Energy & Buildings 323 (2024) 114853

Contents lists available at ScienceDirect

Energy & Buildings

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

Physics-informed ensemble learning with residual modeling for enhanced building energy prediction

Zhihao Ma, Gang Jiang, Jianli Chen *

Department of Civil and Environmental Engineering, The University of Utah, Salt Lake City, UT 84102, USA

A R T I C L E I N F O

Keywords:
Residual modeling
Physics-informed machine learning
Building energy modeling
Time series analysis
Recurrent neural network

A B S T R A C T

Accurate modeling of building energy use is important to support a diverse spectrum of its downstream applications, such as building energy efficiency assessment, resilience analysis, smart control, *etc*. As mainstream approaches of building energy modeling, physics-based modeling builds on different fidelities of physics rules yet are usually compromised in modeling accuracy due to insufficiency of physics rules to capture real-world dynamics and incomplete input information. Data-driven approaches are computationally efficient, but black box (uninterpretable) in nature. For improved modeling of building energy use, this work proposes a physics-informed ensemble learning approach in building energy prediction through residual modeling. Specifically, we first analyze the components of building energy use data. Evidence suggests that the building energy use data can be decomposed into physics-driven part, occupant-driven part, and white noise. Second, high-fidelity physics-based building models (EnergyPlus) and low-fidelity ones (RC models) are developed to capture the physics-driven part while time series methods are explored as the residual modeling approach to capture the occupant-driven discrepancies between physics-based simulation and measured building energy use (*i.e.*, residuals). Finally, the physics-informed ensemble learning is proposed to integrate physics-based and data-driven models for enhanced accuracy and robustness of building energy modeling. Results demonstrate 40–90% increase of accuracy between modeling and field observations compared to traditional physics-based modeling methods. Moreover, when the training dataset size is small, the proposed ensemble model overperforms the pure data-driven models, demonstrating its higher robustness in extrapolation scenarios. This work makes fundamental contributions to the development of convergent modeling approaches in the building modeling field.

1. Introduction

As global warming intensifies, reducing energy consumption and addressing climate-related challenges become increasingly critical. Buildings account for approximately 36 % of the global energy use and almost 40 % of the greenhouse gas emissions [1]. Therefore, managing building energy use plays a significant role in counteracting climate change and moving towards sustainability. To realize this, it is crucial to develop accurate building energy models for optimal building performance [2–4]. As depicted in ASHRAE Handbook [5], building energy models are predominantly utilized for three purposes: comparison, compliance, and prediction. Specifically, Building Energy Modeling (BEM) has been leveraged for a variety of applications, including evaluating alternative designs [6], allocating annual energy budgets [7], automating demand response [8,9], predicting energy costs [10], and detecting energy use anomalies [11,12]. Within these applications, BEM methodologies can be generally classified as physics-based and data-driven modeling.

Physics-based models are based on input of detailed building parameters and solving the governing equations of mass, momentum, and energy to adhere to physics rules. Commonly used models include EnergyPlus [13,14], Modelica [15,16], and TRNSYS [17,18]. The benefit of physics-based modeling lies in its strong interpretability, as these models elucidate the relationship between inputs and outputs through widely recognized physical rules. In contrast, data-driven models are built on mathematical models and sufficient clean data to uncover hidden relationships between inputs and outputs, thereby reducing or even eliminating the need for detailed physical information of buildings in modeling. Machine learning algorithms widely used in data-driven modeling of buildings include Linear Regression (LR) [19,20], Support Vector Machine (SVM) [21,22], Random Forest (RF) [23,24], Recurrent Neural Network (RNN) [25–27]. However, these

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

https://doi.org/10.1016/j.enbuild.2024.114853
Received 1 July 2024; Received in revised form 9 September 2024; Accepted 26 September 2024
Available online 27 September 2024

models also have notable drawbacks. A primary concern is their dependence on high-quality training data: the presence of missing, incorrect, or biased data can significantly compromise model performance [28].

Physics-based models are highly reliant on detailed input parameters, whereas data driven models lack interpretability and are sensitive to the data quality. To make two approaches optimally complement each other, the concept of physics-informed data-driven models, or Physics-informed Machine Learning (PIML) models, has emerged as a new building modeling approach. PIML integrates physics-based and data-driven models by encoding physics knowledge into traditional data-driven models, thereby these models can be trained with the historical data while adhering to physical principles [29]. Therefore, PIML can achieve improved prediction accuracy and interpretability simultaneously with physically consistent results.

Integrating physics information into architectural design is a popular approach for PIML development in the building energy modeling field, aiming to facilitate more robust training process and enhanced interpretability of developed models. The modified structures are usually based on Artificial Neural Network (ANN), RNN, Graph Neural Network (GNN), and linear models [30–32]. For example, Wang *et al.* [33,34] proposed a partially connected neural network to predict indoor temperature thermal dynamics through directly modifying neurons connections. In this design, neurons within each layer are organized into different blocks, hence, only current and historical information are considered in predictions of future time steps. The implementation of this control-oriented model has realized substantial energy savings of over 35 % while maintaining thermal comfort and air quality standards. Xiao *et al.* [35,36] proposed a novel physics-informed RNN structure, enhanced from the traditional Long Short-term Memory (LSTM) with additional RNN cells, to enforce physically consistent dynamics in modeling. The physical information is incorporated by constraining the sign of the partial derivatives of indoor temperature T_k and relative humidity RH_k with respect to various inputs. For example, heating power should logically increase indoor temperature ($\partial T_k / \partial U_{heat,k-i} > 0$), while cooling power should decrease it ($\partial T_k / \partial U_{cool,k-i} < 0$). Other than neural network-based models, physical knowledge can be integrated into linear models as well, whose linear terms can be assigned with particular physical information. Mirfin *et al.* [31] added a solar gain term into the traditional linear regression Time-Of-Week Temperature (TOWT) model and proposed a Time-Of-Week, Solar, and Temperature (TOWST) model for building energy consumption modeling, considering the window area and building orientation. This TOWST model demonstrates 30 %-72 % reduction in Mean Square Error (MSE) compared to the TOWT model. Physics-informed ensemble models for joint prediction is another promising approach of PIML. The rationale behind this method is that the physics-based models are capable of modeling the physical component of a system, whereas the data-driven models excel in uncovering other hidden dynamics [37]. Dong *et al.* [38] combined five data-driven models (*e.g.* ANN and SVM) with the 2R1C model to predict energy consumption in residential buildings, where the 2R1C model was used for Heating, Ventilation, and Air Conditioning (HVAC) energy prediction, while the data-driven models addressed non-HVAC energy (plug load, lighting, *etc.*). The results indicated that this model outperformed data-driven models, showing improvements in the coefficient of variance by 6–10 % and 2–15 % for hour ahead and day ahead forecasting, respectively. These studies illustrate the considerable potential of PIML models to leverage the strengths of both physics-based and data-driven models, thereby enhancing accuracy and interpretability of building energy modeling.

Properly designed PIML models can improve interpretability, accuracy, adaptability, and computational efficiency of BEM simultaneously, hence, promising to further promote applications of BEM in practice. However, there remain research gaps to be bridged, including:

(1) **Understanding discrepancies between modeling and observations with decomposition of building energy use data.** Although physics-based models are widely used to predict building energy consumption, there always exist discrepancies between simulated and observed energy data. These discrepancies sometimes are more than 100 % in high resolution modeling (*e.g.*, hourly or sub-hourly), even after the models are detailly calibrated [39]. It is still unclear what factors explicitly contribute to these discrepancies. Former arguments suggest that the stochastic nature of occupants could be the major factor since activities patterns are difficult to be easily captured by physical governing equations in physics-based modeling [40]. However, there is no evidence or analysis to confirm and demonstrate this.

(2) **Model robustness in extrapolation scenarios**. Sufficient training data for BEM is curial but also challenging due to the labor-intensive data collection process and privacy concerns [41]. Obtaining data from buildings under or newly built buildings is particularly difficult [42]. However, a sufficient training dataset is essential for training data-driven models, as inadequate data prevents the model from learning effectively. Using data-driven models that are not well-trained would ultimately result in prediction failures, which is a significant concern in BEM [43]. Integrating physics information into machine learning models is a potential solution, but rarely explored in existing literature.

(3) **Computational efficiency vs. accuracy.** In scenarios where iterative computation of building energy models are essential, such as in smart building operation [44], the computational efficiency of models becomes crucial. Thus, simplified building energy models are needed due to their high computational efficiency. However, such simplification compromises accuracy, *e.g.*, RC models are lower fidelity models [45] with inferior accuracy compared to high fidelity models (*e.g.*, EnergyPlus) in capturing building dynamics. The compromised level of accuracy could increase as the time resolution of modeling increases, *i.e.*, simplified physics-based models are less accurate in sub-hourly forecasting since these models usually adopt linear interpolation (or other formats of simplification) in high resolution modeling [46,47]. How to effectively address and balance the trade-off between computational efficiency and model complexity is a critical challenge in building modeling.

Motivated by these research gaps, this work aims to understand and demonstrate the decomposition of the building energy use data, then correspondingly develops an accurate and robust physics-informed ensemble model to align modeling with observed energy use across multiple time resolution through residual modeling. This work is organized as follows. Section 2 outlines the employed methodology, including the target buildings to model and energy use dataset, decomposition of building energy use, development of physics-based and data-driven models for time series residual modeling, and model ensemble for load prediction. Section 3 will present results of this work, focusing on the results of energy use decomposition and the outcomes of the proposed ensemble modeling. Then, we will discuss the insights from energy use data decomposition, the trade-off between model complexity and accuracy, together with the enhancement of BEM practical application (Section 4). Finally, Section 5 will conclude this study.

2. Methodology

In this section, the principal methodology of the study is presented, including data collection, building energy use decomposition, physics-based building energy modeling, time series analysis for residual modeling, and building energy prediction with the proposed ensemble model. The workflow of this work is shown in Fig. 1. First, the observed cooling and heating energy data of target buildings are collected (Section 2.1). Then, the decomposition of energy use data is

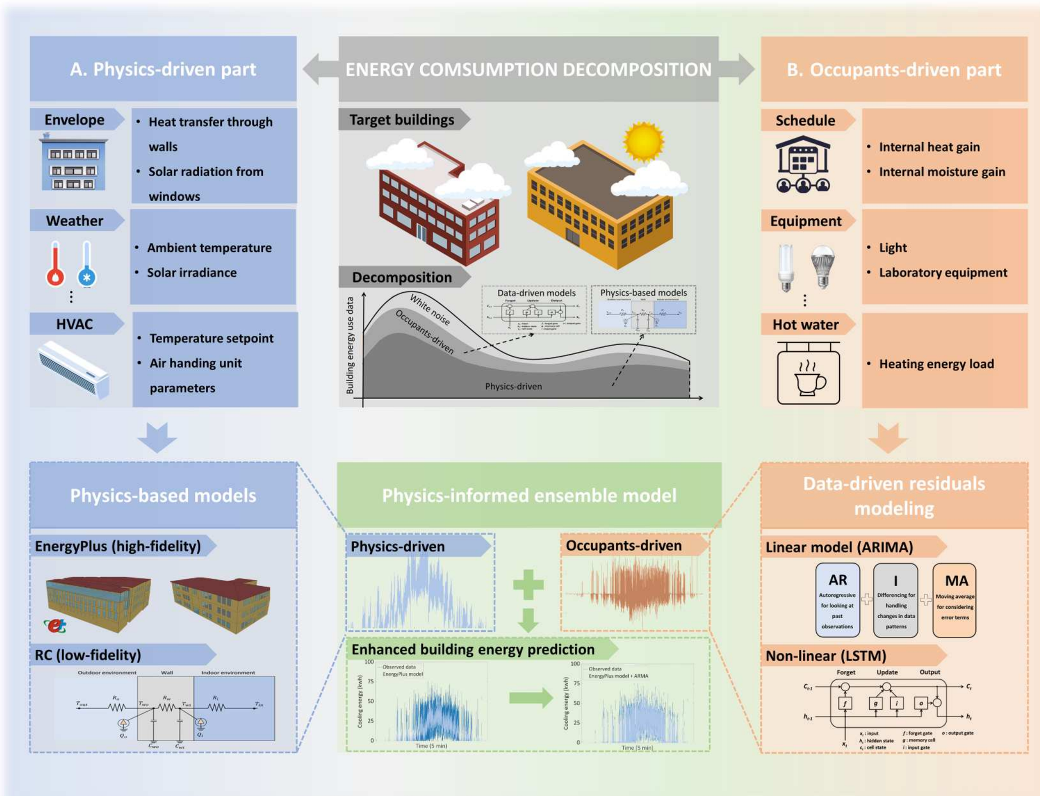

Fig. 1. Overall workflow for the proposed physics-informed ensemble model.

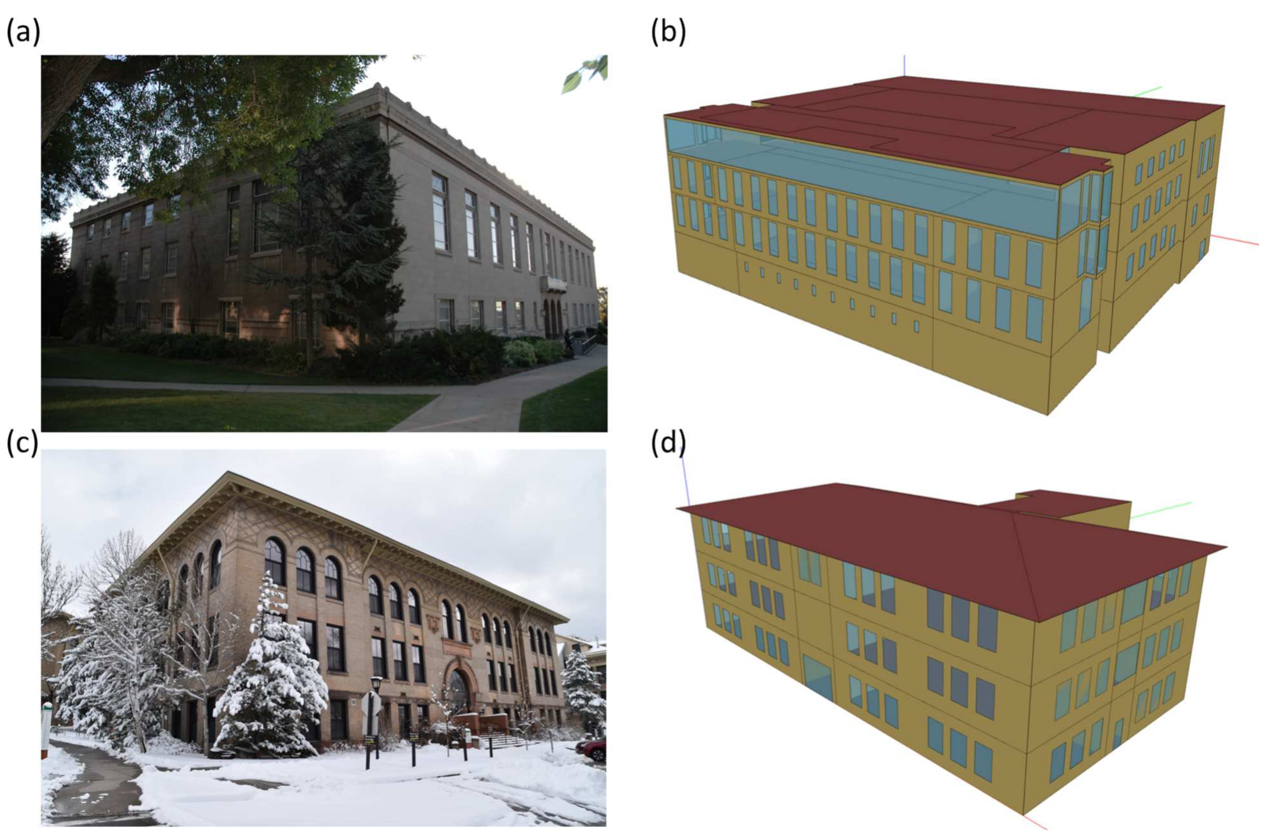

Fig. 2. Target buildings. (a) Field picture and (b) geometry model of CSC. (c) Field picture and (d) geometry model of AEB.

demonstrated in Section 2.2. Generally, building energy use data is expected to be decomposed into three parts: physics-driven part, occupant-driven part, and white noise. The following sections will address each of these aspects separately. First, the different fidelities of physics-based models are developed for capturing the physics-driven part of energy use (Section 2.2.1). Second, data-driven timeseries models are used for modeling residuals, *i.e.*, the discrepancies between physics-based modeling outcomes and field observations (Section 2.2.2). Subsequently, the correlation and white noise analysis presented in Sections 2.2.3-2.2.4 provides evidence and understanding of the building energy use data decomposition. Finally, the ensemble learning framework is proposed to enhance modeling performance by integrating physics-based modeling and the data-driven residual modeling for predicting building energy use in different time resolutions (Section 2.3).

2.1. Target buildings and observed dataset

The target buildings in the case study include the Crocker Science Center (CSC) and Alfred Emery Building (AEB), both of which are campus buildings at the University of Utah (Fig. 2). The four-story CSC building spans a total area of roughly 11,437 square meters and underwent retrofitting in 2016. It encompasses various spaces, including offices, conference rooms, mechanical rooms, an auditorium, and laboratories. The AEB is a three-story building with a total floor area of 4,101 square meters, including classrooms and offices. These buildings are served by a Variable Air Volume (VAV) system for air conditioning, comprising several Air Handling Units (AHUs). The cooling energy for both buildings and heating energy for the AEB are supplied from a central plant in campus, while the heating energy for the CSC is generated by four standalone boilers inside the building. The observed cooling and heating energy data are recorded in the SkySpark system (January to December 2021 for CSC and January to December 2022 for AEB). Specifically, cooling energy data involves space cooling, all-time cooling of the mechanical room, and the chemistry laboratory (*e.g.*, continuous cooling to support protein production in cell culture). Heating energy data includes energy use for domestic hot water and space heating. This dataset encompasses multiple sub-hourly time resolutions, including 5-minute, 15-minute, 30-minute, and 60-minute intervals. There are approximately two months of missing data of CSC (mainly in July and August 2021). Other than this period, the missing values in other few data gaps are estimated based on data on similar days nearby. In addition, the weather data files of 2021 and 2022 are provided by White Box Technologies [48].

2.2. Building energy data decomposition and modeling

This section presents the methods of building energy data decomposition analysis for CSC and AEB. Building energy consumption is highly influenced by various factors, including but not limited to weather conditions, building envelope status, and HVAC and other system parameters. These could be well-described by physical laws such as heat transfer and thermal dynamics. As the internal heat source, occupants also play a pivotal role in the building energy usage, yet this part is challenging to be captured by physics rules due to its stochasticity and heterogeneity. As occupant activities are typically time-related with strong periodicity but non physics relevance, time series-based data-driven models are suitable for capturing this. Therefore, it is hypothesized that the building energy use could be composed of three components. *i.e.*, physics-driven component, occupant-driven component, and the white noise component (*e.g.*, due to data or sensing errors) as remaining. Based on the reasoning to understand this decomposition and further enhance building modeling performance, this section is organized as follows. First, different fidelities of physics-based building models are presented, including high fidelity EnergyPlus model and low fidelity RC model (Section 2.2.1). Then, the residuals between physics-based modeling and observed data are analyzed and modeled using time series-based data-driven models (Section 2.2.2). Finally, the correlation coefficients and white noise are introduced to understand building energy use data with decomposition (Sections 2.2.3 and 2.2.4).

2.2.1. Physics-based models

Physics-based models are first developed to capture the physics-driven component in building energy use in various levels, including high-fidelity EnergyPlus models and low-fidelity RC models. The target variable being simulated is the cooling/heating load of the buildings. Since the heating and cooling for the target buildings are provided by the campus central mechanical system, rather than by standalone systems within the buildings, it is only possible to estimate the heating/cooling load based on the hot/chilled water flow rate and temperature change. In a certain sense, system efficiency metrics, *e.g.*, coefficient of performance, are not relevant and we are only focusing on modeling the physics of buildings in this study.

High-fidelity physics-based building energy models of this research are established based on EnergyPlus (version 23.1) according to the construction plans of target buildings to model. To ensure that the high-fidelity physics-based models accurately reflect the true building physics, we engaged in extensive communication with the campus facility managers to gain a comprehensive understanding of the operational conditions. This allowed us to further calibrate the models in detail using Bayesian calibration based on all collected data at a high resolution [49] (See Supplementary Information S1.1). With calibrated modeling input parameters, the HVAC system load can be calculated by solving the ordinary differential equations pertinent to the heat balance of the zone air:

$$C_Z \frac{dT_Z}{dt} = \sum_{i=1}^{N_i} q_{si} + \sum_{i=1}^{S_i} h_{ci} A_i (T_{si} - T_z) + q_{inf} + q_{ven} + q_{HVAC}, \tag{1}$$

where C_Z is the heat capacitance of zone air [J/K]; T_z is the temperature of zone air [K]; q_{si} is the internal loads from occupants, appliances, and lights, *etc.* [W]; h_{ci} is the convective heat transfer coefficient between surface i and indoor air [W/(m^2 K)]; A_i is the surface area [m^2]; T_{si} is the surface temperature [K]; q_{inf} and q_{ven} are the heat flux from the outside air infiltration and ventilation, respectively [W]; q_{HVAC} is the air-conditioning loads [W]. In addition to the high-fidelity EnergyPlus models, low-fidelity RC models are also developed to describe building physics of energy use (Fig. 3). As representative low-fidelity models, RC models are illustrated as xRyC networks with lumped parameters in analogy to an electrical circuit with x as the number of thermal resistances and y of thermal capacities. The differential equations of the building model are formulated according to the chosen model structure. For this research, we developed two types of typical RC models with a single external wall, *i.e.*, the 3R2C model (three resistances and two capacities) and the 2R1C model (two resistances and one capacities). Then, the cooling and heating energy use are calculated by solving partial derivative equations of building thermal dynamics (See Supplementary Information S1.2).

2.2.2. Data-driven residual modeling

After development of physics-based models, data-driven approaches are explored to model the residuals between physics-based energy models and observed data. Residuals (as the terminology) in this paper refer to the discrepancies between modeling outcomes of physics-based models and actual field observations. These residuals are crucial indicators of the inherent limitations of physics-based models. In this research, even the established high-fidelity physics-based models are detailly calibrated using the Bayesian method, there still exist discrepancies between physics-based modeling outcomes and field observations. In the past, it is suspected that these deviations are attributable to factors such as occupant behaviors, although no evidence was presented to affirm this. The persistence of residuals also underscores a fundamental aspect of BEM: while physics-based approaches can approximate

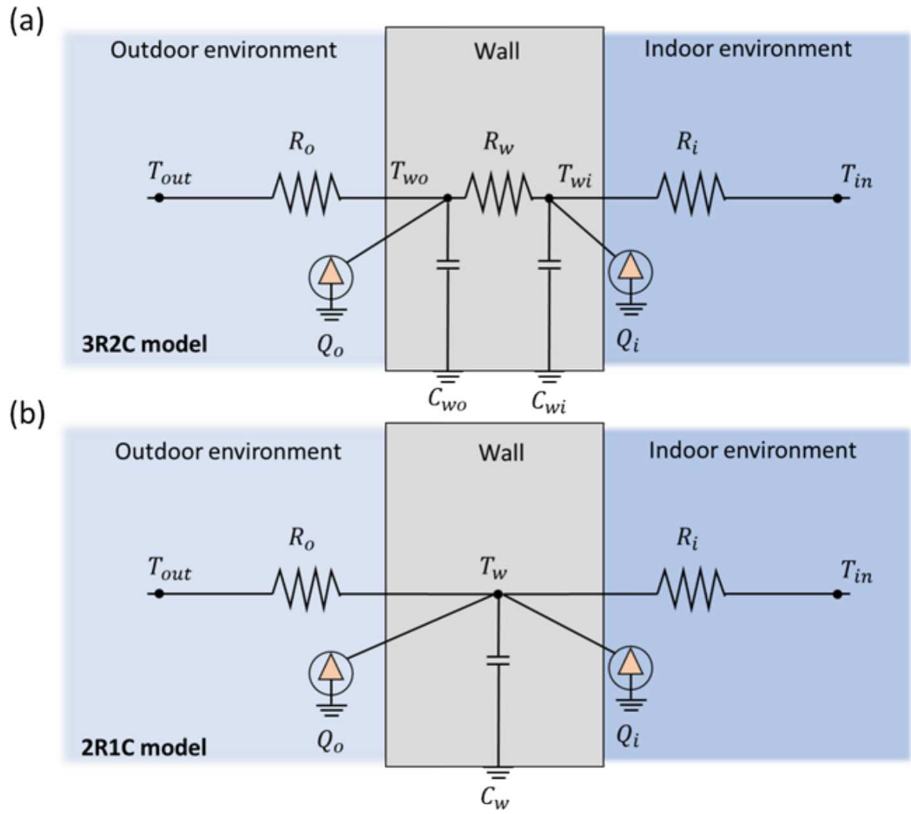

Fig. 3. Thermal network of (a) 3R2C and (b) 2R1C models of the target buildings. T_{in}: indoor temperature; T_{out}: outdoor temperature; T_{wo}: outside layer temperature; T_{wi}: inside layer temperature; R_o: thermal resistance between the wall and the outdoor environment; R_w: thermal resistance within the wall;R_i: thermal resistance between the wall and the indoor environment; Q_o: outside heat source; Q_i: inside heat source.

the building energy consumption, there are always hidden dynamics beyond the scope of designed physics rules to capture. Hence, approaches to complement physics-based modeling are expected to further enhance modeling accuracy of energy use of real buildings.

The inherent periodicity in building operation makes time series analysis effective in understanding and capturing temporal correlations of historical data. Hence, data driven models arise to be appropriate approaches in modeling these residuals [50]. Prominent time series analysis techniques used in building energy consumption forecasting include ANN [51], ARIMA [52], SVM [53], and RNN [54], *etc.* Therein, the loop-based structure of RNN is designed to describe temporal dynamics of modelled systems. Therefore, the utilization of RNNs in energy prediction has attracted increasing research interests in recent years [1]. In this section, both linear autoregressive model (AR, ARMA, and ARIMA) and non-linear LSTM model are explored to fit with the residuals between physics-based modeling and observed data.

The Autoregressive (AR) model is a fundamental and widely used technique for time series forecasting [52,55,56]. An extension of the AR model is the Autoregressive Moving Average (ARMA) model that incorporates terms of forecasted errors. However, both AR and ARMA models are less effective to model data with large fluctuations. For non-stationary data, the Autoregressive Integrated Moving Average (ARIMA) model is usually used, extending the ARMA model by integrating a differential form of the data:

$$y_t^* = \alpha_0 + \alpha_1 y_{t-1}^* + \cdots + \alpha_k y_{t-k}^* + \beta_1 \epsilon_{t-1} + \cdots + \beta_k \epsilon_{t-k} + \epsilon_t, \tag{2}$$

where $y_t^* = y_t - y_{t-1}$ is first-order differential. In model fitting, it is crucial to determine relevant lags for the stationary time series data. The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) are utilized to identify related lag terms by calculating covariance or conditional covariance. In addition, although adopting historical data with more lags is possible to improve model accuracy but unnecessarily increase model complexity (affecting model application in practice). Hence, Bayesian Information Criterion (BIC) is then used to determine the optimal number of lags considering both accuracy and complexity. Detailed order finding process for residual modeling are shown in Supplementary Information S1.3.

While linear models could be insufficient to fully capture the autocorrelation of a time series sequence, in this work, the deep learning networks are also developed to capture the nonlinear relationships in residuals. As a predominant variant of deep RNN network, LSTM is used in our residual modeling. To start with, the max–min data preprocessing is conducted, in which all training data and test data are normalized to increase the calculation and convergence speed. After that, the structure and detailed parameters of LSTM networks are shown, together with the method for order finding.

LSTM networks are extensively used as the deep learning algorithm, particularly for tasks involving sequential data in language modeling, speech recognition, and time series prediction. As shown in Fig. 4a, a typical LSTM network has value and state communication among units. The input of a LSTM unit includes not only feature space parameters x but also the cell state and the hidden state from previous unit. The cell state C acts as the “memory” of the network, carrying information across the sequence, while the hidden state h transfers information to subsequent LSTM cells and to the outputs. Each LSTM unit operates through three specialized gates: Forget Gate, Input Gate, and Output Gate. These gates collectively determine the updated cell state and hidden state, which then serve as inputs for the subsequent unit. The Forget Gate decides which information from the previous unit should be retained or discarded and is calculated as follows:

$$f_t = \sigma\left(w_f \cdot [h_{t-1}, x_t]\right), \tag{3}$$

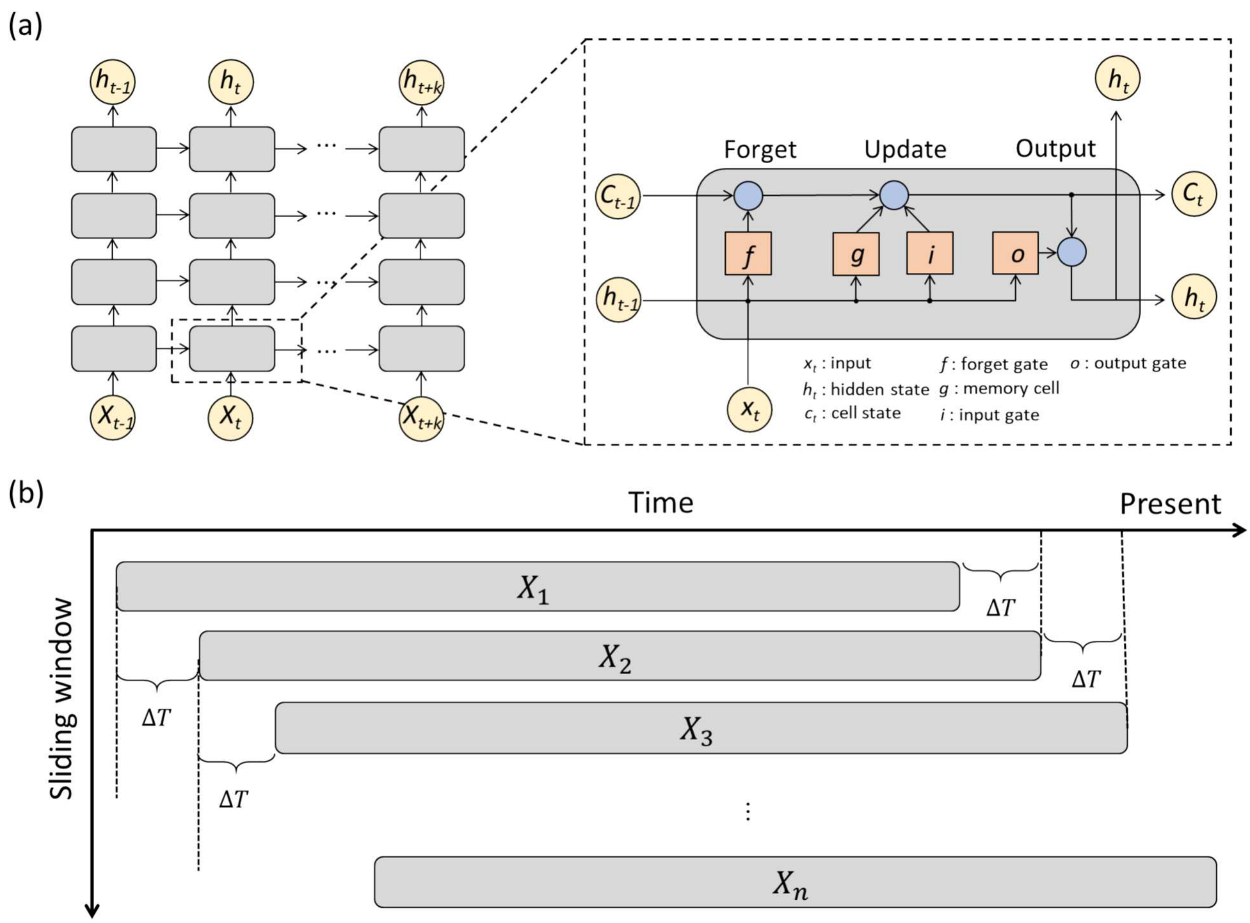

Fig. 4. (a) LSTM network and its internal logic gates. (b) Transformation to sequential data framed by sliding window.

where σ represents the sigmoid function, w_f is the augmented wights. $[h_{t-1}, x_t]$ is the matrix formed by splicing the hidden state at time point t-1 (h_{t-1}) and the input feature space at time t (x_t). The Input Gate selects new information from x_t to be stored in the cell state, calculated as:

$$i_t = \sigma(w_i \cdot [h_{t-1}, x_t]). \tag{4}$$

The cell state is then updated from C_{t-1} to C_t, where the old state is multiplied by f_t to selectively forget what is unnecessary. The new C_t is given by:

$$C_t = f_t C_{t-1} + i_t \tanh(w_C \cdot [h_{t-1}, x_t]). \tag{5}$$

Eventually, based on C_t and x_t, Output Gate yields the new hidden state h_t as:

$$o_t = \sigma(w_o \cdot [h_{t-1}, x_t]), h_t = o_t \cdot \tanh C_t. \tag{6}$$

In constructing a deep LSTM network, multiple LSTM units are stacked. The network structure, therefore, is a critical factor in the LSTM model. The RNN model in this work includes four LSTM layers, a fully connected layer, and a regression layer, employing the ReLU activation function. The training process involves a maximum of 1000 epochs, with the Adam optimizer for objective function optimization. Detailed model structures are defined according to Table 1. Lags are selected in descending order of PACF, similar to the linear model. The number of lag sequences in LSTM networks is determined by evaluating the loss function on the test dataset, balancing network complexity. The criterion for order selection is defined as:

$$L = \frac{1}{m}\sum_{i}^{m}(y_i - \hat{y}_i)^2 |L_{50} - L_n| \le 3\%, \tag{7}$$

Table 1
Detained parameters in LSTM network training process.

Parameters	Value
Training method	Adaptive moment estimation
Normalization method	Max-min normalization
LSTM layer	4
Minimum batch size	20
Maximum epochs	1000
Initial learning rate	0.0001
Learning rate drop factor	0.5
Learning rate drop period	500

where L_n is the loss function of the testing dataset with input parameters number of n. For short-term building energy prediction, the dataset is transformed by sliding window to represent the sequential characteristics. As shown in Fig. 4b, datasets are reconstructed into sequential data according to the size of sliding window, and then used for training and testing of LSTM models.

2.2.3. *Correlation coefficient analysis*

To valid the proposed hypothesis of the building energy decomposition, *i.e.*, building energy consumption encompasses 1) physics-driven part, 2) occupants-driven part, and 3) white noise, this section presents the method to demonstrate the composition of provides part 1 and part 2 through weather-correlation (correlation coefficients between energy use/residuals and weather data) and autocorrelation (correlation coefficients between current and history energy use/residuals data) analysis. The observed data should be strongly related to the weather data, such as temperature, wind velocity, solar radiation intensity, and humidity level [57,58] if governed by the physics rules such as heat transfer laws and air thermal dynamics. Therefore, if the residuals between the physics-based model and observation still contain such physics components, it should be weather variables correlated. Contrarily, the autocorrelation coefficients measure the correlation among time series data, providing insights into the time-related patterns within the data. The occupant schedule or activities of a building exhibits strong time relevant patterns and periodicity [59,60]. Hence, if the residual is not physics related but only autocorrelated, it indicates that this component is driven by occupant behaviors or activities. In this work, the Pearson correlation and the Spearman's Rank correlation are used to examine the linear and non-linear correlation between residuals and weather data. The Pearson correlation coefficients (ρ_p) and Spearman's Rank correlation coefficients (ρ_s) are calculated by equations below:

$$\rho_p(y_1, y_2) = \frac{cov(y_1, y_2)}{\sigma_{y_1}\sigma_{y_2}}, \rho_s(y_1, y_2) = 1 - \frac{6\sum d_i^2}{n(n^2-1)}, \quad (8)$$

where the $cov(y_1, y_2)$ is the covariance between sequences y_1 and y_2; σ is the standard deviation; d is the difference between the two ranks of each sequence; n is the number of observations in the sequence.

2.2.4. White noise analysis

After decomposing the building energy use data into physics-driven and occupants-driven parts, this section further demonstrates the components as the remaining part (after subtracting physics-based and data-driven components) of energy use, *i.e.*, the observation remaining term, via white noise analysis. The reasoning is that if the physics-based component can be captured by physics-based models while occupant driven component can be captured by residual modeling (data-driven component), the observation remaining term should be a white noise that refers to a sequence of uncorrelated random variables [61]. A key feature of a white noise process is that the covariance between any variable Y_t and its historical data at any time lag k (Y_{t-k}) is zero. *i.e.*, $cov(Y_t, Y_{t-k}) = 0$. Consequently, in a standard white noise process, both the ACF and PACF are identically equal to zero. After fitting a time series model, if the ACF and PACF of the observation remaining fall inside the confidence interval, one can consider that this process behaves like white noise, leaving behind this part that are random and not worth of any further investigation. In this work, we calculate the ACF and PACF of the observation remaining term, which is obtained by subtracting results of physics-based and data-driven models from the observation, to dictate whether this observation remaining term is white noise or not.

2.3. Ensemble learning with residuals modeling

Following analysis and decomposition of building energy use data, ensemble learning by integrating data-driven residual models with physics-based models is proposed to enhance BEM prediction accuracy, following the logic and hypothesis that building energy use is composed of physics-driven, occupant-driven, and white noise components. This section presents the model ensembling process (Section 2.3.1), together with the model performance evaluation (Section 2.3.2).

2.3.1. Ensemble learning framework

The detailed model ensembling process for building energy prediction is illustrated in Fig. 5. For demonstration with an example, one can consider the scenario that physics-based data are available and the data-driven residuals model (discrepancy between physics-based models and observations) incorporates lagged data from the previous n timesteps (lag 1 to lag n) as the models to be ensembled. The detailed steps for ensemble learning is as follows, 1) the estimated building energy consumption $\hat{y}$ is simulated using EnergyPlus or RC models; 2) based on the observed data during time point $t-n$ (y_{t-n}) to t (y_t), the residuals for these time point (r_{t-n} to r_t) are calculated by subtracting the simulated values from the observed values ($r_t = y_t - \hat{y}_t$); 3) The calculated residuals are then used as the inputs to the residual model $R(r_{t-n}, \cdots, r_t)$, and the predicted residual at the next timestep ($t+1$) can be given as r_{t+1}; 4) Finally, the predicted energy consumption data at time point $t+1$ is derived by adding the predicted residual to the simulated value $\hat{y}_{t+1} + r_{t+1}$. By implementing this procedure, the model is able to forecast building energy consumption ahead by a time increment ΔT, where ΔT is the time resolution defined as $\Delta T = T_{t+1} - T_t$.

2.3.2. Performance evaluation

The superiority of the proposed ensemble model encompasses two aspects: 1) higher prediction accuracy compared to physics-based

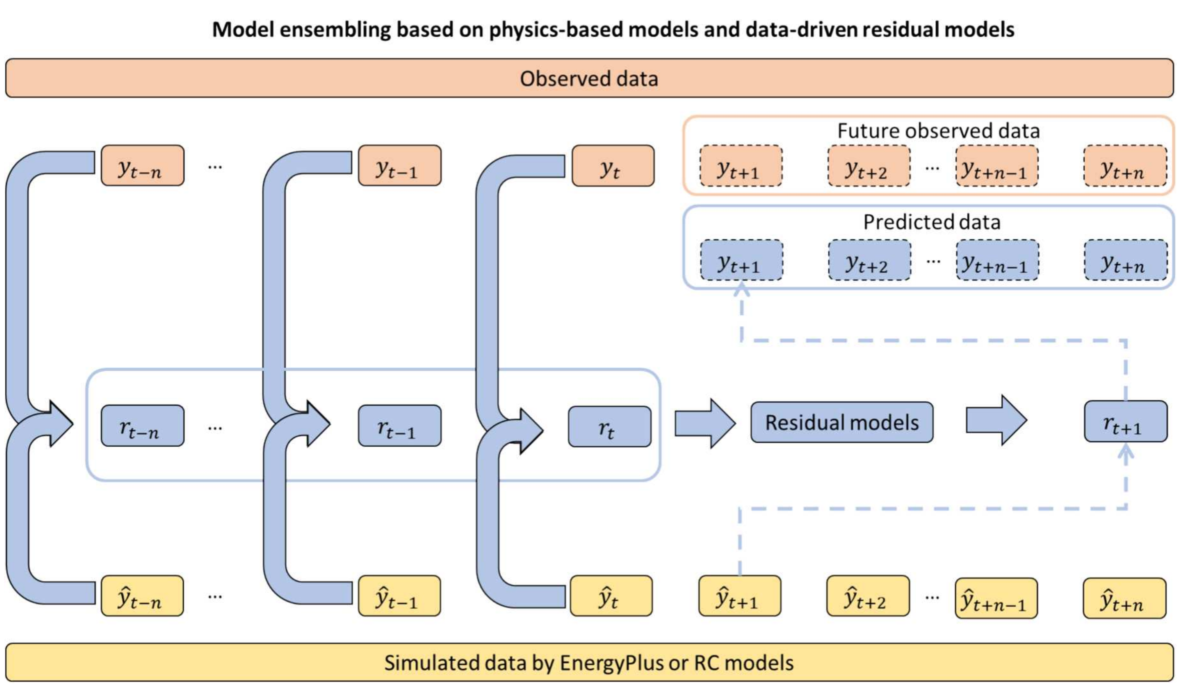

Fig. 5. Model ensembling framework based on the physics-based models and data-driven residual models.

models, and 2) greater robustness in small training dataset scenarios compared to the pure data-driven models. Therefore, the performance evaluation includes two parts. First, two widely used metrics in statistics are employed for assessing the prediction accuracy, namely Mean Absolute Error (MAE) and the Coefficient of Variation of Root Mean Square Error (CV-RMSE). Lower MAE and CV-RMSE indicates that the model has a higher accuracy. These two indices are calculated as:

$$MAE = \frac{1}{n}\sum_{1}^{n}|\widehat{y}_i - y_i|, CV - RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\widehat{y}_i - y_i)^2}{n}} \Bigg/ \frac{\sum_{i=1}^{n} y_i}{n}, \tag{9}$$

where y_i and $\widehat{y}_i$ are the observed and simulated values, and n is the total observation number. Notably, the CV-RMSE is highly influenced by the magnitude of average. Therefore, to avoid the bias introduced by average in CV-RMSE calculation and evaluate modeling outcomes of cooling energy use more accurately, we only incorporate cooling energy use in cooling season data (excluding periods with mostly 0 across the year) for CV-RMSE calculation. Second, the robustness of the models is examined by training them with different dataset sizes. In this study, we selected two state-of-the-art data-driven models, namely the LSTM model and SVM model, as benchmarks [62–64]. The testing dataset consists of the annual energy consumption data, while the training dataset ranging from 10 % to 100 % of the annual data in time series. For example, a 10 % training dataset indicates that the model is trained using 10 % of the annual energy consumption data, approximately equivalent to the energy consumption in January. The trained models are then tested with the entire annule dataset to assess their extrapolation behavior.

3. Results

This section presents the modeling and analysis results. First, Section 3.1 briefly presents the modeling outcomes of physics-based building energy models. Then, the analysis of building energy use data decomposition is presented with correlation and white noise analysis in Section 3.2. Finally, Section 3.3 discusses the overall performance of the proposed ensemble models, including the prediction accuracy and robustness.

3.1. Physics-based modeling outcomes

This section shows the simulation results of physics-based models. Considering modelled cooling and heating energy use of AEB is similar with CSC, we only present the EnergyPlus and 3R2C modeling results of CSC. The observed and simulated energy use of CSC is shown in Fig. 6, together with the discrepancies of these data, *i.e.*, the residuals. These figures predominantly feature data with a 60-minute time resolution, and sub-hourly time resolution data (30-minute, 15-minute, and 5-minute) are available in the Supplementary Information (Figs. S1-S2). Detailed MAE and CV-RMSE indices are presented in Fig. 7. These figures and performance evaluation metrics highlight several important aspects. First, even though the EnergyPlus model is capable of modeling in the sub-hourly level, the resulting simulated outcomes are essentially linear interpolations of hourly data (Fig. S1). This interpolation stems from the fact that the weather data for EnergyPlus is interpolated over one-hour periods. The similar limitation also applies to the 3R2C model (Fig. S2) because the ordinary differential equations in 3R2C models are linear as well. Secondly, both MAE and CV-RMSE show that the 3R2C model has a lower modeling accuracy compared to the EnergyPlus model as high-fidelity models. Especially regarding modeling of heating energy use, both MAE and CV-RMSE are significantly higher for predictions from the 3R2C model compared to EnergyPlus. This is possibly introduced by simplifications of the 3R2C model in capturing building operation dynamics.

3.2. Building energy data decomposition

In this section, the analysis results of building energy data decomposition are presented. To test our hypothesis that building energy use is comprised of three components (physics-driven, occupant driven, and white noise), we start by subtracting the physics-based modeling results (including different fidelities of modeling) from the measured building energy use to obtain residuals. Whether building energy use data can be decomposed into three components could be reflected by two aspects: 1) correlation of residuals with weather variables: the residuals should exhibit no or small correlation with weather data if there is no physics-driven component within the residuals; 2) the existence of white noise after subtracting physics-driven and occupants-driven components from measured energy use: the observation remaining part is essentially white noise, indicating that all systematic variations are captured by ensemble modeling (explained in Section 2.2.4).

The calculated Pearson correlation coefficient $|\rho_p|$ and the Spearman's rank correlation coefficients $|\rho_s|$ are demonstrated in Fig. 8. Results of the weather correlation coefficients analysis of residuals show that there is no obvious physics information uncaptured in residuals, *i.e.*, the physics-driven part is well-captured by physics-based models. As general rules of thumb, a threshold of $|\rho| < 0.3$ is employed to filter out features with no or weak correlation with a target variable [65]. Taking the cooling energy use of CSC as example (Fig. 8a), the original measured energy use is strongly correlated with weather factors such as

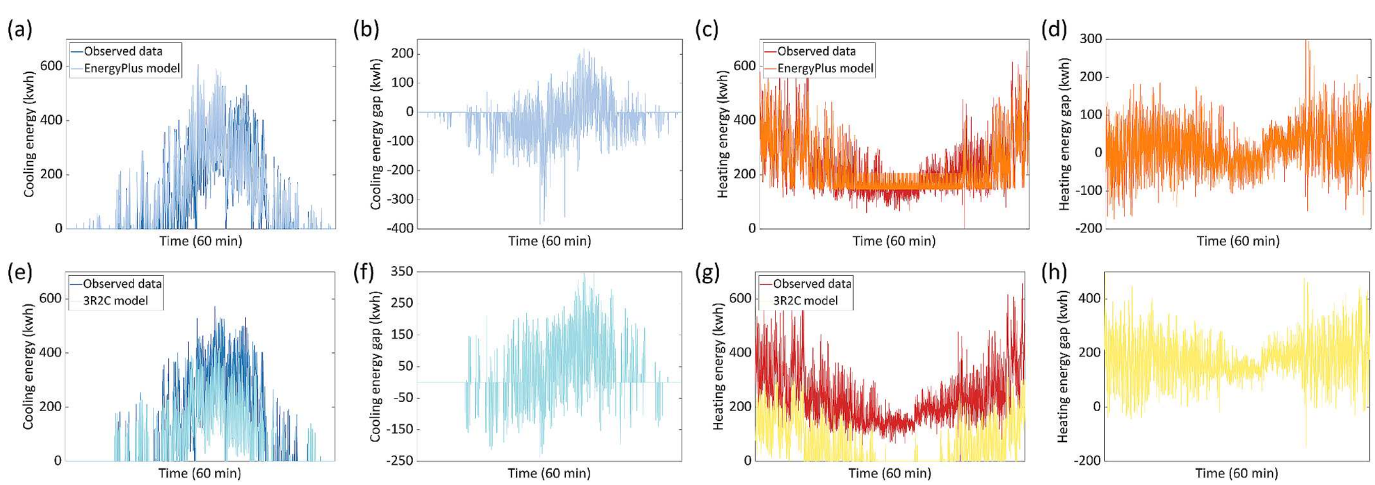

Fig. 6. Observed cooling/heating energy data, physics-based model simulated data, and their residuals. (a-d) EnergyPlus model and their residuals. (e-h) 3R2C model and their residuals.

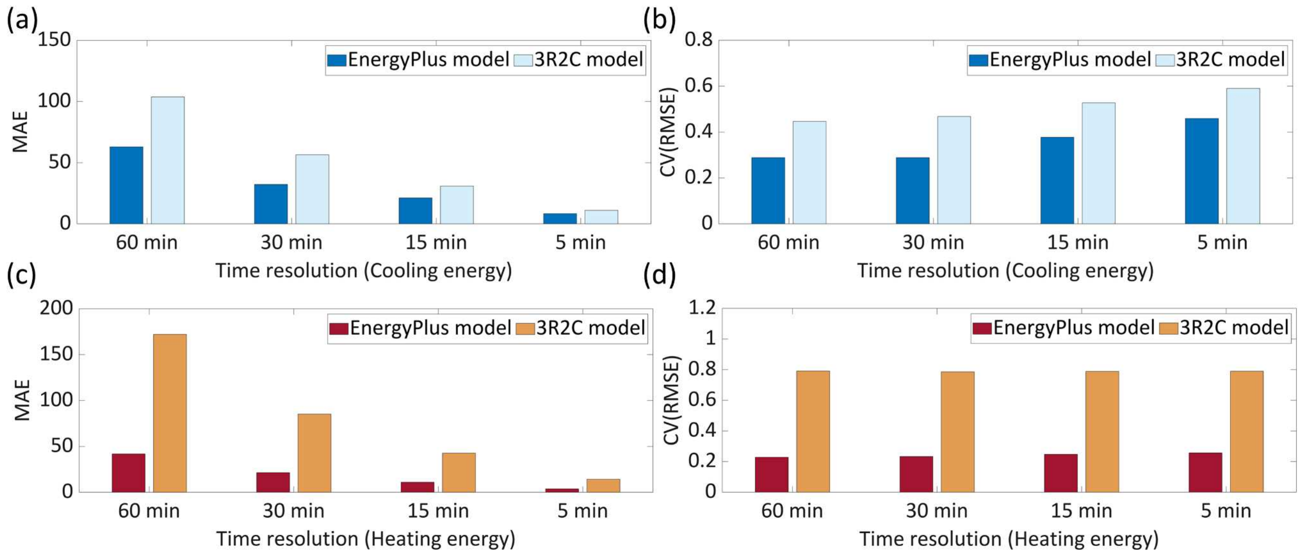

Fig. 7. (a) MAE and (b) CV-RMSE of the cooling energy simulated by physics-based model with multi-time resolution (c) MAE and (d) CV-RMSE of the heating energy simulated by physics-based model with multi-time resolution.

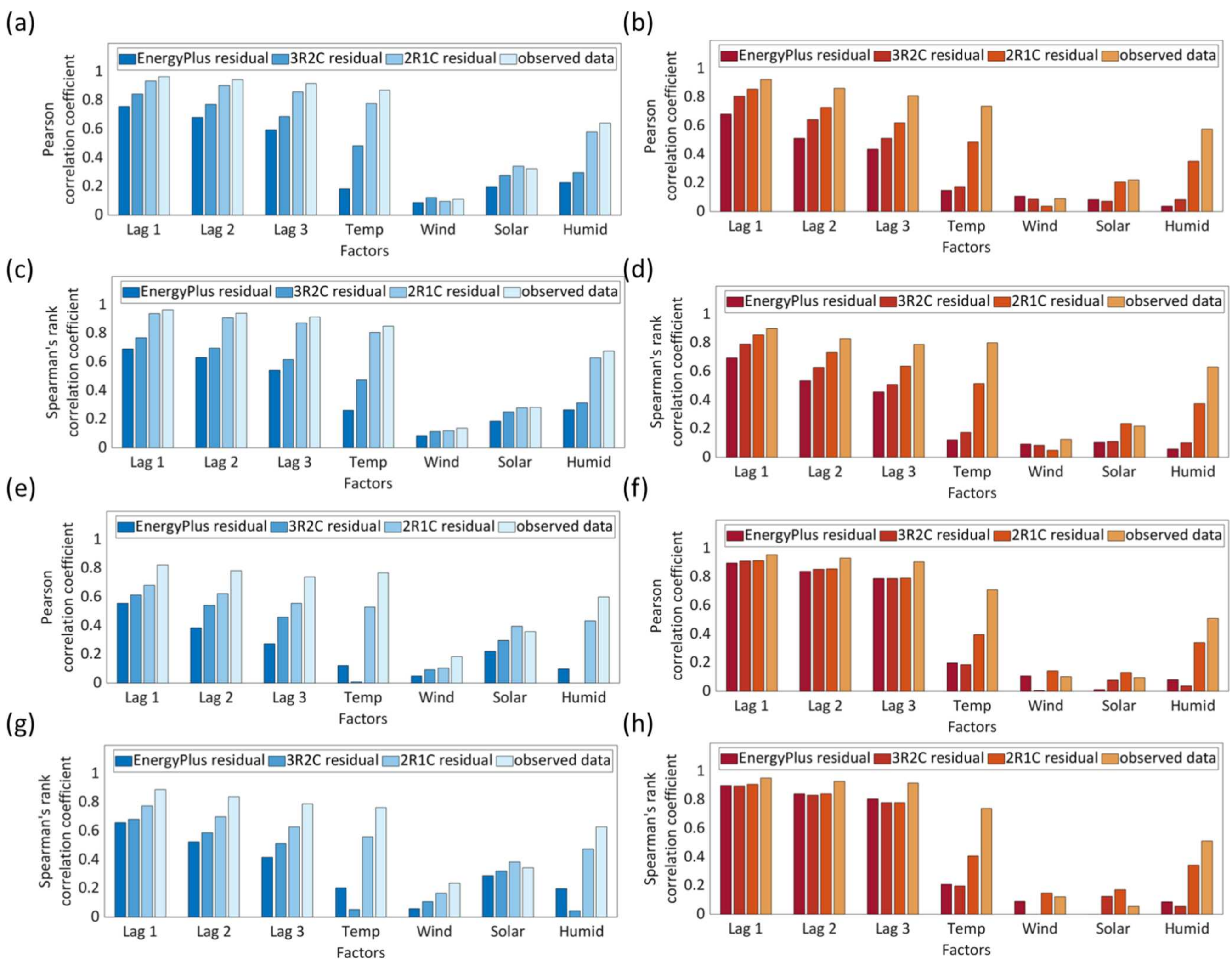

Fig. 8. Pearson and Spearman's rank correlation coefficient (absolute value) between cooling/heating energy and parameters for (a-d) CSC and (e-h) AEB.

outdoor temperature, relative humidity, and solar radiation, indicating that the original energy use involves the physics processes in buildings. This physical component is then captured by physics-based models, resulting in the significantly reduced correlation coefficients (ranges from 0 to 0.2 for EnergyPlus models) between weather data and the residuals after subtracting physics-based modeling outcomes from total energy use. However, the autocorrelation coefficients between residuals of consecutive timesteps are still high (over 0.6), indicating that the physics-based models are insufficient to capture this periodicity and the seasonality information in the building energy use data. The low physics correlation and high autocorrelation in residuals synergistically demonstrate the existence of the occupant driven part in residuals. Timeseries based residual modeling process (presented in Section 2.2.2) is intended as part of the ensemble learning approach to capture this occupant driven part. Another observation is that correlations between residuals and weather data increase as the fidelity of physics-based modeling for residual calculation decreases (*i.e.*, the residuals from 2R1C models are more correlated with weather data compared to residuals from EnergyPlus models). This also clearly indicates the existence of physics-driven component in total energy use. As more physics-driven energy use (from high-fidelity model) is sufficiently extracted from total energy use, the residuals become auto-correlated (occupant driven) instead of weather correlated.

Even though the residuals are not correlated with the current weather data, there is a concern that autocorrelative residuals may contain model errors in the heating and cooling system, such as delays in system response. If the residuals encompass these system delays, the correlation coefficients between the residuals and the historical weather data should be significantly high. To test this, we calculated the Peason correlation coefficients between the residuals and the weather data (temperature and relative humidity data) at three different time lags (1 h, 2 h, and 3 h) for both the CSC and AEB. For this analysis, we used residuals based on the EnergyPlus model and the 3R2C model with a time resolution of 60 min. The results, shown in Tables 2 and 3, indicate that the correlation coefficients between the residuals and the historical weather data are low, suggesting no obvious relationship between the residuals and the delayed energy consumption data. However, the residuals of the 3R2C model for the cooling energy data of the CSC exhibit slightly high correlation coefficients with the historical temperature data. This could be caused by the low fidelity of the 3R2C model, as the correlation coefficients between the residuals and the current temperature data are also high (Fig. 8a) while the correlations between residuals of EnergyPlus model and weather data are much lower. Therefore, the autocorrelation factors in the residuals should be attributed to the occupant activities rather than the delays in system response.

After the validation of physics-driven part and occupants-driven part in building energy use data, the observation remaining term is demonstrated as the white noise. The white noise analysis can be verified through ACF and PACF analysis of the remain part of data beside the physics-based and data-driven models. Fig. 9 and Fig. S3-6 demonstrate that the observation remaining term sequences from these models with hourly and sub-hourly time resolution exhibit negligible autocorrelation (less than 0.1), behaving as white noise. Thus, these analysis results collectively demonstrate that building energy use data can be decomposed into three parts (Fig. 10): 1) physical component well-described by physics rules (*e.g.*, external building thermal load), 2) stochastic component contributed by occupant activities, and 3) white noise. In detail, the quantitative proportions of the residuals and the white noise are shown in Fig. 11. The proportion of white noise is calculated by dividing the CV-RMSE after model ensembling by the CV-RMSE of the physics-based model. For example, in the case of cooling energy based on EnergyPlus model with a time resolution of 60 min, the percentage of residuals is 40 % while that of white noise is 60 %, indicating that the CV-RMSE of the physics-based model would be reduced by 40 % after model ensembling and residual modeling to capture the occupant driven components. The proportion of residuals based on the EnergyPlus model is lower than that of the 3R2C model because more physics information is contained in the residuals of the low-fidelity model. Additionally, the proportion of heating energy residuals increases with reduced time resolution, demonstrating that the residuals model can capture more occupant behavior information within heating energy dataset with smaller time resolutions. Overall, such decomposition allows us to develop an ensemble learning framework where the physics-based models capture the physical information while the data-driven residual models capture the time-series information in building energy use data. Results presented in Fig. 8e-f for AEB lead to the same conclusions as this decomposition analysis for the CSC building.

3.3. Performance evaluation for ensemble models

With the determined optimal orders for both linear and nonlinear models in residual modeling (See Supplementary Information S2.3), the physics-informed ensemble model has been developed for modeling building energy use. This section presents the performance of ensemble model integrating EnergyPlus and 3R2C physics-based models with residual models for the CSC. The results of AEB are not presented to save space since these are similar with CSC.

First, the prediction accuracy improvement is demonstrated, compared to physics-based models. The residual models used in the ensemble framework are trained using annual cooling and heating energy data from 2021. The CV-RMSE for the trained ensemble models generally falls within the range of 0.1–0.3. The predicted building energy consumption is depicted in Figs. S20-24 and the detailed MAE and CV-RMSE are presented in Tabs. S7-10 and Fig. S25. After the model training, the cooling and heating energy data in 2022 are used for testing. Results with time resolution of 60 min are shown in Fig. 12, where the fitting performance of linear model is presented by the ARIMA model for illustration. Results of AR and ARMA model fitting and sub-hourly time resolution of modeling (30-minute, 15-minute, and 5-minute) are similar to hourly modeling results based on ARIMA (Supplementary Information (Figs. S26-29)). The outcomes of the ensemble models fit observed data well, with detailed MAE and CV-RMSE provided in Tabs. S11-S15. Fig. 13 illustrates the relationship of MAE and CV-RMSE with the time resolution and model type. Results reveal that the CV-RMSE of these models generally falls between 0.1 and 0.4, and there is a significant reduction in both MAE and CVRMSE by 40–90 % compared to the outputs from physics-based models, indicating that the time series based residual modeling (integrated with physics-based modeling through model ensembling) substantially enhances the accuracy of building energy modeling. Furthermore, the performance of ensemble modeling is robust regardless of fidelity of physics-based models involved as part of model. Despite the lower accuracy of

Table 2
Pearson correlation coefficients between residuals and the historical weather data for CSC.

Correlation Coefficients	Temp 1	Temp 2	Temp 3	Humid 1	Humid 2	Humid 3
Residuals_EnergyPlus_cooling	0.175	0.163	0.150	0.214	0.199	0.182
Residuals_EnergyPlus_heating	0.150	0.155	0.164	0.051	0.064	0.083
Residuals_3R2C_cooling	0.459	0.429	0.400	0.274	0.242	0.211
Residuals_3R2C_heating	0.180	0.178	0.177	0.085	0.078	0.074

Note: Temp n/Humid n represents the n time-lags temperature/relative humidity data.

Table 3
Pearson correlation coefficients between residuals and the historical weather data for AEB.

Correlation Coefficients	Temp 1	Temp 2	Temp 3	Humid 1	Humid 2	Humid 3
Residuals_EnergyPlus_cooling	0.110	0.092	0.074	0.084	0.063	0.041
Residuals_EnergyPlus_heating	0.202	0.205	0.209	0.086	0.091	0.099
Residuals_3R2C_cooling	0.044	0.079	0.111	0.039	0.077	0.114
Residuals_3R2C_heating	0.189	0.191	0.194	0.035	0.031	0.031

Note: Temp n/Humid n represents the n time-lags temperature/relative humidity data.

3R2C model compared to the EnergyPlus model, the accuracy level of the ensemble models is similar after model integration (integrating EnergyPlus or 3R2C with residual models). This implies that enhancing the fidelity of physics-based models might not be critical when applying the ensemble modeling approach with residual models in building energy modeling. Ensemble model performance varies across different time resolution and the chosen data-driven models. Models with smaller time resolutions typically exhibit lower MAE. This trend can be attributed to the fact that MAE is highly dependent on the magnitude of the difference between simulated and observed data, which tends to be smaller at finer time resolutions. For the chosen data-driven models, the nonlinear LSTM models show no significant advantages over linear models such as AR, ARMA, and ARIMA, except at a 5-minute resolution. Compared to MAE, CV-RMSE is independent with the magnitude of the data due to the normalization by dividing the average of observed data. More details will be discussed in the Section 4 below.

Then, the enhanced prediction robustness of the proposed ensemble model is evident when compared to purely data-driven models (LSTM and SVM as the baselines). For this study, we selected the CSC building and time resolution of 60 min as an example. The physics-informed ensemble model incorporates the EnergyPlus model as the physical component and the LSTM model as the data-driven component, while the LSTM and SVM are used as the baseline data-driven models for comparison. For a fair comparison, the features used in the pure data-driven models include not only the three time-lags of historical data but also weather data, such as outdoor temperature, wind velocity, solar radiation, and outdoor relative humidity. In contrast, the LSTM model within the ensemble framework uses only the three time-lags of historical data. The results for cooling and heating energy predictions are shown in Fig. 14a-b and Fig. 14c-d, respectively. The proposed ensemble model consistently outperforms the LSTM and SVM models, especially with smaller training datasets. Taking cooling energy prediction as instance, when using the entire dataset for training, the CV-RMSE of the three models are similar: 0.254 for the LSTM model, 0.225 for the SVM model, and 0.210 for the physics-informed ensemble model. However, when using just 10 % of the annual data for training, the CV-RMSE of both the LSTM and SVM model is 1.231, whereas that of physics-informed ensemble model remains acceptable as 0.300. Therefore, although the pure data-driven models show similar prediction accuracy as the proposed ensemble model with a full dataset, their performance significantly deteriorates with smaller training datasets, indicating that pure data-driven models struggle to learn useful information from limited data, while the ensemble model remains robust. Furthermore, even without any training data, the baseline performance of the ensemble model would approximately converge to that of the physics-based model (with a CV-RMSE of 0.289 for cooling energy perdition using EnergyPlus model alone), showcasing the great potential of the physics-informed ensemble model for application in extrapolation scenarios. Overall, this physics-informed ensemble learning by residuals modeling shows effectiveness for improving both the accuracy and robustness of building energy modeling.

4. Discussion

- **Building energy use data decomposition and ensemble modeling**

Physics-based models have been widely used for BEM since the 1960s, yet significant discrepancies between model predictions and ground observations still exist while being challenging to explain and close. In this work, we firstly decompose the building energy use data by correlation and white noise analysis, and correspondingly demonstrate that the building energy consumption can be composed of three parts: 1) a physics-driven component governed by physics rules, such as heat transfer and thermodynamics; 2) a stochastic component driven by occupant activities that exhibits certain periodicity and uncertainty; 3) a white noise component caused by unexpected or unknown sources (*e.g.*, anomalies in the data collection process). The low correlations between residuals (difference between physics-based modeling and field observations) and weather variables and the high auto-correlation of residuals suggest that it is the occupant behaviors contribute to these residuals, which are hard to be captured by physics-based models. The validation of white noise as observation remaining components provides further evidence. This decomposition aligns with Norford *et al.* [66], attributing 88 % of the prediction discrepancies to unexpected occupant energy use and HVAC systems operation. Another similar idea was proposed by Chen *et al.* [67], *i.e.*, the building energy use data can be divided into a linear component obtained by prior physical knowledge and a nonlinear component from unpredictable factors such as environment noise. However, no strict evidence was provided for such decomposition. According to correlation analysis, we find the autocorrelation coefficients of the residuals between physics-based modeling and observation are still as high as over 0.5, indicating that it is necessary to further describe the nonlinear component with the occupants-relevant (data-driven) model and white noise.

Given these insights, it is natural to use the ensemble modeling approach for building energy modeling. On one hand, using physics-based model would result in inevitable accuracy compromise since the stochastic component is hard to capture by physics-based model even with dedicated model calibration. On the other hand, compared with the pure data-driven model, such ensemble approach ensures the reliability and robustness when the data for training a model is limited.

- **Model complexity for ensemble modeling**

Using the physics-informed ensemble learning approach to enhance BEM accuracy, selecting appropriate model complexity for both physics-based models and data-driven residual modeling is worthy of discussion. The similar MAE and CV-RMSE of ensemble models based on EnergyPlus and RC models as physics-based models suggests that enhancing the fidelity of physics-based models as part of ensemble models may not be necessary to realize decent model performance if the ensemble modeling approach is adopted. Instead, employing low-fidelity physics-based models is sufficient to encapsulate basic physics dynamics, while data-driven models could address the remaining gaps introduced by occupant-driven dynamics without compromising modeling accuracy. Furthermore, based on our tests, the highly complex data-driven models for residual modeling are also unnecessary to realize a decent building energy modeling performance in the proposed ensemble learning approach. For the analyzed dataset, linear models such as AR, ARMA, and ARIMA (with CV-RMSE of ~ 0.17–0.57) generally outperform the nonlinear LSTM model (with CV-RMSE of ~ 0.18–0.6), except at a 5-minute resolution, possibly due to the predominantly linear nature of

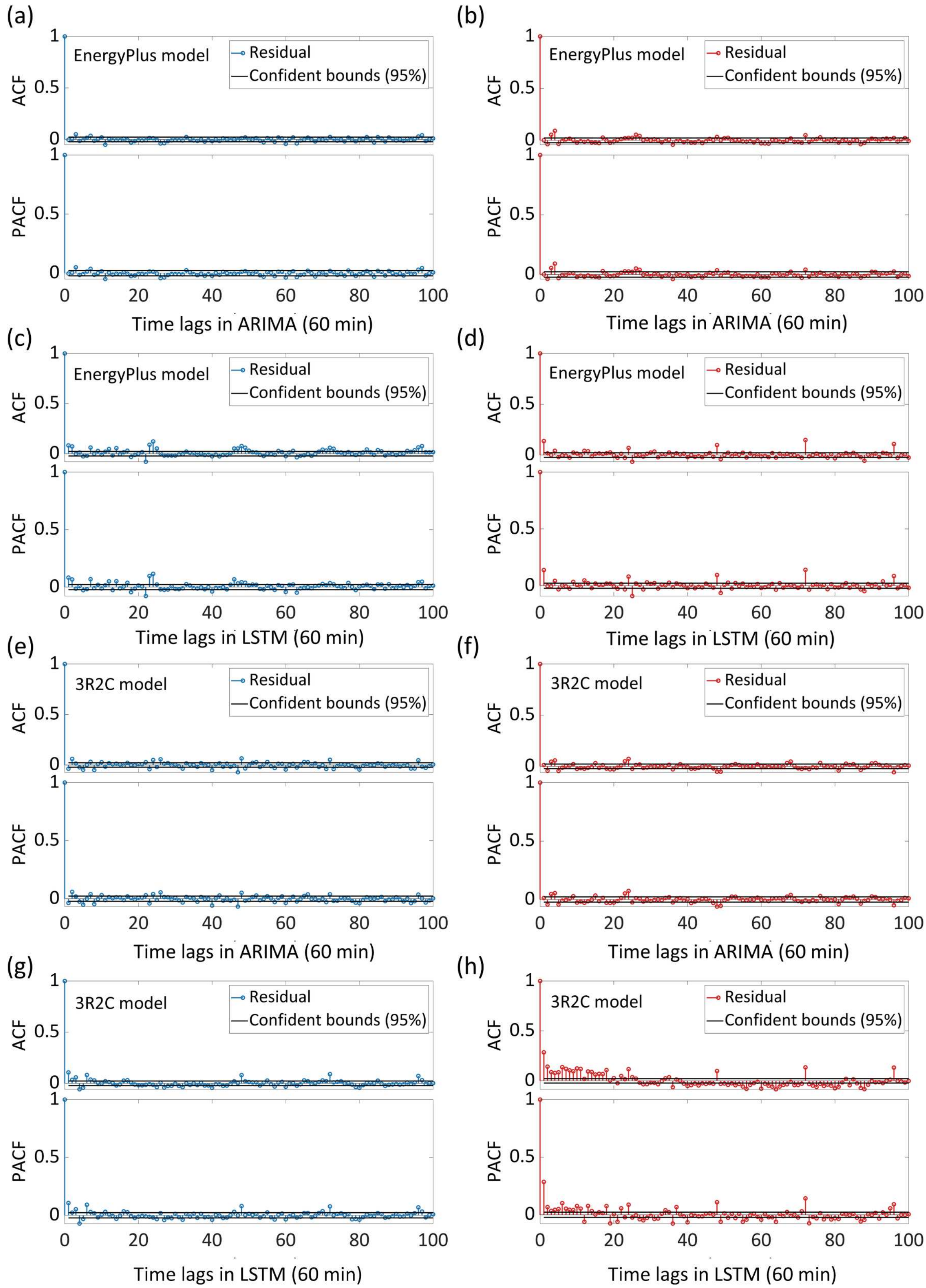

Fig. 9. ACF and PACF of the residuals (observation remaining term) of these hybrid models based on EnergyPlus model. (a) Cooling and (b) heating energy by ARIMA model. (c) Cooling and (d) heating energy by LSTM model. The time resolution is 60 min.

autoregression in the residuals. Additionally, the volume of data influences residual model performance. While nonlinear LSTM models perform better with large datasets, linear models achieve accuracy with smaller data volumes. This finding aligns with research about the impact of data volume on performance of ARIMA and LSTM models [68–70]. Consequently, for modeling of low temporal resolutions (60, 30, and 15 min), linear models are not only sufficient to capture information in residual data, but also offer superior computational efficiency.

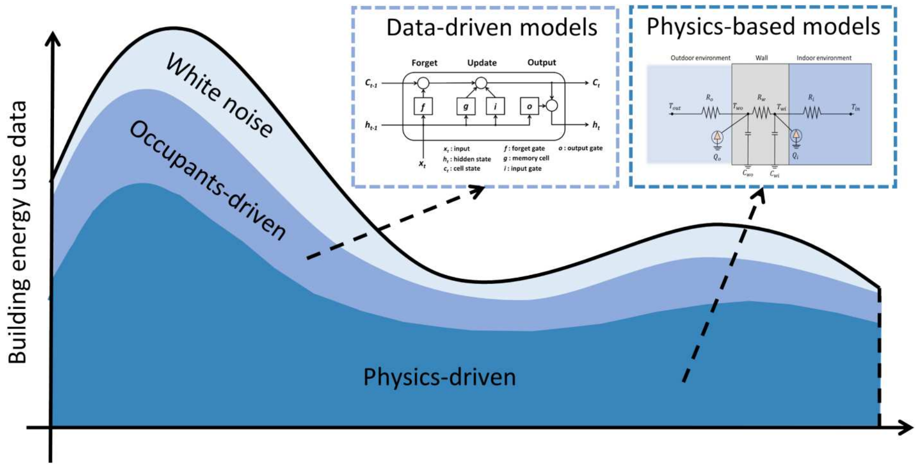

Fig. 10. Building energy use data decomposition.

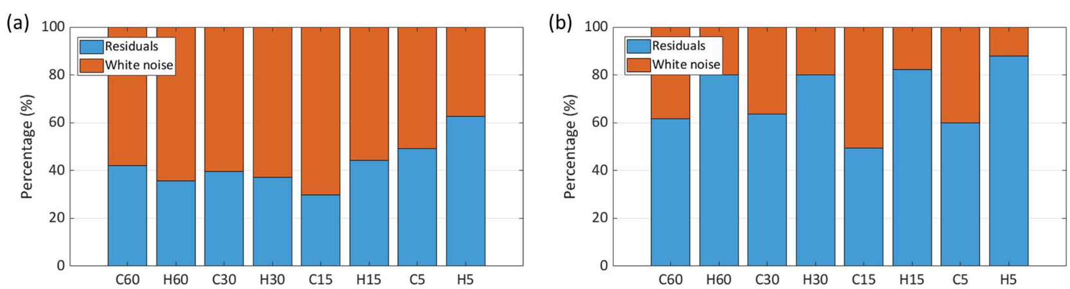

Fig. 11. The percentage of error factors, *i.e.*, residuals (occupant-driven part) and white noise, for cooling and heating energy with different time resolution. (a) Residuals (occupant-driven part) based on EnergyPlus model. (b) Residuals (occupant-driven part) based on 3R2C model. Note that the data-driven model for training the residuals is LSTM. Note that Cn/Hn represents cooling/heating energy data with n-minute time resolution.

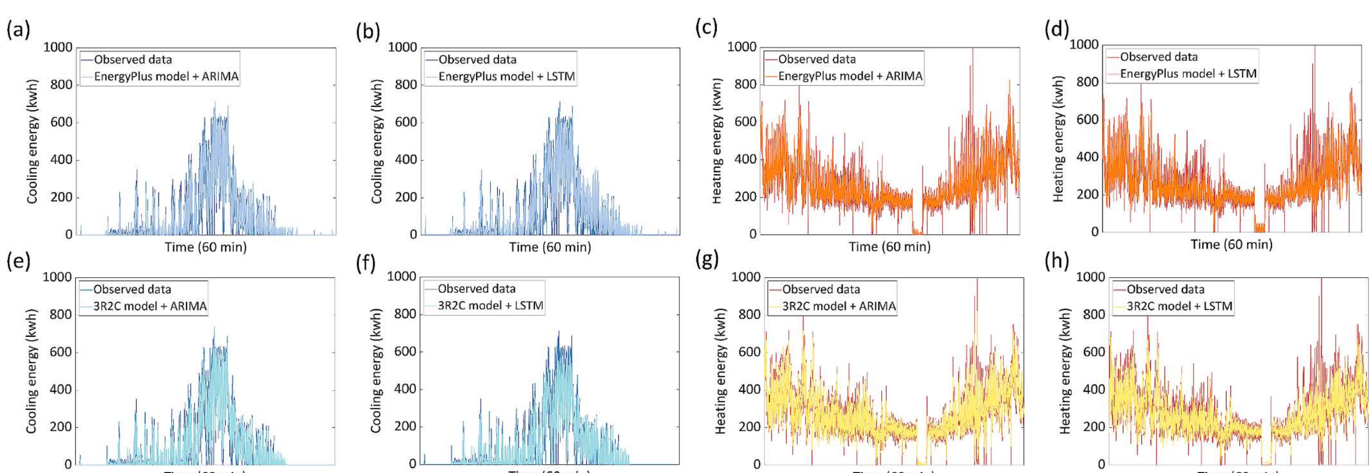

Fig. 12. Comparing observed data and prediction data based on EnergyPlus models. (a) Cooling energy predicted by ARIMA and (b) LSTM model. (c) Heating energy predicted by ARIMA and (b) LSTM model.

In detail, for a regular computer equipped with 12-core Intel(R) Core (TM) i5-10400F CPU @ 2.90 GHz and NVIDA GeForce RTX 2060 GPU, the computational efficiency of each model is listed in Table 4. The EnergyPlus model (high fidelity model) requires significant more computation time than simple RC models (low fidelity model), especially when the time resolution is small (164 s for the EnergyPlus model and 44.8 s for the 2R1C model), not to mention the time-consuming parameter collection process and complex model development processes required for a high-fidelity EnergyPlus model. The accuracy evaluation indicates that using a low-fidelity model as the physics component in the ensemble model does not compromise the prediction performance. Thus, the proposed ensemble learning framework could have much higher computational efficiency (with time cost of 5.6 s to 46.2 s) compared to EnergyPlus models (with time cost of 32 s to 164 s) by combining 2R1C and ARIMA models. This approach not only enhances prediction accuracy but also reduces computation time, which is particularly advantageous for building energy optimization, typically involving heavy iterative tasks. Therefore, in building energy modeling,

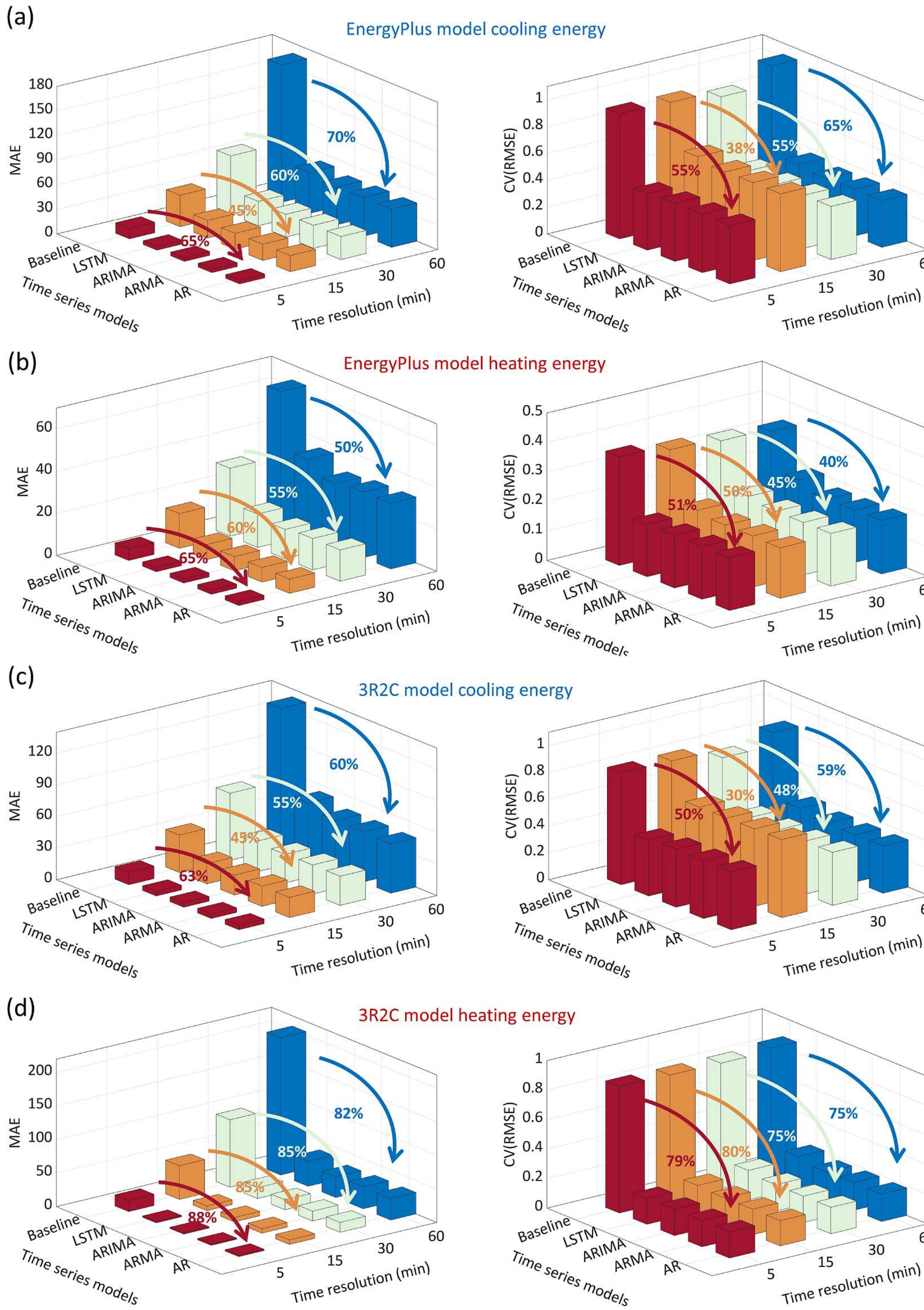

Fig. 13. MAE and CV-RMSE of the ensemble models prediction results versus time resolution and type of time series models. (a-b) EnergyPlus model. (c-d) 3R2C model. Baselines are the prediction of physics-based models.

it is reasonable to use ensemble models composed of simple physics-based and data-driven models, avoiding excessive time and efforts in modeling instead of high-fidelity physics-based modeling or complex deep learning-based algorithms in data-driven residual modeling. The combination of these simple models could realize even higher accuracy, as demonstrated in Section 3.3.

- **Performance of ensemble modeling of different time resolution**

When BEM is widely used in building energy use prediction, analysis of its performance in prediction of different time resolution is usually overlooked. Coarse time resolution, such as monthly, weekly, daily, or even hourly ahead prediction are inadequate for precise demand control and system optimization [71], calling for a need of sub-hourly ahead prediction. However, building energy use in smaller time resolution typically exhibits increased randomness and uncertainties. The significantly increased CV-RMSE in the 15-minute and 5-minute prediction for

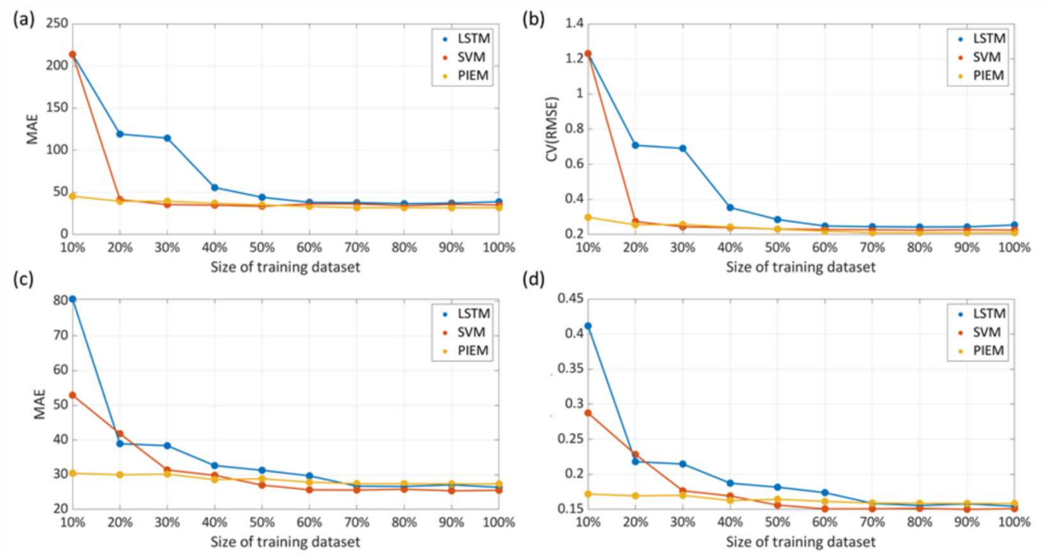

Fig. 14. MAE and CV-RMSE of the ensemble models prediction results versus size of training dataset. (a-b) Cooling energy prediction. (c-d) Heating energy prediction. Baselines are the outcomes from LSTM model and SVM model.

Table 4
Computational efficiency of different physics-based and data-driven models.

Time resolutionModel	60 min	30 min	15 min	5 min
EnergyPlus	32 s	41 s	64 s	164 s
3R2C	6.8 s	11.5 s	20.7 s	59.1 s
2R1C	4.9 s	8.3 s	15.7 s	44.8 s
ARIMA	0.7 s	0.7 s	0.9 s	1.4 s

Note: The inputs for the ARIMA models include three time-lags of historical data.

both EnergyPlus and 3R2C modeling (in Section 3.3) implies that it is challenging for physics-based models to effectively capture such uncertainties. The proposed physics-informed ensemble model addresses these challenges by effectively capturing stochastic component through residual modeling, demonstrating robustness against the inherent randomness and uncertainty in high-time resolution data of building energy use. For example, in prediction of 5-minute time resolution, this ensemble model approach significantly reduces the coefficient of variation of CV-RMSE by 50–60 % for cooling energy and 60–90 % for heating energy prediction. This underscores the effectiveness of high-resolution ensemble modeling in enhancing the performance of building energy use predictions in fine time resolution, suggesting significant impacts on further pushing BEM to support applications demanding accuracy in fine temporal resolution, such as smart control, demand response, and fault detection and diagnosis [72–74].

- **Insights for practical application**

Although there has been significant advancement in the research on Internet of Things (IOT), data collection for BEM is still time-consuming due to the installation and maintenance of sensors. Additionally, privacy concerns [41] and data limitation for buildings in the design phase and newly built buildings [42] present significant challenges. Data-driven models for BEM are usually restricted to the specific building operating conditions for which they are trained, leading to prediction failures when the training data is limited in certain situations [75]. Therefore, the limited availability of data hinders the application of machine leaning models in BEM. As demonstrated in Section 3.3, the performance of the proposed physics-informed ensemble model overperforms that of pure data-driven models, particularly when the training dataset size is small. For instance, when dealing with a building with limited observed energy use data, data-driven models are not the first choice in such situations. While physics-based models can be used, they also suffer from compromised prediction accuracy. Hence, in this circumstance, the proposed ensemble model could serve as a more accurate and robust alternative to both physics-based and data-driven models, effectively avoiding prediction failures when data is limited.

- **Limitations and future studies**

Finally, we summarize the limitations of this study and suggest directions for future research. First, the generalizability of the physics-informed ensemble model can be further addressed. This study was conducted on two specific educational buildings at the University of Utah. As building operation could vary significantly across different types of buildings and climate conditions, it is important to examine the generalizability of the proposed model for other types of buildings such as residential and commercial buildings in more climate zones. Evaluating the performance of the model across different climate zones is also worthy investigating. Then, the proposed method was evaluated using an HVAC system that operates 24 h per day. Future study could focus on the performance of physics-informed ensemble models for buildings with HVAC systems that operate intermittently. Finally, the case study presented in this work demonstrated that the residuals between physics-based simulation data and observations are primarily attributed to occupant activities. However, in other building modeling scenarios without detailed building operation information and calibrated building energy models, other factors may contribute to the modeling residuals in different case studies, such as system delays, time-related COP for electricity use prediction, and HVAC operation schedules. These autocorrelated terms should be further explored in future research.

5. Conclusion

This study introduces a novel physics-informed ensemble modeling that integrates physics-based modeling and data-driven residual modeling as the joint model for building energy use prediction. First, we demonstrate the decomposition of building energy use into physics-driven component, occupant-driven component, and white noise based on correlation and white noise analysis. Then, physics-based models such as EnergyPlus and RC models are developed to capture the physics-driven component in building energy use. Subsequently, the residuals between physics-based simulated and observed data are modeled by linear and nonlinear data-driven time series-based models. Finally, the ensemble modeling approach demonstrates excellent performance for enhanced building energy modeling. Compered to physics-based models, the proposed ensemble model achieves a CV-RMSE of less than 0.3 for cooling energy prediction and less than 0.2 for heating energy prediction, showcasing accuracy improvements of 40–90 % in both MAE and CV-RMSE. The proposed ensemble model also shows greater robustness than pure data-driven models. The CV-RMSE of cooling and heating energy prediction based on the proposed ensemble model are 0.300 and 0.172, respectively, when using just 10 % of the annual dataset for model training, whereas those for LSTM model are 1.231 and 0.412, respectively. Such robustness enhancement highlights the great potential of applying the physics-informed ensemble model in extrapolation scenarios.

The significant insight from this work is that the field observations of building energy use data can be decomposed into different parts, suggesting ensemble modeling as an effective and natural approach for enhanced modeling performance. Moreover, the ensemble approach based on different fidelities physics-based models show similar prediction performance, indicating that in certain scenarios, it may not be necessary to develop a high-fidelity physics-based model. The study also finds that linear models, which are computationally efficient, tend to be more accurate in ensemble modeling compared to LSTM models, except at a 5-minute time resolution. Overall, this physics-informed ensemble learning framework based on residual modeling is effective for enhancing the accuracy and robustness of building energy predictions, offering valuable insights for both research and practical applications in the field of building energy modeling.

CRediT authorship contribution statement

Zhihao Ma: Writing – review & editing, Writing – original draft, Methodology, Investigation. **Gang Jiang:** Validation, Software, Methodology. **Jianli Chen:** Writing – review & editing, Supervision, 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

Data will be made available on request.

Acknowledgement

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).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.enbuild.2024.114853.

References

[1] Y. Sun, F. Haghighat, B.C.M. Fung, A review of the-state-of-the-art in data-driven approaches for building energy prediction, Energy Build. 221 (2020) 110022.
[2] H. Gao, C. Koch, Y. Wu, Building information modelling based building energy modelling: A review, Appl. Energy 238 (2019) 320–343.
[3] E. Fabrizio, V. Monetti, Methodologies and advancements in the calibration of building energy models, Energies 8 (4) (2015) 2548–2574.
[4] V.S.K.V. Harish, A. Kumar, A review on modeling and simulation of building energy systems, Renewable Sustainable Energy Rev. 56 (2016) 1272–1292.
[5] Heating, A.S.o. and A.-C. Engineers, ASHRAE handbook. 2021: American Society of Heating, Refrigerating and Air-conditioning Engineers.
[6] N.-S. Truong, D.L. Luong, Q.T. Nguyen, BIM to BEM Transition for Optimizing Envelope Design Selection to Enhance Building Energy Efficiency and Cost-Effectiveness, Energies 16 (2023), https://doi.org/10.3390/en16103976.
[7] M. Dong, T. Lan, L. Zhong, Rethink energy accounting with cooperative game theory. Proceedings of the 20th Annual International Conference on Mobile Computing and Networking, 2014.
[8] A. Amin, et al., Demand response in buildings: Unlocking energy flexibility through district-level electro-thermal simulation, Appl. Energy 305 (2022) 117836.
[9] Y. Gao, et al., Energy management and demand response with intelligent learning for multi-thermal-zone buildings, Energy 210 (2020) 118411.
[10] R. Missaoui, et al., Managing energy smart homes according to energy prices: analysis of a building energy management system, Energy Build. 71 (2014) 155–167.
[11] A. Capozzoli, F. Lauro, I. Khan, Fault detection analysis using data mining techniques for a cluster of smart office buildings, Expert Syst. Appl. 42 (9) (2015) 4324–4338.
[12] M. Wrinch, T.H. El-Fouly, S. Wong, Anomaly detection of building systems using energy demand frequency domain analysis. 2012 IEEE Power and Energy Society General Meeting, IEEE, 2012.
[13] U.S. Department of Energy, Energy Efficiency and Renewable Energy Office, Building Technology Program. 2023; Available from: https://energyplus.net/.
[14] D.B. Crawley, et al., EnergyPlus: creating a new-generation building energy simulation program, Energy Build. 33 (4) (2001) 319–331.
[15] Open source library for building energy and control systems. 2023; Available from: http://simulationresearch.lbl.gov/modelica/index.html.
[16] Nouidui, T.S., Validation and application of the room model of the Modelica Buildings library. 2012.
[17] A transien system simulation program. 2023; Available from: http://www.trnsys.com/.
[18] M. Wetter, C. Haugstetter, Modelica versus TRNSYS–A comparison between an equation-based and a procedural modeling language for building energy simulation, Proc. SimBuild 2 (1) (2006).
[19] M. Al Gharably, J.F. DeCarolis, S.R. Ranjithan, An enhanced linear regression-based building energy model (LRBEM+) for early design, J. Build. Perform. Simul. 9 (2) (2016) 115–133.
[20] G. Ciulla, A. D'Amico, Building energy performance forecasting: A multiple linear regression approach, Appl. Energy 253 (2019) 113500.
[21] B. Dong, C. Cao, S.E. Lee, Applying support vector machines to predict building energy consumption in tropical region, Energy Build. 37 (5) (2005) 545–553.
[22] Z. Ma, et al., Applying support vector machines to predict building energy consumption in China, Energy Procedia 152 (2018) 780–786.
[23] Y. Liu, et al., Enhancing building energy efficiency using a random forest model: A hybrid prediction approach, Energy Rep. 7 (2021) 5003–5012.
[24] Z. Wang, et al., Random Forest based hourly building energy prediction, Energy Build. 171 (2018) 11–25.
[25] S. Abedi, S. Kwon, Rolling-horizon optimization integrated with recurrent neural network-driven forecasting for residential battery energy storage operations, Int. J. Electr. Power Energy Syst. 145 (2023) 108589.
[26] C. Fan, et al., Assessment of deep recurrent neural network-based strategies for short-term building energy predictions, Appl. Energy 236 (2019) 700–710.
[27] A. Rahman, V. Srikumar, A.D. Smith, Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks, Appl. Energy 212 (2018) 372–385.
[28] M. Bourdeau, et al., Modeling and forecasting building energy consumption: A review of data-driven techniques, Sustainable Cities Soc. 48 (2019) 101533.
[29] G.E. Karniadakis, et al., Physics-informed machine learning, Nat. Rev. Phys. 3 (6) (2021) 422–440.
[30] L. Di Natale, et al., Physically Consistent Neural Networks for building thermal modeling: Theory and analysis, Appl. Energy 325 (2022) 119806.
[31] A. Mirfin, X. Xiao, M.W. Jack, TOWST: A physics-informed statistical model for building energy consumption with solar gain, Appl. Energy 369 (2024) 123488.
[32] Y. Hu, et al., Times series forecasting for urban building energy consumption based on graph convolutional network, Appl. Energy (2022) 307.
[33] X. Wang, B. Dong, Physics-informed hierarchical data-driven predictive control for building HVAC systems to achieve energy and health nexus, Energy Build. 291 (2023).
[34] X. Wang, B. Dong, Long-term experimental evaluation and comparison of advanced controls for HVAC systems, Appl. Energy 371 (2024).
[35] T. Xiao, F. You, Building thermal modeling and model predictive control with physically consistent deep learning for decarbonization and energy optimization, Appl. Energy 342 (2023).
[36] T. Xiao, F. You, Physically consistent deep learning-based day-ahead energy dispatching and thermal comfort control for grid-interactive communities, Appl. Energy 353 (2024).
[37] B. Huang, J. Wang, Applications of physics-informed neural networks in power systems - a review, IEEE Trans. Power Syst. 38 (1) (2023) 572–588.
[38] B. Dong, et al., A hybrid model approach for forecasting future residential electricity consumption, Energy Build. 117 (2016) 341–351.
[39] C. Turner, M. Frankel, U. Council, Energy performance of LEED for new construction buildings, New Buildings Institute 4 (4) (2008) 1–42.
[40] S. Carlucci, et al., Modeling occupant behavior in buildings, Build. Environ. 174 (2020) 106768.
[41] C. Fan, et al., Personalized federated learning for cross-building energy knowledge sharing: Cost-effective strategies and model architectures, Appl. Energy 362 (2024) 123016.

[42] N. Forouzandeh, et al., Room energy demand and thermal comfort predictions in early stages of design based on the Machine Learning methods, Intell. Build. Int. 15 (1) (2023) 3–20.
[43] L. Zhang, et al., A review of machine learning in building load prediction, Appl. Energy 285 (2021) 116452.
[44] D. Mariano-Hernández, et al., A review of strategies for building energy management system: Model predictive control, demand side management, optimization, and fault detect & diagnosis, J. Build. Eng. 33 (2021) 101692.
[45] Y. Li, et al., Grey-box modeling and application for building energy simulations - A critical review, Renew Sustain Energy Rev. 146 (2021).
[46] P.G. Ellis, Development and Validation of the Unvented Trombe Wall Model in EnergyPlus, University of Illinois at Urbana-Champaign, 2003.
[47] Z. Yuan, et al., Combination of cuckoo search and wavelet neural network for midterm building energy forecast, Energy 202 (2020) 117728.
[48] White Box Technologies Weather Data. 2023; Available from: http://weather.whiteboxtechnologies.com/.
[49] G. Jiang, et al., A deep learning-based Bayesian framework for high-resolution calibration of building energy models, Energy Build. 114755 (2024).
[50] C. Deb, et al., A review on time series forecasting techniques for building energy consumption, Renewable Sustainable Energy Rev. 74 (2017) 902–924.
[51] Y.T. Chae, et al., Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings, Energy Build. 111 (2016) 184–194.
[52] P. Chujai, N. Kerdprasop, K. Kerdprasop, Time series analysis of household electric consumption with ARIMA and ARMA models. Proceedings of the International Multiconference Of Engineers And Computer Scientists, IAENG Hong Kong, 2013.
[53] S. Kapp, J.-K. Choi, T. Hong, Predicting industrial building energy consumption with statistical and machine-learning models informed by physical system parameters, Renewable Sustain. Energy Rev. 172 (2023) 113045.
[54] A. Petrucci, et al., Modelling of a multi-stage energy management control routine for energy demand forecasting, flexibility, and optimization of smart communities using a Recurrent Neural Network, Energy Convers. Manage. 268 (2022) 115995.
[55] B. Nepal, et al., Electricity load forecasting using clustering and ARIMA model for energy management in buildings, Japan Architect. Rev. 3 (1) (2019) 62–76.
[56] J. Zhuang, et al., Building cooling load prediction based on time series method and neural networks, Internat. J. Grid Distribut. Comput. 8 (4) (2015) 105–114.
[57] S. Karatasou, M. Santamouris, V. Geros, Modeling and predicting building's energy use with artificial neural networks: Methods and results, Energy Build. 38 (8) (2006) 949–958.
[58] H. Lu, et al., A multi-source transfer learning model based on LSTM and domain adaptation for building energy prediction, Int. J. Electr. Power Energy Syst. 149 (2023) 109024.
[59] J. Chen, et al., Stochastic simulation of occupant-driven energy use in a bottom-up residential building stock model, Appl. Energy 325 (2022) 119890.
[60] S. Pan, et al., Cluster analysis for occupant-behavior based electricity load patterns in buildings: A case study in Shanghai residences, Build. Simul. 10 (6) (2017) 889–898.
[61] I.U. Moffat, E.A. Akpan, White noise analysis: a measure of time series model adequacy, Appl. Math. 10 (11) (2019) 989.
[62] L.-Y. Chen, et al., Applicability of energy consumption prediction models in a department store: A case study, Case Stud. Therm. Eng. 49 (2023) 103380.
[63] F. Dinmohammadi, Y. Han, M. Shafiee, Predicting energy consumption in residential buildings using advanced machine learning algorithms, Energies 16 (9) (2023) 3748.
[64] W. Li, M. Wang, Office building energy consumption prediction based on LSTM and SVR. 2020 IEEE 3rd International Conference of Safe Production and Informatization (IICSPI), IEEE, 2020.
[65] R.A. Armstrong, Should Pearson's correlation coefficient be avoided? Ophthal. Physiol. Opt. 39 (5) (2019) 316–327.
[66] L.K. Norford, et al., Two-to-one discrepancy between measured and predicted performance of a 'low-energy' office building: insights from a reconciliation based on the DOE-2 model, Energy Build. 21 (2) (1994) 121–131.
[67] X. Chen, et al., A hybrid-model forecasting framework for reducing the building energy performance gap, Adv. Eng. Inf. 52 (2022) 101627.
[68] M. Elsaraiti, A. Merabet, A comparative analysis of the ARIMA and LSTM predictive models and their effectiveness for predicting wind speed, Energies 14 (20) (2021).
[69] T.J. Mbah, et al., Using LSTM and ARIMA to simulate and predict limestone price variations, Min Metall Explor 38 (2) (2021) 913–926.
[70] P.T. Yamak, L. Yujian, P.K. Gadosey, A Comparison between ARIMA, LSTM, and GRU for Time Series Forecasting, in: Proceedings of the 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence, 2019, pp. 49–55.
[71] C. Yan, et al., A novel air-conditioning system for proactive power demand response to smart grid, Energy Convers. Manage. 102 (2015) 239–246.
[72] B. Dong, et al., A review of smart building sensing system for better indoor environment control, Energy Build. 199 (2019) 29–46.
[73] Y. Chen, et al., Measures to improve energy demand flexibility in buildings for demand response (DR): A review, Energy Build. 177 (2018) 125–139.
[74] J. Chen, et al., A review of computing-based automated fault detection and diagnosis of heating, ventilation and air conditioning systems, Renewable Sustain. Energy Rev. 161 (2022) 112395.
[75] X. Li, J. Wen, Review of building energy modeling for control and operation, Renewable Sustain. Energy Rev. 37 (2014) 517–537.