Build Simul
https://doi.org/10.1007/s12273-026-1411-6

A VRF zonal thermal load prediction method based on transfer learning

Junyu Chen[1], Peng Xu[1] (✉), Yi Zhu[1], Jiefan Gu[2], Kan Chen[3], Yunxiao Ding[3], Leqi Zhu[3], Renrong Ding[3]

1. School of Mechanical Engineering, Tongji University, Shanghai 201804, China
2. College of Architecture and Urban Planning, Tongji University, Shanghai 200092, China
3. GD Midea Heating & Ventilating Equipment Co., Ltd., Foshan 528311, China

Research Article

Abstract

In the context of global efforts to enhance building energy efficiency, variable refrigerant flow (VRF) systems are recognized for their high performance and flexible control, achieving widespread adoption, particularly in East Asia. This proliferation necessitates accurate thermal load prediction, which is essential for right-sizing systems and meeting performance guarantees. However, conventional whole-building forecasting fails at the outdoor unit (ODU) zone level because it overlooks the distinct zonal characteristics and the granular, user-driven operational dynamics that primarily govern the load. This paper addresses this critical gap by introducing a physics-guided transfer learning framework centered on a novel methodology for creating a high-fidelity, physics-based source domain. The methodology first empirically establishes the cooling-capacity-weighted indoor unit (IDU) activation ratio as a key determinant, and then develops a dynamic psychrometric blending method to integrate this metric into EnergyPlus. This physics-guided simulation approach enables the creation of a large-scale simulation database. Building on this foundation, a Long short-term memory (LSTM) network is pre-trained to learn general thermal principles, and a transfer learning strategy is then used to adapt this knowledge to data-scarce, real-world scenarios. The framework's efficacy was demonstrated through three distinct transfer strategies that systematically evaluated its performance using non-target data (for zero-shot prediction), limited target-specific data, and a hybrid of both. All strategies markedly outperformed a model pre-trained solely on simulation data, with the optimal hybrid strategy achieving a final R^2 of 0.866 and reducing the mean absolute error (MAE) by 18.9%. This approach demonstrates a promising pathway toward reliable prediction for ODU zones, offering valuable support for more efficient VRF system design and operation.

Keywords

zonal load prediction
variable refrigerant flow (VRF) system
transfer learning
EnergyPlus

Article History

Received: 06 August 2025
Revised: 04 December 2025
Accepted: 29 December 2025

Advances in Modeling and Simulation Tools

1 Introduction

The building sector consumes about 30% of global energy and 60% of global electricity (UNEP 2014; IEA 2023a), with heating, ventilation, and air conditioning (HVAC) systems being a major contributor (Lee et al. 2015). Accurate thermal load forecasting during the design phase is therefore essential, as it directly informs HVAC system sizing (Abdolvand et al. 2024). Proper sizing reduces energy consumption, avoids excessive capital investment, and enhances both operational and economic efficiency (Pan et al. 2023). Additionally, HVAC manufacturers are increasingly expected to provide predictive energy performance guarantees to clients, further elevating the need for reliable load estimation (IEA 2023b).

Among modern HVAC technologies, variable refrigerant flow (VRF) systems have gained widespread adoption due to their high energy efficiency and flexible zone-level control (Lin et al. 2015; Wang et al. 2024), now comprising approximately 50% of the central air-conditioning market in China (Electromechanical Information 2021). Achieving this promised efficiency fundamentally depends on accurate load prediction at the outdoor unit (ODU) zone level. This necessity is amplified by intense market pressure on manufacturers to meet energy performance guarantees with minimal budgets. However, this task presents unique

E-mail: xupeng@tongji.edu.cn

challenges not found in conventional systems. Traditional building-level prediction methods are rendered inadequate by their inability to account for significant variations in zonal parameters. More importantly, they fail to capture the specific operational patterns within each zone that ultimately dictate the thermal load, leading to impractical predictions.

This paper posits that this pattern is best quantified by a crucial yet often overlooked metric termed the hourly cooling-capacity-weighted activation ratio of indoor units (IDUs). This metric captures the inherent, user-driven variability that poses the primary challenge to zonal-level forecasting for both cost-effective system design and energy-efficient operation. However, translating this dynamic, real-world user behavior into a physics-based simulation framework like EnergyPlus presents a significant methodological challenge, as conventional setpoint schedules cannot adequately represent the partially activated thermal conditions of a zone. This study addresses this core challenge by developing a novel dynamic psychrometric blending approach, which is described in detail in the methodology section.

While several studies have begun to explore zonal information to enhance building-level prediction—e.g., Hu et al. (2021) incorporated zonal temperature data into artificial neural networks (ANNs) to enhance whole-building accuracy, and Machado et al. (2023) utilized EnergyPlus-generated metamodels for training—these approaches primarily use zonal data as a means to refine a larger-scale forecast. In contrast, dedicated efforts to predict VRF ODU zonal loads as the primary target remain scarce. This reveals a critical research gap: accurate and scalable thermal load forecasting at the ODU zone level for VRF systems has not been adequately addressed, despite its importance for system selection and energy-efficient design.

To address this gap, it is essential to first evaluate the suitability of existing building load forecasting approaches for this specific zonal-level task. Current building load forecasting approaches can be broadly classified into physics-based white-box models, data-driven black-box models, and hybrid gray-box methods (Li and Wen 2014; Wang et al. 2020). White-box models, such as EnergyPlus and DOE-2, simulate building energy dynamics using detailed physical parameters (e.g., envelope properties, occupancy schedules, weather data). While these tools are widely adopted for whole-building simulations, their accuracy heavily depends on precise input parameters and modeling expertise. Even minor discrepancies in building geometry or material properties can lead to significant prediction errors (Imam et al. 2017). Gray-box methods simplify physical processes, such as resistor-capacitor (RC) thermal networks (Wang and Xu 2006; Blum et al. 2019), offering computational efficiency but often lacking behavioral realism. The rise of smart buildings and Internet of things (IoT) technologies has enabled data-driven black-box models to gain prominence. Black-box approaches, including statistical learning (e.g., linear regression (Forrester and Wepfer 1984), autoregressive integrated moving average (ARIMA) (Leiprecht et al. 2021)) and machine learning models (e.g., support vector machine (SVM) (Li et al. 2009), artificial neural network (ANN) (Gunay et al. 2017), extreme learning machine (ELM) (Gao et al. 2022) and light gradient boosting machine (LightGBM) (Ke et al. 2017)) extract patterns from data without requiring explicit physical knowledge.

More recently, the performance of black-box models has been significantly propelled by deep learning. Deep learning architectures, characterized by multiple layers, excel at both autonomously learning hierarchical features from raw data and capturing the complex, long-term temporal dependencies crucial for time-series forecasting. Long short-term memory (LSTM) networks have demonstrated strong performance in modeling building load time series (He et al. 2019). Sendra-Arranz and Gutiérrez (2020) used LSTM to achieve high-accuracy, day-ahead HVAC load prediction for demand-side management, while Gao et al. (2024) proposed a BAS-optimized GRNN&LSTM hybrid model that improved predictive robustness. However, such models require substantial historical data, which may be unavailable in many real-world scenarios.

To address data scarcity, transfer learning has emerged as a viable strategy to adapt knowledge from data-rich source domains to data-limited targets. Pinto et al. (2022) demonstrated its potential in adapting pre-trained building energy models to new contexts, highlighting that the success of transfer learning hinges on the quality and relevance of the source-domain data. This dependency presents a significant hurdle for VRF systems. For instance, while Park et al. (2025) successfully developed a transfer learning framework for VRF energy prediction, the challenge of acquiring or generating a high-fidelity source domain that captures dynamic, user-driven zonal activation persists, especially for new buildings in the design phase where no historical data exists. The work of Zhou et al. (2020) affirms the viability of using synthetic data to overcome scarcity. Their use of a BiGAN to generate data directly inspired our strategy of creating a large-scale synthetic dataset for pre-training. However, purely data-driven generative models may lack the physical consistency and interpretability required for robust engineering design. These limitations reveal a critical need for a method capable of generating a large-scale, physically-grounded synthetic dataset that accurately models the unique user-driven dynamics of ODU zones.

Despite the demonstrated potential of these advanced

methods, their application to ODU zonal load prediction in VRF systems remains scarce. This specific research gap persists primarily due to the critical challenge of data availability. While obtaining sufficient measured thermal load data is a common hurdle for any building-level forecasting, the problem is exacerbated at the ODU zone level. This data scarcity has historically hindered the development of dedicated models and thus limited support for crucial design decisions like outdoor unit selection. However, this challenge also presents an opportunity for innovative data generation methodologies. Fortunately, modern VRF systems increasingly support cloud-connected data acquisition, enabling the retrieval of zone-level activation ratios and load traces. These data streams provide a viable basis for accurate zonal forecasting. While still limited in scale, this emerging real-world data provides a vital foundation—not for direct model training, but for calibrating and validating physics-based simulation frameworks capable of generating the large-scale datasets required for advanced modeling.

This study addresses the aforementioned limitations by proposing a novel physics-guided, transfer learning–based framework for VRF ODU zonal load prediction. The core of this framework is a systematic methodology for overcoming data scarcity. It involves creating a large-scale, physically realistic synthetic dataset for pre-training, which is then adapted to real-world scenarios via transfer learning. This hybrid approach combines the scalability of simulation with the predictive strength of deep learning, demonstrating a viable method for achieving accurate load prediction to support better-informed VRF system design and operation.

2 Methodology

To bridge the gap between the need for large-scale data and its practical scarcity, the proposed methodology integrates physics-based simulation with data-driven learning through a transfer learning approach. The framework is structured into the following key stages, as illustrated in Figure 1. First, the methodology defines the core challenge by identifying and quantifying the crucial driver of ODU zonal load—the user-driven activation ratio. The second stage presents a novel white-box modeling approach, centered around dynamic psychrometric blending method, to generate a high-fidelity source domain database that accurately reflects this driver. Third, a deep learning framework, using an LSTM network, is pre-trained on this database to learn generalized physical patterns. Finally, transfer learning strategies are employed to adapt this pre-trained knowledge to real-world, data-scarce scenarios. This structured approach systematically bridges the simulation-to-reality gap, yielding a robust and accurate prediction model.

2.1 White-box framework for simulating activation-driven zonal loads

The foundation of the entire transfer learning approach is

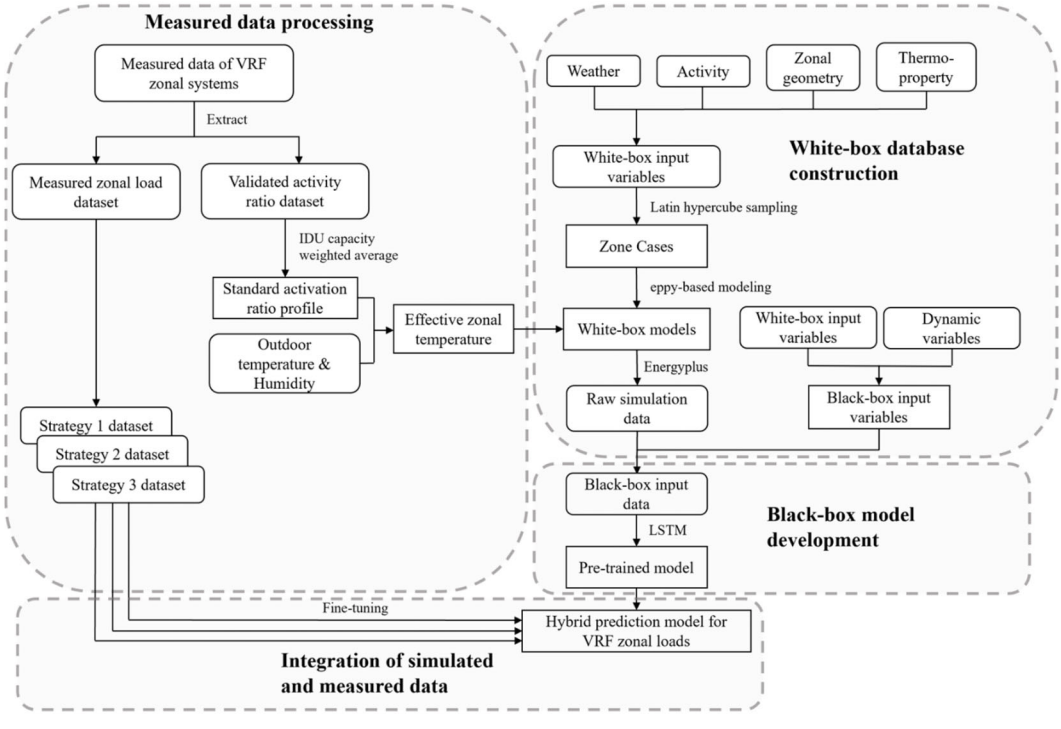


Fig. 1 Transfer learning framework for ODU zonal load prediction

a robust and diverse source domain database. This section details the development of a comprehensive white-box framework, from identifying the key load determinants in VRF ODU zones to generating a 5,000-sample synthetic database that accurately captures these dynamics.

2.1.1 The IDU activation ratio as a key load determinant

Unlike building-level models where occupancy is often generalized, the load in an ODU zone is directly and dynamically governed by the operational behavior of its own individual indoor units (IDUs). The common scenario of partial and intermittent IDU activation means that the effective zonal load fluctuates significantly based on user schedules (Figure 2). Therefore, quantifying this behavior is the primary challenge and the most critical component for accurate zonal load prediction.

In response to this challenge, this study posits that the hourly activation ratio—defined as the cooling-capacity-weighted proportion of active IDUs—is a powerful predictor of instantaneous thermal demand. The hourly activation ratio α_t, is formally defined as:

$$\alpha_t = \frac{\sum_{i=1}^{n} \left(S_{i,t} \cdot C_{\text{cool},i} \right)}{\sum_{i=1}^{n} C_{\text{cool},i}} \tag{1}$$

where $S_{i,t}$ is the binary status (0 = off, 1 = on) of the i-th IDU and $C_{\text{cool},i}$ is its nominal cooling capacity.

To validate this foundational hypothesis, empirical operational data was analyzed from three distinct ODU zones within a real-world office building. The analysis revealed a consistently strong and statistically significant positive correlation between the hourly activation ratio and the measured zonal cooling load. As shown in Figure 3, the Pearson correlation coefficients were 0.958, 0.954, and 0.935 for the three test cases, respectively. This empirical evidence confirms that the activation ratio is not merely an influencing factor but a dominant driver of the thermal load.

Consequently, any high-fidelity model for ODU zones must be built around a robust mechanism to handle these crucial behavioral patterns. The following section details the framework developed to achieve this by first creating standardized activation profiles.

2.1.2 Parameterizing activation dynamics: standardized profiles

While the activation ratio is a primary driver of zonal load, its raw, hour-by-hour form is inherently high-dimensional and stochastic. Directly using such noisy, day-to-day data as an input for a predictive model would require an impractically

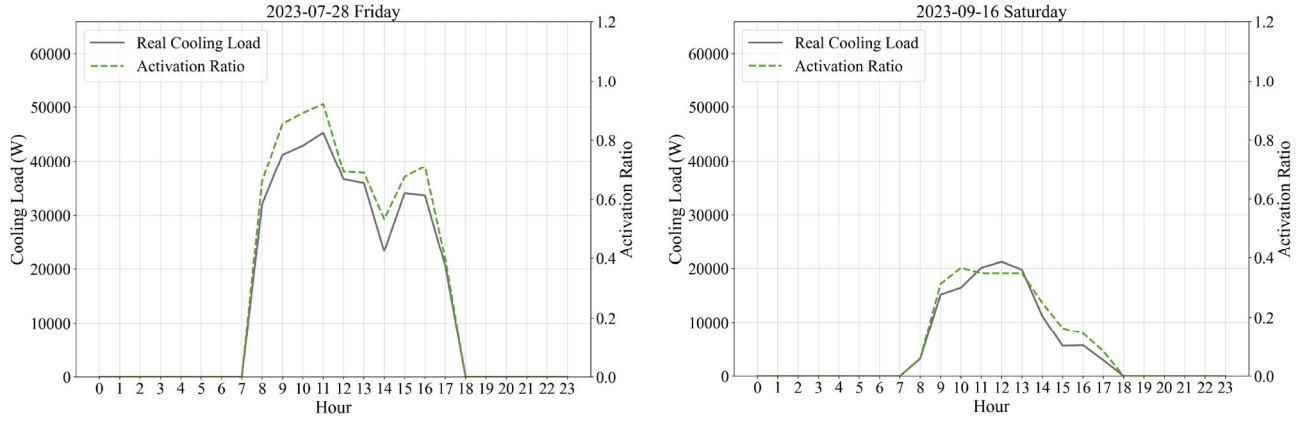


Fig. 2 Hourly activation ratio and corresponding zonal load

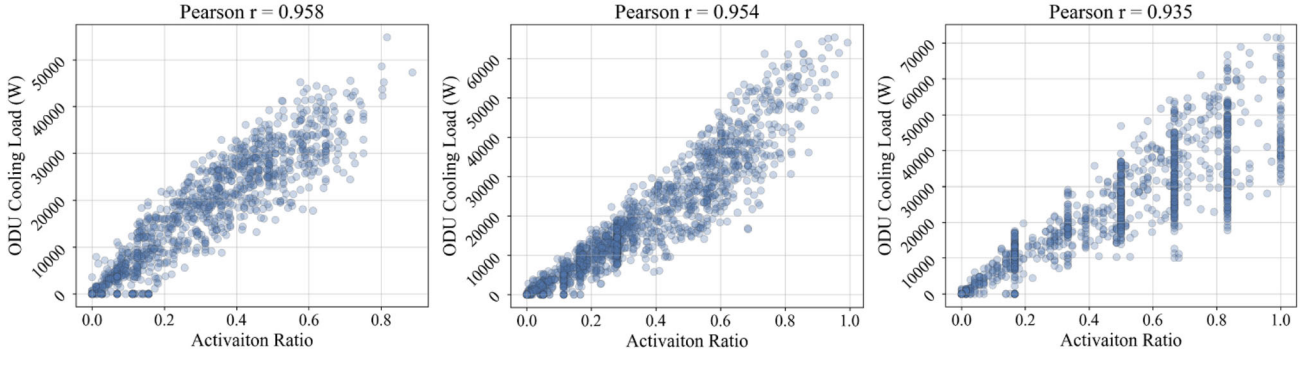


Fig. 3 Correlation between activation ratio and measured load

large number of variables and a massive training dataset to learn the underlying patterns.

To create a more efficient and generalizable modeling framework, this complex behavior is first distilled into a set of standardized, parameterized profiles. These profiles serve as reusable "archetypes" of user behavior. This approach offers several key advantages: for the white-box simulation, it provides a structured way to generate diverse yet realistic operational schedules; for the black-box model, it transforms a complex time series into a simple, low-dimensional feature set, enhancing learning efficiency; and for practical application, it offers designers a streamlined and realistic way to define operational scenarios.

Based on the previously introduced hourly activation ratio α_t, measured data from 21 ODU zones was processed to derive the standardized profiles. To account for significant operational differences across a typical week, three characteristic 24-hour daily profiles were defined:

- Weekday ratios: to represent typical operational patterns from Monday to Friday.
- Weekend ratios: to capture reduced activation levels observed on Saturdays and Sundays.
- Peak ratios: to represent extreme conditions, derived from the three weekdays with the highest loads.

To illustrate these distinct patterns, Figure 4 presents the extracted Weekday, Weekend, and Peak activation profiles for 3 representative, normally operating ODU zones. This visual example clarifies the typical temporal signatures that the framework aims to capture.

This classification framework was applied to field-measured data from 21 initial ODU zones. For each zone, the three types of daily profiles were extracted. A quality control analysis was then performed on these extracted profiles. This process identified seven ODU zones corresponding to persistently underutilized spaces (e.g., reserve conference rooms, auxiliary functional areas). For these zones, the bound IDUs remained predominantly dormant, rendering their derived daily activation profiles statistically insignificant and flat (Figure 5). Consequently, based on the criterion of establishing reliable and dynamic daily profiles, these seven non-representative cases were excluded from the aggregation step.

The preprocessing thus retained 14 ODU zones with valid and meaningful activation information. The final, standardized activation profiles for office buildings were then derived through a weighted aggregation of the corresponding curves (Weekday, Weekend, and Peak) from these 14 validated zones (Figure 6). The weighting factor for each ODU was proportional to the total cooling capacity of its associated IDUs. This aggregation process distills the collective user behavior into three standard profiles, serving as the foundational schedules for the subsequent white-box

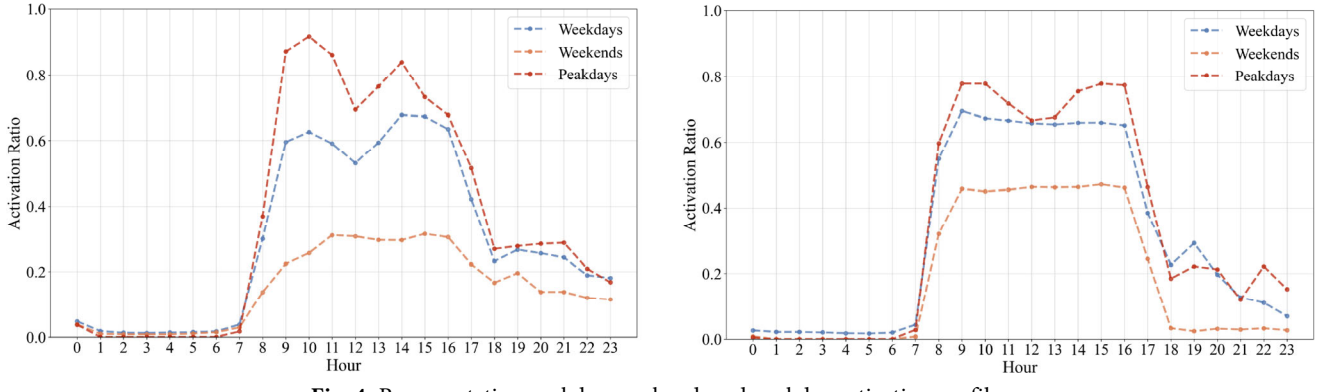


Fig. 4 Representative weekday, weekend, and peakday activation profiles

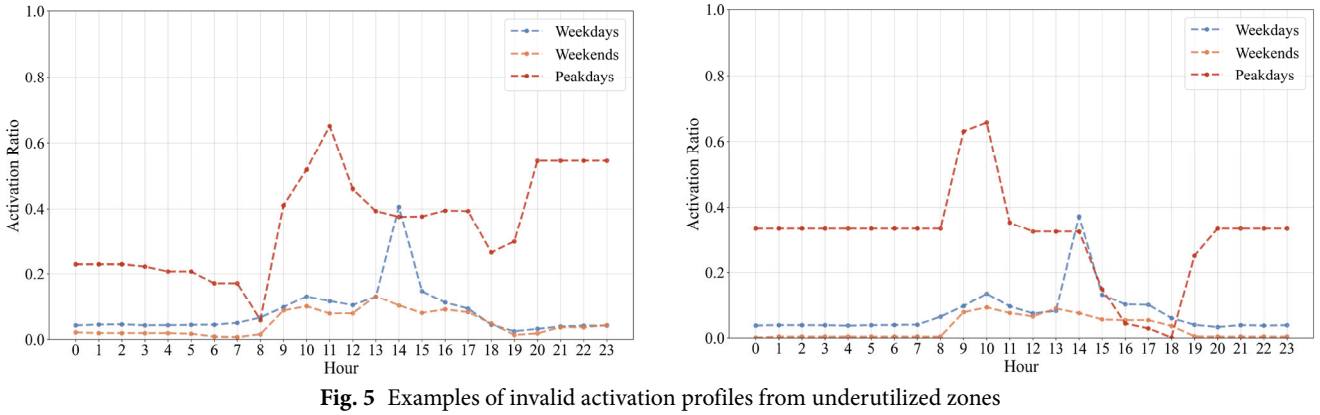


Fig. 5 Examples of invalid activation profiles from underutilized zones

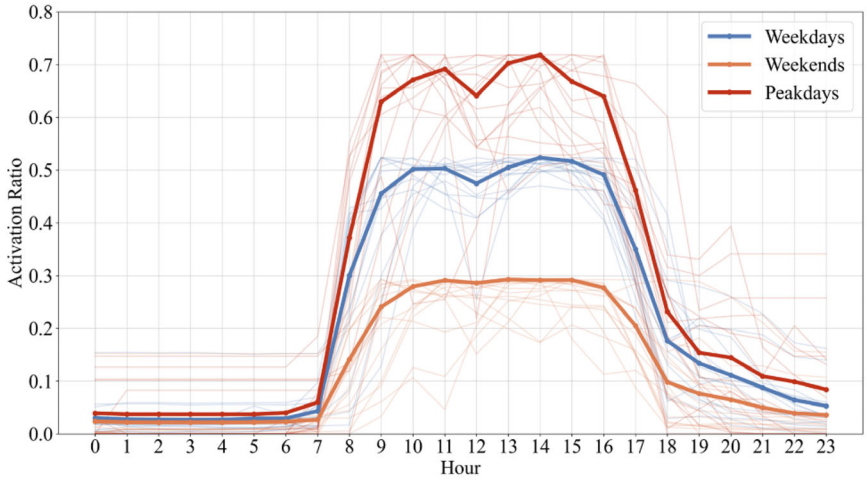


Fig. 6 Aggregated activation profiles and source distribution

simulations. Figure 6 visually presents these archetypes against the distribution of the 14 source profiles. To enhance visual comparability of their temporal shapes, the original profiles for each day type were scaled to match the peak value of the final corresponding aggregated standard profile.

2.1.3 Dynamic zone condition modeling via psychrometric blending

With the activation ratio profiles established, the central challenge becomes implementing the dynamic activation ratio α_t within the EnergyPlus simulation. Conventional fixed temperature setpoints are incapable of representing the blended air conditions of a partially activated ODU zone, especially because determining both the optimal number of setpoint segments and their respective temperature values presents significant challenges. Moreover, a fixed set of temperature setpoints is inadequate when confronted with diverse outdoor weather conditions—particularly in VRF systems where activation ratios αt substantially influence system performance.

To address these limitations, this study develops a novel dynamic psychrometric blending method. This method applies thermodynamic principles to calculate the effective zone conditions by treating the zone air as a mixture of conditioned indoor air (at the state defined by setpoints) and unconditioned outdoor air, weighted by α_t. Moving beyond simplistic linear interpolation approaches, the proposed methodology rigorously determines effective zone conditions through mass and energy conserving air blending.

Humidity ratio blending maintains water vapor mass balance:

$$W_{\text{zone}} = \alpha_t \cdot W_{\text{indoor}} + (1-\alpha_t) \cdot W_{\text{outdoor}} \tag{2}$$

Enthalpy blending satisfies energy conservation:

$$h_{\text{zone}} = \alpha_t \cdot h_{\text{indoor}} + (1-\alpha_t) \cdot h_{\text{outdoor}} \tag{3}$$

Derived effective zonal temperature is calculated using thermodynamic relationships:

$$T_{\text{zone}} = \frac{h_{\text{zone}} - h_{\text{fg}} \cdot W_{\text{zone}}}{c_{\text{p,a}} + c_{\text{p,v}} \cdot W_{\text{zone}}} \tag{4}$$

where $c_{\text{p,a}} \approx 1.006$ kJ/(kg·K) (dry air specific heat), $c_{\text{p,v}} \approx 1.86$ kJ/(kg·K) (water vapor specific heat), and $h_{\text{fg}} \approx 2501$ kJ/kg (latent heat of vaporization).

Relative humidity is computed as a state function:

$$\text{RH}_{\text{zone}} = \frac{p_{\text{v}}}{p_{\text{sat}}(T_{\text{zone}})} \times 100\% \tag{5}$$

where p_{v} represents vapor pressure derived from W_{zone} and atmospheric pressure.

As shown in Figure 7, this approach realistically represents blended air conditions in partially activated zones, thereby determining the effective zonal temperature for the simulation. Compared to conventional fixed-setpoint simulations, the proposed framework can provide effective temperature setpoints for zonal load calculations through its adaptive mixing algorithm with outdoor parameters. Crucially, this algorithm generates physically reasonable indoor air setpoints across diverse outdoor air conditions. In essence, this method provides a significant leap in physical realism compared to conventional static or multi-step setpoint approaches. By dynamically coupling the indoor state with both external weather conditions and internal activation patterns based on first principles, this method ensures that the generated simulation data foundation is not only diverse but also physically robust across a wide range of weather conditions.

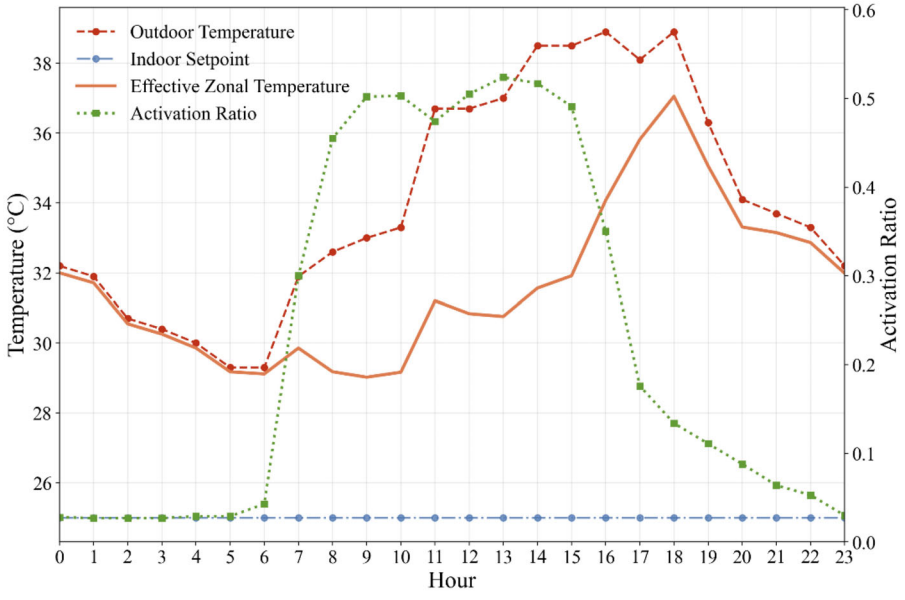


Fig. 7 Graph illustrating the dynamic psychrometric blending method

2.1.4 Parameter space definition and automated database generation

With the critical activation ratio profiles empirically derived and the simulation method defined, the next step is to define the complete and structured set of input variables that will drive the large-scale simulation process. This comprehensive variable set serves a dual purpose: first, to programmatically configure the 5,000 white-box simulations in EnergyPlus, and second, to form the foundational static features for the subsequent LSTM prediction model.

The variables for the white-box simulations are organized into the following categories, defining the physical, operational, and environmental context for each case:

Activation schedule parameters

The empirically derived profiles (Weekday, Weekend, Peakday) maintain fixed temporal shapes. Their intensity and timing are controlled by five key parameters that allow for programmatic variation in the simulations:

- WDR, WER, and PKR: these scalar parameters serve as amplitude scaling factors for the Weekday, Weekend, and Peakday profiles, respectively, enabling adjustment of the overall activation strength, with the Peakday profile being applied to the three weekdays with the highest outdoor temperatures.
- OPN (activation onset time) and SHT (shutdown time): these parameters regulate the daily operational window, with dynamic ramp-up logic for OPN and a direct cutoff for SHT.

Together, these five parameters provide a concise yet powerful mechanism to represent a wide variety of operational schedules.

Geographic and weather parameters

To ensure simulations reflect regional climatic influences, this category includes:

- Location and latitude: the location variable is an encoded integer representing one of seven typical climate zones across China, specifically Beijing, Xi'an, Shanghai, Wuhan, Chongqing, Kunming, and Guangzhou. The corresponding latitude for each location governs the solar geometry.
- Weather data: hourly data from corresponding EPW files serve as the primary drivers of the simulation.

Zonal geometry and thermophysical properties

These parameters define the physical characteristics of the simulated zone:

- Geometric properties: these define the zone's 3D shape, orientation, and exposure. They include fundamental dimensions (storey height (*H*), horizontal length (*X*), vertical length (*Y*)), building rotation (orientation offset (AZI)), facade exposure ratios (EOWR, SOWR, WOWR, NOWR), glazing ratios (EWWR, SWWR, WWWR, NWWR), and vertical position flags for roof/ground contact (is top floor (TOP), is bottom floor (BOT)).
- Thermophysical properties: these govern the envelope's thermal performance. They include solar radiation interaction properties (solar heat gain coefficient (SHGC), wall/roof solar absorption (WSA/RSA), internal shading transmittance (ST)), insulation performance via *U*-values for walls, windows, roof, and floor (WALLU, WINU, RU, FU), and thermal inertia (wall specific heat (WSP)).

Internal gains and operational parameters

These variables represent internal heat generation and control targets:

- Internal gains: loads from occupants (occupant density (OPD)) and plug loads (equipment power density (EQP)) are dynamically linked to the activation ratio (α_t). Lighting power density (LGT) is modeled as a constant during operational hours.
- Air exchange: outdoor air exchange is quantified by the air infiltration rate (INFIL).
- Operational setpoints: the cooling setpoint temperature (SPC) and heating setpoint temperature (SPH) define the target indoor air state for the psychrometric blending method.

To generate the diverse database, this study employs Latin hypercube sampling (LHS) to create 5,000 unique variable combinations, ensuring uniform coverage of the parameter space (Ding et al. 2024). The complete list of these parameters and their respective sampling ranges is detailed in Table 1. An automated workflow using Python and the Eppy library was then created to batch-generate and execute the 5,000 annual EnergyPlus simulations. The output—a comprehensive database of hourly zonal thermal loads—forms the foundational dataset for the data-driven modeling stages.

2.2 Deep learning framework for zonal load prediction

With the high-fidelity white-box database established, the framework transitions to data-driven modeling. This section describes the deep learning model used to learn thermal patterns from the data and the pre-training process.

2.2.1 Long short-term memory (LSTM) networks

The data-driven modeling process is built upon a long short-term memory (LSTM) network, selected for its proficiency in time-series forecasting. Neural network models are broadly categorized into feedforward neural networks (FNNs) and recurrent neural networks (RNNs). FNNs propagate information unidirectionally from the input layer to the output layer, exhibiting no inherent memory of past inputs. In contrast, RNNs incorporate recurrent connections through feedback loops within their neuron units, endowing them with internal memory and enabling information sharing across sequential time steps (Somu et al. 2020). This architecture makes RNNs well-suited for sequential data processing.

However, conventional RNNs suffer from difficulties in learning long-term dependencies within sequences, often manifesting as the vanishing or exploding gradient problem due to the iterative propagation of local errors through many time steps. To overcome these critical limitations inherent in basic RNNs, Hochreiter and Schmidhuber (1997) introduced a novel architecture termed LSTM, which builds upon the recurrent framework by incorporating specialized units and gating mechanisms. Specifically, LSTM utilizes memory cells, which are self-connected units designed to maintain state

Table 1 Variables and value ranges for LHS

Variable	Abbreviation	Unit	Value range	Variable	Abbreviation	Unit	Value range
Storey height	H	m	2–8	Internal shading transmittance	ST	—	0.1–1
Horizontal length	X	m	1–50	Solar heat gain coefficient (glazing)	SHGC	—	0.1–0.9
Vertical length	Y	m	1–50	Solar absorption coefficient (walls)	WSA	—	0.1–0.9
Orientation offset	AZI	°	−45 to 45	Solar absorption coefficient (roof)	RSA	—	0.1–0.9
East exterior wall ratio	EOWR	—	0–1	Wall heat transfer coefficient	WALLU	W/(m²·K)	0.2–2
South exterior wall ratio	SOWR	—	0–1	Wall specific heat	WSP	J/(kg·K)	800–2000
West exterior wall ratio	WOWR	—	0–1	Window heat transfer coefficient	WINU	W/(m²·K)	0.5–6
North exterior wall ratio	NOWR	—	0–1	Roof heat transfer coefficient	RU	W/(m²·K)	0.3–2
East window-wall ratio	EWWR	—	0–1	Floor heat transfer coefficient	FU	W/(m²·K)	0.3–2
South window-wall ratio	SWWR	—	0–1	Peak day maximum ratio	PKR	—	0.5–1
West window-wall ratio	WWWR	—	0–1	Weekday maximum ratio	WDR	—	0.3–1
North window-wall ratio	NWWR	—	0–1	Weekend maximum ratio	WER	—	0–1
Cooling setpoint temperature	SPC	°C	18–28	Location	LOC	—	0 to 6
Heating setpoint temperature	SPH	°C	18–24	Latitude	LAT	°	—
Occupant density	OPD	person/m²	0–0.5	Is top floor	TOP	—	0 or 1
Equipment power density	EQP	W/m²	0–25	Is bottom floor	BOT	—	0 or 1
Lighting power density	LGT	W/m²	0–20	Startup time	OPN	—	6, 7, 8, 9
Air infiltration rate	INFIL	time/h	0.5–3	Shutdown time	SHT	—	17, 18, 19, 20

information over extended durations, thereby facilitating the capture of long-range temporal dependencies. Crucially, the flow of information into, out of, and within these memory cells is regulated by three adaptive gating units: the input gate (controlling the writing of new information into the cell), the forget gate (controlling the erasure of outdated information from the cell), and the output gate (controlling the reading of information from the cell state to the hidden output). This gated structure effectively mitigates the vanishing/exploding gradient issues by providing regulated pathways for gradient flow during backpropagation, while the memory cells allow for persistent storage of relevant context. The detailed internal structure of an LSTM cell is illustrated in Figure 8.

Due to these architectural advantages, LSTM is selected as the core learning algorithm for our time-series forecasting task. Its ability to capture long-term temporal dependencies (spanning hours or days) addresses critical load persistence effects and diurnal patterns. Furthermore, the adaptive gating mechanism (input, forget, output gates) enables the model to selectively retain relevant historical states while integrating new inputs and filtering transient noise inherent in operational data. This combination allows LSTMs to effectively model the sequential, interconnected processes driving HVAC loads with greater accuracy than static methods or basic RNNs (Srivastava and Lessmann 2018). Therefore, this study adopts LSTM as the core modeling approach to leverage its unique advantages.

2.2.2 *Input feature engineering for the LSTM model*

To ensure the LSTM network has a complete understanding of each case, its input feature set is designed to be comprehensive. It inherits all static input variables used in the white-box simulations (as detailed in Section 2.1.4 and Table 1), which define the unique physical, geographical, and operational context of each zone. In addition to these static context variables, the feature set includes two categories of dynamic inputs:

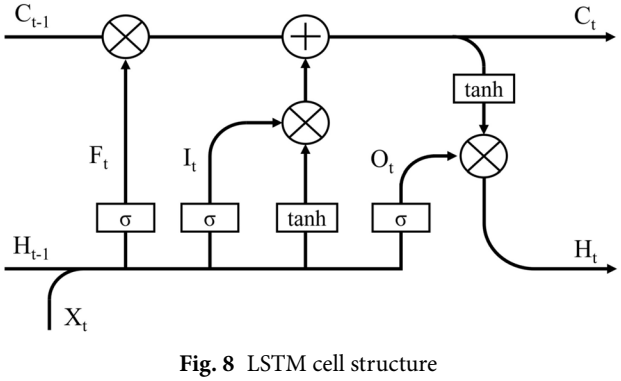


Fig. 8 LSTM cell structure

- Weather variables: this category comprises the four principal hourly weather variables that were used as inputs for the EnergyPlus simulations. These variables serve as the direct physical drivers of thermal load at each time step for the LSTM model, namely dry-bulb temperature (DryT), relative humidity (RH), direct normal irradiance (DNI), and Diffuse horizontal irradiance (DHI).
- Temporal variables: this category consists of a set of engineered indices created exclusively to provide a complete dynamic context for the LSTM network, enabling it to recognize cyclical patterns and operational states that are not explicitly represented by the physical weather data. These features are further divided into two types. The first type, cyclic features, captures natural seasonal and diurnal rhythms through variables such as month in year (MinY), day in month (DinM), day in week (DinW), and hour in day (HinD). The second type, binary classifiers, explicitly defines the operational context of each day. These classifiers, including Is weekday (Weekday) and Is peakday (Peakday), indicate which of the three standardized activation profiles was used for a given day, thus allowing the model to directly associate input conditions with the distinct load patterns characteristic of each day type.

2.2.3 *Pre-training on white-box database*

The first stage of the data-driven modeling involves training the LSTM network on the synthetic data generated in Section 2.1. This pre-training endows the model with a generalized understanding of building thermal dynamics across a wide range of designs and conditions.

Prior to training, the input features were prepared for the model. For each zone's time series, the static features were repeated at every time step and concatenated with the corresponding dynamic inputs to form a complete feature vector. Given the LSTM's inherent robustness to multicollinearity, this physically-grounded feature set was used directly. Subsequently, all input features were normalized to a range of 0 to 1 using min-max scaling to ensure consistent feature scaling and improve model convergence.

The LSTM model architecture comprises two hidden layers, each with 64 units, followed by a fully connected output layer. The model employs hyperbolic tangent (tanh) activation functions in the recurrent units to capture nonlinear temporal dynamics, while a linear activation function is used in the final output layer to predict the thermal load intensity. Model optimization is performed using the Adam algorithm with an initial learning rate of 0.002, coupled with an exponential decay rate of 0.99 to ensure gradual convergence. To improve training stability and prevent

gradient explosion, gradient clipping is applied by constraining the global norm of the gradients to a threshold of 3.0.

The model is trained for 1,500 epochs using a batch size of 64. To ensure generalization and mitigate overfitting to synthetic patterns, the entire dataset of 5,000 ODU zone samples is randomly partitioned into training, validation, and test subsets following an 8:1:1 split ratio.

A load-scaled weighted mean absolute error (WMAE) is adopted as the loss function:

$$L_{\mathrm{WMAE}} = \frac{1}{N}\sum_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{|y_i| + \varepsilon} \tag{6}$$

where y_i and $\hat{y}_i$ are the actual and predicted thermal loads for sample, respectively.

This formulation normalizes the absolute prediction error of each sample by its current actual load value, functioning similarly to a mean absolute percentage error (MAPE). However, the inclusion of a small constant (ε = 0.001) prevents division-by-zero issues when the actual load is near or at zero. This ensures that the model provides balanced attention across all load conditions, from low to high, avoiding the tendency of traditional MAE to be dominated by high-load scenarios. As a result, the pre-trained model learns more generalized and robust thermal dynamics applicable across a wide variety of zone configurations.

2.3 Knowledge adaptation via transfer learning

The final stage of the framework addresses the "simulation-to-reality" gap by adapting the pre-trained model to real-world operational data using transfer learning. This process aligns the generalized model with the specific nuances of a real-world environment.

2.3.1 Principles of transfer learning in zonal load prediction

Transfer learning is a powerful technique for adapting the knowledge from a pre-trained model to a new, related task, thereby addressing critical challenges related to data scarcity and computational constraints in modeling applications. Its core principle involves identifying and leveraging underlying similarities between a data-rich source domain and a data-limited target domain to facilitate knowledge transfer (Yan et al. 2023). This approach effectively bridges the gap between the need for extensive labeled datasets and the practical difficulties in acquiring sufficient high-quality real-world data, making it invaluable where such data is challenging or expensive to obtain (e.g., in building energy monitoring).

In this study, transfer learning is applied to fuse simulation outputs with operational measurements for HVAC load modeling. Specifically, the source domain consists of a comprehensive dataset generated via EnergyPlus white-box simulations, providing well-characterized synthetic building thermal load profiles.The target domain consists of measured ODU zonal load data from operational systems.

To operationalize this within the proposed framework, the LSTM network described previously is first pretrained on the abundant synthetic data to learn fundamental temporal patterns and load dynamics. Subsequently, the pretrained model undergoes domain adaptation by fine-tuning on limited real-world measurements. This process allows the model to retain the generalized thermal load patterns learned from the broad and diverse simulation dataset, while gradually adapting to the specific operational behaviors and data characteristics of the target domain. As a result, prediction performance on real-world data is improved without discarding the foundational knowledge acquired from simulations. This approach combines the scalability and controllability of simulation data with the realism and reliability of measured data, enabling effective and efficient cross-domain knowledge transfer.

2.3.2 Fine-tuning strategies and experimental setup

To transfer the pre-trained knowledge while preserving its robust temporal feature extraction capabilities, the recurrent layers of the LSTM are frozen during fine-tuning. This strategy is based on the principle that the recurrent layers act as feature extractors for general physical and temporal patterns. Learning these complex, universal features requires the vast and diverse dataset provided by the simulations. In contrast, the final fully connected layer primarily serves to map these extracted features to the specific load magnitudes of the target zone, a simpler task for which the limited measured data is sufficient. This design therefore restricts weight updates to only this final layer, allowing it to adapt to the target's characteristics while minimizing the risk of overfitting. A reduced learning rate of 0.0001 and the mean squared error (MSE) loss function are employed during this stage.

To validate the proposed framework, a case study was conducted on a multi-storey office building located in Hangzhou, China (30.3°N, 120.2°E). This region is characterized by a hot-summer-cold-winter climate, comparable to ASHRAE Climate Zone 3A. The focus of this study is a single floor within this building, which is served by four distinct outdoor units (ODUs), designated as ODU1 through ODU4. As shown in Figure 9, these four ODU zones possess unique geometric attributes, orientations, and internal load profiles, representing a realistic and diverse set of operational conditions. The specific properties of these zones are detailed in Table 2. This multi-zone setup provides an ideal testbed for evaluating different data utilization strategies.

Table 2 Properties of the target ODU zones

ODU zone	*X* (m)	*Y* (m)	AZI (°)	EOWR	SOWR	WOWR	NOWR	Room types
1	42.5	13	0	0	1	0.3	0	Meeting room/streaming studio
2	30	16	20	0	0	0	1	Office
3	26.5	17	10	0	0	0.5	1	Office
4	35	44	45	0	0	0.5	1	Office

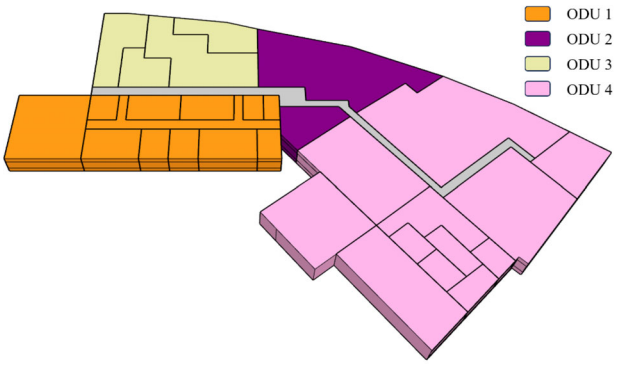


Fig. 9 Layout of the target office floor and ODU zoning

To investigate the optimal use of limited real-world data, three fine-tuning strategies are proposed and validated:

- Strategy 1: uses only July–September measurements from non-target zones within the same building (ODU1–ODU3). This simulates a zero-shot transfer scenario where no data from the specific target zone is available.
- Strategy 2: uses only one month of data (July) from the target zone (ODU4). This represents a minimal fine-tuning scenario with very limited target-specific data.
- Strategy 3: a hybrid approach that combines data from Strategy 1 and Strategy 2, integrating July–September data from ODU1–ODU3 and July data from ODU4. This leverages both cross-zone and target-specific knowledge.

All fine-tuned models were evaluated against unseen time-series data from the target ODU4 (August–September measurements) to assess their prediction generalizability and the effectiveness of each transfer strategy. Table 3 summarizes the setup and data allocation for these three strategies.

To provide a comprehensive overview of the entire experimental workflow, including the relationship between the pre-training, fine-tuning strategies, and the benchmark models, a schematic diagram is presented in Figure 10. This figure visually details the data flow from the initial sources to the final model evaluation.

To rigorously evaluate the effectiveness of the proposed transfer learning framework, its performance was also compared against two conventional data-driven benchmark models: a pure LSTM network and a light gradient boosting machine (LightGBM). As depicted in Figure 10, these benchmarks were trained from scratch using only the limited target-specific data (one month from ODU4), identical to the data used in Strategy 2. This dataset solely consists of dynamic time-series variables, as the static features remain constant for a single zone.

For clarity, the configurations of the LSTM-based models used throughout this study encompassing the pre-training, fine-tuning, and benchmark stages are summarized in Table 4.

For the LightGBM benchmark, a robust configuration was established to provide a strong baseline. The model was trained with up to 2000 estimators and a learning rate of 0.05, utilizing an MAE loss objective. Key structural parameters included num_leaves = 10 and max_depth = 8. Crucially, an early stopping strategy was implemented with a patience of 100 rounds on a validation set to determine the optimal number of boosting rounds and prevent overfitting.

Table 3 Data allocation for transfer learning strategies

Transfer strategy	Source domain	Train data	Test data
1	White-box data from EnergyPlus	ODU 1,2,3 (Jul to Sep)	ODU 4 (Aug to Sep)
2		ODU 4 (Jul)	
3		ODU 1,2,3 (Jul to Sep) & ODU 4 (Jul)	

Table 4 Configuration of LSTM-based models

Parameter	Pre-training stage	Fine-tuning stage	Pure LSTM (benchmark)
Input features	Static & dynamic	Static & dynamic	Dynamic only
LSTM hidden layers	2	2 (Frozen)	2
Units per layer	64	64	32
Recurrent activation	tanh	tanh	tanh
Output layer activation	Linear	Linear	Linear
Optimizer	Adam	Adam	Adam
Learning rate	0.002 (with 0.99 decay)	0.0001	0.0002
Loss function	WMAE	MSE	MAE
Epochs	1500	300	750

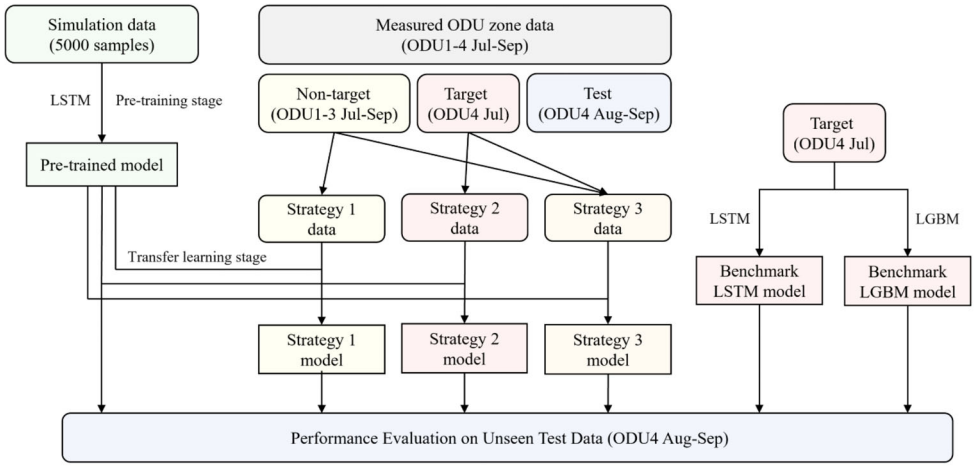


Fig. 10 Schematic diagram of the experimental framework

2.4 Prediction performance metrics

To comprehensively evaluate the performance of the proposed VRF zonal load prediction framework, this study adopts a triad of widely accepted evaluation metrics: the coefficient of determination, mean absolute error, and root mean squared error. These metrics jointly characterize model fidelity, operational accuracy, and sensitivity to extreme deviations.

In the following equations, y_i represents the actual measured value, $\hat{y}_i$ is the model's predicted value, $\overline{y}$ is the mean of the actual values, and n is the total number of samples.

- Coefficient of determination (R^2)
  R^2 quantifies the proportion of the variance in the actual load that is predictable from the model. A value approaching 1 indicates a strong fit, suggesting the model successfully captures the underlying physical dynamics driving the thermal load. Mathematically, it is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{7}$$

- Mean absolute error (MAE)
  MAE quantifies the average magnitude of prediction errors in absolute terms (units: W/m^2). As a non-directional metric, it treats all deviations equally regardless of sign, offering a transparent measure of typical prediction deviation. In energy forecasting, MAE is particularly relevant for assessing control robustness and the impact of prediction error on operational cost. The formula is:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{8}$$

- Root mean squared error (RMSE)
  RMSE is the square root of the average of squared errors. By squaring the errors before averaging, it disproportionately penalizes large deviations, making it particularly sensitive to the prediction of peak loads and outliers that could compromise system reliability or trigger inefficient operation. RMSE is computed as:

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

Together, these three metrics provide a complementary and comprehensive evaluation. R^2 assesses the model's ability to capture the overall trend and variance. MAE offers an intuitive measure of the average operational error in physical units. Meanwhile, RMSE highlights the model's susceptibility to large, potentially critical errors. The combined use of these metrics ensures a robust assessment of both statistical performance and practical engineering reliability.

3 Results

3.1 White-box model performance

The EnergyPlus-based white-box simulation framework was first validated against measured operational data from four distinct ODU zones. The performance metrics are summarized in Table 5.

Table 5 Performance metrics of white-box simulation

ODU zone	R^2	MAE (W/m^2)	RMSE (W/m^2)
1	0.656	10.54	15.42
2	0.745	6.59	12.61
3	0.828	8.14	15.12
4	0.836	7.26	11.17

Overall, the simulations demonstrated a strong capability to replicate the thermal load profiles. Zones with consistent usage patterns, such as ODU3 and ODU4, achieved high R^2 values (0.828 and 0.836) and low MAE (8.14 and 7.26 W/m^2), respectively. As illustrated in the load profile comparison (Figure 11), this indicates an accurate capture of both the timing and magnitude of loads in these zones.

However, the simulations revealed limitations in capturing non-routine behavior, highlighting the "simulation-to-reality" gap. ODU1, a zone subject to irregular high-load events like meetings, exhibited the lowest performance. Conversely, ODU2 showed a low MAE (6.59 W/m^2) but a high RMSE (12.61 W/m^2), indicating that the model struggled to predict large deviations caused by unexpected operational behaviors.

These results confirm that the white-box model provides a solid foundational understanding of the zone's thermal behavior. However, its accuracy in capturing real-world operational nuances can be further enhanced. The observed deviations highlight an opportunity for data-driven methods to uncover latent patterns and stochastic dynamics hidden within the measured data. This motivates the subsequent use of transfer learning to build upon the physical model's foundation and achieve a higher level of predictive fidelity.

3.2 Performance of cross-domain transfer learning framework

An LSTM model was pre-trained on the 5,000 synthetic data samples. The training process converged optimally, achieving a validation R^2 of 0.9859, with MAE and RMSE reaching 2.29 W/m^2 and 6.65 W/m^2, respectively. These strong results indicate that the model effectively learned the complex thermal dynamics from the diverse simulation scenarios, establishing a powerful knowledge base for subsequent transfer to the real-world domain.

To establish a strong point of comparison, two benchmark models (pure LSTM and pure LGBM) were first trained from scratch using only the limited target data (one month from ODU4). As shown in the top section of Table 6, both benchmarks exhibited poor predictive performance. The pure LGBM model achieved an R^2 of only 0.613, while the pure LSTM reached an R^2 of 0.741. These results demonstrate the inherent challenge for conventional data-driven methods when faced with limited datasets. With scarce data, such models may struggle to learn the underlying generalizable patterns of the thermal dynamics, leading to poor generalization on unseen data.

In contrast, the proposed transfer learning framework demonstrated significantly superior performance across all configurations. The pre-trained model was then fine-tuned and evaluated on measured data from the target zone (ODU4), using three progressive transfer learning strategies. The detailed results for the proposed framework are also presented in Table 6 and analyzed below.

- Baseline (no transfer):
  The performance of the pre-trained model, trained solely on the synthetic white-box database, was first evaluated directly on the real-world test data from ODU4 to establish a baseline. This "zero-shot" application, without any exposure to measured data for fine-tuning, yielded a respectable R^2 of 0.811, an MAE of 7.42 W/m^2, and an RMSE of 11.11 W/m^2, which already surpassed the performance of both benchmark models trained on real

Table 6 Performance comparison of all models and strategies

Model/strategy	R^2	MAE (W/m^2)	RMSE (W/m^2)
Benchmark models			
Pure LGBM	0.613	10.90	14.87
Pure LSTM	0.741	9.36	12.18
Proposed framework			
Baseline (no transfer)	0.811	7.42	11.11
Strategy 1	0.832	6.95	10.54
Strategy 2	0.864	6.03	9.40
Strategy 3	0.866	6.02	9.35

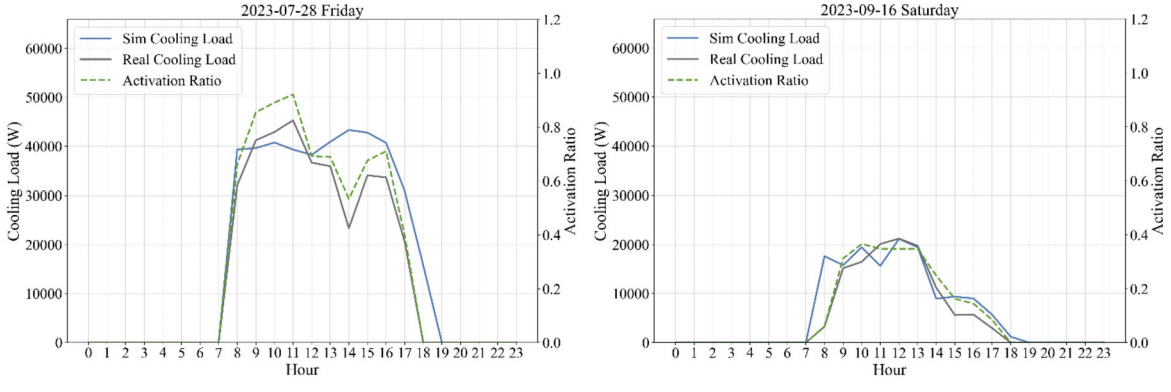


Fig. 11 White-box load prediction curves

data. These metrics indicate that the knowledge learned from the diverse simulation scenarios provides a strong foundational understanding of general thermal load dynamics. A visual comparison of the predicted and measured load distributions (Figure 12) further corroborates this, showing a general agreement. However, the histogram also reveals noticeable discrepancies in the frequency of peak and low-load occurrences, highlighting the "simulation-to-reality" gap.

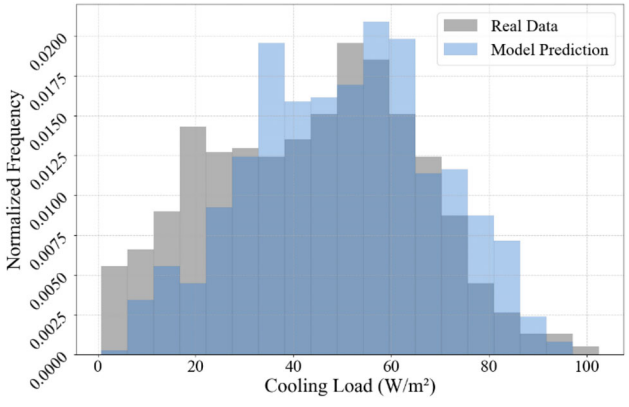
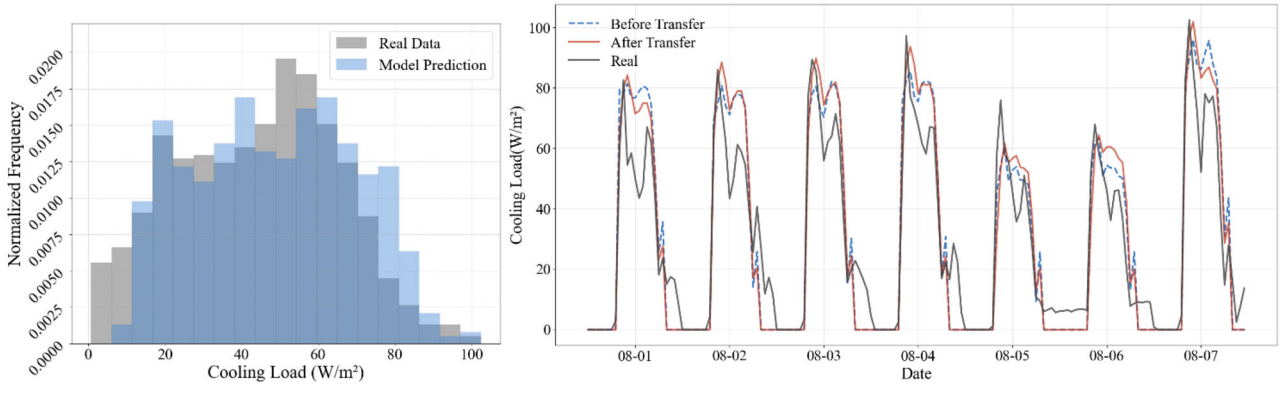


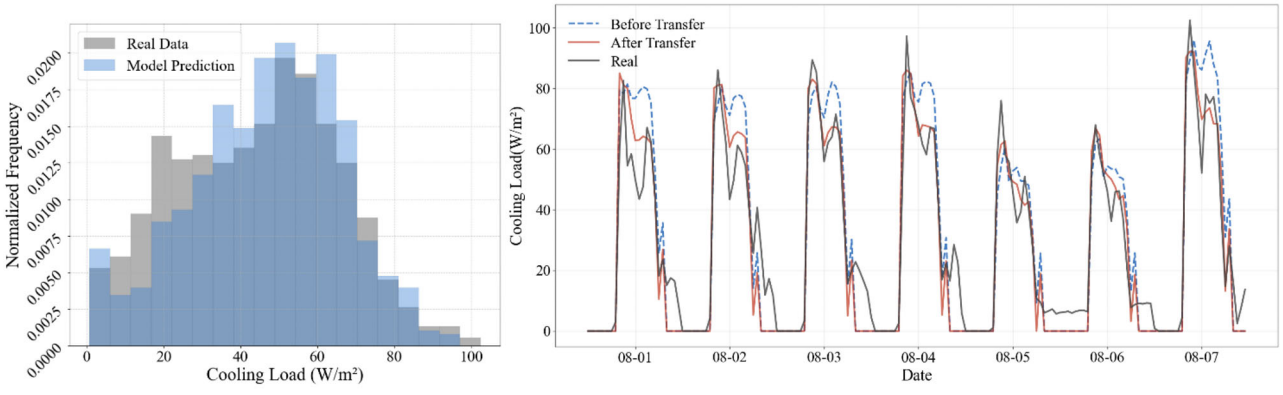
Fig. 12 Performance of the baseline model

- Transfer Strategy 1 (fine-tuning on non-target zones):
  Fine-tuning the model solely on data from adjacent, non-target zones (ODU1–ODU3) validated the framework's zero-shot prediction capability for a target zone lacking its own historical data. This strategy yielded a notable improvement, with R^2 increasing to 0.832 and MAE decreasing by 6.3% to 6.95 W/m^2. The model successfully learned to bridge the simulation-to-reality gap at a building level by adapting to the specific microclimate and overall thermal behavior. This is reflected in the distribution plot, which shows a tighter match of data points compared to the baseline (Figure 13(a)). However, as the model was not exposed to the target's unique operational signature, the load curve shows only marginal improvement in tracking the specific dynamics of ODU4 (Figure 13(b)). The primary value of this strategy lies in demonstrating that data from monitored zones can be effectively transferred to generate a more accurate forecast for a new, unmonitored zone.
- Transfer Strategy 2 (fine-tuning on limited target data):
  A significant improvement in performance was achieved by using just one month of data from the target zone (ODU4). This approach drove R^2 to 0.864 and reduced the MAE by 18.7% to 6.03 W/m^2, markedly outperforming the pure LSTM benchmark (R^2 = 0.741), which was trained from scratch using the exact same amount of limited measured data. The improvement is visually striking: the distribution plot shows a much stronger alignment between predicted and measured values (Figure 14(a)), and the load curve demonstrates a marked increase in fidelity, accurately tracing the zone's specific temporal dynamics, including peak magnitudes and ramp-up/down times (Figure 14(b)). By adapting to a small set of real operational data, the model effectively learned the zone's unique "fingerprint"—encompassing its specific occupancy schedule and internal gain patterns. This result strongly validates that minimal target-domain data is effective to bridge the simulation-to-reality gap, offering a highly practical solution for newly commissioned systems.
- Transfer Strategy 3 (hybrid fine-tuning):
  The hybrid approach, combining data from both non-target (ODU1–ODU3) and target (ODU4) zones, yielded the best overall performance, with an R^2 of 0.866, an MAE of 6.02 W/m^2, and the lowest RMSE of 9.35 W/m^2. This represents an 18.9% reduction in MAE compared to the baseline (no-transfer) model. The distribution plot for this strategy shows the tightest data distribution (Figure 15(a)), while the load curve maintains the high



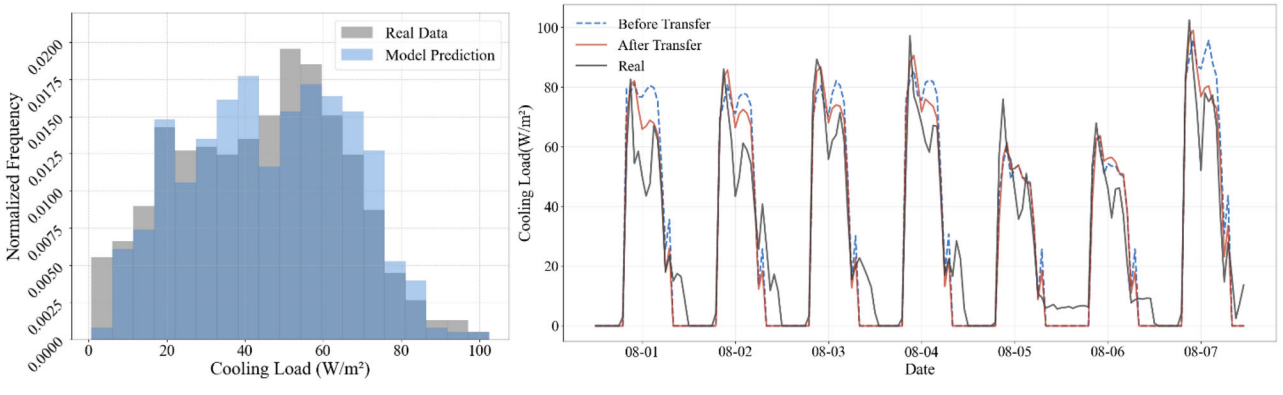
(a) (b)

Fig. 13 Performance of Strategy 1 (fine-tuned on non-target data)



(a) (b)

Fig. 14 Performance of Strategy 2 (fine-tuned on limited target data)



(a) (b)

Fig. 15 Performance of Strategy 3 (hybrid fine-tuning)

fidelity (Figure 15(b)). This superior performance can be attributed to a synergistic effect: the target data provides the essential zone-specific patterns, while the data from other ODUs offers a broader source of measured operational contexts, which simultaneously helps prevent overfitting.

In summary, the results demonstrate a clear and progressive improvement as the model was fine-tuned with real-world data. The pre-trained model provided a solid baseline that already surpassed the benchmark models trained on limited data, yet still showed an observable gap when applied to measured data. Each transfer learning strategy then systematically improved prediction accuracy. The results validated the model's zero-shot capability using non-target data (Strategy 1) and showed that a small amount of target data yielded substantial performance gains (Strategy 2). The hybrid approach (Strategy 3) delivered the best overall performance, achieving the highest R^2 of 0.866 and an 18.9% reduction in MAE compared to the baseline (no-transfer) model. This progression confirms that the proposed framework can successfully adapt simulation-based knowledge to real-world data, demonstrating its potential to deliver increasingly accurate and reliable predictions.

4 Conclusion and discussion

This study confronted the critical yet uniquely challenging task of VRF ODU zonal load prediction—a problem fundamental to efficient system design but hindered by unpredictable user behavior and data scarcity. To address this challenge, a physics-guided, transfer learning-based framework was introduced and its viability was validated through a real-world case study. The results demonstrate that by synergizing physics-based simulations with data-driven learning, it is possible to significantly improve the accuracy of load predictions, particularly for zones lacking extensive historical data. The core contributions are threefold: establishing the capacity-weighted activation ratio as a key predictor of zonal load; developing a novel psychrometric blending mechanism to integrate this dynamic user behavior into EnergyPlus; and validating a cross-domain transfer learning methodology that significantly enhances prediction accuracy.

The empirical results clearly demonstrated the framework's efficacy within the tested case. While the foundational white-box model provided a solid predictive baseline, the

introduction of transfer learning systematically bridged the "simulation-to-reality" gap by learning to model the unpredictable operational dynamics not captured by the physics-based simulation. This highlights a key strength of our hybrid approach: it leverages simulation to establish a robust physical foundation and uses real-world data to capture the stochastic "personality" of a specific zone, thereby effectively mitigating the simulation-to-reality gap. The fine-tuning process revealed a clear and compounding benefit from different data sources. Notably, fine-tuning the model solely on data from non-target zones was sufficient in this case to achieve a crucial improvement in zero-shot predictions for an unseen zone. The subsequent introduction of even minimal target-specific data then proved highly valuable, allowing the model to capture a zone's unique operational "fingerprint". Ultimately, the framework's greatest strength emerged from synergistically combining these two data types. In this hybrid approach, the broader auxiliary data provided a robust operational context that grounded the model and prevented overfitting, while the focused target data performed the high-fidelity calibration. This dual-pronged data strategy—where one source provides robustness and breadth while the other offers accuracy and depth—is the primary driver of the model's superior performance.

The implications of this research are significant for both HVAC system design and operation. For designers and engineers, this framework provides a more reliable tool for ODU sizing, moving beyond oversized, conservative estimates to right-sized systems that reduce initial capital investment and improve operational efficiency. For HVAC manufacturers and energy service companies, it offers a pathway to providing more accurate and dependable energy performance guarantees. In the operational phase, the model's high-fidelity forecasts can serve as a baseline for operational performance assessment, helping to identify deviations from expected behavior and inform long-term energy management strategies.

Despite the promising results, this study has several limitations. First, the generalizability of the framework should be further validated, as the current validation was conducted on a single building type (office) in a specific climate zone, with data exclusively from one cooling season. Although the proposed methodology is designed to be generalizable, its performance across diverse contexts—including different building types, climatic conditions, and operational seasons—has not yet been established. This data concentration also limited the ability to fully assess the impact of training and test set similarity on the results. Second, the standardized activation profiles, while effective, are deterministic. This simplification is well-suited for annual load profile generation for design purposes, where typical patterns are paramount, but its deterministic nature limits the framework's direct applicability for short-term forecasting, which is highly sensitive to stochastic user behavior. Finally, while LSTM proved effective, this study did not explore more advanced time-series architectures.

These limitations, in turn, suggest clear directions for future work. Future efforts should focus on three key areas. First, the framework's generalizability should be validated across a wider range of scenarios, including diverse building types, climate regions, and operational seasons to assess its performance under more heterogeneous conditions. Second, to address the issue of deterministic profiles, research could proceed on two fronts: enhancing source domain diversity by employing generative modeling techniques to create richer user schedules, and improving short-term prediction accuracy by augmenting the feature set with real-time behavioral indicators. Finally, exploring more advanced time-series architectures, such as Transformers, is recommended to potentially capture complex long-range temporal dependencies more effectively.

In conclusion, this paper presents a robust and practical framework, offering a viable strategy to significantly improve the accuracy of ODU zonal load prediction. By effectively fusing the strengths of physical modeling and deep learning, this research provides a methodology with the potential for broad applicability for more energy-efficient building design, more reliable system performance, and smarter operational control in the growing landscape of VRF applications.

Acknowledgements

This work was funded by Guangdong Key Laboratory of Thermal Energy Storage Technology for Buildings.

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

Author contribution statement

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Junyu Chen, Peng Xu and Renrong Ding. The first draft of the manuscript was written by Junyu Chen and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

References

Abdolvand M, Nezhad A, Bambach M, et al. (2024). Integrated

climate-responsive thermal load ML model and cost/embodied energy estimate from a preliminary building design. *Energy and Buildings*, 304: 113837.

Blum DH, Arendt K, Rivalin L, et al. (2019). Practical factors of envelope model setup and their effects on the performance of model predictive control for building heating, ventilating, and air conditioning systems. *Applied Energy*, 236: 410–425.

Ding T, Liu S, Wang Z, et al. (2024). A novel mixture sampling strategy combining Latin hypercube sampling with optimized one factor at a time method: A case study on mixtures of antibiotics and pesticides. *Journal of Hazardous Materials*, 461: 132568.

Electromechanical Information (2021). Summary report of China's central air-conditioning market in 2020: Chapter 2: Analysis of industry operating conditions. *Electromechanical Information*, 2021(4): 7–20. (in Chinese)

Forrester JR, Wepfer WJ (1984). Formulation of a load prediction algorithm for a large commercial building. *ASHRAE Transactions*, 90(2): 536–551.

Gao Z, Yu J, Zhao A, et al. (2022). A hybrid method of cooling load forecasting for large commercial building based on extreme learning machine. *Energy*, 238: 122073.

Gao Z, Yang S, Yu J, et al. (2024). Hybrid forecasting model of building cooling load based on combined neural network. *Energy*, 297: 131317.

Gunay B, Shen W, Newsham G (2017). Inverse blackbox modeling of the heating and cooling load in office buildings. *Energy and Buildings*, 142: 200–210.

He F, Zhou J, Feng Z, et al. (2019). A hybrid short-term load forecasting model based on variational mode decomposition and long short-term memory networks considering relevant factors with Bayesian optimization algorithm. *Applied Energy*, 237: 103–116.

Hochreiter S, Schmidhuber J (1997). Long short-term memory. *Neural Computation*, 9: 1735–1780.

Hu J, Zheng W, Zhang S, et al. (2021). Thermal load prediction and operation optimization of office building with a zone-level artificial neural network and rule-based control. *Applied Energy*, 300: 117429.

IEA (2023a). World Energy Outlook 2023. Paris: International Energy Agency.

IEA (2023b). Energy Efficiency 2023. Paris: International Energy Agency.

Imam S, Coley DA, Walker I (2017). The building performance gap: Are modellers literate? *Building Services Engineering Research and Technology*, 38: 351–375.

Ke G, Meng Q, Finley T, et al. (2017). LightGBM: A highly efficient gradient boosting decision tree. In: Proceedings of 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

Lee YM, Horesh R, Liberti L (2015). Optimal HVAC control as demand response with on-site energy storage and generation system. *Energy Procedia*, 78: 2106–2111.

Leiprecht S, Behrens F, Faber T, et al. (2021). A comprehensive thermal load forecasting analysis based on machine learning algorithms. *Energy Reports*, 7: 319–326.

Li Q, Meng Q, Cai J, et al. (2009). Applying support vector machine to predict hourly cooling load in the building. *Applied Energy*, 86: 2249–2256.

Li X, Wen J (2014). Review of building energy modeling for control and operation. *Renewable and Sustainable Energy Reviews*, 37: 517–537.

Lin X, Lee H, Hwang Y, et al. (2015). A review of recent development in variable refrigerant flow systems. *Science and Technology for the Built Environment*, 21: 917–933.

Machado RMES, Geraldi MS, Bavaresco M, et al. (2023). Metamodel to predict annual cooling thermal load for commercial, services and public buildings: A country-level approach to support energy efficiency regulation. *Energy and Buildings*, 301: 113690.

Pan Y, Zhu M, Lv Y, et al. (2023). Building energy simulation and its application for building performance optimization: A review of methods, tools, and case studies. *Advances in Applied Energy*, 10: 100135.

Park C, Kim I, Kim W (2025). Transfer learning–based energy consumption prediction for variable refrigerant flow system in buildings. *Applied Thermal Engineering*, 267: 125811.

Pinto G, Wang Z, Roy A, et al. (2022). Transfer learning for smart buildings: A critical review of algorithms, applications, and future perspectives. *Advances in Applied Energy*, 5: 100084.

Sendra-Arranz R, Gutiérrez A (2020). A long short-term memory artificial neural network to predict daily HVAC consumption in buildings. *Energy and Buildings*, 216: 109952.

Somu N, M R GR, Ramamritham K (2020). A hybrid model for building energy consumption forecasting using long short term memory networks. *Applied Energy*, 261: 114131.

Srivastava S, Lessmann S (2018). A comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data. *Solar Energy*, 162: 232–247.

UNEP (2014). UNEP-SBCI Annual Report 2013/2014. United Nations Environment Programme, Sustainable Buildings and Climate Initiative.

Wang S, Xu X (2006). Simplified building model for transient thermal performance estimation using GA-based parameter identification. *International Journal of Thermal Sciences*, 45: 419–432.

Wang Z, Hong T, Piette MA (2020). Building thermal load prediction through shallow machine learning and deep learning. *Applied Energy*, 263: 114683.

Yan R, Zhao T, Rezgui Y, et al. (2023). Transferability and robustness of a data-driven model built on a large number of buildings. *Journal of Building Engineering*, 80: 108127.

Wang J, Lu X, Adetola V, et al. (2024). Modeling Variable Refrigerant Flow (VRF) systems in building applications: A comprehensive review. *Energy and Buildings*, 311: 114128.

Zhou D, Ma S, Hao J, et al. (2020). An electricity load forecasting model for Integrated Energy System based on BiGAN and transfer learning. *Energy Reports*, 6: 3446–3461.