SimBuild
2024

Eleventh National Conference of IBPSA-USA
Denver, Colorado
May 21 – 23, 2024

Advancing Building Energy Modeling: An Open-Source Bayesian Calibration Framework for Non-Residential Buildings

Katalin Fülep[1], Siling Chen[1], Stefan Brandt[2] and Rita Streblow[1]
[1]Technische Universität Berlin, Einstein Center Digital Future, Berlin, Germany
[2]Technische Universität Berlin, Hermann-Rietschel-Institut, Berlin, Germany

Abstract

This study presents a comprehensive approach to calibrating a building energy model using Bayesian calibration. The calibration process involves identifying the most influential parameters, defining prior distributions, and developing a meta-model for efficiency. Furthermore, a likelihood function is defined to quantify the model's ability to explain observed data, and Bayesian calibration is performed, resulting in posterior probability distributions for both parameters and simulation results. The analysis demonstrates that this calibration framework is efficient in fine-tuning a building energy model, resulting in simulation outputs closely aligned with metered data. Concurrently, it assigns a specific level of uncertainty to the obtained results. However, when the calibrated model is tested against a separate dataset, it exhibits challenges in predicting the observed data. The particular Case Study presented highlights the importance of high-quality and consistent observed data for better prediction ability. Furthermore, it also highlights additional potential influencing factors currently under investigation, contributing to the development of an effective working tool.

Introduction

The Paris Agreement seeks to limit global warming to 2 degrees Celsius above pre-industrial levels (UNFCCC, 2018). According to the International Energy Agency, in 2021 the operation of buildings accounted for 27 % of global energy-related emissions and for 30 % of the world´s total final energy consumption (IEA 2022). The improvement of building energy efficiency is critical in mitigating climate change. Retrofitting existing buildings can have a significant impact on reducing energy consumption and greenhouse gas emissions, given their large share of the building stock and the energy they consume (Mukhtar et al., 2021).

Building energy models are essential for the estimation of energy savings for a proposed retrofit. Yet, simulation results from building energy models often deviate from actual observed or metered energy consumption of buildings. Therefore, calibration of these models is a common practice (Hong et al., 2018), yet it requires expert knowledge and is often a time consuming and labor-intensive process. In addition, these methods do not incorporate any estimation for the uncertainty in the results (Riddle and Muehleisen, 2014). This complicates the process for decision-makers when it comes to assessing the financial viability of investing in building retrofit projects.

A method is examined in this paper which enables the calibration of building energy models and concurrently quantifies the uncertainty of the results. The method is based on Bayes Theorem and is strongly influenced by the framework that Kennedy and O´Hagan formulated (Kennedy and O'Hagan, 2001).

There has been a growing interest in this approach, which is also reflected by the number of generated research papers (Hou et al., 2021). Calibration has been conducted for individual buildings involving various time resolutions such as hourly (Chong et al, 2017) and monthly intervals (Heo et al., 2012).

Similarly, calibration of building stocks has been assessed for various time resolutions, including hourly (Remmen, 2021) and annual intervals (Wang et al., 2020). Yet, there is still an absence of a well-defined framework and an easily applicable tool for calibrating building energy models. This restricts the accessability, applicability and transparability to other case studies. (Chong and Menberg, 2018) (Heo et al., 2012). Addressing this, we aim to develop a practical open-source modular calibration framework. Its integrated building energy model with a reduced number of input parameters can support decision-makers without a deep scientific background to attain more reliable simulation results. Furthermore, the framework can be adopted by planners integrating their own building energy model but without the need of understanding the mathematical details of the calibration process.

Methodology

Bayes´ Theorem

Bayes' Theorem serves as a fundamental tool for refining our understanding of specific events or parameters in light of new evidence or data. This theorem enables us to adjust a parameter, considering its capacity to align with observed data and the inherent uncertainty associated with that parameter. By integrating these two critical aspects, we enhance our ability to estimate reality effectively (Riddle and Muehleisen, 2014).

Bayes' Theorem states that the posterior parameter distribution, denoted as P(θ ∣ y), is directly proportional to the likelihood function, P(y ∣ θ), and the prior probability distribution of the parameter, P(θ):

$$P(\theta|y) \propto P(y|\theta) \times P(\theta). \qquad (1)$$

Posterior parameter distribution: This signifies the updated beliefs regarding the parameter, shaped by the observed data.

Likelihood function: It quantifies the probability of the observed data given a specific parameter value.

Prior probability distribution: This reflects the initial beliefs about the parameter.

Framework for Bayesian calibration

In this study, we propose a framework consisting of eight steps to construct a Building Energy Model, apply Bayesian Calibration, and evaluate the model efficacy. Based on this framework, we build up a toolchain composed of open-source libraries written in Python. The proposed framework, together with the key Python libraries implemented in our toolchain are illustrated in Figure 1. Upon full implementation and testing, this toolchain will be released open source.

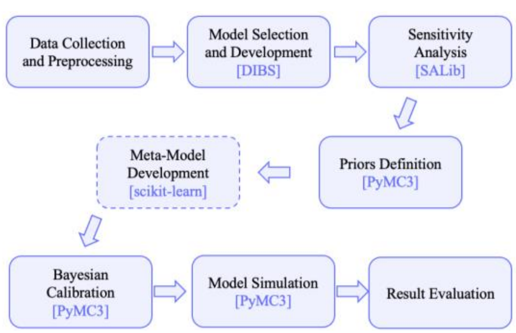


Figure 1 The proposed framework for Bayesian calibration of a building energy model. The related Python packages implemented into the toolchain include DIBS (Bischof et al., 2022), SALib (Herman and Usher, 2017), scikit-learn (Pedregosa et al., 2011), and PyMC3 (Salvatier et al., 2016)

More specifically, the eight steps are described below:

Step 1: Data Collection and Preprocessing

The initial step entails gathering all pertinent information related to the building, including (1) information about the building envelope; (2) information about the technical systems; (3) metered energy data; (4) weather data.

Step 2: Model Selection and Development

A building energy model, to be later calibrated, is selected depending on aspects such as input and output parameters, computational effort, etc. Upon model selection, the model is preliminarily set up with the required variables estimated.

In this study, an open-source model “Dynamic ISO Building Simulator (DIBS)” (Bischof et al., 2022) is employed. DIBS is a dynamic building energy simulator that utilizes a resistance-capacity model to compute heating, cooling and electricity demand for non-residential buildings with an hourly resolution. The consideration of the German building energy ordinance, ISO 13790:2008 for hourly user schedules and simplified modeling of building energy makes DIBS an ideal model for our case study.

Step 3: Sensitivity Analysis

A sensitivity analysis is conducted to identify the variables that have the most significant impact on the model's output. The identified variables are selected as input variables for the Bayesian Calibration to maintain manageable computational time and prevent overparameterization (Chong and Menberg, 2018).

In this study, a variance-based Sobol sensitivity analysis is conducted, employing the total Sobol Sensitivity Index. This index not only captures the sensitivity of individual parameters but also considers the interactions between all parameters. Depending on the settings, we select three to five variables with the highest total Sobol Sensitivity Index and only these will be included in the model calibration.

Step 4: Priors Definition

In this step, prior probability distributions are defined and assigned to uncertain model parameters selected in Step 3. These distributions encapsulate our beliefs about parameter values prior to any data observation. Usually the prior probability distributions can be categorized into informative (Normal distribution with low standard deviation), weakly informative (Normal distribution with high standard deviation) and uninformative (Uniform distribution) (Chong and Menberg, 2018). These can be defined by collected information from literature, expertise knowledge, and/or extracted from a database.

In our study, we assume a normal distribution for all parameters, with their means extracted from the case study database. Additionally, lower and upper bounds are defined to encapsulate the possible parameter value range. The standard deviation is initially determined so that the 90% confidence interval entails this range.

Step 5: Meta-Model Development

The Bayesian Calibration process contains thousands to millions of simulation iterations. Consequently, a model of shorter runtime can enormously reduce the computational and time requirement. To this purpose, implementing meta-models that shorten the simulation process can greatly enhance the calibration efficiency (Chong and Menberg, 2018). Implementing a meta-model is not compulsory, but recommended due to its potential.

In our proposed toolchain, a meta-model is created using Gaussian regression. A Latin Hypercube sampling is employed, encompassing the calibration parameter intervals established by the lower and upper bounds identified in Step 4. Each parameter combination undergoes simulation in DIBS, resulting in a sample of the corresponding simulation outcomes. Subsequently, a Gaussian regression is applied to these parameter combinations and their associated simulation results. This meta-model is preserved in a pickled format, allowing for convenient access.

Step 6: Bayesian Calibration

In the Bayesian calibration process, prior beliefs are combined with the likelihood function to revise our understanding of parameter values, resulting in a posterior probability distribution. As the most important process within our proposed framework, the following procedures are conducted in this step:

(1) Define the likelihood function

Kennedy and O'Hagan (2001) originally formulated a mathematical framework for Bayesian calibration. The likelihood calculation is based on a model that describes the relationship between the observed data (y), the calibration parameters (θ) and the observable inputs (x). The likelihood function quantifies how well the model with its parameters explains the observed data:

$$y(x) = \eta(x, \theta) + \delta(x) + \varepsilon(x). \qquad (2)$$

x: Represents known input parameters, such as climate data.
θ: Represents the unknown input parameters, i.e., the calibration parameters.
y(x): Denotes the observed data.
η(x, θ): Indicates the model's output, given both the known and unknown parameters.
δ(x): Accounts for model inadequacy, addressing errors inherent in the model.
ε(x): Reflects measurement error, encompassing inaccuracies in the observed data.

In other words, the likelihood function is introduced to quantify the likelihood of obtaining the measured data given the model and parameter values. This function connects the model's predictions to the observed data, accounting for measurement errors.

(2) Calibrate and update the prior probability distributions of the identified variables by the likelihood and measured data

This iterative procedure employs the Markov Chain Monte Carlo method, drawing samples from the prior parameter distributions. Through model simulations, distributions of corresponding model results are acquired, facilitating the calculation of the conditional probability of observing metered data given the parameter values. Based on predefined acceptance criteria, the mean of the prior parameter distribution is either retained or replaced.

(3) Acquire posterior probability distributions of the variables

Accepted mean values lead to the formation of the posterior probability distribution for the parameters once the calibration process concludes. In this step, the posterior probability distributions are obtained for the identified variables.

Step 7: Posterior Model Simulation

In Step 6, the posterior parameter probability distributions are obtained. Following a similar procedure, a probability distribution of the calibrated model results can be attained with Monte Carlo simulations. With this, the certainty of the calibrated model results can be drawn.

Step 8: Result Evaluation

The calibrated model is validated, and the proposed framework is evaluated in this step. First, evaluation metrics for model validation are established. Second, additional, independent data is used to evaluate the model prediction. This crucial step is essential for assessing the model's accuracy and generalizability. Last, the efficacy of the proposed framework is evaluated.

Case Study

This study is based on a database containing information on 183 buildings in the vicinity of Berlin, Germany, Figure 2. This database provides essential data concerning attributes, location, technical systems, and typology of buildings with eight various usage types. For each building, yearly metered heat energy data spanning from 2010 to 2017 are available. Additional specifics regarding each building, such as building geometry and internal loads, are unavailable. Consequently, in step 2 of the framework we opted for a simplified building energy model with reduced input data requirements, DIBS (Bischof et al., 2022). For instance, DIBS determines internal loads by considering factors like Building type category and Occupier category, utilizing corresponding hourly schedules from ISO 13790:2008. Another feature of DIBS is its assumption of a cubic geometry, wherein only the Façade area north needs to be provided, and the remaining façade areas are calculated based on this input. In our preliminary analysis, an office building located in Berlin is used for the demonstration and evaluation of the framework. A list of attributes as model input variables for this building are shown in Table 1 as an example.

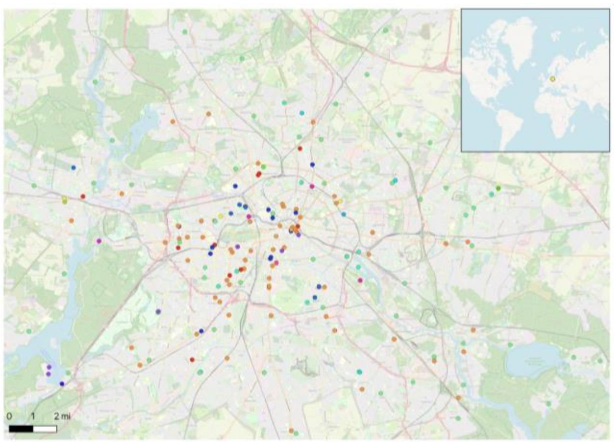


Figure 2 A map with 183 non-residential building objects in the database located in Berlin, Germany, together with a smaller map indicated the location of Berlin in relation to the USA. The markers of different colors represent different building types. Background map: OSM ©OSM

The primary aim of the final tool is to cater to individuals without expertise in building energy simulations, and therefore one of the predefined criteria was to minimize user input. Since acquiring specific parameters for climate data in particular years can be challenging, the concept of using normalized climate data was introduced. This approach involves embedding design climate data (Test Reference Year) and climate factors as hardcoded components of the tool, eliminating the need for users to search for suitable climate data. The design climate data represents a normal year for the location of Potsdam, Brandenburg in Germany. Location based climate factors are provided for the whole of Germany, where the climate factor for each building can be retrieved based on the zip code and year. The heat

Table 1 Processed attributes from the database and the values from the example building

Parameter	Value	Unit
Construction year	1979 - 2009	-
Building type category	Office, administrative or official buildings	-
Occupier category	Public administration or office building, town hall	-
Heating source	ElectricHeating	-
Cooling source	WaterCooled-PistonScroll	-
Heating distribution	AirConditioning	-
Cooling distribution	AirConditioning	-
Hot water distribution	CentralDHW	-
Number of floors under ground	0	-
Air change rate infiltration	Unknown	1/h
Night flushing flow	Unknown	1/h
Exterior wall area under ground	0	m^2
Energy losses in conversion and distribution	1.25	-
Max occupancy	148	nr
Exterior wall area over ground	1254.71	m^2
Window/wall-area ratio	0.20602	-
Facade area	1580.28	m^2
Facade area north	616.582	m^2
Roof area	842.58	m^2
Net room area	2963.39	m^2
Energy reference area	3370.32	m^2
Number of floors above ground	4	-
Height	11.27	m
U-value windows	2.83	$W/m^2,K$
U-value exterior walls	0.5	$W/m^2,K$
U-value roof	0.25	$W/m^2,K$
U-value base	0.4	$W/m^2,K$
Heat recovery efficiency of AHU	0	-
Thermal capacitance	260000	J/K
Heating coefficient of the system	1	-
Glass solar transmittance	0.78	-
Glass solar shading transmittance	1	-
Lighting power density	0.039	$W/m^2,lx$
Lamp type adjustment factor	1.09	0

energy demand for a typical year of a building can be calculated according to Equation (3) (Halbig, G. and Namyslo, J., 2014).

$$G = G(TRY, P) \,/\, KF. \tag{3}$$

G: Heating degree days for a specific zip code
G(TRY, P): Heating degree days for a typical year for Potsdam measuring station
KF: Climate factor for a specific zip code

As the metered data of each year is converted to the heat energy demand of a typical year, the simulation must only be performed for the Potsdam Test Reference Year throughout the calibration process, rather than for a time-series encompassing each year. This method streamlines the calibration process, requiring only one simulation output regardless of the calibration period's duration. Given the high number of simulations necessary for Bayesian calibration, this can significantly reduce computational costs.

Figure 3 illustrates the normalized metered heat energy demand compared to the prior DIBS simulation results using weather data of a Test Reference Year.

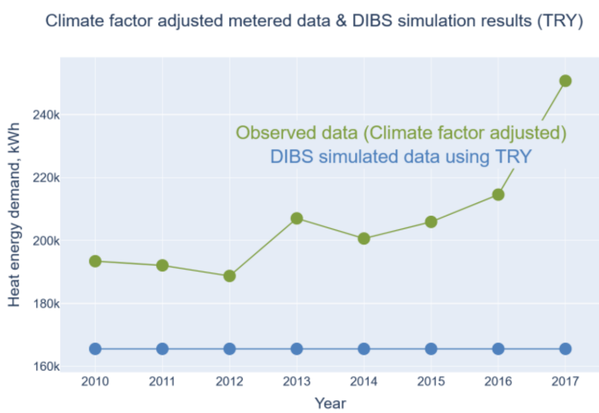


Figure 3 Comparison of normalized metered data with DIBS simulation data. The metered data is normalized by the climate factor, whereas the simulation data is obtained using DIBS with TRY weather data as input

Considering the amount of parameters and the computational time required by the DIBS simulation the Sensitivity Analysis is performed using 2816 samples to identify the variables that exert the most significant impact on the model output. The analysis utilizes the SALib Python library, employing the built-in Saltelli method for sampling and the Sobol method for analysis (Herman and Usher, 2017).

Utilizing comparable parameter criteria to those employed in the Sensitivity Analysis, the meta-model is trained with 1000 samples derived from the most influential parameters, generated through the Latin hypercube sampling method from the SALib library. Subsequently, it is constructed utilizing the Gaussian process technique, employing the Gaussian process regressor from the Python library scikit-learn (Pedregosa et al., 2011).

The Bayesian calibration is carried out in accordance with the guidelines outlined in the framework. A total of 40,000 accepted samples are requested, along with an additional 8,000 samples for tuning.

To validate the calibrated model, the metered yearly heat energy data (n=8) is divided into a training dataset and a test dataset. Three different splitting of the training/test data are assessed:

(Version 1) Training dataset with 4 points and test dataset with 4 points.

(Version 2) Training dataset with 5 points and test dataset with 3 points.

(Version 3) Training dataset with 6 points and test dataset with 2 points.

The computer employed for the entire calibration process features an AMD Ryzen 7 PRO 5850U processor, accompanied by 32 GB of RAM and a 64-bit architecture.

Results

Identified variables from the Sensitivity Analysis

The parameters with the highest Total Sobol Sensitivity Index are shown in Figure 4. Performing the sensitivity analysis required a computational time of 5823 seconds. The parameter Lamp type adjustment factor showed the highest influence on the model output according to the performed Sensitivity Analysis. This is followed by the U-value windows, the Lighting power density, the U-value exterior walls and the Thermal capacitance. These parameters were subsequently integrated into the calibration process.

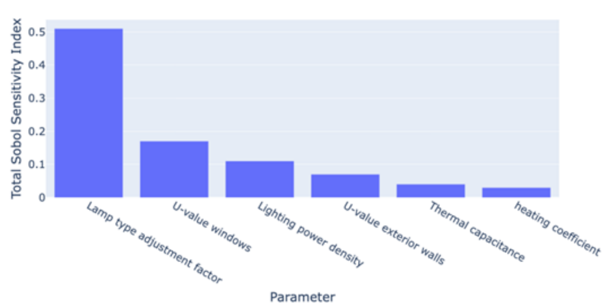


Figure 4 Total Sobol Sensitivity Index for the six most influential parameters

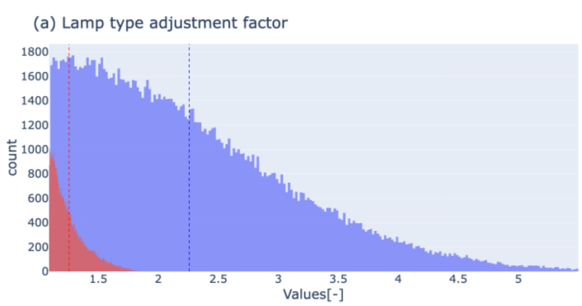
Meta-model

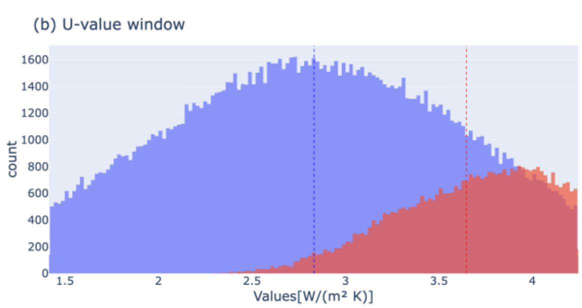
Generating the samples for the training of the meta-model took 1704 seconds and the training of the model required 55 seconds. To assess the meta-model's performance, the coefficient of determination statistical method was utilized, revealing a result of 0.9995 for 200 newly generated test samples.

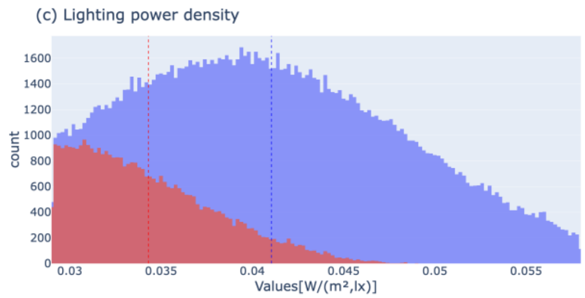
Bayesian calibration

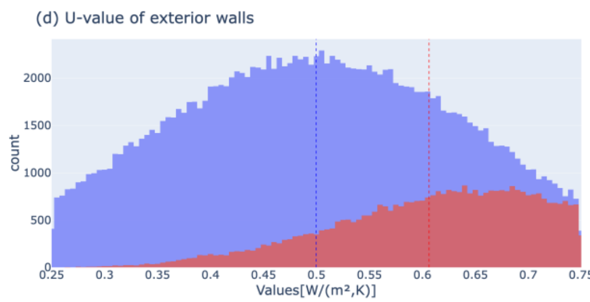
The calibration process using the meta-model required 805 seconds. In this section, we provide an overview of the posterior distributions of the parameters for Version 3, along with a summary of the posterior simulation result compared to both the training and the test datasets for all versions. A complete report containing posterior distribution of variables are attached in the Appendix.

Figure 5 visualizes the prior and posterior probability distributions for the calibrated parameters. As per the calibration process, the (b) U-value window and the (d) U-value of exterior walls should be adjusted to higher values in the model. In contrast, the (a) Lamp type adjustment factor, the (c) Lighting power density and the (e) Thermal capacitance should be set to lower values. Notably, the posterior standard deviations demonstrate a reduction following the calibration process.









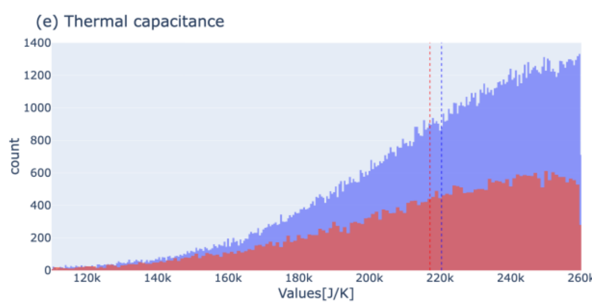


Figure 5 Illustration of prior and posterior distribution of variables for Version 3 (training/testing data points=6/2). The prior and posterior distributions are colored in blue and red, respectively. The blue and red dash lines indicate the prior and posterior means

In general, the posterior distribution derived through Bayesian inference not only furnishes a point estimate for the parameters and the simulation result but also encapsulates the associated uncertainty by the shape and spread of the distributions. This is visually represented in Figure 5 for the parameters and in Figure 6, which depict the distributions of simulation results based on posterior parameter probabilities for each of the three training-test versions.

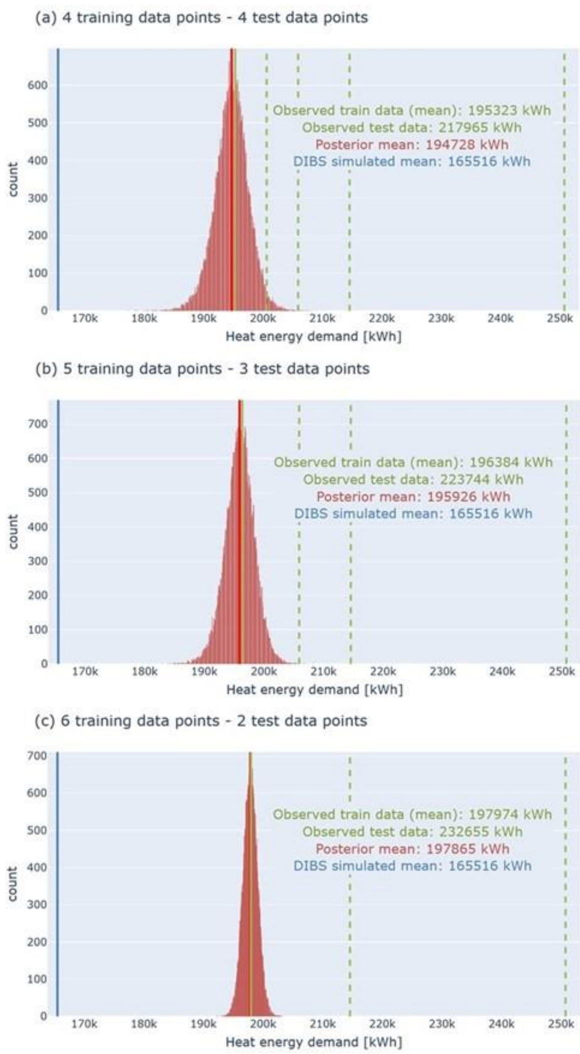


Figure 6 Comparison of metered and simulated heat energy demand, as well as probability densities of posterior simulations with different training data inputs – subplots (a), (b), (c) with 4, 5, 6 training data points, respectively

Posterior simulation results

Before calibration, the heat energy demand simulated for the Test Reference Year using the DIBS tool was 166 MWh.

Version 1: Consisting four training points and four test points out of the seven yearly energy demand data from 2010 - 2017, the posterior mean of the simulation results stands at 195 MWh, which portrays effectively the mean of the metered heat energy demand for the training points, 195 MWh. The mean of the metered heat energy demand for the test years (2014-2017) is 218 MWh, meaning that the calibrated model is performing better than the uncalibrated model for the test years.

Version 2: Consisting five training points and three test points, the posterior mean of the simulation results stands at 196 MWh, which again portrays effectively the mean of the metered heat energy demand for the training points, 196 MWh. The mean of the metered heat energy demand for the test years (2015-2017) is 224 MWh, meaning that the calibrated model is also in this case performing better than the uncalibrated model for the test years.

Version 3: Consisting six training points and two test points, the posterior mean of the simulation results stands at 198 MWh, which portrays effectively the mean of the metered heat energy demand for the training points, 198 MWh. The mean of the metered heat energy demand for the test years (2016-2017) is 233 MWh, meaning that the calibrated model is once again performing better than the uncalibrated model for the test years.

In Table 2 the resulting 90 % confidence intervals for the simulation results are summarized, based on each training version.

Table 2 Mean value, Lower and Upper bound for the 90 % confidence interval for the posterior simulation results

Posterior simulation results			
	Number of training points		
	Four	Five	Six
Mean value, MWh	195	196	198
Lower bound of a 90% confidence interval, MWh	192	194	196
Upper bound of a 90% confidence interval, MWh	198	199	200

Results and evaluation

The calibration process yields a probability distribution of the simulated results, estimating uncertainty. In the course of this calibration, a 90 %-confidence interval emerges, ranging between 196 MWh and 200 MWh with the six training data points. This signifies that, based on the calibration conducted with the DIBS building energy simulator tool, there is a 90 % likelihood that the simulated heat energy demand falls within the range of 196 to 200 MWh for the specific building under consideration. Prior to calibration, the DIBS simulation indicated a heat energy demand of 166 MWh. This result not only lacked information regarding its confidence level, but also exhibited a considerable deviation when compared to the metered energy data.

Yet, when the calibrated model is applied to predict the heat energy demand and compared with the test dataset, its performance is not particularly strong. As evident in

Figure 6, the "Posterior mean," which represents the mean of simulated results after calibration, significantly deviates from the "Observed test data". This divergence can be attributed to the notable fluctuations in observed heat energy data across the years. For Version 3, the mean of "Train 3" observed data stands at 198 MWh, matching the calibrated model's mean of 198 MWh, while the mean of "Test 3" observed data amounts to 233 MWh. Similar patterns emerge for version 1 and version 2.

Discussion

As outlined earlier, the presented Bayesian calibration framework is efficient in fine-tuning a building energy model that yields simulation outputs closely aligned with metered data, while it simultaneously facilitates an estimation of uncertainty. However, when the calibrated model is applied to predict the heat energy demand and is compared with the test dataset, its current performance is not particularly strong. The metered heat energy test data showed a considerable deviation from the posterior mean of the simulation results. The calibrated model's limited predictive capability can be attributed to significant fluctuations in observed heat energy data over the years, particularly noticeable in the last two years used for the test data. A consistent measurement of the heat demand with more accessible information can assist further analysis in this regard and enhance the predictive performance of the calibration.

Relying solely on climate factors does not fully address the annual variations in weather, as evidenced by the significant fluctuations in normalized metered heat energy demand. Exploring calibration with Actual Meteorological Year instead of Test Reference Year warrants further investigation. This approach may reveal climate parameters capable of accommodating fluctuations, ultimately leading to the development of a calibrated model that can elucidate these variations through refined parameter values.

In addition to the input data, the designated prior probability distributions play a pivotal role in shaping the resulting posterior parameter probability distributions, as illustrated by Bayes's Theorem (Eq. 1). The attributes of these prior probability distributions underscore the initial beliefs regarding parameter values and significantly influence the updated beliefs in the posterior distribution. The impact of the prior probability distributions is intriguing and requires closer scrutiny.

In this particular case study, calibration was performed for five parameters, yet metered data was limited to eight consecutive years. The number of calibration parameters could also affect the quality of the prediction and needs further investigation. Another intriguing aspect to explore is determining the optimal number of yearly metered data for accurate predictions. Here, an increased time resolution could offer valuable insights, facilitating more frequent feedback within the calibration process.

In summary, numerous factors contribute to the calibration results, and additional experiments are necessary to develop a general tool that can be applied on a broader scale and on different types of building energy models.

Conclusion and Outlook

This paper presents a Bayesian calibration framework tailored for a building energy model, outlining established connections among essential steps for creating a comprehensive automated tool. While Bayesian calibration has demonstrated its potency in calibrating building energy models, additional investigations are required to establish its reliability for prospective users in the future. Currently, a thorough investigation is being conducted to explore calibration settings, prior probability distributions, and hyperparameter configurations in order to ensure a robust framework with broader applicability. Additionally, further assessments will be conducted on various case study buildings and time resolutions.

Acknowledgments

We gratefully acknowledge the funding from Forschungsinitiative Zukunft Bau des Bundesinstitutes für Bau-, Stadt- und Raumforschung (reference number SWD-10.08.18.7-21.37)

References

Amasyali, K. and El-Gohary, N.M., 2018. A review of data-driven building energy consumption prediction studies. Renewable and Sustainable Energy Reviews, 81, pp.1192-1205.

Bischof, J., Knoll, S. and Duffy, A., 2022. Development of a Python-based simplified hourly building model for non-domestic building stock operational energy simulations. In the proceedings of the BauSIM 2022 in Weimar, 20. – 22. September 2022. In: BauSIM2022 2022.

Chong, A., Lam, K.P., Pozzi, M. and Yang, J., 2017. Bayesian calibration of building energy models with large datasets. Energy and Buildings, 154, pp.343-355.

Chong, A. and Menberg, K., 2018. Guidelines for the Bayesian calibration of building energy models. Energy and Buildings, 174, pp.527-547.

DIN EN ISO 13790:2008-09. Energieeffizienz von Gebäuden - Berechnung des Energiebedarfs für

Heizung und Kühlung (ISO 13790:2008). Berlin: Beuth Verlag.

Halbig, G. and Namyslo, J., 2014. Neue Witterungsbereinigung für Energieausweise auf der Basis des Referenzklimas Potsdam. EneV aktuell, IV, 2014, S. 11-13. (published in German)

Heo, Y., Choudhary, R. and Augenbroe, G.A., 2012. Calibration of building energy models for retrofit analysis under uncertainty. Energy and Buildings, 47, pp.550-560.

Herman, J. and Usher, W., 2017. SALib: An open-source Python library for sensitivity analysis. Journal of Open Source Software, 2(9), p.97.

Hong, T., Langevin, J. and Sun, K., 2018. Building simulation: Ten challenges. Building Simulation, 11 (5), 871–898.

Hou, D., Hassan, I.G. and Wang, L., 2021. Review on building energy model calibration by Bayesian inference. Renewable and Sustainable Energy Reviews, 143, p.110930.

IEA (2022), Global Energy Review: CO2 Emissions in 2021, IEA, Paris https://www.iea.org/reports/global-energy-review-co2-emissions-in-2021-2, License: CC BY 4.0

Kennedy, M.C. and O'Hagan, A., 2001. Bayesian calibration of computer models. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63(3), pp.425-464.

Mukhtar, M., Ameyaw, B., Yimen, N., Zhang, Q., Bamisile, O., Adun, H. and Dagbasi, M., 2021. Building retrofit and energy conservation/efficiency review: A techno-environ-economic assessment of heat pump system retrofit in housing stock. Sustainability, 13(2), p.983.

OpenStreetMap contributors. (2017). Planet dump retrieved from https://planet.osm.org.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E., 2011. Scikit-learn: Machine Learning in Python, JMLR 12, pp. 2825-2830, 2011.

Remmen, P., Automated calibration of non-residential urban building energy modeling (Doctoral dissertation, Dissertation, RWTH Aachen University, 2021).

Riddle, M. and Muehleisen, R.T., 2014, September. A guide to Bayesian calibration of building energy models. In Building simulation conference.

Salvatier, J., Wiecki, T.V. and Fonnesbeck, C., 2016. Probabilistic programming in Python using PyMC3. PeerJ Computer Science, 2, p.e55.

Wang, C.K., Tindemans, S., Miller, C., Agugiaro, G. and Stoter, J., 2020. Bayesian calibration at the urban scale: a case study on a large residential heating demand application in Amsterdam. Journal of Building Performance Simulation, 13(3), pp.347-361.

UNFCCC. (2018). What is the Paris agreement? Retrieved 12 Jan 2019 from United Nations Climate Change: https:// unfccc.int/process-and-meetings/the-paris-agreement/ what-is-the-paris-agreement

Appendix

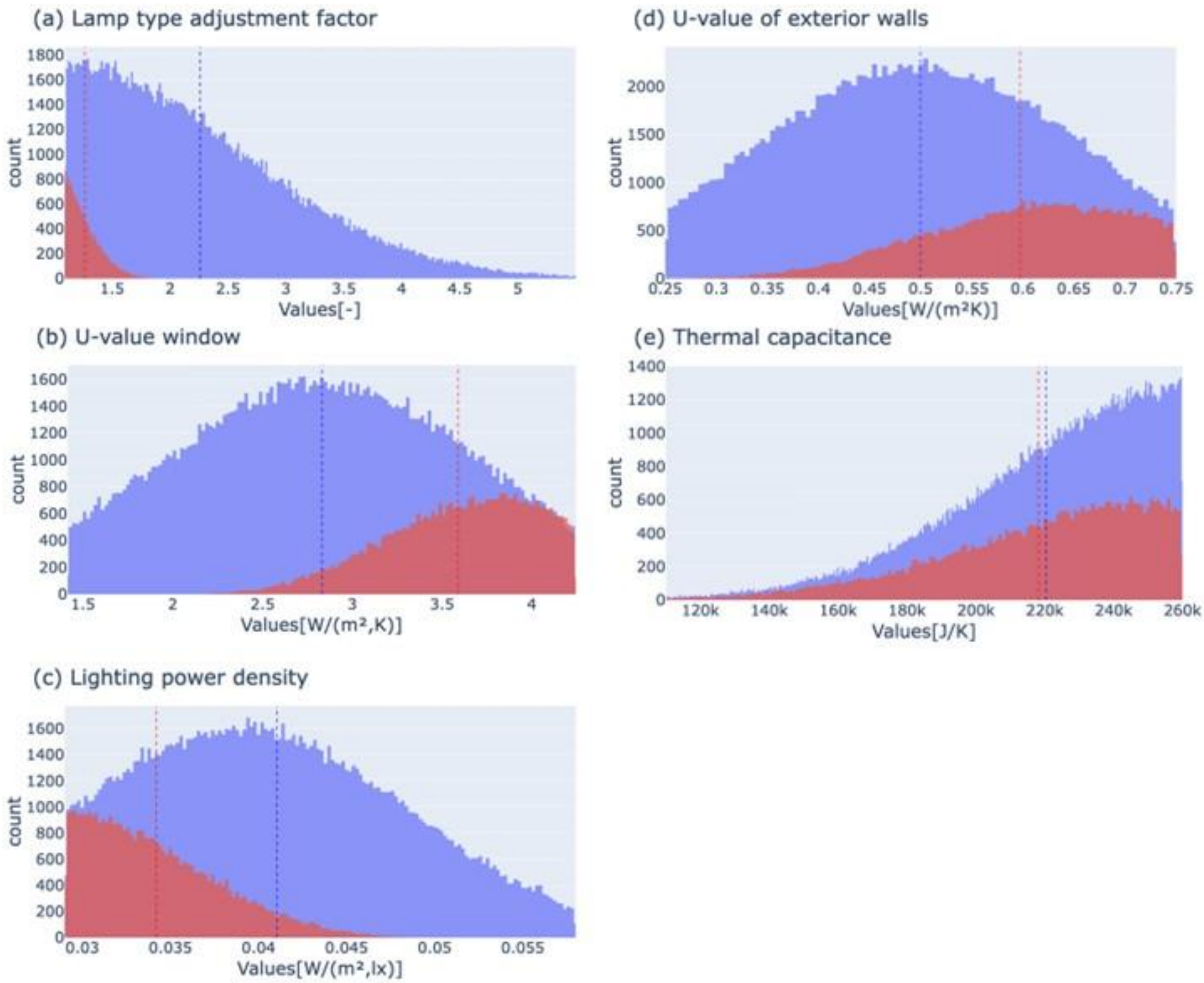


Figure A-1 Illustration of prior and posterior distribution of variables for Version 1 (training/testing data points=4/4). The prior and posterior distributions are colored in blue and red, respectively. The blue and red dash lines indicate the prior and posterior means

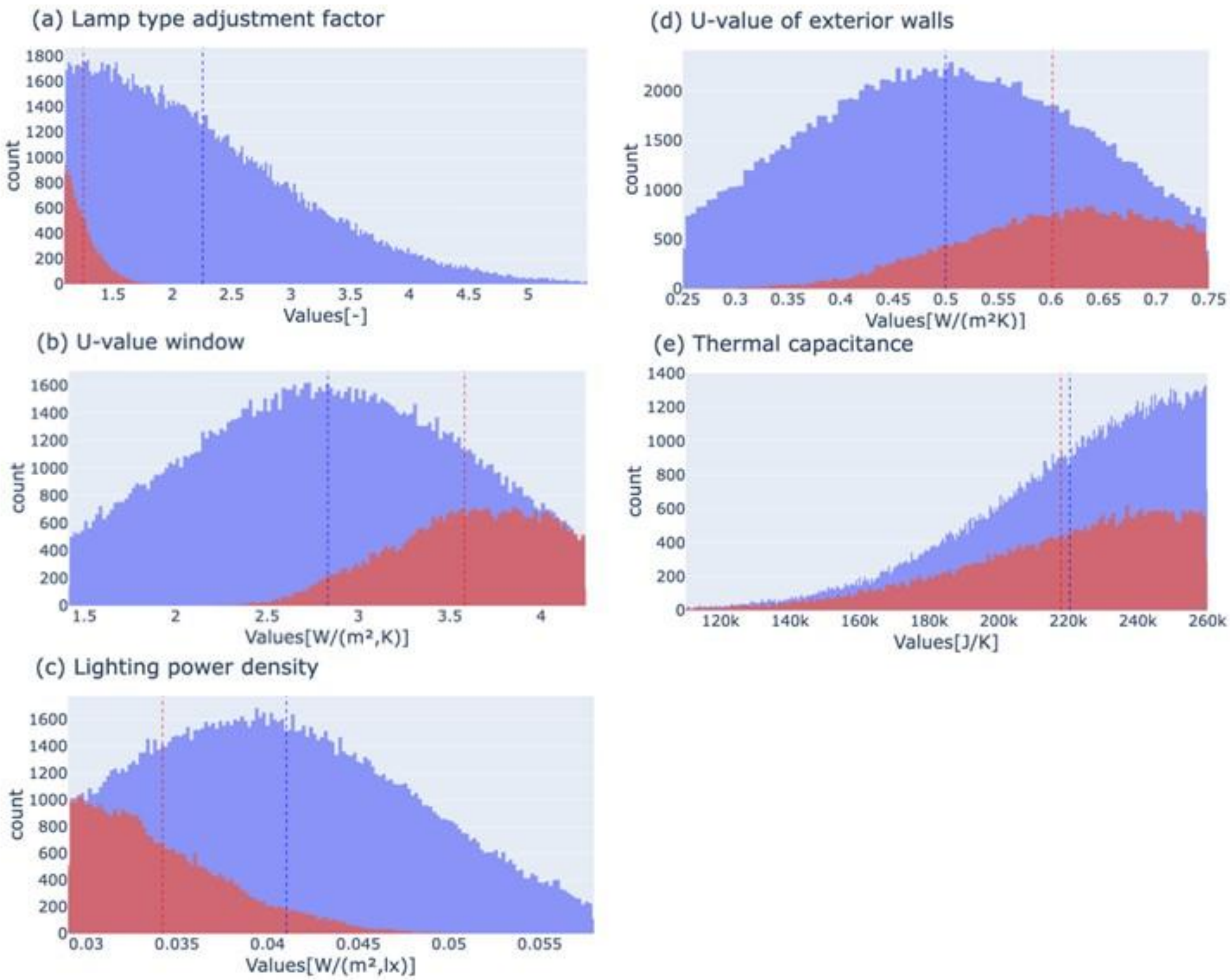


Figure A-2 Illustration of prior and posterior distribution of variables for Version 2 (training/testing data points=5/3). The prior and posterior distributions are colored in blue and red, respectively. The blue and red dash lines indicate the prior and posterior means