int build perform simul assoc; 2013. p. 2217–24.
[118] Lunn D, Spiegelhalter D, Thomas A, Best N. The BUGS project: evolution, critique and future directions. Stat Med 2009;28:3049–67. https://doi.org/10.1002/sim.
[119] Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. Stat Comput 2000;10:325–37. https://doi.org/10.1093/ohr/oht039.
[120] Surhone LM, Tennoe MT, Henssonow SF. OpenBUGS. Betascript Publ; 2010.
[121] Stan n.d. https://mc-stan.org/(accessed April 7, 2020).
[122] Stan Development Team. RStan: the R interface to Stan. R News; 2019.
[123] Thomas A, O'Hara B, Ligges U, Sturtz S. Making BUGS open. R News 2006;6:12–7.
[124] Greta introduction [n.d].
[125] Martyn P, Nicky B, Kate C, Karen V. CODA: convergence diagnosis and output analysis for MCMC. R News 2006;6:7–11.
[126] Guideline. ASHRAE. Guideline 14-2015, measurement of energy and demand savings. 2015. Atlanta, Georgia.
[127] EVO. International performance measurement & verification protocol. Effic Valuat Organ 2007.
[128] U.S. Department of Energy. M&V guidelines: measurement and verification for performance-based contracts. 2015., Version 4.0.
[129] Gneiting T, Raftery AE. Strictly proper scoring rules, prediction, and estimation. J Am Stat Assoc 2007;102:359–78. https://doi.org/10.1198/016214506000001437.
[130] Sun Y. Closing the building energy performance gap by improving our predictions. Georgia Institute of Technology; 2014.