An Open-Source Framework: Bayesian Building Energy Model Calibration with Uncertainty Quantification

Katalin Fülep[1], Mary Lidiya Kalathiparambil Kennedy[1], Stefan Brandt[2], Rita Streblow[1]

[1]Technische Universität Berlin, Einstein Center Digital Future, Berlin, Germany

[2]Technische Universität Berlin, Hermann-Rietschel-Institut, Berlin, Germany

ABSTRACT

Building energy models are crucial in predicting and optimising energy consumption in buildings, but their simulation results often deviate from actual metered energy data, necessitating calibration. However, conventional calibration methods are labour-intensive and lack uncertainty estimation, making financial assessments and quantification of the broader environmental implications of the energy savings of building retrofit projects a challenge. To address these challenges, this study proposes a framework for calibrating building energy models and quantifying simulation uncertainties. Drawing on Bayes' Theorem and influenced by Kennedy and O'Hagan's article on Bayesian model calibration, the approach has been evaluated in various studies but lacks a widely accessible and user-friendly application, limiting its applicability across diverse case studies. In response, we introduce an open-source modular calibration framework that integrates building energy models with simplified input parameters, allowing decision-makers without expert knowledge in modelling and calibration to obtain reliable simulation results. Developed in Python, this user-friendly framework is set for open-source publication on the Git Repository. It encompasses the essential sub-processes and is tested and validated using a dataset from 183 buildings.

Keywords: Building Energy Models, Bayesian calibration, Energy savings estimation, Uncertainty estimation, Open-source framework

INTRODUCTION

In 2022, buildings alone constituted 30% of the global final energy consumption and 26% of the global energy-related emissions, according to the International Energy Agency (IEA, 2022). Given the increasing importance of energy efficiency, building performance simulation (BPS) is becoming prominent for assessing building performance and identifying strategies to reduce energy consumption, thereby contributing to carbon mitigation within the building sector. On the other hand, for BPS predictions to be accurate, the simulation model must closely align with the building's actual physical characteristics and operations (Heo et al., 2012). However, this is often challenging due to the lack of detailed information about buildings, resulting in discrepancies between BPS outputs and actual metered energy consumption (Hong et al., 2018).

To reconcile simulations with real-world scenarios, it is necessary to calibrate the model parameters. This calibration process is often constrained by inadequate building information, requiring comprehensive audits and expert judgement for non-observable and uncertain parameters. Traditional calibration methods are computationally demanding and depend substantially on extensive data collection and expert knowledge, making them inefficient (Riddle & Muehleisen, 2014). Furthermore, these methods are subjective, and the chosen parameter combinations may not accurately represent the true characteristics of the building (Hong et al., 2018). These uncertainties, arising from limited data availability, subjectivity, and deterministic

parameter estimation, are not accounted for in the calibrated models. As a result, they pose challenges for making retrofit decisions related to both financial and carbon mitigation strategies.

In the current study, a Bayesian statistics-assisted calibration approach is investigated to address the forementioned challenges. This approach not only aligns model outputs with metered data, but it also quantifies the inherent uncertainties associated with both the model parameters and the model itself. As a result of this process, a calibrated model emerges, featuring assigned probabilities for its outputs. These probabilities are instrumental in simulating various retrofit measures and calculating potential energy savings with associated uncertainties. Bayesian calibration integrated building energy models enhances the reliability of cost-benefit analyses, enabling decision-makers to make informed choices with heightened awareness of risks.

In recent years, this approach has become ubiquitous in model calibration (Chong & Menberg, 2018; Heo et al., 2012). Nevertheless, the underlying mathematics remains highly complex, devoid of a generalised, adaptable open-source framework, limiting its broader application (Hou et al., 2021). The proposed user-friendly Bayesian Calibration Framework delineates the essential steps required for effective calibration and integrates them into an automated process, enabling its application without the need for expertise in Bayesian statistics or building energy modelling.

METHOD

Bayes´ Theorem

Bayes' Theorem offers a robust framework for updating and refining the probability of a specific hypothesis given new evidence. By combining new evidence with existing knowledge, it modifies and improves the probability of a hypothesis. The likelihood of the evidence reduces the number of potential scenarios based on the evidence and prescribes prior beliefs about the hypothesis to be revised. This can be very helpful in model calibration, as it enables us to modify a parameter based on how well it fits the observed data while also taking its inherent uncertainty into consideration. By integrating these two critical aspects into building energy simulation models, the capability of the model to estimate real-world scenarios can be enhanced (Riddle & Muehleisen, 2014).

Bayes' Theorem establishes that the posterior parameter distribution, denoted as $P(\theta|y)$, is proportional to the likelihood function, $P(y|\theta)$, and the prior probability distribution of the parameter, $P(\theta)$:

$$P(\theta|y) \propto P(y|\theta) * P(\theta) \quad (1)$$

The prior probability distribution represents the prior knowledge regarding the hypothesis, which in the current case corresponds to the model input parameter defined by an initial estimation of its possible values. The likelihood function quantifies the probability of the evidence, given the hypothesis. The posterior probability distribution then represents the refined, updated estimates of the input parameter values given the observed data. Thus, calibrating the model involves iteratively comparing the simulated data produced by the model with the observed data, ascertaining a distribution of possible values that yield the most accurate values.

Framework for Bayesian Calibration

The current section presents a comprehensive overview of the implemented framework for the Bayesian calibration of a building energy model. A Python-based automated calibration tool was developed, comprising data pre-processing, a building energy simulation model, sensitivity analysis, and a gaussian process based meta model leading to Bayesian calibration tailored with multiple optional choices for the user to conduct analysis on data of their choice. The implemented toolchain will be released as open source upon further testing and validation.

Step 1: Data Collection and Preprocessing

The initial step entails gathering all pertinent information related to the building as required by the building energy model. The presumed values and the degree of uncertainty of the measured or collected parameters should be defined, as these parameters will be further evaluated throughout the framework.

Step 2: Building Energy Model Development

The building energy model to be subjected to calibration is implemented and integrated into the toolchain at this

stage. In the proposed framework, the DIBS (Dynamic ISO Building Simulator) building energy model has been incorporated as an optional component. DIBS facilitates the calculation of the total heat energy demand and corresponding global warming potential (GWP), along with other outputs, specifically for German non-residential buildings. This resistance-capacity model is founded on the simplified hourly method outlined in ISO 13790:2008 (Bischof et al., 2022). Owing to its minimal input parameter requirements and high computational efficiency — capable of generating hourly energy output for an entire year in a mere 1.3 seconds on a business laptop — DIBS emerges as an optimal choice for the calibration process necessitating numerous model iterations. Additionally, this generalizable approach can be adapted to work in conjunction with a multitude of building energy models, provided they are integrated into the toolchain and primed for the subsequent sensitivity analysis phase.

Step 3: Sensitivity Analysis

Following the model integration, a sensitivity analysis is carried out to assess the input parameters that exhibit the most substantial influence on the model's output. Sensitivity analysis of the input parameters is crucial because if a model has numerous unknown parameters, it might lead to over-parameterization. In such instances, the parameter values become indistinguishable, as different parameter settings can produce similar model outputs. This, in turn, could lead to unacceptably high uncertainty in the posterior distributions. Another advantage of the sensitivity analysis is that it helps to identify parameters that do not affect model performance or that have little effect on the model and which, during the calibration process, can be set to their presumed values, thereby reducing the computational costs significantly (Chong & Menberg, 2018). In the current study, a variance-based Sobol sensitivity analysis has been employed.

A critical prerequisite to begin with sensitivity analysis is defining an interval of feasible values for each uncertain parameter. Utilising the Saltelli sampling technique, a uniform and comprehensive set of parameter combinations is generated and subsequently employed to simulate the model output. The ensuing step includes the calculation of the Sobol Sensitivity Index, which serves to determine the sensitivity of each parameter. Those parameters with the highest sensitivity indices are then selected for further calibration.

Ascertaining the appropriate sample size to accurately capture sensitivity is a pivotal aspect of this process. Due to the computationally intensive nature of the simulation, the analysis commences with a limited number of samples. The sample size is then systematically increased, and the simulations are repeated until the ranking of the Sobol Sensitivity Index stabilises, indicating that the sensitivity order of the parameters is consistent.

Step 4: Meta-model Development

Bayesian calibration, as a computationally demanding process, involves conducting numerous iterations of simulation across an array of different parameter values drawn from the parameter space, with the need for substantial simulation trials. Due to the increased computational effort and time limitations in the case of complex systems, a meta-model can be implemented as a surrogate model for the original simulator.

In the given framework, the meta-model that maps the inputs to the desired output is implemented using Gaussian regression. This method is used extensively for surrogate modelling because of its non-parametric nature, enabling uncertainty quantification in the model while exhibiting smooth and continuous behaviour (Monari & Strachan, 2017). Besides, the Gaussian process regressor provides a normal distribution for each model output rather than a single value accounting for model discrepancy, which is the central premise of Bayesian inference.

The process begins with generating a set of parameter combinations and their corresponding outputs derived from the building energy model. In accordance with the intervals established in Step 3, the Latin hypercube sampling technique is used to adequately sample across the parameter space. Realisations of these parameter combinations and their corresponding outputs are used to perform Gaussian regression and build the meta-model. The meta-model is subsequently preserved in a pickled format for convenient access.

As in Step 3, determining the appropriate number of samples required to create a well-fit meta-model is a key consideration. In the current model, this aspect is addressed through a systematic approach, starting with a small number of samples. The meta-model is then initially trained using fewer samples, and, at a later stage, its outputs are compared to the original model outputs using quality metrics such as the coefficient of determination (R^2) and the mean average percentage error (MAPE). This is achieved iteratively, adding more samples to the training set with each iteration in an effort to realise the R^2 value of near-unity and the corresponding MAPE of less than 0.1%. At this point, the meta-model is deemed sufficiently accurate as a surrogate model.

Step 5: Priors Definition

The characterization of prior probability distributions is required for incorporating prior beliefs into the calibration of sensitive parameters. Chong and Menberg (2018) have shown in their study that an informative prior is approximated by a normal distribution with a small standard deviation, a weakly informative prior by a normal distribution with a high standard deviation, and an uninformative prior by a uniform distribution. These priors can be formulated using information derived from existing literature, expert knowledge, or database extractions. Such distributions encapsulate our initial beliefs about parameter values before observing any data.

Step 6: Bayesian Calibration

In the Bayesian calibration process, prior beliefs are integrated with the likelihood function to fine-tune the parameter values, resulting in a posterior probability distribution, which then serves as the input for the model representing the calibrated model. This pivotal process in the current framework comprises two essential components:

Definition of the Likelihood Function

Kennedy and O'Hagan, in their study (Kennedy & O'Hagan, 2001), introduced a foundational mathematical framework for Bayesian calibration. This framework involves calculating a likelihood function that describes the relationship between observed data (y), uncertain parameters (θ), and known inputs (x). The likelihood function assesses the adequacy of the model parameters in accounting for the data that has been observed.

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

η(x, θ): Model output given both known and unknown parameters
δ(x): Model discrepancy
ε(x): Measurement error

The likelihood function thus quantifies the probability of obtaining the observed data given the model and its parameters, thereby linking model predictions to observed data while duly accounting for measurement uncertainties.

Calibration Process

During the calibration process, the prior probability distributions of the parameters are updated using the likelihood function and the observed data. This iterative step employs the Markov Chain Monte Carlo (MCMC) approach, with samples drawn from prior distributions. A probability distribution based on these parameter values is obtained through the meta-model simulation, which then facilitates calculating the conditional probability of the observed data, given the model. In the subsequent step, a new parameter combination is chosen, and a new model output distribution emerges, from which the conditional probability is calculated again. If this parameter combination has a higher conditional probability, the parameter values are saved for further iterations. This iterative process continues until the number of parameter samples accepted meets the predefined threshold. Once the calibration process is completed, the accepted mean values serve as the parameters' posterior probability distributions.

The determination of the optimal number of samples remains a challenge in this step as well. To this end, a small sample size is initiated and incrementally increased until the calibration process converges. This state of convergence is assessed by sampling multiple chains simultaneously and estimating the Gelman-Rubin statistics, which relate variance within each chain to variance across all chains, estimating how well chains have assimilated or converged to a common distribution

indistinguishable from each other. A Gelman-Rubin statistic close to 1.0 ± 0.1 for all the five parameters indicates satisfactory convergence, signifying that the MCMC sampling process has achieved a stationary distribution, and the resulting estimates are reliable (Calama-González et al., 2021).

Step 7: Calibrated Model

The posterior parameter values, obtained in Step 6, qualify as inputs to the model, resulting in calibrated model outputs. Given that the posterior parameter values are characterised by a probability distribution, so will the calibrated model output. Consequently, the calibrated model does not yield a single output value; instead, it provides a distribution that also accounts for the model's uncertainties. When simulating a specific retrofit, the results include associated uncertainties, indicating the probability that the retrofit will achieve a certain amount of energy savings. These results are presented with confidence intervals, offering a quantifiable measure of the retrofit's effectiveness.

Step 8: Calibrated Model Evaluation

In evaluating the calibrated model, the observed time series data is divided into two sets: one for calibration and the second for evaluation. The model is validated using R^2 and MAPE as evaluation metrics, which underscores the model's accuracy and generalizability.

Case Study

The current study relies on a dataset comprising information on 183 non-residential buildings situated in the vicinity of Berlin, Germany. The dataset includes comprehensive details regarding the location, building envelope, technical systems, and usage type of each building. Additionally, it provides annual metered heat energy consumption data spanning from 2010 to 2017. However, information on internal loads is not available. As a result, in step 2 of the framework, a simplified building energy model with minimal input data requirements was employed, referred to as DIBS (Bischof et al., 2022). For instance, DIBS uses certain data, such as building type and occupancy category, to estimate inner loads based on corresponding hourly schedules from ISO 13790:2008. For the demonstration of the framework, an office building located in Berlin was chosen. Originally constructed in 1911, the heating energy for the building under study is supplied by gas.

Table 1 presents a list of attributes used as input variables for the DIBS model, that are exclusive to the chosen building for demonstration. The data pertaining to these attributes was retrieved from the forementioned database. Based on whether a corresponding parameter value was directly available in the database or inferred logically from the provided data, these attributes had been categorised according to their degree of uncertainty.

Table 1 Variable input parameters to the simulation tool DIBS

Parameters	Value	Lower bound	Upper bound	Degree of uncertainty
Max occupancy	840	630	1050	Weakly Informative
Exterior wall area over ground	2472 m^2	2224	2719	Informative
Window/wall-area ratio	0.57	0.51	0.62	Informative
Total facade area	5712 m^2	5141	6283	Informative
Facade area north	2161 m^2	1081	3242	Uninformative
Roof area	3397 m^2	3057	3736	Informative
Net room area	16808m^2	15127	18489	Informative
Energy reference area	20381m^2	18343	22419	Informative
Number of floors above ground	6	5.40	6.60	Informative
Height	21 m^2	18.93	23.13	Informative
U-value windows	3.24 $W/m^2,K$	1.62	4.86	Weakly Informative
U-value exterior walls	2.00 $W/m^2,K$	1.00	3.00	Weakly Informative
U-value roof	0.29 $W/m^2,K$	0.15	0.44	Weakly Informative
U-value base	0.55 $W/m^2,K$	0.28	0.83	Weakly Informative
Thermal capacitance	260000 J/K	110000	260000	Uninformative
Energy expenditure coefficient	1.09	0.98	1.19	Uninformative
Glass solar transmittance	0.78	0.53	0.87	Uninformative
Glass solar shading transmittance	1.00	0.22	1.00	Uninformative
Lighting power density	0.04 $W/m^2,lx$	0.03	0.06	Uninformative
Lamp type adjustment factor	1.09	0.80	1.50	Uninformative

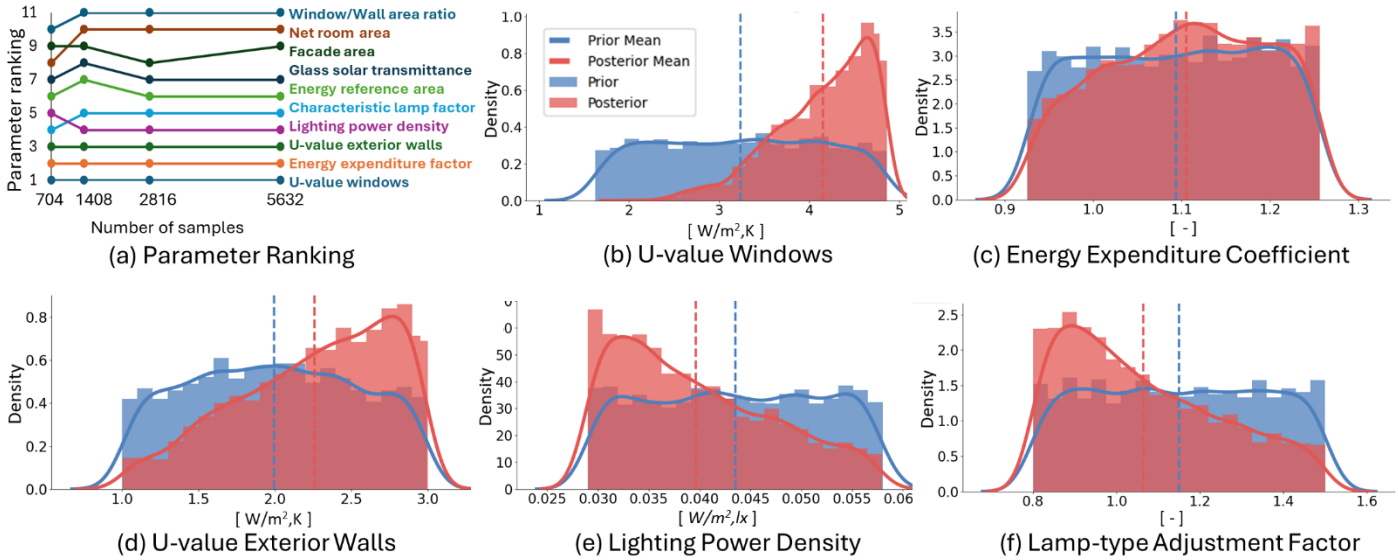

Figure 1 (a) Parameter ranking based on sensitivity. (b-f) Prior and posterior parameter probability distribution with the mean values

RESULTS

As evident from Figure 2(a), the simulated output diverges from the metered data. This discrepancy likely stems from the fact that models are simplifications of reality and from the input parameters used in the simulation. Since the parameter data was obtained from energy providers rather than a controlled experimental setup, it may not accurately reflect real-world conditions due to factors such as data availability and quality, as well as variations in reporting methods among different providers. This can lead to a mismatch between the observed and the simulated energy demand, highlighting the necessity for model calibration.

To refine the input parameters and create a more representative model, a calibration process is therefore undertaken. This involves utilising the data from the initial five-year period to calibrate the model, followed by model evaluation, where the refined model is tested against the subsequent three-year data.

The forthcoming step involved the identification of the most influential parameters through sensitivity analysis. The defined intervals for each input parameter are presented in Table 1. The minimum number of samples, N, was set at 16, resulting in 704 parameter combinations (N×(2D+2)). The Total Sobol Sensitivity Index converged at N = 64, corresponding to 2816 samples, indicating stabilisation in the sensitivity ranking of the first five parameters, see Figure 1(a). Achieving this convergence required approximately 2.8 hours. The computations were performed on a desktop with an AMD Ryzen 7 PRO 5850U processor and a 32 GB of RAM.

Based on these results, the five most sensitive parameters were selected for further calibration.

In the next step, a meta-model is constructed using the forementioned parameter value intervals as in the sensitivity analysis. The process generates five parameter sample combinations, allocating 70% and 30% of the total samples for the model's training and evaluation, respectively. As illustrated by Figure 2 (b), the number of samples for the training is raised until the predefined performance prerequisites for the meta-model are met, which occurs at 34 samples. The convergence was attained within a span of 7 minutes. The trained model was subsequently utilised in the calibration process.

The essence of this framework lies in defining prior probability distributions for the parameters reflecting their inherent uncertainty, as shown in Table 1. Additionally, priors were defined for model discrepancy and observational error. The model discrepancy is modelled using a normal distribution with a standard deviation derived from the meta-model. Observational error follows an exponential distribution, where the median of the distribution corresponds to a 5%

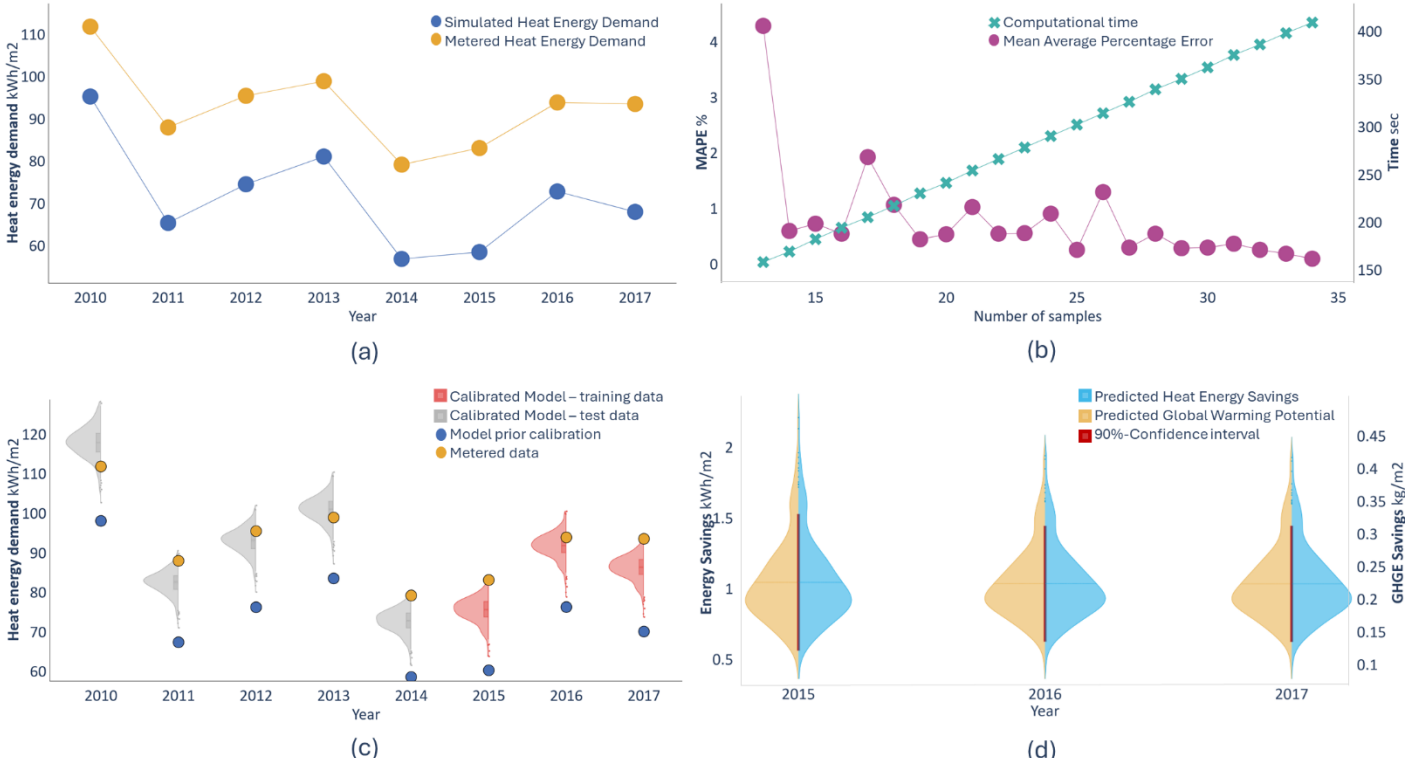

Figure 2 (a) Comparison between simulated and metered heat energy demand prior calibration. (b) Mean percentage error and computational time for different sample sizes for the training of the meta-model. (c) Posterior simulation results of the heat energy demand. (d) Predictive scenario for heat energy demand and global warming potential savings

observational error relative to the mean of the measured data. Initially, 100 posterior samples were requested. However, if the chains fail to converge, the sample size is doubled for each iteration. In the current case, calibration converges at 800 samples, requiring 4 minutes.

The obtained parameter samples define the posterior probability distributions for the calibrated parameters, illustrated in Figure 1(b-f). These distributions show noticeable shifts in their mean values and reduced standard deviations, indicating enhanced confidence in the posterior estimates compared to the initial assumptions. This is in line with anticipated results since parameter estimates are refined iteratively using observed data throughout the calibration phase.

Calibrated Model

Calibrated model predictions are derived through sampling from the obtained posterior parameter distributions and subsequent model simulations. In Figure 2(c), the training data utilised for calibration is indicated in grey, while the predictions from the calibrated model are depicted in red. The input parameter values sampled from the posterior distributions lead to simulation outputs that are in good agreement with the observed data (yellow), in contrast to the simulation outputs (blue) prior to calibration. However, because the calibration process is constrained by the prior probability definitions of the parameters, it may not fully align with the observed data. These comparisons serve as an evaluation of the calibrated model's performance.

Predictive Scenario

A predictive scenario was generated and simulated to highlight the potential impact of a retrofit measure such as the replacement of windows with lower U-values, with this being the most influential parameter. The U-values were adjusted from the posterior mean of 4.1 to 1.3, meeting the renovation standards for non-residential buildings in Germany under the GEG regulations (GEG, 2020). The calibrated model was then utilised to estimate the resulting savings in heat energy and global warming potential, as illustrated in Figure 2(d). As indicated by the predictive scenario results, the proposed retrofit measure would achieve energy savings

between 0.56 and 1.52 kWh/m^2 and global warming potential savings between 0.12-0.33 kg/m^2 with a 90% confidence interval.

DISCUSSION

The Bayesian calibration framework introduced in this study effectively calibrates building energy models, yielding simulation results that closely align with metered data while incorporating uncertainty quantification. Additionally, the methodology proves successful in estimating energy savings from retrofit measures, while also providing uncertainty estimates for risk assessment.

A major advantage of this framework is its ability to autonomously determine optimal solutions for sensitivity analysis, Gaussian process convergence, and Bayesian calibration, eliminating the necessity for user intervention.

The implemented framework has been tested on diverse datasets, in contrast to studies where data is derived from controlled environmental test setups (Calama-González et al., 2021; Heo et al., 2012; Monari & Strachan, 2017). This approach showcases its applicability to a larger and more diverse set of buildings and realistic measurement scenarios with no interruption to occupants. It is also designed to handle various time resolutions—such as hourly, monthly, or yearly—demonstrating its applicability to a wider range of buildings and realistic measurement scenarios. Additionally, this framework is implemented in Python, which is one of the four most widely used programming languages, contrasting with similar frameworks that are typically developed in R (Chong & Menberg, 2018; Orlowska et al., 2021).

CONCLUSION AND OUTLOOK

In conclusion, the framework offers a comprehensive automated solution for calibrating building energy models, integrating uncertainty estimation. It enhances user accessibility and operational efficiency, optimising computational speed and enabling informed decision-making on retrofit measures. Future research will expand evaluations across diverse buildings to validate its broader applicability.

ACKNOWLEDGEMENT

We gratefully acknowledge the funding from Forschungsinitiative Zukunft Bau des Bundesinstitutes für Bau-, Stadt- und Raumforschung.

REFERENCES

Bischof, J., Knoll, S., & Duffy, A. (2022, September 20). Development Of a Python-Based Simplified Hourly Building Model for Non-Domestic Building Stock Operational Energy Simulations. BauSim Conference 2022.

Calama-González, C. M., Symonds, P., Petrou, G., Suárez, R., & León-Rodríguez, Á. L. (2021). Bayesian calibration of building energy models for uncertainty analysis through test cells monitoring. *Applied Energy*, *282*, 116118.

Chong, A., & Menberg, K. (2018). Guidelines for the Bayesian calibration of building energy models. *Energy and Buildings*, *174*, 527–547.

GEG (2020). *Building Energy Act—Gebäudeenergiegesetz*. Federal Republic of Germany.

Heo, Y., Choudhary, R., & Augenbroe, G. A. (2012). Calibration of building energy models for retrofit analysis under uncertainty. *Energy and Buildings*, *47*, 550–560.

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

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

IEA. (2022). *Technology and Innovation Pathways for Zero-carbon-ready Buildings by 2030*. IEA.

Kennedy, M. C., & O'Hagan, A. (2001). Bayesian Calibration of Computer Models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *63*(3), 425–464.

Monari, F., & Strachan, P. (2017, August 7). *CALIBRO: An R Package for the Automatic Calibration of Building Energy Simulation Models*. 2017 Building Simulation Conference.

Orlowska, A., Chrysoulas, C., Jaroucheh, Z., & Liu, X. (2021). Programming Languages: A Usage-based Statistical Analysis and Visualization. *2021 The 4th International Conference on Information Science and Systems*, 143–148.

Riddle, M., & Muehleisen, R. (2014). *A Guide to Bayesian Calibration of Building Energy Models*. 276--283.