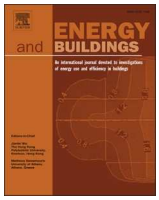
Energy & Buildings 323 (2024) 114755

Contents lists available at ScienceDirect

Energy & Buildings

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



A deep learning-based Bayesian framework for high-resolution calibration of building energy models

Gang Jiang [a], Yixing Chen [b], Zhe Wang [c], Kody Powell [a], Blake Billings [d], Jianli Chen [a,*]

[a] The University of Utah, United States
[b] Hunan University, China
[c] The Hong Kong University of Science and Technology, Hong Kong, China
[d] Oak Ridge National Laboratory, United States

ARTICLE INFO

Keywords:
Building energy modeling
Model calibration
Bayesian calibration
Deep learning
Machine learning
Bayesian optimization

ABSTRACT

Calibrating building energy models (BEMs), i.e., closing discrepancy between modeling and field measurements, is of significance to support its applications in building sustainability and resilience analysis. However, as being widely used in practice, current Bayesian calibration is mostly performed in low-resolution (annual or monthly), instead of high-resolution (hourly or sub-hourly), which is crucial to support emerging BEM applications, such as building-renewable energy integration (demand response) and smart control. This is attributable to the gaps in current Bayesian calibration process, including (1) difficulty in supporting reliable high-resolution calibration with over-parameterization and multi-solution issues, (2) inadequacy of *meta*-model to capture temporal building dynamics in high-resolution, and (3) excessive computational burdens of covariance matrix calculation in Bayesian inference. Therefore, to close these gaps, this research proposes a novel deep learning-based Bayesian calibration framework, involving pre-calibration mechanism, Long Short-Term Memory as surrogate models, and simplified covariance matrix calculation, to calibrate BEMs in high temporal resolution (i.e., hourly) with enhanced accuracy and computational efficiency. The case study demonstrates its effectiveness to match modeling outcomes with measurements and realize CV-RMSE of < 30 % and NMBE of < 6 % in hourly resolution, as well as a significant reduction of calibration time (by > 99 %, from > 600 h to ~ 1.5 h).

1. Introduction

1.1. Background

The energy consumption of buildings comprises a significant proportion of the overall societal energy usage, accounting for approximately 36 % of global energy consumption [1]. For improved building performance, building energy modeling (BEM) has emerged as a pivotal tool for simulating and forecasting energy consumption, serving various purposes such as analysis of building retrofitting and enhancement of energy efficiency [2]. In recent years, new applications of BEM arise, such as demand response (DR), fault detection and diagnosis (FDD), and smart control [3–5]. These emerging applications requires building simulation to be able to capture building dynamics with higher accuracy and resolutions (e.g., hourly and sub-hourly), compared to application of building modeling in design scenarios that low resolution modeling (e.g., monthly or yearly predictions) are deemed sufficient. Therefore, the development of accurate BEM in high-resolution has become particularly important to further promote the application of BEM in practice. However, the escalating complexity of effectively capturing building operation dynamics in high resolution led to an increasing disparity between simulation outcomes and actual measurements in high-resolution prediction scenarios [6,7]. To address this, the inputs of BEM need to be meticulously adjusted. This process, known as model calibration, involves adjusting various inputs of BEM to ensure the close matching between modeling and measurements (i.e., field observations) [8].

1.2. Manual and automated calibration

Calibration methods in BEMs can be broadly categorized into manual calibration and automated calibration based on the techniques employed. Manual calibration is a commonly used approach in BEM calibration. This method involves a “trial and error” process that relies

* Corresponding author.
E-mail address: jianli.chen@utah.edu (J. Chen).

https://doi.org/10.1016/j.enbuild.2024.114755
Received 19 June 2024; Received in revised form 13 August 2024; Accepted 1 September 2024
Available online 6 September 2024

on iterative manual tuning of model input parameters [9]. Most of the earlier methods are based on graphic comparisons, including plot and chart-based analysis [10–12], end-use disaggregation and analysis [13–18]. These manual calibrations can reduce the monthly modeling deviations of building models for energy efficiency and retrofits leveraging engineering experience. However, manual calibration is time-consuming, heavily relying on expert judgments and comparisons, hence, challenging to be applied in complex calibration scenarios (e.g., calibration involving large number of parameters or in high-resolution). On the other hand, the advancement of computing powers and algorithms contribute to the development of automated calibration, realized through optimization methods. Chong et al. [19] conducted a systematic review of automated model calibration in building simulation with a synthesis analysis and classification of simulation inputs and outputs, data types and resolutions, key calibration methods, and evaluations. The conclusions show that the existing calibration practice mostly uses monthly data and it is difficult to consider building schedules in calibration due to computational costs and over-parametrization issues, although the importance of schedule adjustment in model calibration has been demonstrated as important to build up accurate building energy models [20,21]. Vera-Piazzini et al. [22] also emphasized that low-resolution (monthly and annually) calibrations are prevalent in practice, while occupancy behavior relevant parameters are identified as the pivotal parameter for precise building modeling in high resolution (but difficult to obtain through calibration). Using high-resolution data is expected to enhance calibration quality (to identify ground truth values of calibration parameters as well as improve model performance), ultimately leading to more reliability building model in application and better understanding of building operation dynamics [7,23–25]. Coakley et al. [9] indicate that due to the vast number of inputs required for detailed building energy simulations and limited available measurements, calibration in high resolution is typically an uncertain process with over-parameterization, i.e., tuning input parameters to make modeling outcomes of the calibrated building model match actual measurements is a highly under-determined problem that involves multiple non-unique solutions. Given the challenges in building calibration, multiple tools, e.g., data analysis, *meta*-model, and calibration techniques, are utilized to ensure reliable calibration practice [22].

1.3. Bayesian calibration

Among different automated calibration methods, the Bayesian calibration [26], proposed by Kennedy and O'Hagan, emerges as a prominent approach and becomes widespread used in various domains, including physics [27], materials science [28], biomedical engineering [29], energy storage [30], and ecology [31]. The Bayesian calibration is advantageous in (1) making full use of prior knowledge to close discrepancies between observed data and model predictions, reducing system uncertainty; (2) not relying on specific functions or assumptions, hence, being flexible to be applied in various complex scenarios and problems; and (3) providing probability distributions for calibration parameters instead of a single point estimate in inference, enabling Bayesian calibration to attain comprehensive parameter information to evaluate the reliability and uncertainty of parameters. As one of the pioneering works, Heo et al. [32] employed Bayesian calibration in BEM to assess building retrofitting strategies and quantify associated risks. They utilized Gaussian Process to represent various uncertainty relationships between the building energy model and observation data. The Bayesian rule was applied to determine the likelihood of calibration parameters, followed by the use of Markov chain Monte Carlo (MCMC) sampling to explore the posterior distribution of parameters.

Although Bayesian calibration has advantages in the calibration as mentioned, it is limited in calibrating BEM in the monthly resolution in the current practice. Current Bayesian calibrations typically only select a small number of calibration parameters and use a small amount of data (e.g., 12 months of energy use) to calibrate these parameters to avoid excessive computational burdens. It could take hours to days to complete the model calibration in the monthly resolution, let alone calibrating parameters using 8,760 h of data in the hourly resolution. To balance the computational cost and calibration performance [33–35], researchers utilized sensitivity analysis to select the influential parameters affecting building energy consumption, hence, reducing the number of parameters in calibration. By using techniques such as correlation analysis and clustering methods [35–38], reducing data redundancy can further mitigate computation time (though using fewer data may affect the accuracy of calibration [35]).

Additionally, *meta*-models (or surrogate models) are used to save calibration time by employing reduced-order or data-driven building models to approximate BEM outputs in parameter evaluation, without compromising calibration accuracy [39]. Various experiments demonstrate that employing *meta*-models accelerate the calibration process compared to iteratively running traditional physics-based models (e.g., EnergyPlus) while maintaining sufficient accuracy in calibration [34]. Lim [40] compared five *meta*-models to determine the impact of *meta*-model accuracy on Bayesian calibration. For monthly calibration, *meta*-models only need to capture the monthly building energy usage, hence, even using the simplest linear regression as the surrogate model is sufficient to achieve satisfactory calibration results [41]. However, for high-resolution (e.g., hourly) calibration, the situation is much more complex with a significantly growing number of outputs (8,760 h per year) by *meta*-models. The high-dimensional measurement data (hourly use of cooling, heating, and electricity across the year) and increasing number of parameters to calibrate (e.g., thermal properties, control parameters, occupant relevant parameters) not only increase the computational burden, but also enhance the difficulty of calibration. Gu et al. [6] developed a multi-output Gaussian surrogate model and compared monthly-resolution and hourly-resolution calibrations. For monthly calibration, the CV-RMSEs of calibrated energy use for 7 test buildings were below 10 %. However, these calibrated models present ~ 50 %-70 % deviations of modelled energy use in the hourly resolution, failing to meet the ASHRAE requirements [42] for successful hourly-calibration. Moreover, the authors mentioned that, even with GPU acceleration, these hourly calibrations still take several weeks. Kristensen et al. [7] conducted calibration of an ISO 13790 BEM based on the Bayesian approach, investigating the calibration performance in different temporal resolutions (6-hour, daily, weekly, and monthly). The validation results indicate that the reliability and applicability of calibrated models increase with higher resolution of calibration. Researchers also attempt to make the Bayesian calibration a more efficient process through sampling techniques (e.g., No-U-Turn [43] and HMC [44]), simplified physics models (e.g., reduced-equations [45], RC model [46], and ISO13790 [7,47,48], or approximate Bayesian inference [49,50] and *meta*-learning [51]. Nevertheless, in current practice, the Bayesian calibration can still only process a small number of building thermal performance parameters (typically 2–6) and targeting on one aspect of building energy uses, e.g., cooling or heating energy usage [7,43,45,49–54]. This significantly reduces the reliability of BEM to fully capture actual building dynamics in high resolution, hence, more broadly applicable in emerging applications such as DR, FDD, and control.

1.4. Existing gaps

Despite these efforts and advancements in building calibration, there remain limitations to the existing Bayesian approach to calibrate building energy models in high resolution, including challenges to deal with increasing number of data and calibration parameters (e.g., building operation schedules, control settings) along with the over-parameterization and multi-solution issue in high resolution calibration, insufficient surrogate modeling to capture temporal dependencies of energy use in actual building operation, and computational burden (inefficiencies) when dealing with high-dimensional parameters and

large datasets (hourly data) in high-resolution calibration, as explained below.

(1) Difficulty of high resolution (hourly) calibration: Although hourly calibration can better support a broader range of advanced building applications, e.g., DR, FDD, and control, it also presents greater challenges. When conducting hourly calibration, careful consideration of building occupant behaviors and schedules becomes important, as they to some extent reflect the underlying operation patterns of buildings (hence, the energy usage of buildings). This requires collection of more measurement data (8,760 h in one year), calibration of more model input parameters (e.g., occupant relevant schedules), and deepened building system analysis to capture the thermodynamic process and temporal correlations in building operation, hence, ensuring matching between modeling and observations in calibration. However, as the number of calibration parameters increases, the issue of over-parameterization and multiple solutions arises, i.e., different combinations of parameters are likely to produce similar modeling outcomes that match observations. In such cases, it becomes ambiguous for the calibration algorithm to determine the optimal sets of parameters, hence, uncovering ground-truth parameters that reflect the actual operating conditions of the building. This will directly affect the validity of high-resolution calibration. Fig. 1 provides a schematic diagram of multi-solutions in calibration due to over-parameterization. Although the values of parameter combinations 1 and 2 for lighting and equipment differ significantly, a similar calibration outcome (estimated Building Cooling Load) is obtained.

(2) Limitations of surrogate models: While surrogate models are employed to enhance the efficiency of Bayesian calibration, they face challenges in capturing the temporal dependency and complexity of building operation in high-resolution calibration of building energy models. Building systems experience time-dependent influences (e.g., thermal inertia) from human behaviors, weather conditions, and system operations. Moreover, interactions between different sub-systems contribute to the complexity of system operation (e.g., the equipment load not only increases the electricity consumption, but also triggers a correlated increase in heating and cooling load). The use of simple surrogate models (e.g., Gaussian Process or Linear regression models) ignore the temporal dependency as well as simplify the complexity of building operation, raising concerns about its reliability and validity to support BEM calibration in high resolution. Models considering temporal dependency and multi-output can provide more accurate and consistent results [25,55]. No surrogate model has yet considered capturing these temporal dependencies and multi-output complexities for high resolution BEM calibration.

(3) Calculation efficiency of Bayesian calibration: The computational burden of the Bayesian calibration method originates from necessity to compute covariance matrices and likelihoods in parameter evaluation. In cases of high-resolution calibration involving a multitude of parameters or substantial data volumes, the size of these covariance matrices grows exponentially. As a result, solving high-dimensional covariance matrices becomes problematic, leading to difficulty in high-resolution calibration with excessive computation burdens.

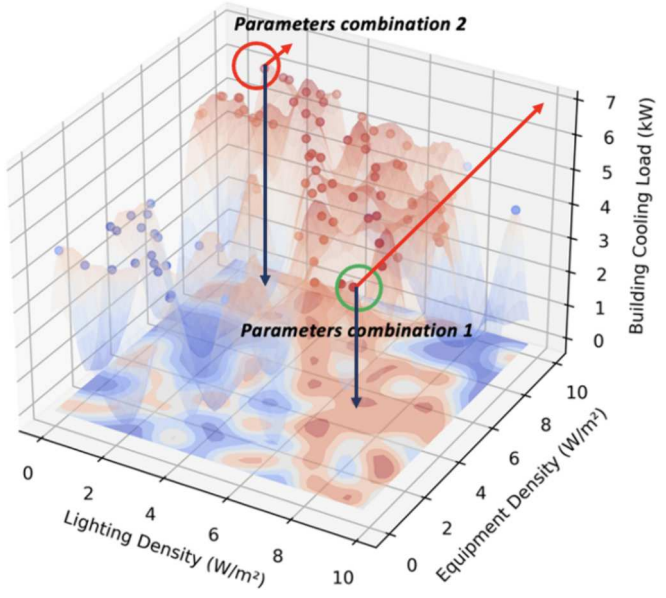


Fig. 1. Over-parameterization phenomenon: Parameters combination 1: Lighting is 7 W/m^2, Equipment is 5 W/m^2; Parameters combination 2: Lighting is 1 W/m^2, Equipment is 9 W/m^2.

1.5. Proposed framework

To address the issues mentioned above, we propose a novel deep learning-based Bayesian calibration framework specifically designed for high-resolution BEMs. This framework is novel in (1) involving a pre-calibration mechanism to derive informative priors as well as facilitate parameter selection and building operation schedule analysis that were not considered in current methods. This mechanism also helps alleviate over-parameterization issues, i.e., high posterior parameter identifiability [38], achieving reliable Bayesian calibration results in hourly resolution; (2) leveraging deep learning techniques, i.e., Long Short-Term Memory network (LSTM), as the surrogate model to capture thermo-dynamics and temporal-dependencies of energy use in high-resolution modeling and realize multi-channels of outputs (i.e., heating, cooling, and electricity), addressing limitations of surrogate model mentioned above; (3) simplifying the covariance matrix calculation to significantly reduce the computational burden. The proposed framework aims to enhance reliability, applicability, computational efficiency, and calibration resolution of the current Bayesian-based calibration approach, to make automated high-resolution calibration be possible to produce BEMs usable in broader applications (e.g., demand response and smart control).

2. Methodology

2.1. Overview of the proposed framework

The proposed framework consists of two primary phases: Pre-Calibration and Rapid Auto-Calibration, as shown in Fig. 2. During the Pre-Calibration phase, the first step is Data Collection & Disaggregation, aiming to gather and disaggregate data to create more reliable datasets for subsequent schedule analysis and parameter selection. After disaggregation, schedule analysis and parameter selection steps (Steps 2 and 3) are performed. These 2 steps focus on identifying specific information to better define informative priors for schedules and important building physical parameters in calibration. After identifying calibrated parameters, we establish the high-fidelity physics-based model (EnergyPlus) as the initial building model to calibrate in this research (Step 4). Moving on to the Rapid Auto-Calibration phase, our framework employs a sampling method to generate the simulation dataset (Steps 5) for training the surrogate model to support calibration parameter evaluation in the following step (Step 6). Subsequently, optimization steps for rapid auto-calibration can be conducted using the developed novel Bayesian structure integrating the surrogate model (Step 7). Finally, validation is performed to ensure the calibrated model meeting requirements of ASHRAE guideline in hourly resolution (Step 8).

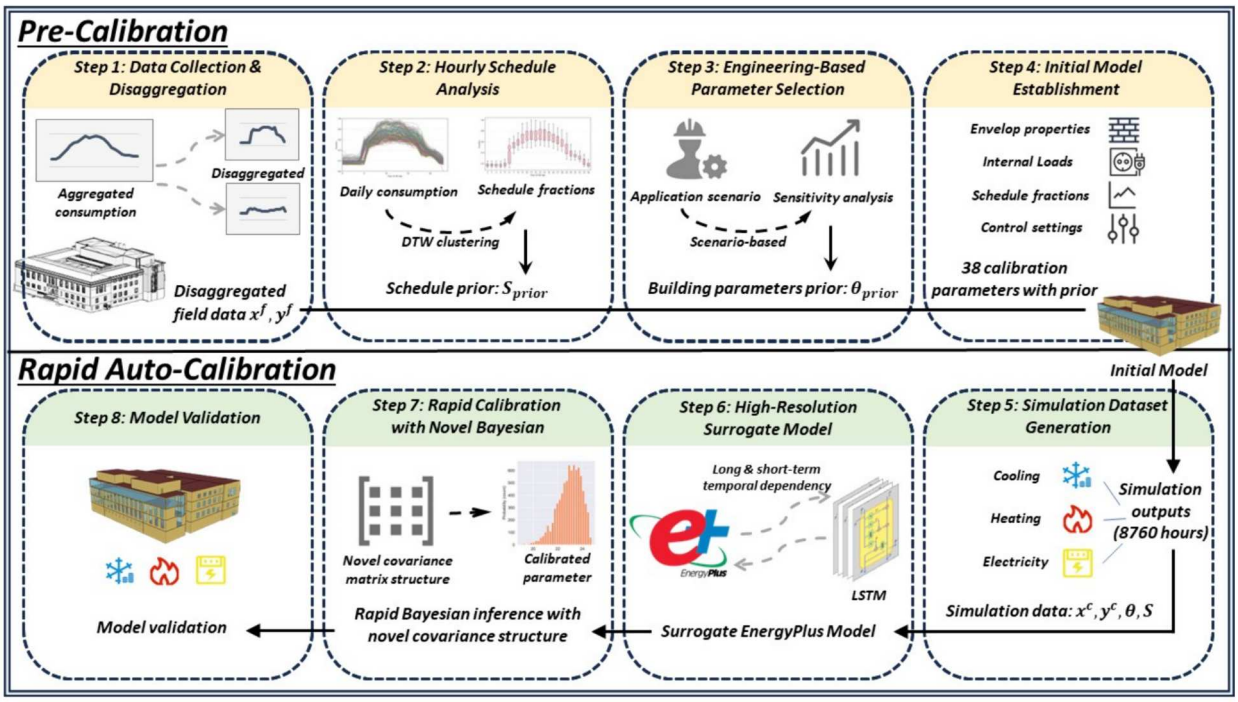


Fig. 2. Proposed novel deep learning-based Bayesian calibration framework for high-resolution.

2.2. Pre-calibration

In optimization-based auto-calibration, modelers firstly specify the number and range of parameters to be calibrated. In former practice, modelers typically don't detailly analyze the building operation patterns and energy usage, e.g., related with the control strategy, operation schedule, and occupant behaviors while all calibration works (parameter adjustment) are automatically handled by the optimization algorithms. This is acceptable for low-resolution calibration (e.g., monthly), as monthly calibration usually incorporates limited measurement data (e.g., 12 months for a year) with fewer calibration parameters (typically 4–8). Therefore, it is relatively easy to auto-identify the optimal set of parameters in model calibration that accurately reflects monthly building energy use. However, for high-resolution calibration with hourly measurements (8,760 h of heating, cooling, and electric), calibration of larger number of parameters related to building construction, control, and operation schedules are required to comprehensively capture the building dynamics. Defining informative priors for these parameters with appropriate initial ranges is challenging due to the large volume of data and parameters involved, but important to facilitate the converging process during calibration and ensure reliable calibration performance. This is why high-resolution calibration necessitates pre-calibration with a detailed analysis of building operation patterns and energy use in the first place.

2.2.1. Data collection and disaggregation

This task aims to gather field data and decompose it to extract detailed building energy consumption patterns to support high-resolution calibration. In general, the types of building consumption to analyze include cooling, heating, and electricity. Cooling

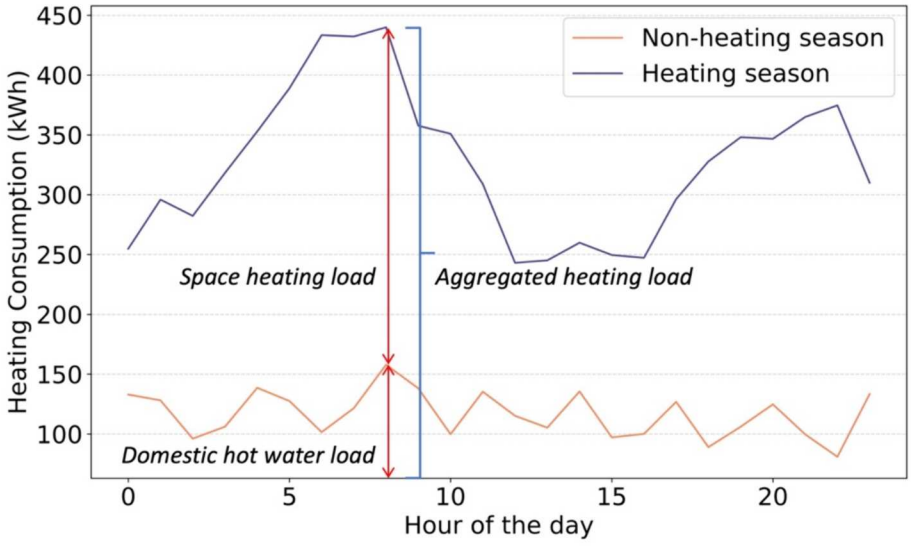


Fig. 3a. Disaggregation of heating consumption.

consumption involves space cooling, continuous cooling of the mechanical room, and the chemistry laboratory (e.g., continuous cooling to support protein culture). Heating consumption includes energy use for domestic hot water and space heating. Electrical consumption covers equipment, appliances, and lighting usage. To improve calibration accuracy, it is crucial to disaggregate and process these consumptions separately: (1) Disaggregation of Heating Consumption: In our case study, heating consumption persist even during the non-heating season, e.g., summer (Fig. 3a). This steady heat demand is attributed to hot water usage. Subtracting non-heating season heating from total heating yields space heating energy use. (2) Disaggregation of Cooling Consumption: In our case study, there is a base cooling load throughout the year for mechanical rooms and chemistry laboratories (Fig. 3b). If this base load is not separated, it might lead to a significant overestimation of the cooling load (for space cooling), affecting calibration accuracy. Subtracting this constant base load from total cooling reveals actual energy use for space cooling. (3) Disaggregation of Electrical Consumption: Without sub-metering, differentiating electricity consumption between lighting and equipment could pose a challenge. Therefore, it is crucial to improve the auditing process and gather more comprehensive information on usage of sub-systems (e.g., lighting and equipment usage), to effectively attribute aggregate energy consumption to different sectors and facilitate high-resolution calibration. Effective disaggregation of energy use provides accurate prior information on heating, cooling, and electricity usage, supporting the calibration process. Particularly, separating the base load for laboratory cooling is crucial for accurate cooling load calibration and avoiding overestimation of cooling energy use (i.e., the base cooling load for mechanical room and laboratory is added to the regular load for space cooling).

2.2.2. Hourly schedules analysis

One of the barriers to realize high-resolution calibration of BEMs is failing to consider operation schedules and occupant behaviors in the calibration process [56]. Occupants and schedules have been recognized as important and influencing factors affecting the accuracy of a calibrated BEM [19–22,57,58]. Consequently, reliable estimate of schedule fraction parameters of occupant behaviors (e.g., occupancy, plug load usage) are crucial to achieve accurate calibration of BEM in high resolution. Even though the building energy usage (including lighting, equipment, plug-in loads, occupant, etc.) is fluctuating throughout its operation periods, there remains an underlying pattern. To explore this pattern, we analyzed the historical electricity data to derive these schedules as informative priori of these input parameters in high resolution Bayesian calibration of BEM.

In this step, our objective is to obtain the prior of schedule fraction parameters (S_{prior}) in calibration by analyzing the uncertainty of schedule fractions ($S = [s_1, s_2, \cdots, s_q]$, where q is the number of schedule fractions). This step mainly consists of three small steps to effectively approximate the prior of hourly schedule fraction parameters: (1) Clustering and Normalization of Daily Electricity Profiles: By clustering the normalized daily electricity consumption profiles using the Dynamic Time Warping (DTW) algorithm [59], we derive energy use patterns for different types of the day (e.g., weekdays or weekends) and obtain their respective daily schedules (Fig. 4). The strength of DTW algorithm lies in its insensitivity to local changes, making it robust in handling noise and deformation in time series data analysis, especially in complex scenarios [60]; (2) Quantifying the Range of Schedule Fractions: After completing the clustering and normalization of daily electricity profiles, we employ Box plots [61] to represent the uncertainty range of schedules. The utilization of the third quantile indicates the upper limit of hourly schedule fractions, while the first quantile represents the lower limit of hourly schedule fractions (Fig. 5). This approach provides a precise prior range for hourly schedule fractions; (3) Merging Hourly Schedule Fractions: Adjacent hourly schedules sometimes have similar operation modes, e.g., 1 AM to 5 AM all have similar schedule ranges. Therefore, we merge the hourly schedules of neighboring schedule modes, and the number of hourly schedule fractions is reduced from 48 to 22 interval schedule fractions (Fig. 6). The benefits of this merging include reducing calibration parameters and avoiding over-parameterization, aiding in surrogate model training and the posterior distribution sampling. Following these steps, we can effectively approximate the uncertainty range of hourly schedule fraction parameters and derive meaningful interval schedule fractions for further analysis and modeling.

2.2.3. Engineering-based parameter selection

After determining the schedules of building operation, the next step is to select other building calibration parameters that, in addition to schedules, need to be calibrated and are related to building thermal properties and control ($\theta = [\theta_1, \theta_2, \cdots, \theta_k]$, where k is the number of calibration parameters) and determine parameters prior (θ_{prior}). The conventional parameter selection method typically involves calibrating a few parameters identified through sensitivity analysis that have a significant impact on the calibration results [36–38,62]. Due to the computational burden of the Bayesian calibration in high-dimensional parameter spaces, using fewer parameters can reduce the

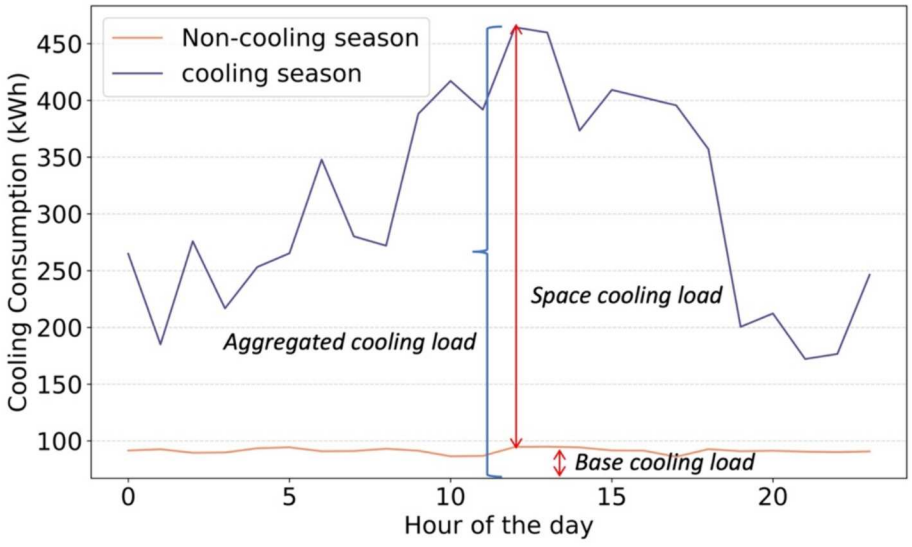


Fig. 3b. Disaggregation of cooling consumption.

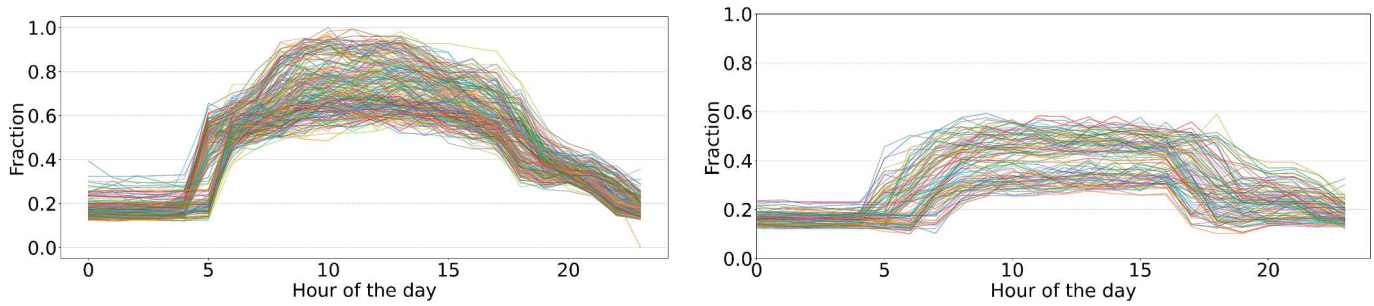


Fig. 4. Clusters of electricity consumption profiles (weekdays: left, weekends: right).

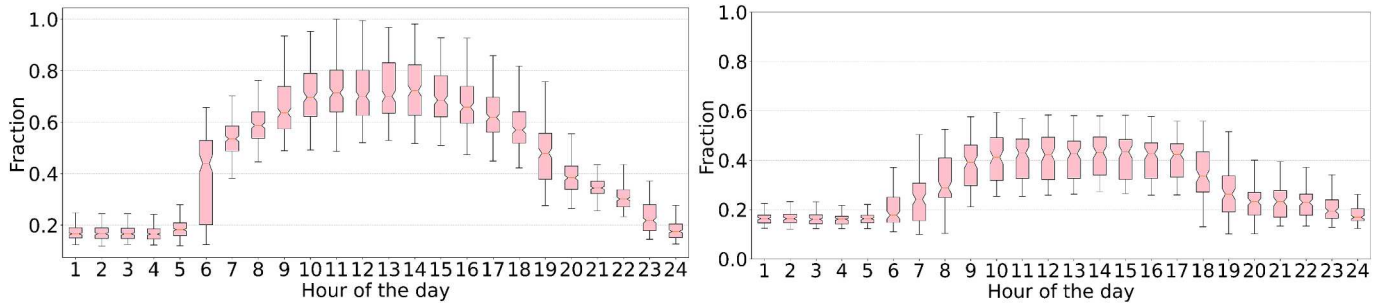


Fig. 5. Daily schedules (weekdays: left, weekends: right).

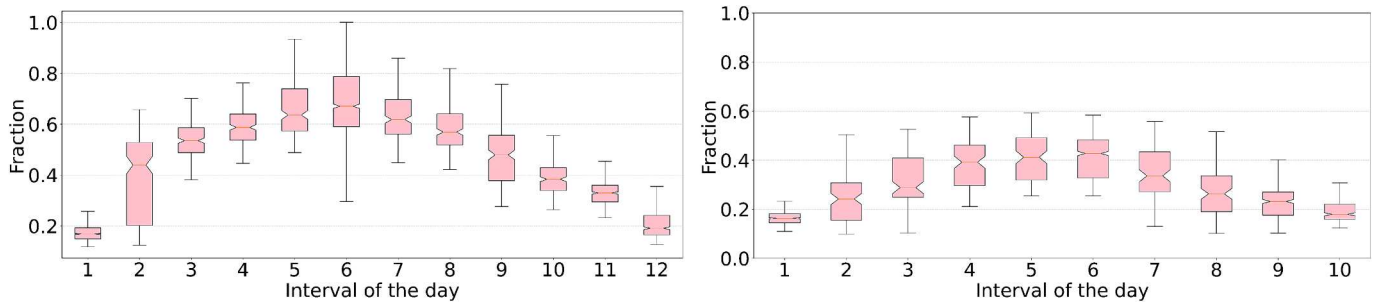


Fig. 6. Interval schedules (weekdays: left, weekends: right).

computational cost [37,38]. Sensitivity analysis is feasible for monthly calibration because a small number of parameters are sufficient to fit monthly building behavior, while too many parameters may lead to overfitting since 12 points of monthly measurements in a year is a small dataset. However, selecting only a few important parameters are insufficient to realize satisfactory performance in high-resolution building model calibration. Calibration of more parameters are needed to accurately describe the temporal dependencies of building dynamics in operation. Our experiments demonstrated that simply selecting representative parameters through sensitivity analysis proves challenging in meeting the requirements of high-resolution calibration (by ASHRAE guideline 14 [42]). Consequently, in order to capture the building operation status, more input parameters are selected in calibration based on engineering experience, i.e., an engineering-based parameter selection. This typically involves model input parameters such as envelope properties, control settings, and schedule parameters reflecting building dynamics and operation patterns.

2.2.4. Initial model establishment

In this step, we create an initial EnergyPlus model based on selected calibration parameters (including schedules). This initial model serves as the basis for calibrating parameters to train the surrogate model and conduct auto-calibration. Table 1 illustrates the 38 calibration parameters determined through the engineering-based selection.

2.3. Rapid auto-calibration

2.3.1. Simulation dataset generation for surrogate modeling

In the auto-calibration process, iterating physics-based simulations (required to evaluate how different combinations of parameters could result in a matching between modeling and measurements) can be computationally intensive. To mitigate computational burdens in calibration, a surrogate model or *meta*-model is typically used to emulate physics-based modeling (e.g., EnergyPlus) for evaluation of sampled combination of building parameters. To establish the surrogate model, the first step is to sample different sets of parameters and correspondingly perform simulations to generate simulation datasets (D^c) for training of surrogate models.

To select design points in parameter sampling and surrogate model training, Latin hypercube sampling (LHS) [63] methodology is employed. The goal of using LHS is to comprehensively explore the multi-dimensional parameter space, covering a wide range of building operation scenarios possible in practice. Through Python scripts, we automatically sample and feed different parameter combinations into physics-based simulation programs (EnergyPlus) to generate corresponding simulation outputs, as the training dataset for surrogate

Table 1
Calibration parameters.

Building information parameters	Control parameters	Schedule parameters	
		Weekdays	Weekends
Conductivity of wall insulation (W/m • K)	Cooling set-point at occupied hours (°C)	Interval 1	Interval 1
Conductivity of roof insulation (W/m • K)	Cooling set-point at unoccupied hours (°C)	Interval 2	Interval 2
Conductivity of window glass (W/m • K)	Heating set-point at occupied hours (°C)	Interval 3	Interval 3
SHGC	Heating set-point at unoccupied hours (°C)	Interval 4	Interval 4
Electric equipment definition (W/m^2)	Chilled water supply temperature for AHU (°C)	Interval 5	Interval 5
Lights definition (W/m^2)	Supply air temperature of each AHU (°C)	Interval 6	Interval 6
People definition (m^2/person)	Outdoor air flow at occupied hours (1/h)	Interval 7	Interval 7
	Outdoor air flow at unoccupied hours (1/h)	Interval 8	Interval 8
	Hot water peak flow rate (m^3/s)	Interval 9	Interval 9
		Interval 10	Interval 10
		Interval 11	
		Interval 12	

modeling in the subsequent step. This training datasets consist of 38 calibration parameters form pre-calibration and 5 weather relevant parameters (including dry bulb, humidity, wind speed, solar radiation, and day type) as inputs, and heating, cooling, and electricity use as outputs. Automating the simulation process and collecting simulation datasets expedite the calibration process while minimizing the needs of manual intervention.

2.3.2. *High-resolution surrogate model*

After generating training data for the surrogate model from the last step, we employ the Long Short-Term Memory (LSTM) algorithm [64] as the deep learning-based surrogate model for high-resolution calibration.

Auto-calibration work typically requires the use of optimization tools (e.g., Bayesian inference) to find the optimal sets of model parameters that describe actual building operation. However, the process of auto-identifying optimal sets of parameters always involves parameter evaluation, i.e., inputting candidate parameter sets into building models to determine if the sampled parameters make modeling outcomes match observations. This triggers significant computational burdens for iterative running of models, especially when high fidelity physics-based models (e.g., EnergyPlus) are involved. The use of surrogate (or meta) models in auto-calibration aims to simplify this iterative computing process during optimization [39]. The efficiency and effectiveness of using surrogate models in Bayesian calibration have been extensively demonstrated through various methods such as multiple linear regression (MLR), Gaussian process (GP), multilayer perceptron (MLP), etc [34,40]. However, current surrogate models struggle to capture the nonlinear and time-dependent relationships in building operation, making them difficult to be applied in high-resolution calibration.

To address these challenges and support high resolution calibration, a capable surrogate model is needed to effectively capture building dynamics in high resolution with temporal dependencies. Hence, the Long Short Term Memory (LSTM), as the widely used model for processing time series data [55,65], is used as the surrogate model in this research. Fig. 7 is the high-level structure of one LSTM unit. Each LSTM unit has outputs (h) and a cell state (C). At step t, the input (h_{t-1}, C_{t-1}, x_t) consists of the output (h_{t-1}, C_{t-1}) from the previous step $t-1$ and the input parameters (x_t) for step t. By passing through the forgetting gate f_t, updating gate i_t, and output gate o_t, the new unit output h_t and cell state

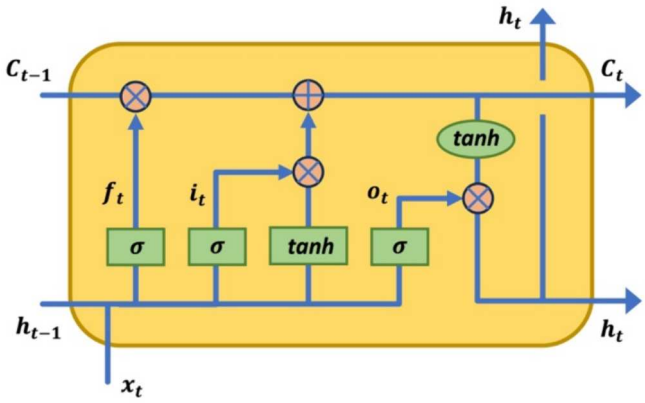


Fig. 7. LSTM unit structure.

C_t are obtained. This mechanism enables continuous forward propagation and captures long-term dependencies, addressing the coupling of sub-systems, human behavior, and thermal inertia in the building thermal processes.

The proposed surrogate model structure, as shown in Fig. 8, consists of an input layer, LSTM layers, a ReLU layer, and a fully connected (FC) layer. The input layer is derived from the raw dataset through reshaping and splitting, resulting in three dimensions: input size, time step, and batch size. Input size represents the number of input parameters for the model, time step represents the length of each time series data, and batch size represents the total number of time series data. The output layer has three outputs: cooling, heating, and electricity consumption. This deep LSTM architecture aims to capture the complex thermodynamics and temporal dependency of buildings in operation (i.e., 8760 h of building operation). As suggested in the reference [66], a FC layer is typically added after the LSTM layer to map all the predicted sequence to the desired output size. Additionally, we innovatively introduced the ReLU [67] layer after the LSTM layer to eliminate negative values, as energy consumption cannot be negative. This innovation enhances both the performance and robustness of the model. The ReLU function, defined as $y = max(0, y)$, where negative values are transformed to zero.

2.3.3. *Proposed Bayesian framework for calibration*

The proposed framework incorporates a novel Bayesian calibration method that optimizes the calibration process in a computationally efficient manner. This method reduces the size of the covariance matrix and simplifies the covariance structure, achieving time efficient computation.

Within the framework of KOH Bayesian calibration [26] for BEMs, the Bayesian inference can be expressed as the relationship between the building observation y_i (measurement data, e.g., cooling, heating, and electricity), the true building operation process $\zeta(x_i)$, and the physics-based model $\eta(x_i, \theta)$ (the framework adjusts the initial parameters in EnergyPlus model to align with ground truth building operation), as described by Eq. 1:

$$y_i = \zeta(x_i) + e_i = \eta(x_i, \theta) + \delta(x_i) + e_i \quad (1)$$

where e_i represents the observation error for the i th observation and $\delta(x_i)$ is a model inadequacy function. We usually assume that the e_i s are independently distributed as $N\left(0, \sigma_y^2\right)$. $\delta(x_i)$ can be seen as modeling discrepancy between true building operation process $\zeta(x_i)$ and physics-based simulation $\eta(x_i, \theta)$. x_i denotes the field weather data, and θ represents the model parameters to be calibrated.

In order to minimize computation time in parameter calibration, we employed a surrogate model (Section 2.3.2) to substitute iterative computing of physical-based model $\eta(x_i, \theta)$, as described by Eq. 2.

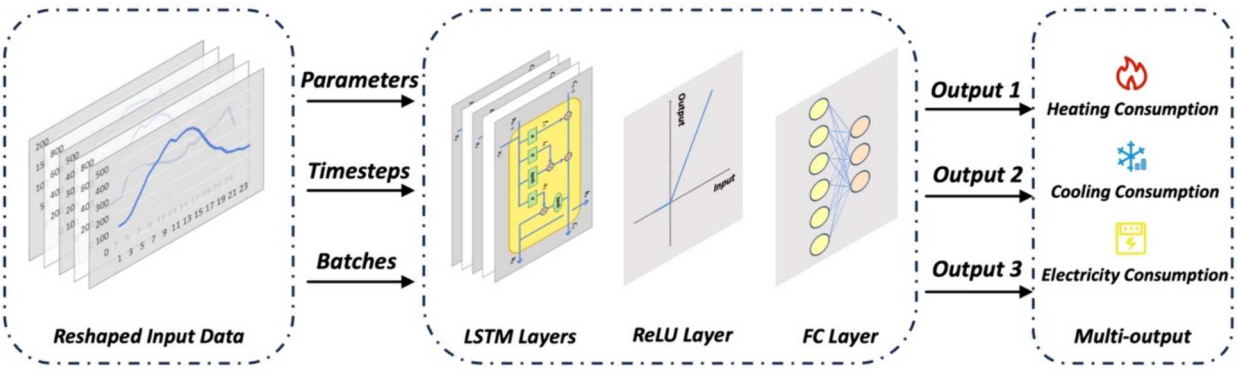


Fig. 8. High-resolution surrogate model framework.

$$\eta(x_i,\theta) = f_{sur}(x_i,\theta), \eta(x_i,\theta) \sim N\left(f_{sur},\sigma_\eta^2\right) \tag{2}$$

where σ_η is the RMSE of the surrogate model that is the discrepancy between the surrogate model $f_{sur}(x_i,\theta)$ and the physics-base model $\eta(x_i,\theta)$, i.e., EnergyPlus model in our case study.

By specifying the structure of the covariance matrix, a typical Gaussian process (GP) (Eq. 3) can flexibly represent the model behavior and achieve an exact fit on the given observation samples y_i.

$$y_i \sim N(\mu,\Sigma), \mu = \eta(x_i,\theta), \Sigma = \Sigma_y + \Sigma_\delta + \Sigma_\eta \tag{3}$$

where, Σ_y is a $n \times n$ covariance matrix used to present for observation error e_i, given n field observations. Σ_δ is a $n \times n$ covariance matrix used to present for model inadequacy with true building process $\delta(x_i)$. Σ_η is a $(m+n)\times(m+n)$ covariance matrix used to present the error of physics-based simulation $\eta(x_i,\theta)$, given m simulation sample points. The calibration parameters include the model parameters θ in the mean matrix μ and the hyperparameters ϕ in the covariance matrix Σ. Following the Bayes' rule, the posterior distribution $p\left(\theta,\phi|y_i\right)$ can be derived from their priors $p(\theta), p(\phi)$, and the likelihood of the observations $p\left(y_i|\theta,\phi\right)$ as follows:

$$p(\theta,\phi|y_i) \propto p(y_i|\theta,\phi)\cdot p(\theta)\cdot p(\phi) \tag{4}$$

The likelihood $p\left(y_i|\theta,\phi\right)$ can be expressed as:

$$p(y_i|\theta,\phi) = |\Sigma|^{-\frac{1}{2}} exp\left\{-\frac{1}{2}\left[(y_i-\eta)^T\Sigma^{-1}(y_i-\eta)\right]\right\} \tag{5}$$

where priors $p(\theta)$ and $p(\phi)$ can be determined based on engineering experience and practical situations, with details provided in the engineering-based parameter selection and hourly schedule analysis sections.

As Eq. 1, the current traditional Bayesian calibration of physics-based simulation $\eta(x_i,\theta)$ and model inadequacy $\delta(x_i)$ typically assume a joint multivariate GP [32,38,41], requiring the fitting of field observations and simulation outputs. For hourly calibration, the dimensionality $(m+n)^*(m+n)$ of the covariance matrix Σ_η becomes enormous given n (n = 8,760 for hourly data) field observations and m physics-based simulation samples (easily exceeding 10,000 samples in practice). This leads to a substantial computational burden to compute Σ^{-1} in calculating likelihood $p\left(y_i|\theta,\phi\right)$ (Eq. 5), making hourly calibration unfeasible. Rather than employing GP as in the current traditional Bayesian calibration framework, we use a deep learning method (LSTM, in Section 2.3.2), as a high-resolution surrogate model effective in capture temporal correlations in hourly building modeling, to emulate $\eta(x_i,\theta)$ (Eq. 6):

$$\eta(x_i,\theta) = f_{LSTM}(x_i,\theta), \eta(x_i,\theta) \sim N\left(f_{LSTM},\sigma_\eta^2\right) \tag{6}$$

where σ_η is the RMSE of the surrogate model that is the discrepancy between the LSTM surrogate model $f_{LSTM}(x_i,\theta)$ and the computer simulation $\eta(x_i,\theta)$. This means that the Bayesian process in our approach no longer requires direct fitting to the m simulation sample points; instead, pre-trains the simulation data using a surrogate model. This strategy reduces the size of the covariance matrix Σ from $m+n$ to n.

However, the computational burden remains unacceptably high due to the modeling of model inadequacy term $\delta(x_i)$ as a GP (with large covariance matrix). The advantage of our high-resolution surrogate model, combined with an effective pre-calibration analysis, lies in its ability to handle the $\delta(x_i)$ between the true building thermal process and physics-based simulation. As a result, instead of employing a multivariate joint GP to fit the model inadequacy term $\delta(x_i)$, we simplify the approach and assume that $\delta(x_i)$ follows a Gaussian distribution: $N\left(0,\sigma_\delta^2\right)$.

In conclusion, by (1) reducing the size of Σ to $n \times n$ by fitting an LSTM surrogate model, and (2) assuming $\delta(x_i)$ $N\left(0,\sigma_\delta^2\right)$ with our pre-calibration analysis to ensure the accuracy of the initial model. This simplification leads to the entire covariance matrix (Eq. 7) as follows:

$$\Sigma = \Sigma_y + \Sigma_\delta + \Sigma_\eta = (\sigma_y^2+\sigma_\delta^2+\sigma_\eta^2)\cdot I_n = \sigma^2\cdot I_n \tag{7}$$

In Bayesian calibration, after deriving the likelihood function, it is common to employ sampling to obtain posterior distributions of calibration parameters, accelerating the Bayesian inference process. The Metropolis-Hastings (MH) method [68], as a common algorithm in the MCMC for sampling from intricate probability distributions, generates samples by introducing a proposal distribution of candidate parameter and utilizes an acceptance-rejection criterion to determine whether the candidate parameter should be accepted. Detailed elucidation and the procedural steps of the MH algorithm are presented in Algorithm 1. The validity and efficacy of our previous simplification of covariance in the likelihood evaluation lies in the fact that this simplification does not significantly affect the acceptance ratio α. Consequently, our simplification preserves the integrity of subsequent inference and posterior distribution sampling processes.

Algorithm 1 MCMC Sampling (Metropolis-Hastings)

Require: Target distribution $P(\theta|y_i)$, proposal distribution $Q(\theta|y_i)$, number of samples N
1: Initialize sample set $X = \{\}$
2: **for** $n = 1$ to N **do**
3: Sample θ from proposal distribution $Q(\theta|y_i)$
4: Sample u form uniform distribution $U(0,1)$
5: Compute acceptance ratio $\alpha = \frac{P(\theta_n|y_i)Q(\theta_n|y_i)}{P(\theta_{n-1}|y_i)Q(\theta_{n-1}|y_i)}$
6: **if** $u \le \alpha$ **then**
7: Add θ_n to sample set X
8: **else**
9: Add θ_{n-1} to sample set X
10: **end if**
11: **end for**
12: **return** Sample set X

2.3.4. Model validation

The evaluation of BEMs often adheres to the ASHRAE guideline 14 [42], as a standard for assessing the performance of BEM calibration. One of the metrics used is the coefficient of variation root mean square error (CV-RMSE) that quantifies the percentage error between the simulated and measured data. Additionally, the normalized mean bias error (NMBE) indicates the deviation percentage of actual data in calibration, with either underestimate (NMBE>0) or overestimate (NMBE<0). According to the guideline, for hourly calibration results, NMBE should be below 10 %, and CV-RMSE should be less than 30 %. The CV-RMSE is calculated using the following:

$$CV-RMSE=\frac{\sqrt{\frac{\Sigma_{i=1}^{n}\left(y_i-\widehat{y}_i\right)^2}{(n-p)}}}{\overline{y}}\times 100 \tag{7}$$

The NMBE is calculated as follows:

$$NMBE=\frac{\Sigma_{i=1}^{n}(y_i-\widehat{y}_i)}{(n-p)\times\overline{y}}\times 100 \tag{8}$$

where:

y_i represents the value of the field observation.

$\overline{y}$ denotes the arithmetic mean of the sample comprising n observations.

$\widehat{y}_i$ signifies the predicted value of y obtained through computer simulation.

n corresponds to the number of data points or periods in the baseline period.

p refers to the number of parameters or terms in the baseline model, which is developed through mathematical analysis of the baseline data. For building calibration, $p = 1$.

3. Case study

In this section, we present a case study to demonstrate and validate the effectiveness of the proposed Bayesian calibration framework. Firstly, we provide an overview of a campus building (the model of which we calibrated) in this study (Section 3.1). Subsequently, we describe the pre-calibration process to obtain accurate priors of parameters and schedules for constructing an initial model capable to capture the actual thermal processes of building operation (Section 3.2). Finally, in Section 3.3, we present the process of the rapid auto-calibration.

3.1. Building description

As depicted in Fig. 9a, the case study focuses on the Crocker Science Center (CSC) building, located in the main campus of the University of Utah. This four-story building spans a total area of roughly 11,437 square meters. It encompasses various spaces, including offices, conference rooms, mechanical rooms, an auditorium, and laboratories. The building employs a variable air volume (VAV) system for air conditioning, comprising two air handling units (AHUs). The cooling is supplied by a central plant on the campus, while heating is generated by four boilers within the building. A BEM of the case study was constructed using EnergyPlus (Fig. 9b), leveraging the building information model (BIM) (Fig. 9c) and adhering to the building design code [69]. The weather data files are provided by White Box Technologies [70].

The calibration process is based on hourly energy consumption data, including 8,760 h of heating, cooling, and electricity data collected from January to December 2021 (with ~ 800 missing data points in mainly July and August).

Fig. 9a. CSC Building.

3.2. Pre-calibration process

The purpose of pre-calibration is to acquire accurate parameters and prior ranges, particularly schedule fraction parameters, in order to establish a high-resolution initial model that effectively captures dynamics of building operation. In this case, an hourly schedule analysis was conducted, resulting in the extraction of two types of daily schedules with 22 schedule parameters (representing the dynamics of building operation and human behavior on weekdays and weekends as comprehensively as possible without over-parameterization, as Section 2.2.2). Additionally, by applying the engineering-base parameter selection method, a comprehensive set of 16 parameters were selected, encompassing thermal properties and control parameters of the building. The initial range of these parameters are set as triangular distributions in calibration, as detailed in Table 2.

3.3. Rapid auto-calibration process

In the automated Bayesian calibration process, an LHS approach is employed to generate 38 × 20 design points (Parameters = 38, LH design points = 20) based on the uncertainty range of the input parameters. These points are used to simulate hourly energy consumption, resulting in a simulation dataset (D^c) comprising 175,200 points with 38 input building parameters, 5 weather parameters, and 3 outputs (cooling, heating, and electricity use). The dataset is normalized and used as training data for the surrogate model. A two-layered multi-output LSTM model is fitted as the surrogate model to establish the relationship between the model inputs and the outputs of interest. The LSTM model parameters are specified in Table 3. It is important to note that for each unique building, the LSTM model needs to be retrained to reflect building dynamics. An MCMC sampling of 10,000 runs is implemented to derive the posterior distribution of calibrated parameters, with the first 2,000 sample points discarded as burn-in. Towards the end, this case study is also compared with traditional Bayesian calibration methods (GP) [32] and lightweight Bayesian calibration methods

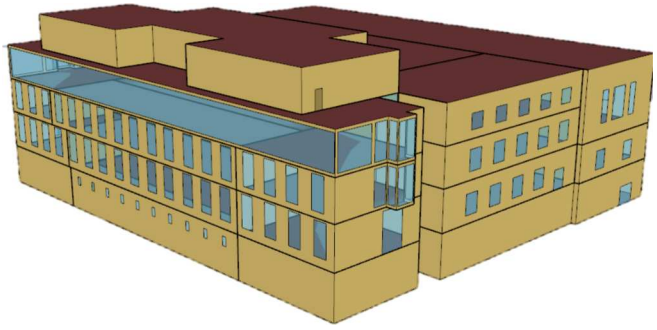


Fig. 9b. Building energy model in EnergyPlus.

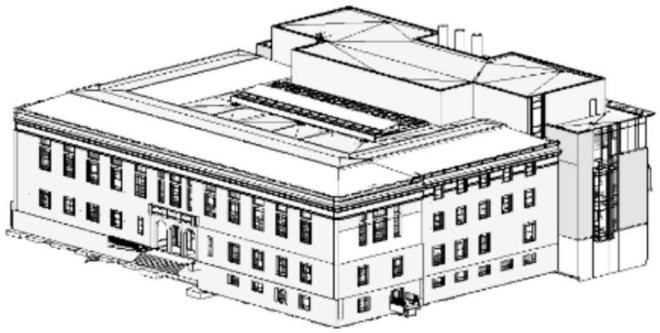


Fig. 9c. Building information model in REVIT.

(Linear Regression) [41] under the same configuration.

4. Results

4.1. Calibrated parameters

This section presents the acquisition of posterior distributions of calibration parameters through Bayesian inference. All prior distributions assigned to the calibration parameters are modeled as triangular distributions. Fig. 10 illustrates an example posterior distribution of calibrated parameters. The light blue triangle represents the uncertainty range (prior) of the calibration parameter listed in Table 2, and the dark "*" point denotes the estimated value of the calibrated parameter (Table 4) derived from the posterior distribution as inputs into the physics-based model for simulation. The posterior distribution demonstrates that the calibration results for all the parameters are identifiable, each exhibiting a distinct peak. This substantiates that our hourly calibration results constitute a unique solution and do not suffer from the over-parametrization issue, and the calibrated parameters have strong validity [38].

4.2. Model validation

Fig. 11 compares hourly energy consumption observations with simulation outputs from the calibrated model. The cooling, heating, and electricity consumption simulation results closely match field observation data. Heating, cooling, and electricity exhibit NMBE values of 4.5 %, −2.9 %, and 5.5 %, respectively (Table 5). The CV-RMSE for heating, cooling, and electricity are 23.9 %, 28.4 %, and 26.9 %, respectively. All energy consumption types demonstrate satisfactory performance according the ASHRAE guideline [42]. We can find that our weekdays and weekends schedule work for most of the time. Due to data missing (~800 points, mainly around Jul 17th-Aug 17th), we have excluded this period from validation. As a result, there is a discontinuity around this missing period in Fig. 11. In Fig. 11a, since heating consumption includes domestic hot water and space heating. There is a period with repetitive patterns in summer (May-Sep). In Fig. 11b, the trend of cooling consumption of calibrated model aligns well with field observations of actual building operation. However, due to the complexity of building functions and uncertainties, there are still some deviations, particularly around the period of data missing. Regarding the electricity (Fig. 11c), setting typical schedules (one for weekday and one for weekend) is sufficient to realize decent calibration results for the majority of days. However, due to the multifunctionality (e.g., lab) of the

Table 3
LSTM model parameters.

Training data size	140,158
Test data size	35,040
Input dimension	43
Lookback	3
Output dimension	3
Hidden dimension	128
LSTM Layers	2
Epochs	200
Learning rate	0.001

Table 2
Uncertainty range of calibration parameters.

Thermal parameters	Min	Mode	Max	**Control parameters**	Min	Mode	Max	**Schedule parameters**							
								Weekdays	Min	Mode	Max	**Weekends**	Min	Mode	Max
Conductivity of wall insulation (W/m • K)	0.03	0.04	0.05	Cooling set-point at occupied hours (°C)	20	24	26	Interval 1	0.1	0.2	0.3	Interval 1	0.1	0.15	0.2
Conductivity of roof insulation (W/m • K)	0.08	0.09	0.10	Cooling set-point at unoccupied hours (°C)	24	26	28	Interval 2	0.1	0.35	0.6	Interval 2	0.1	0.25	0.4
Conductivity of window glass (W/m • K)	0.01	0.015	0.02	Heating set-point at occupied hours (°C)	18	21	24	Interval 3	0.4	0.5	0.6	Interval 3	0.2	0.35	0.5
SHGC	0.3	0.5	0.8	Heating set-point at unoccupied hours (°C)	12	18	20	Interval 4	0.4	0.55	0.7	Interval 4	0.2	0.35	0.5
Electric equipment definition (W/m^2)	10	17.5	25	Chilled water supply temperature for AHU ()	3	6	9	Interval 5	0.5	0.65	0.8	Interval 5	0.3	0.45	0.6
Lights definition (W/m^2)	5	10	15	Supply air temperature of each AHU (°C)	10	14	18	Interval 6	0.5	0.7	0.9	Interval 6	0.3	0.45	0.6
People definition (m^2/person)	5	10	15	Outdoor air flow at occupied hours (1/h)	0.8	1.1	1.4	Interval 7	0.5	0.65	0.8	Interval 7	0.2	0.35	0.5
				Outdoor air flow at unoccupied hours (1/h)	0.4	0.595	0.79	Interval 8	0.4	0.55	0.7	Interval 8	0.1	0.25	0.4
				Hot water peak flow rate (m^3/s)	0	0.0015	0.003	Interval 9	0.3	0.45	0.6	Interval 9	0.1	0.25	0.4
								Interval 10	0.3	0.4	0.5	Interval 10	0.1	0.2	0.3
								Interval 11	0.2	0.3	0.4				
								Interval 12	0.1	0.2	0.3				

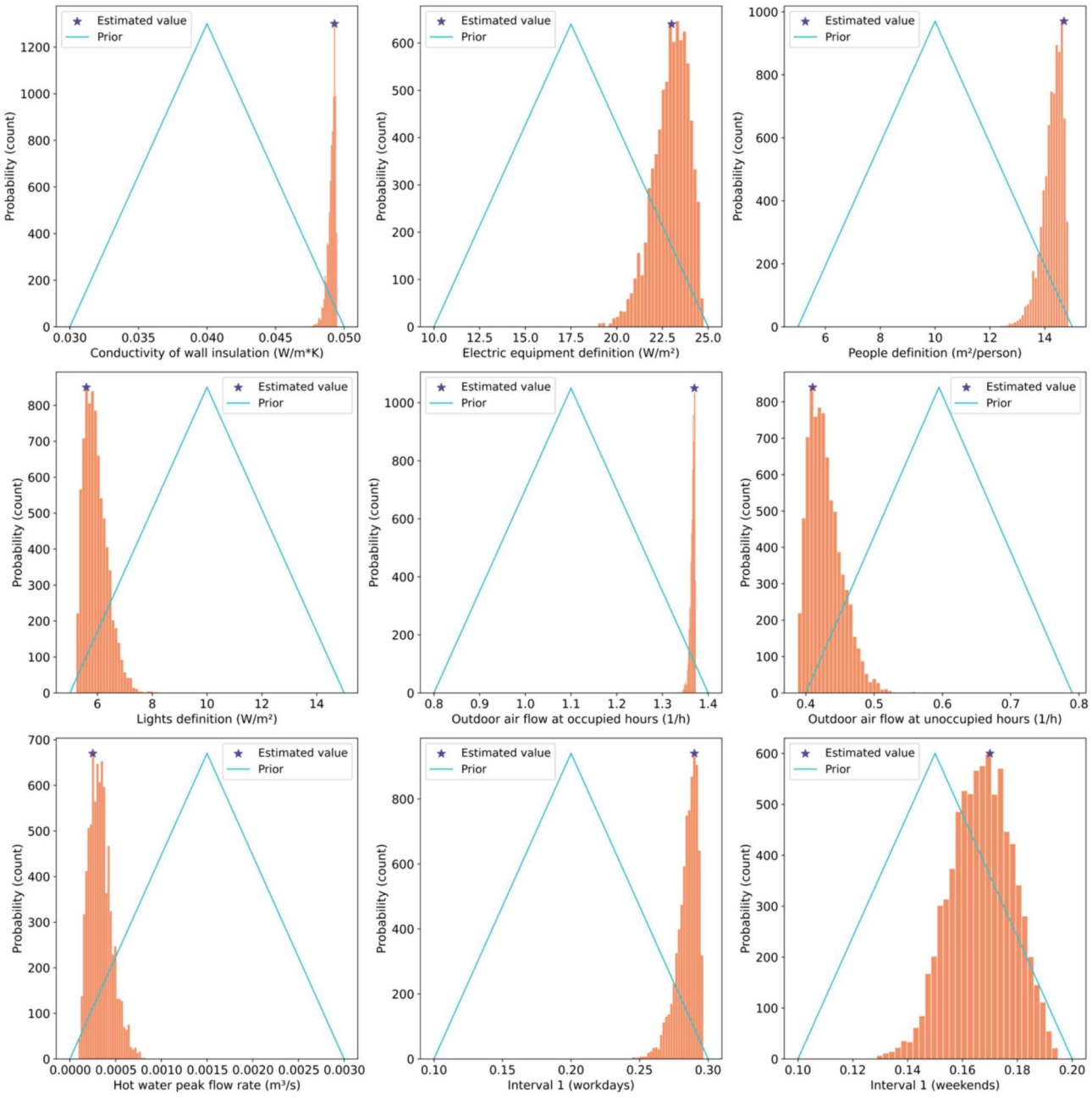


Fig. 10. Example posterior distribution of calibrated parameters.

building, deviations between modeling and measurements are more obvious in certain days than others.

Notably, for the cooling consumption, it exhibits the lowest NMBE (−2.9 %), while the CVRMSE is the highest, reaching 28.4 %. Conversely, for the electricity, it shows the highest NMBE (5.5 %), while the CVRMSE is relatively better at 26.9 %. This distinction is clearly evident in Fig. 11b and Fig. 11c. We will further explore the potential reasons behind this phenomenon in the discussion, Section 5.2.

4.3. Computational time comparisons

For comparison of computational efficiency, the computational time of different Bayesian calibration approaches in calibration (assuming that the calibration approaches converge with 10,000 iterations) is estimated by multiplying the computing time of a limited number of iterations (10 in this case) with a factor of 1,000 to save computational efforts on comparison. Table 6 presents the comparison results, depicted by the computation times of our Proposed Bayesian Approach (using LSTM surrogate and simplifying covariance matrix), the Traditional Bayesian Approach (using Gaussian process model as surrogates) [32], and the Lightweight Bayesian Approach (using Gaussian process and linear regression model as surrogates) [41]. The Proposed Bayesian Approach is able to perform hourly calibration automatically and rapidly, completing the process within 1.67 h, compared to $>$ 3,600 h using the traditional Bayesian method and $>$ 600 h for lightweight Bayesian methods in performing the same calibration task. Additionally, the former methods only support a single output, while the proposed method can handle calibration of multiple outputs (heating, cooling, and electricity) simultaneously. The computational efficiency of our proposed approach is almost independent of the simulation data dimensions and the number of parameters because we simplified the structure of covariance matrix, making the covariance size just relate to

Table 4
The estimated values of the calibrated parameters.

Thermal parameters	Estimated value	Control parameters	Estimated value	Schedule parameters			
				Weekdays	Estimated value	Weekends	Estimated value
Conductivity of wall insulation (W/m • K)	0.0493	Cooling set-point at occupied hours (°C)	22.22	Interval 1	0.29	Interval 1	0.17
Conductivity of roof insulation (W/m • K)	0.0885	Cooling set-point at unoccupied hours (°C)	26.67	Interval 2	0.55	Interval 2	0.12
Conductivity of window glass (W/m • K)	0.0194	Heating set-point at occupied hours (°C)	20	Interval 3	0.58	Interval 3	0.256
SHGC	0.65	Heating set-point at unoccupied hours (°C)	12.78	Interval 4	0.63	Interval 4	0.45
Electric equipment definition (W/m^2)	23	Chilled water supply temperature for AHU (°C)	5.5	Interval 5	0.73	Interval 5	0.425
Lights definition (W/m^2)	5.6	Supply air temperature of each AHU (°C)	11.5	Interval 6	0.81	Interval 6	0.485
People definition (m^2/person)	14.7	Outdoor air flow at occupied hours (1/h)	1.37	Interval 7	0.68	Interval 7	0.42
		Outdoor air flow at unoccupied hours (1/h)	0.41	Interval 8	0.64	Interval 8	0.2
		Hot water peak flow rate (m^3/s)	0.00025	Interval 9	0.56	Interval 9	0.18
				Interval 10	0.34	Interval 10	0.12
				Interval 11	0.34		
				Interval 12	0.18		

Table 5
CV-RMSE and NMBE values.

Error	Heating (kWh)		Cooling (kWh)		Electricity (kWh)	
	NMBE (%)	CV-RMSE (%)	NMBE (%)	CV-RMSE (%)	NMBE (%)	CV-RMSE (%)
Initial model	138.9	145.8	13.2	56.9	8.7	27.7
Calibrated model	4.5	23.9	−2.9	28.4	5.5	26.9

the number of field observations n during Bayesian inference (Eq. 7).

5. Discussion

5.1. Significance of pre-calibration for high-resolution

With trials and errors to calibrate BEM in high-resolution, we find the pre-calibration to derive accurate priors is of paramount importance, as it not only defines the informative priors to inform the parameter sampling and tuning process, but also facilitates the construction of a reliable initial model and surrogate model to help calibration. Pre-calibration is not a conventional manual trial-and-error process; rather, it is driven by knowledge mined from collected building data in general (operation data, construction drawings, etc.), to uncover the hidden building operation pattern from data. Automated calibration can be viewed as the process of identifying optimal parameters from a range of potential (candidate) combinations to achieve calibration objectives. However, the successful implementation of this process hinges on the existence of such a set of calibration parameters as candidate parameters and better definition of initial parameter set that accurately align modeling outcomes with field observations (ground truth). A successful pre-calibration process helps identify the potential set of parameters that make modeling closely match observations while defining better initial calibration parameters to initiate the calibration process (similar to defining better initial parameters in training of machine learning algorithms), hence, facilitating the calibration process with enhanced accuracy and convergence speed in high resolution calibration. When pre-calibration was not carried out, although we embarked on a series of

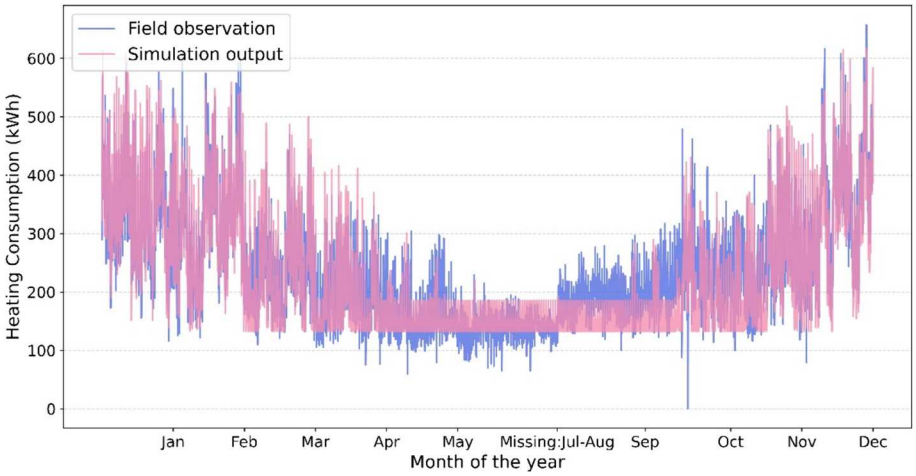


Fig. 11a. Comparison of heating consumption.

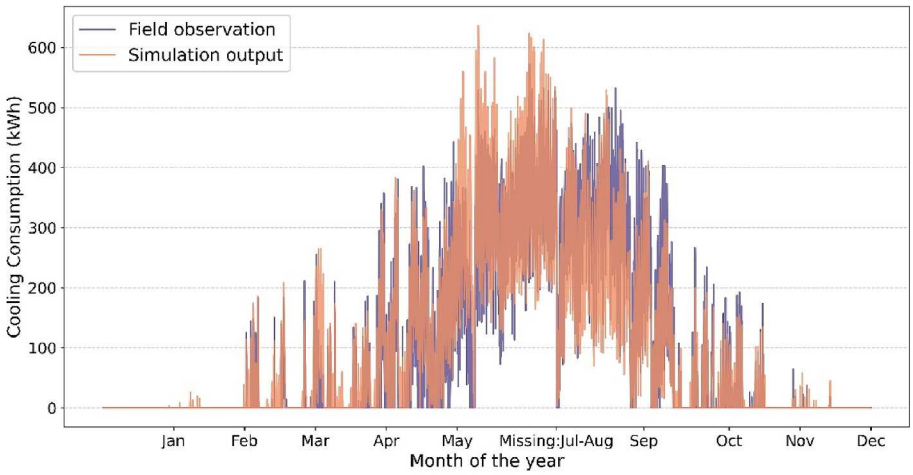


Fig. 11b. Comparison of cooling consumption.

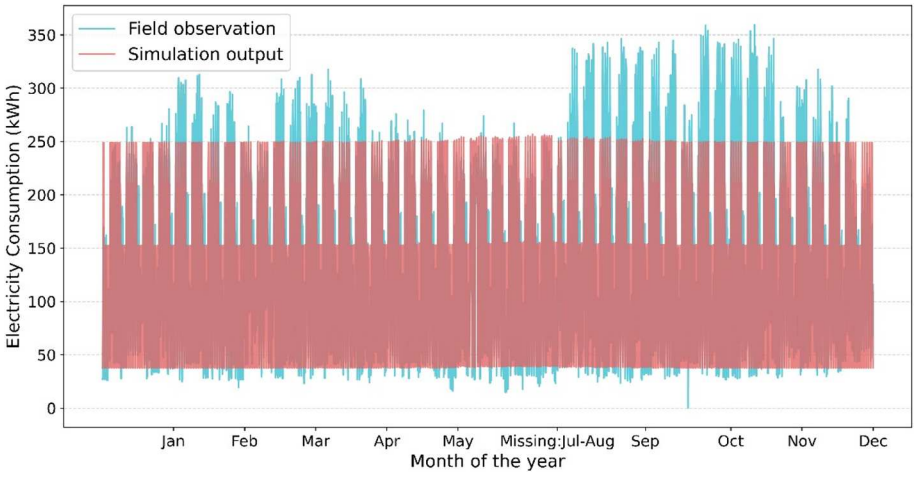


Fig. 11c. Comparison of electricity consumption.

Table 6
Comparison of time results.

Method	Time
Traditional Bayesian Approach [32]	3,646.39 h
Lightweight Bayesian Approach [41]	615.14 h
The Proposed Bayesian Approach	1.67 h

parameter fitting process, it is likely that the outcomes of the high-resolution calibration were not ideal, i.e., did not meet the requirements of ASHRAE guideline 14 [42]. Previous research [6,7,71] also confirmed this. The calibrated model without pre-calibration as the preliminary step is hard to capture of intricate building dynamics (e.g., energy consumption trends, fluctuations, peak and off-peak periods) within the day, with one example presented in Fig. 12 (calibration using informative priors defined by vs. calibration using the default schedules of large office operation from ASHRAE).

In the context of monthly calibration, the objective function only encompasses 12 months of energy consumption. As such, the task of matching monthly energy consumption with actual field observed data is relatively straightforward, without the need to ensure matching hourly or daily operating records. On the contrary, the complexity of calibrating high-resolution models is significantly enhanced, considering a much larger observation dataset with hourly measurements of heating, cooling, and electricity use data across the year (8,760 h) and increasing number of parameters to calibrate. The selection of incorrect parameters or in appropriate boundaries of parameters (e.g., with a far deviated schedules) can result in the persistent mismatch between simulation and ground truth (easily stuck at local minima in objective function) and un-convergence.

Hence, to address the challenges outlined earlier, we find a knowledge and engineering-driven pre-calibration process significant for high resolution building model calibration. As introduced in Section 2.2, this involves a systematic disaggregation of loads, meticulous analysis of schedules, and an engineering-based parameter selection process. Furthermore, we leveraged cluster analysis and schedule reduction methodologies, to prevent over-parameterization and mitigate the emergence of multi-solution issue in the calibration process.

5.2. *Evaluation indicators applicable to hourly calibration*

Through the case study, we also recognize a potential issue of current evaluation indicators, i.e., CV-RMSE, in quantifying the accuracy of calibration. It is possible that CV-RMSE could result in an exaggerated deviation of calibration results since its calculation is highly influenced

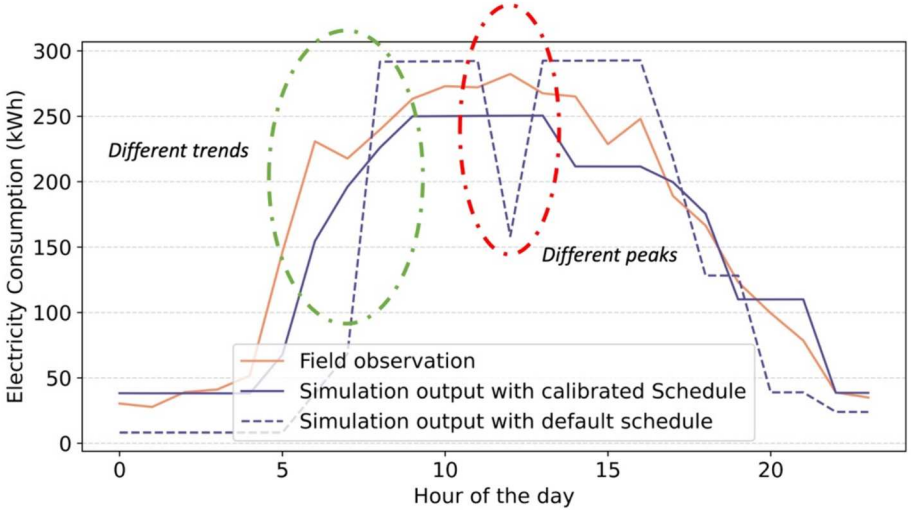


Fig. 12. Influences of the schedule on daily load profile pattern.

by the average of field data $\bar{y}$ (Where $CV-\text{RMSE} = \frac{\sqrt{\sum_{i=1}^{n}(y_i-\widehat{y_i})^2}}{\bar{y}}$). Hence, a small average field measurement $\bar{y}$ could directly lead to high CV-RMSE for cooling load calibration in cold climate or heating load in hot climate, considering a significant portion of the hourly cooling/heating load is at or near zero throughout the year because of climate conditions. In our case study, since the average annual cooling consumption is relatively small, it is easier to observe more significant deviations of modeling to observations in calibration of cooling energy use compared to heating and electricity usage, using CV-RMSE as a measure.

As demonstrated in Fig. 13a, although the trend of modeled and measured cooling is well-matched (with an MBE of only −2.9 %), the CV-RMSE value for cooling energy use was relatively high, i.e., 28.4 %. In contrast, although more significant deviations of heating and electricity energy use trends are observed (as demonstrated in Figs. 13b and 13c), higher average heating and electricity usage across the year result in smaller CV-RMSE values (23.9 % and 26.9 %, respectively), considering the fact that in the cold climate zone of our test building, heating and electricity energy use could be much higher compared to the cooling energy use. Hence, the deviation of modelled building energy use could possibly be exaggerated using CV-RMSE as the indicator, depending on the climate the calibrated building is located in.

6. Conclusion

This research presents a novel deep learning-based Bayesian calibration framework, involving Pre-Calibration and Rapid Auto-Calibration, to calibrate BEMs in high resolution (hourly level). Compared to the traditional Bayesian calibration framework, the proposed approach utilizes deep learning to construct high-resolution surrogate models, capturing complexity of building in operation. Additionally, we simplify the calculation of covariance matrix, hence, significantly reducing the computational burden of the Bayesian calibration process. To enhance robustness and reliability of calibration, we implement a pre-calibration mechanism, including data disaggregation, hourly schedule analysis, engineering-based parameter selection, and initial model establishment. These steps lead to informative priors to avoid over-parameterization and enhance calibration accuracy. Then, a case study is presented to demonstrate satisfactory calibration

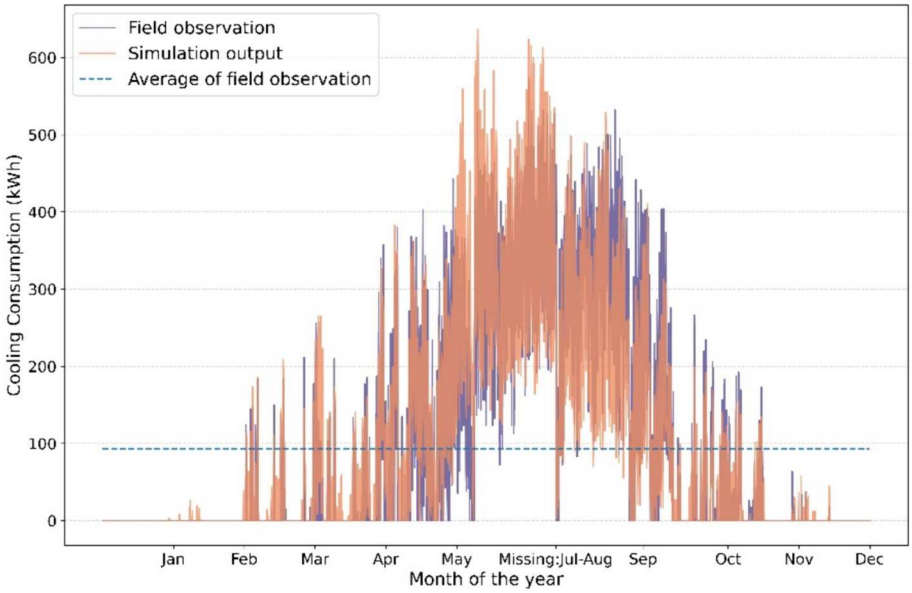


Fig. 13a. Hourly cooling consumption distribution.

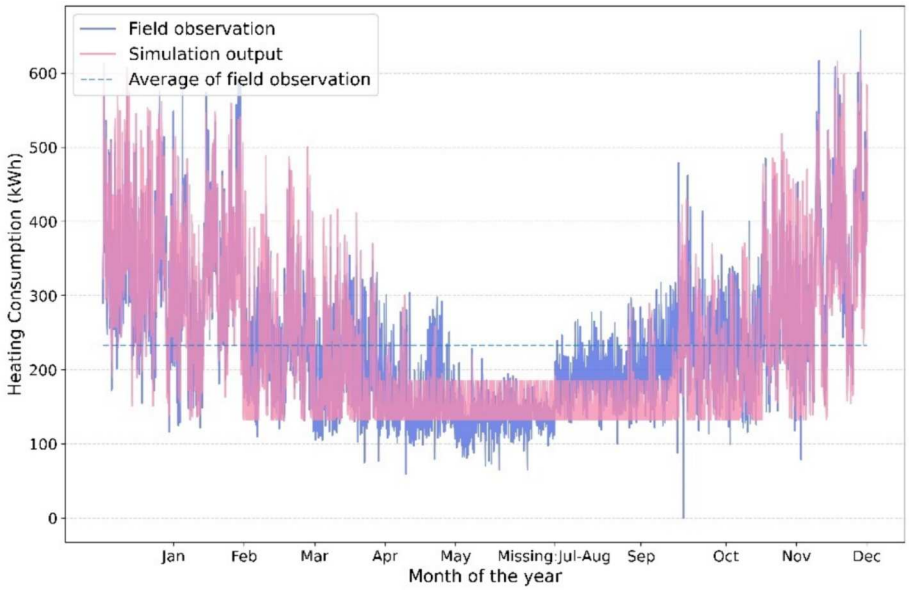


Fig. 13b. Hourly heating consumption distribution.

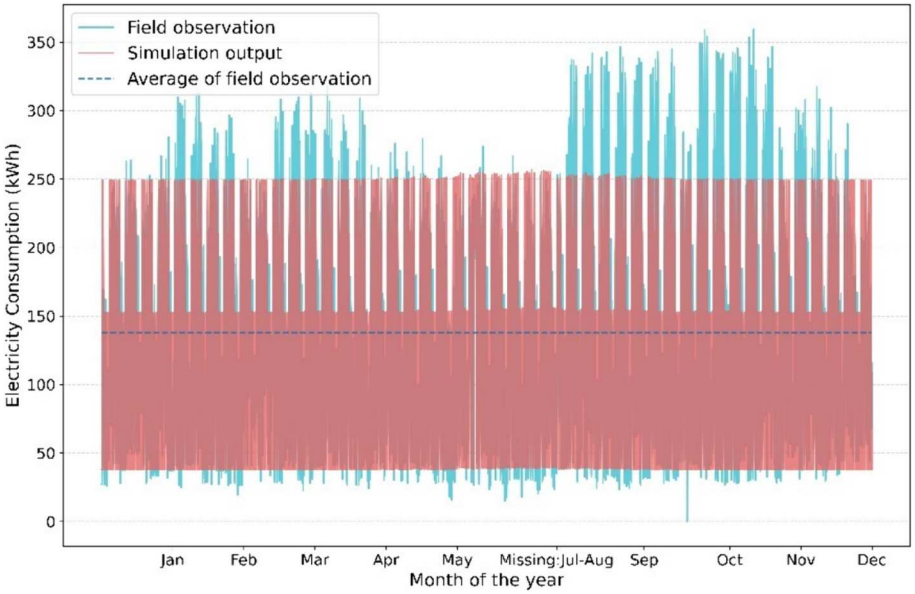


Fig. 13c. Hourly electricity consumption distribution.

performance of applying the developed high-resolution BEM calibration approach in practice. The calibration results indicate that cooling, heating, and electricity consumption all meet the requirements specified in ASHRAE guideline 14. Moreover, our framework exhibits the capability and efficiency to handle high-dimensional parameters and large datasets by reducing $>$ 99 % computational burden compared to traditional Bayesian methods. Through case study and discussion, we recognize the importance of pre-calibration for initial specification of calibration parameter and insufficiency of current indicators to describe the model calibration performance due to strong reliance on average values.

For the limitations of this study, as discussed earlier in the parameter selection, it is tricky to use traditional parameter selection methods (e.g., sensitivity analysis) to identify certain important parameters in high-resolution calibration, such as schedules and building operation parameters at unoccupied hours. Therefore, we mainly rely on engineering experience and communication with facility managers (for information collection) to choose calibration parameters. This process is case specific, depending on the judgment of engineers and may involve subjectivity. Moreover, more case studies should be conducted on different types of buildings to further verify the effectiveness of the proposed framework. Although we utilized the proposed framework that enables Bayesian calibration to be applicable in high-resolution calibration with reduced computational burden, there is still potential to further reduce the computational burden of Bayesian calibration (e.g., more efficient calibration methods and more suitable surrogate models). Additionally, we encourage researchers to develop more flexible and customized evaluation indicators to meet the growing demand of high-resolution building model calibration.

CRediT authorship contribution statement

Gang Jiang: Writing – original draft, Visualization, Software, Methodology, Investigation, Conceptualization. **Yixing Chen:** Writing –

review & editing, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Zhe Wang:** Writing – review & editing, Investigation, Conceptualization. **Kody Powell:** Writing – review & editing, Formal analysis, Conceptualization. **Blake Billings:** Writing – review & editing, Investigation, Formal analysis, Conceptualization. **Jianli Chen:** Writing – review & editing, Writing – original draft, Supervision, Resources, Methodology, Funding acquisition, Conceptualization.

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.

Data availability

The authors do not have permission to share data.

Acknowledgement

We would like to acknowledge the funding provided by the University of Utah's SEED2SOIL program, with invaluable support from the Global Change and Sustainability Center (GCSC). Additionally, the research is funded by the US National Science Foundation (NSF). Award title: Elements: A Convergent Physics-based and Data-driven Computing Platform for Building Modeling (# 2311685).

References

[1] P. Cdb, Global Status Report for Buildings and Construction 2022 (2022).
[2] C. Deb, A. Schlueter, Review of data-driven energy modelling techniques for building retrofit, Renew. Sustain. Energy Rev. 144 (2021) 110990, https://doi.org/10.1016/j.rser.2021.110990.
[3] B. Kuang, Y. Shi, Y. Hu, Z. Zeng, J. Chen, Household energy resilience in extreme weather events: an investigation of energy service importance, HVAC usage behaviors, and willingness to pay, Appl. Energy 363 (2024) 123051, https://doi.org/10.1016/j.apenergy.2024.123051.
[4] J. Chen, L. Zhang, Y. Li, Y. Shi, X. Gao, Y. Hu, A review of computing-based automated fault detection and diagnosis of heating, ventilation and air conditioning systems, Renew. Sustain. Energy Rev. 161 (2022) 112395, https://doi.org/10.1016/j.rser.2022.112395.
[5] J. Wang, J. Chen, Y. Hu, A science mapping approach based review of model predictive control for smart building operation management, J. Civ. Eng. Manag. 28 (2022) 661–679, https://doi.org/10.3846/jcem.2022.17566.
[6] Y. Gu, W. Tian, C. Song, A. Chong, Quantifying the effects of different data streams on the calibration of building energy simulation, Energ. Buildings 296 (2023) 113352, https://doi.org/10.1016/j.enbuild.2023.113352.
[7] M.H. Kristensen, R. Choudhary, S. Petersen, Bayesian calibration of building energy models: comparison of predictive accuracy using metered utility data of different temporal resolution, Energy Procedia 122 (2017) 277–282, https://doi.org/10.1016/j.egypro.2017.07.322.
[8] T. Hong, J. Langevin, K. Sun, Building simulation: ten challenges, Build. Simul. 11 (2018) 871–898, https://doi.org/10.1007/s12273-018-0444-x.
[9] D. Coakley, P. Raftery, M. Keane, A review of methods to match building energy simulation models to measured data, Renew. Sustain. Energy Rev. 37 (2014) 123–141, https://doi.org/10.1016/j.rser.2014.05.007.
[10] J.S. Haberl, T.E. Bou-Saada, Procedures for calibrating hourly simulation models to measured building energy and environmental data, J. Sol. Energy Eng. 120 (1998) 193–204, https://doi.org/10.1115/1.2888069.
[11] B.D. Hunn, J.A. Banks, S.N. Reddy, Energy analysis of the texas capitol Restoran on 1992.
[12] S.N. Reddy, B.D. Hunn, D.B. Hood, Determination of retrofit savings using a calibrated building energy simulation model 1994.
[13] J. Yoon, E.J. Lee, D.E. Claridge, Calibration procedure for energy performance simulation of a commercial building, J. Sol. Energy Eng. 125 (2003) 251–257, https://doi.org/10.1115/1.1564076.
[14] G. Liu, M. Liu, A rapid calibration procedure and case study for simplified simulation models of commonly used HVAC systems, Build. Environ. 46 (2011) 409–420, https://doi.org/10.1016/j.buildenv.2010.08.002.
[15] J. Haberl, D. Claridge, C. Culp, ASHRAE's Guideline 14-2002 for Measurement of Energy and Demand Savings: How to Determine What Was Really Saved by the Retrofit 2005.
[16] H. Akbari, S.J. Konopacki, Application of an end-use disaggregation algorithm for obtaining building energy-use data, J. Sol. Energy Eng. 120 (1998) 205–210, https://doi.org/10.1115/1.2888070.
[17] P. Raftery, M. Keane, J. O'Donnell, Calibrating whole building energy models: an evidence-based methodology, Energ. Buildings 43 (2011) 2356–2364, https://doi.org/10.1016/j.enbuild.2011.05.020.
[18] Y. Pan, Z. Huang, G. Wu, Calibrated building energy simulation and its application in a high-rise commercial building in Shanghai, Energ. Buildings 39 (2007) 651–657, https://doi.org/10.1016/j.enbuild.2006.09.013.
[19] A. Chong, Y. Gu, H. Jia, Calibrating building energy simulation models: a review of the basics to guide future work, Energ. Buildings 253 (2021) 111533, https://doi.org/10.1016/j.enbuild.2021.111533.
[20] Y.-S. Kim, M. Heidarinejad, M. Dahlhausen, J. Srebric, Building energy model calibration with schedules derived from electricity use data, Appl. Energy 190 (2017) 997–1007, https://doi.org/10.1016/j.apenergy.2016.12.167.
[21] A. Chong, G. Augenbroe, D. Yan, Occupancy data at different spatial resolutions: Building energy performance and model calibration, Appl. Energy 286 (2021) 116492, https://doi.org/10.1016/j.apenergy.2021.116492.
[22] O. Vera-Piazzini, M. Scarpa, Building energy model calibration: a review of the state of the art in approaches, methods, and tools, J. Build. Eng. (2023:) 108287, https://doi.org/10.1016/j.jobe.2023.108287.
[23] H. Guy, S. Vittoz, G. Caputo, T. Thiery, Benchmarking the energy performance of European commercial buildings with a bayesian modeling framework, Energ. Buildings 299 (2023) 113595, https://doi.org/10.1016/j.enbuild.2023.113595.
[24] F.C. Melo, G. Carrilho da Graça, M.J.N. Oliveira Panão, A review of annual, monthly, and hourly electricity use in buildings, Energ. Buildings 293 (2023) 113201, https://doi.org/10.1016/j.enbuild.2023.113201.
[25] H. Zhang, W. Tian, J. Tan, J. Yin, X. Fu, Sensitivity analysis of multiple time-scale building energy using Bayesian adaptive spline surfaces, Appl. Energy 363 (2024) 123042, https://doi.org/10.1016/j.apenergy.2024.123042.
[26] M.C. Kennedy, A. O'Hagan, Bayesian calibration of computer models, J. R. Stat. Soc. Ser. B Stat Methodol. 63 (2001) 425–464, https://doi.org/10.1111/1467-9868.00294.
[27] Y. Ling, J. Mullins, S. Mahadevan, Selection of model discrepancy priors in Bayesian calibration, J. Comput. Phys. 276 (2014) 665–680, https://doi.org/10.1016/j.jcp.2014.08.005.
[28] P. Robbe, D. Andersson, L. Bonnet, T.A. Casey, M.W.D. Cooper, C. Matthews, et al., Bayesian calibration with summary statistics for the prediction of xenon diffusion in UO2 nuclear fuel, Comput. Mater. Sci 225 (2023) 112184, https://doi.org/10.1016/j.commatsci.2023.112184.
[29] Z. Yan, Y. Hu, H. Shi, P. Wang, Z. Liu, Y. Tian, et al., Experimentally characterizing the spatially varying anisotropic mechanical property of cancellous bone via a Bayesian calibration method, J. Mech. Behav. Biomed. Mater. 138 (2023) 105643, https://doi.org/10.1016/j.jmbbm.2022.105643.
[30] B.W. Billings, P.J. Smith, S.T. Smith, K.M. Powell, Industrial battery operation and utilization in the presence of electrical load uncertainty using Bayesian decision theory, J. Energy Storage 53 (2022) 105054, https://doi.org/10.1016/j.est.2022.105054.
[31] M. Viswanathan, A. Scheidegger, T. Streck, S. Gayler, T.K.D. Weber, Bayesian multi-level calibration of a process-based maize phenology model, Ecol. Model. 474 (2022) 110154, https://doi.org/10.1016/j.ecolmodel.2022.110154.
[32] Y. Heo, R. Choudhary, G.A. Augenbroe, Calibration of building energy models for retrofit analysis under uncertainty, Energ. Buildings 47 (2012) 550–560, https://doi.org/10.1016/j.enbuild.2011.12.029.
[33] J. Yuan, V. Nian, B. Su, Q. Meng, A simultaneous calibration and parameter ranking method for building energy models, Appl. Energy 206 (2017) 657–666, https://doi.org/10.1016/j.apenergy.2017.08.220.
[34] S. Nagpal, C. Mueller, A. Aijazi, C.F. Reinhart, A methodology for auto-calibrating urban building energy models using surrogate modeling techniques, J. Build. Perform. Simul. (2019).
[35] W. Tian, S. Yang, Z. Li, S. Wei, W. Pan, Y. Liu, Identifying informative energy data in Bayesian calibration of building energy models, Energ. Buildings 119 (2016) 363–376, https://doi.org/10.1016/j.enbuild.2016.03.042.
[36] W. Tian, A review of sensitivity analysis methods in building energy analysis, Renew. Sustain. Energy Rev. 20 (2013) 411–419, https://doi.org/10.1016/j.rser.2012.12.014.
[37] K. Menberg, Y. Heo, R. Choudhary, Sensitivity analysis methods for building energy models: comparing computational costs and extractable information, Energ. Buildings 133 (2016) 433–445, https://doi.org/10.1016/j.enbuild.2016.10.005.
[38] A. Chong, K. Menberg, Guidelines for the Bayesian calibration of building energy models, Energ. Buildings 174 (2018) 527–547, https://doi.org/10.1016/j.enbuild.2018.06.028.
[39] B. Eisenhower, Z. O'Neill, S. Narayanan, V.A. Fonoberov, I. Mezić, A methodology for meta-model based optimization in building energy models, Energ. Buildings 47 (2012) 292–301, https://doi.org/10.1016/j.enbuild.2011.12.001.
[40] H. Lim, Z.J. Zhai, Comprehensive evaluation of the influence of meta-models on Bayesian calibration, Energ. Buildings 155 (2017) 66–75, https://doi.org/10.1016/j.enbuild.2017.09.009.
[41] Q. Li, G. Augenbroe, J. Brown, Assessment of linear emulators in lightweight Bayesian calibration of dynamic building energy models for parameter estimation and performance prediction, Energ. Buildings 124 (2016) 194–202, https://doi.org/10.1016/j.enbuild.2016.04.025.
[42] ASHRAE Guideline 14-2002 - Measurement of Energy and Demand Savings n.d. https://webstore.ansi.org/standards/ashrae/ashraeguideline142002 (accessed April 28, 2024).
[43] A. Chong, K.P. Lam, M. Pozzi, J. Yang, Bayesian calibration of building energy models with large datasets, Energ. Buildings 154 (2017) 343–355, https://doi.org/10.1016/j.enbuild.2017.08.069.

[44] A. Chong, K. Lam, A comparison of MCMC algorithms for the Bayesian calibration of building energy models. 2017. 10.26868/25222708.2017.336.
[45] G. Li, J. Xiong, R. Tang, S. Sun, C. Wang, In-situ sensor calibration for building HVAC systems with limited information using general regression improved Bayesian inference, Build. Environ. 234 (2023) 110161, https://doi.org/10.1016/j.buildenv.2023.110161.
[46] S. Rouchier, M. Rabouille, P. Oberlé, Calibration of simplified building energy models for parameter estimation and forecasting: stochastic versus deterministic modelling, Build. Environ. 134 (2018) 181–190, https://doi.org/10.1016/j.buildenv.2018.02.043.
[47] V. Marty-Jourjon, A. Goyal, T. Berthou, P. Stabat, Identifiability study of an RC building model based on the standard ISO13790, Energ. Buildings 276 (2022) 112446, https://doi.org/10.1016/j.enbuild.2022.112446.
[48] R.E. Hedegaard, M.H. Kristensen, T.H. Pedersen, A. Brun, S. Petersen, Bottom-up modelling methodology for urban-scale analysis of residential space heating demand response, Appl. Energy 242 (2019) 181–204, https://doi.org/10.1016/j.apenergy.2019.03.063.
[49] C. Zhu, W. Tian, B. Yin, Z. Li, J. Shi, Uncertainty calibration of building energy models by combining approximate Bayesian computation and machine learning algorithms, Appl. Energy 268 (2020) 115025, https://doi.org/10.1016/j.apenergy.2020.115025.
[50] X. Faure, R. Lebrun, O. Pasichnyi, Impact of time resolution on estimation of energy savings using a copula-based calibration in UBEM, Energ. Buildings 311 (2024) 114134, https://doi.org/10.1016/j.enbuild.2024.114134.
[51] S. Zhan, G. Wichern, C. Laughman, A. Chong, A. Chakrabarty, Calibrating building simulation models using multi-source datasets and meta-learned Bayesian optimization, Energ. Buildings 270 (2022) 112278, https://doi.org/10.1016/j.enbuild.2022.112278.
[52] V. Martinez-Viol, E.M. Urbano, M. Delgado-Prieto, L. Romeral, Automatic model calibration for coupled HVAC and building dynamics using Modelica and Bayesian optimization, Build. Environ. 226 (2022) 109693, https://doi.org/10.1016/j.buildenv.2022.109693.
[53] P. Wang, C. Li, R. Liang, S. Yoon, S. Mu, Y. Liu, Fault detection and calibration for building energy system using Bayesian inference and sparse autoencoder: a case study in photovoltaic thermal heat pump system, Energ. Buildings 290 (2023) 113051, https://doi.org/10.1016/j.enbuild.2023.113051.
[54] A. Bampoulas, F. Pallonetto, E. Mangina, D.P. Finn, A Bayesian deep-learning framework for assessing the energy flexibility of residential buildings with multicomponent energy systems, Appl. Energy 348 (2023) 121576, https://doi.org/10.1016/j.apenergy.2023.121576.
[55] A novel method of creating machine learning-based time series meta-models for building energy analysis. Energy Build 2023; 281:112752. 10.1016/j.enbuild.2022.112752.
[56] H. Yoshino, T. Hong, N. Nord, IEA EBC annex 53: Total energy use in buildings—analysis and evaluation methods, Energ. Buildings 152 (2017) 124–136, https://doi.org/10.1016/j.enbuild.2017.07.038.
[57] N. Li, Z. Yang, B. Becerik-Gerber, C. Tang, N. Chen, Why is the reliability of building simulation limited as a tool for evaluating energy conservation measures? Appl. Energy 159 (2015) 196–205, https://doi.org/10.1016/j.apenergy.2015.09.001.
[58] C. Clevenger, J. Haymaker, The impact of the building occupant on energy modeling simulations, J. Int. Conf. Comput. Decis Mak. Civ. Build. Eng. (2006).
[59] H. Sakoe, S. Chiba, Dynamic programming algorithm optimization for spoken word recognition, IEEE Trans. Acoust. Speech Signal Process. 26 (1978) 43–49, https://doi.org/10.1109/TASSP.1978.1163055.
[60] E. Keogh, C.A. Ratanamahatana, Exact indexing of dynamic time warping, Knowl. Inf. Syst. 7 (2005) 358–386, https://doi.org/10.1007/s10115-004-0154-9.
[61] J. Banfield, W. Esty, The box-percentile plot, J. Stat. Softw. 08 (2003), https://doi.org/10.18637/jss.v008.i17.
[62] M.D. Morris, Factorial sampling plans for preliminary computational experiments, Technometrics 33 (1991) 161–174, https://doi.org/10.1080/00401706.1991.10484804.
[63] M.D. McKay, R.J. Beckman, W.J. Conover, A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics 21 (1979) 239–245, https://doi.org/10.2307/1268522.
[64] A. Sherstinsky, Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network, Phys Nonlinear Phenom 404 (2020) 132306, https://doi.org/10.1016/j.physd.2019.132306.
[65] S. Im, J. Lee, M. Cho, Surrogate modeling of elasto-plastic problems via long short-term memory neural networks and proper orthogonal decomposition, Comput. Methods Appl. Mech. Eng. 385 (2021) 114030, https://doi.org/10.1016/j.cma.2021.114030.
[66] Z. Zhang, Z. Lv, C. Gan, Q. Zhu, Human action recognition using convolutional LSTM and fully-connected LSTM with different attentions, Neurocomputing 410 (2020) 304–316, https://doi.org/10.1016/j.neucom.2020.06.032.
[67] A.F. Agarap, Deep Learning using Rectified Linear Units (ReLU) 2019.
[68] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, Equation of state calculations by fast computing machines, J. Chem. Phys. 21 (2004) 1087–1092, https://doi.org/10.1063/1.1699114.
[69] M. Wilburn, ANSI/ASHRAE/IES Standard 90.1-2016 Performance Rating Method Reference Manual n.d.
[70] White Box Technologies Weather Data n.d. http://weather.whiteboxtechnologies.com/ (accessed August 11, 2023).
[71] A.D. Dilsiz, K.E. Nweye, A.J. Wu, J.H. Kämpf, F. Biljecki, Z. Nagy, How spatio-temporal resolution impacts urban energy calibration, Energ. Buildings 292 (2023) 113175, https://doi.org/10.1016/j.enbuild.2023.113175.