IOP Conference Series:
Earth and
Environmental Science

PAPER • OPEN ACCESS

Data-driven smart control for window and HVAC systems in sports space considering thermal comfort and energy efficiency

To cite this article: Y Li *et al* 2025 *IOP Conf. Ser.: Earth Environ. Sci.* **1500** 012065

Data-driven smart control for window and HVAC systems in sports space considering thermal comfort and energy efficiency

Y Li[1,2*], L Li[1,2] and P Shen[3]

[1] School of Architecture and Design, Harbin Institute of Technology, Harbin, 150001, China
[2] Key Laboratory of Cold Region Urban and Rural Human Settlement Environment Science and Technology, Ministry of Industry and Information Technology, Harbin, 150001, China
[3] Institute for Future Human Habitat, Shenzhen International Graduate School, Tsinghua University, Shenzhen, China

*E-mail: 19B334003@stu.hit.edu.cn

Abstract. Optimizing the control of window and HVAC systems for thermal comfort and energy efficiency is a critical challenge in intelligent building operations and controls. Despite the application of artificial intelligence of things (AIoT) in supporting smart window and HVAC operations, inaccurate forecasting methods in existing systems hinder their widespread adoption. This study develops a data-driven smart control method, utilizing EnergyPlus-MATLAB co-simulation and Long Short-Term Memory (LSTM)-based time-series forecasting, to clarify the impacts of real-time control of window and HVAC systems on thermal comfort and energy efficiency in large sports buildings. A two-way EnergyPlus-MATLAB co-simulation framework is developed leveraging the Building Controls Virtual Test Bed (BCVTB) platform as the middleware. In addition, a pretrained LSTM model is deployed to quickly predict occupant thermal comfort across large sports spaces, which serves as the basis for window and HVAC control decisions. To demonstrate the feasibility of the proposed approach, a case study of a national fitness hall (NFH) in Shenzhen, China is conducted. The results reveal that the proposed method can maintain more stable thermal comfort though it leads to a 12.26% increase in energy consumption due a 27.9% rise in HVAC operation time. Future work will integrate the proposed method with deep reinforcement learning to further enhance thermal comfort and energy efficiency.

Keywords: Data-driven, Smart control, Sports space, Thermal comfort, Energy efficiency

1. Introduction

The building sector is responsible for approximately 40% of global energy consumption and 25% of CO_2 emissions [1]. Among all building facilities, Heating, Ventilation, and Air Conditioning (HVAC) systems are major energy consumers, contributing to 50 % of total building energy usage [2, 3], with this figure expected to triple by 2050 due to climate change impacts [4]. Well-designed and efficient HVAC controls play an essential role in providing a comfortable and energy-efficient building environment. Currently, rule-based control (RBC) and proportional-integral-derivative (PID) controllers are commonly used for HVAC systems. These controllers rely on fixed schedules to set temperature setpoints or use setpoint tracking algorithms, such as PID, to adjust system operations, thereby maintaining comfortable conditions. However, these methods face challenges in adapting to external factors, such as occupancy levels, meteorological conditions, and electricity costs [5].

As artificial intelligence (AI) and digital twins (DT) continue to advance, there has been a growing emphasis on intelligent building control algorithms among researchers. In recent years, Model predictive control (MPC) has emerged as a promising solution for controlling complex systems. MPC, a proven optimal control strategy, leverages physics-based, grey-box, or black-box models to ensure strong and reliable performance. However, the widespread application of MPC is limited by the high demands for modeling, expertise, data, hardware, usability, and computational power [6]. Given the critical role of the building sector in decarbonisation, it is essential to develop adaptive approaches that better align control systems with real environmental conditions while also accelerating the computation of physics-based MPC. Traditional simulation-based control strategies for HVAC systems, which rely on Computational Fluid Dynamics (CFD) or Building Energy Simulation (BES), incur significant computational costs, making them impractical for real-time applications. As a result, recent research has increasingly focused on machine learning methods (ML) for optimal HVAC control. Machine learning enables the transition from physics-based MPC to data-driven predictive control. The rise of big data, enhanced computing power, and advances in algorithms have made it possible to implement ML-based HVAC control approaches[7].

In large sports spaces, the phenomenon of uneven "thermal stratification" presents significant challenges in accurately predicting indoor thermal environment and occupant thermal comfort [8-10], thereby hindering effective window and HVAC controls. Existing studies have employed BES-CFD coupling methods [8, 10, 11] to improve the accuracy of thermal distribution predictions and enhance HVAC system control performance. However, the computational burden of CFD makes it impractical for real-time control applications. Recently, time-series forecasting (TSF) has emerged as a alternative to CFD for estimating occupant thermal comfort, integrating window and HVAC control simulations with building energy simulations. Time-series data, including window and HVAC operation status, indoor environmental conditions, outdoor weather, and occupant behavior, can be extracted to characterize the dynamic environment. Long Short-Term Memory (LSTM) networks, coupled with convolutional neural networks (CNN) and the attention mechanism (AM), are then used for the rapid prediction of future environmental changes, enabling real-time coordination of window and HVAC operations.

To address the gaps identified, this study introduces a data-driven smart control framework for window and HVAC systems in sports spaces, leveraging EnergyPlus-MATLAB co-simulation and LSTM-based TSF. This framework explores the potential for thermal comfort and building energy efficiency improvement through real-time window and HVAC operations.

2. Methodology

The flowchart of the proposed data-driven smart control method is shown in figure 1. It consists of two main stages: Stage 1-LSTM training through parametric studies from BES-CFD co-simulations, and Stage 2-Real-time window and HVAC control via BES-(CNN-LSTM-AM) co-simulations. The method is enabled in a co-simulation environment combining EnergyPlus and MATLAB through the middleware, Building Controls Virtual Test Bed (BCVTB), which facilitates control interactions and data exchange between MATLAB and EnergyPlus via the External Interface. A variable configuration file is written to determine the flow of data exchange among EnergyPlus, BCVTB and MATLAB in this study.

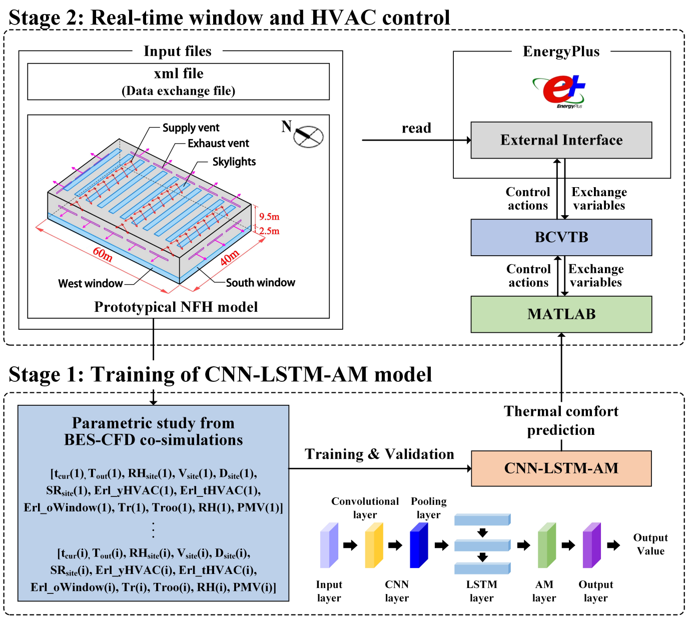


Figure 1. The flowchart of the data-driven smart window and HVAC control method.

A prototypical national fitness hall (NFH) model was established as the target building in this study. It is a large two-story sports building located in Shenzhen, China, with the second-story fitness hall designated as the target area with a dimension of 60m × 40m × 12m (length× width× height). The building is oriented northward, with windows on the south and west elevations and several skylights on the roof for natural ventilation. The HVAC system in the building is a Packaged Terminal Air Conditioner (PTAC) system. Shenzhen, classified as Climate Zone 1A, has a "hot summer and warm winter" climate. For most of the year, the climate is warm and humid, with a prolonged rainy summer spanning May to November. Air conditioning (AC) cooling energy constitutes the primary building energy use in Shenzhen. Meanwhile, natural ventilation remains a viable option in transition seasons, with a duration from March to November. In this study, the AC cooling setpoint was set at 18 °C. The study period, from April 28th to May 21st, was selected for co-simulations and training the CNN-LSTM-AM model (LSTM coupled with a convolutional neural network and attention mechanism) with a 15-minute time step. The weather file used for training was CHN_Guangdong.Shenzhen.594930_SWERA.epw [12].

2.1 Stage 1: Training of CNN-LSTM-AM

Thermal stratification is a common phenomenon in large sports buildings, and it has been reported to affect the prediction accuracy of thermal comfort and energy consumption. At present, a feasible method for increasing the accuracy of estimating the uneven thermal distribution in large sports space is to conduct BES-CFD co-simulations [8]. Hence, this study develops a two-way external EnergyPlus-Fluent coupling method to predict occupant thermal comfort for window and HVAC operations. The stage 1 of the proposed method involves developing the CNN-LSTM-AM model using the dataset from BES-CFD co-simulations, encompassing data source, data preprocessing, model training and validation, as follows:

2.1.1 Data source. The EnergyPlus-Fluent co-simulation was conducted from April 28 to May 21 with a time interval of 15 minutes, leaving a total of 2304 simulation runs, the detailed coupling process can be found in our previous study [10]. The co-simulations were executed on a workstation equipped with a 3.60 GHz Intel® Core™ i9-9900KF CPU, an NVIDIA® Quadro® RTX 5000 GPU, 16 GB of RAM, and a Windows 10 (64-bit) operating system, requiring approximately 14 days of computation time. The simulation results, including indoor and outdoor conditions, window and HVAC control statuses, and thermal comfort metrics, were compiled into a dataset (D) for developing the CNN-LSTM-AM model. We denote those most influencing factors at time step i as In_i, which consists of site outdoor air drybulb temperature (T_i^{out}), site outdoor air relative humidity (RH_i^{site}), site wind speed (V_i^{site}), site wind direction (D_i^{site}), site direct solar radiation rate per area (SR_i^{site}), HVAC status ($yHVAC_i$), HVAC cooling setpoint ($tHVAC_i$), window open area fraction ($oWindow_i$), zone mean radiant temperature (T_i^{mrt}), zone mean air temperature (T_i^{in}), zone air relative humidity (RH_i^{in}). Besides, we denote the outputs of thermal comfort at time step t as Out_i, which is the Adaptive Predicted Mean Vote ($aPMV$) or Predicted Mean Vote (PMV). Those parameters are denoted as:

$$In_i = (T_i^{out}, T_i^{out}, RH_i^{site}, V_i^{site}, D_i^{site}, SR_i^{site}, yHVAC_i, tHVAC_i, oWindow_i, T_i^{mrt}, T_i^{in}, RH_i^{in}) \tag{1}$$

$$Out_i = ((a)PMV_i) \tag{2}$$

$$D = [In, Out] \tag{3}$$

2.1.2 Data preprocessing. The dataset for training CNN-LSTM-AM model undergoes several preprocessing steps. First, missing data due to simulation errors or failures are re-simulated. Next, different data items are normalized to the range [0,1] using min-max normalization. Afterwards, the dataset is shuffled to ensure it is randomly distributed, and finally it was divided into training set and validation set.

2.1.3 CNN-LSTM-AM model training and validation. This study constructs an occupant thermal comfort prediction model based on the CNN-LSTM-AM algorithm. The CNN component captures spatial correlations among various features in the collected dataset (D), addressing LSTM's limitations in handling spatial dependencies. The attention mechanism (AM) allows the LSTM network to differentiate the importance of various input parameters on future forecasting, increasing the robustness of thermal comfort predictions. Besides, the hyperparameter settings of the LSTM were tuned by optimizing the LSTM layers, max epochs and initial learning rate using

Table 1. Architecture parameters of CNN-LSTM-AM.

Parameters	Value
Convolution layer filters	64
Convolution layer kernel size	1
Convolution layer activation function	ReLU
Pooling layer pool size	1
LSTM layers	25
Max epochs	81
Initial learning rate	0.0504154042814
Learning rate schedule	Piecewise
Learning rate drop factor	0.1
Learning rate drop period	90
LSTM optimizer	Adam

the Pelican Optimization Algorithm (POA). The root mean square (RMSE), mean absolute percentage error (MAPE), mean absolute error (MAE), R-square (R^2) were adopted as the evaluation indicators. Table 1 shows the optimal parameter settings of the CNN-LSTM-AM model.

2.2 Stage 2: Real-time window and HVAC control

After training the CNN-LSTM-AM model, stage 2 involves integrating the trained model for real-time window and HVAC control. This section outlines how the trained model is used to predict occupant thermal comfort and adjust the system operations accordingly. The stage 2 leverages a two-way EnergyPlus-MATLAB co-simulation framework, enabling dynamic window and HVAC control according to thermal comfort levels predicted by CNN-LSTM-AM model. BCVTB 1.6.0 was used as the middleware to couple the EnergyPlus 9.6.0 and MATLAB R2021a. BCVTB is a software environment that couples various simulation engines for co-simulation, which has been frequently adopted for building energy assessment and system control. MATLAB code was written to deploy the CNN-LSTM-AM model for the rapid thermal comfort prediction and then to exchange the data with EnergyPlus. Figure 2(a) illustrates the BCVTB model for BES-(CNN-LSTM-AM) co-simulation-based window-HVAC control, where the embedded simulators entitled "*EP_Run*" and "*LSTM_Predict*" are in charge of calling EnergyPlus and CNN-LSTM-AM model (via MATLAB).

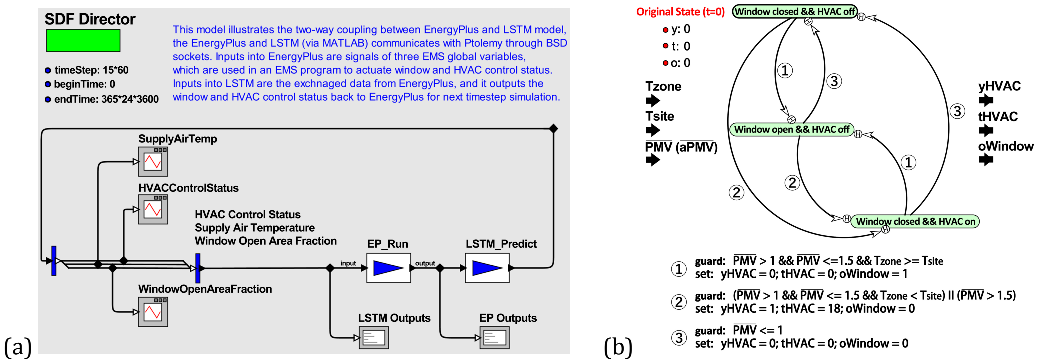


Figure 2. (a) The BCVTB model of coupling EnergyPlus and CNN-LSTM-AM model for controlling window and HVAC systems; (b) The status transition process and corresponding control conditions.

In this study, three scenarios of window and HVAC control were considered: (1) S_1: Open window & turn HVAC off; (2) S_2: Close window & turn HVAC on; (3) S_3: Close window & turn HVAC off. To enable real-time window and HVAC operations, this study employs a state transition approach to adjust their operating status at each time interval. At each timestep, i ($i = 0, 1, 2, \dots T$), once the control condition $C(i)$ is met, the control status transitions to $S(i)$, and the corresponding control signals are output as $Y(i)$. Note that the initial state was set as: $S(0) = S_3$: Close window and turn off HVAC. Figure 2(b) details the window and HVAC control policy as well as the transition process of control status. Hence, the control problem is mathematically formulated as:

$$\text{At each time step i:}$$

$$guard\colon\ C(i) = f\left(T_{zone}(i), T_{site}(i), \overline{(a)PMV}(i)\right) =$$

$$\begin{cases} C_1, & if\ \ \overline{(a)PMV}(i) \leq 1 \\ C_2, & if\ \ 1 < \overline{(a)PMV}(i) \leq 1.5\ and\ T_{zone}(i) \geq T_{site}(i) \\ C_3, & if\ \ \left(1 < \overline{(a)PMV}(i) \leq 1.5\ and\ T_{zone}(i) < T_{site}(i)\right) or\ \left(\overline{(a)PMV}(i) > 1.5\right) \end{cases} \tag{4}$$

$$transfer\ to{:}\ S(i) = \begin{cases} S_1{:}\ Open\ window, turn\ HVAC\ off, & if\ \ C(i) = C_1 \\ S_2{:}\ Close\ window, turn\ HVAC\ on, & if\ \ C(i) = C_2 \\ S_3{:}\ Clsoe\ window, turn\ HVAC\ off, & if\ \ C(i) = C_3 \end{cases} \quad (5)$$

$$set{:}\ Y(i) = \big(yHVAC(i), tHVAC(i), oWindow(i)\big) = \begin{cases} (0,0,1), & if\ \ S(i) = S_1 \\ (1,18,0), & if\ \ S(i) = S_2 \\ (0,0,0), & if\ \ S(i) = S_3 \end{cases} \quad (6)$$

Where: i ($i = 0, 1, 2, \ldots T$) denotes the simulation time steps, starting from 0 and ending at T; $yHVAC(i), tHVAC(i)$ and $oWindow(i)$ are the HVAC and window control signals, where 0 indicates closing window or turning HVAC off, while 1 denotes to opening window or turning HVAC on. $tHVAC(i)$ is switched to 18 °C when the HVAC is turned on ($yHVAC(i) = 1$); The range of $\overline{PMV}$ ($\overline{aPMV}$) for determining three window and HVAC control scenarios was tailored for strenuous sports settings, according to [8, 13].

3. Results and discussion

3.1 Model training results

The trained CNN-LSTM-AM model demonstrated strong predictive performance, with predicted values closely following the trend of observed data. Specifically, the RMSE, MAPE, MAE and R^2 values of the trained model for predicting thermal comfort ($\overline{PMV}$ or $\overline{aPMV}$) are 0.1098, 0.0548, 0.0435 and 0.9399, respectively, which confirms that the prediction errors by the trained model are well accepted compared with similar studies [14, 15]. Possible sources of error in the trained model include biases and limitations in the training datasets, which may hinder the model's generalization across different environmental conditions and time periods. Future work should focus on expanding the training dataset to cover a broader range of time points and durations. Additionally, integrating physics-informed LSTM models with optimized hyperparameters may enhance the model's robustness and reduces its reliance on purely data-driven learning.

3.2 Zone mean air temperature

By applying the trained CNN-LSTM-AM model in the stage 2, the proposed data-driven smart control framework across a whole year in Shenzhen, China was implemented. The zone mean air temperature (T_{zone}) under BES-(CNN-LSTM-AM) co-simulation and standalone EnergyPlus simulation are depicted in figure 3. It indicates that the proposed framework achieved a lower T_{zone} compared with the traditional method throughout the whole year. The discrepancy of the T_{zone} between the proposed framework and traditional method is significant during winter and part of the transition seasons when the T_{zone} of proposed framework is approaching to the outdoor air temperature (T_{site}), while the T_{zone} of traditional method is higher than T_{site}. Another interesting phenomenon is that the daily fluctuations of the T_{zone} under proposed framework are less than that under the traditional method, indicating more stable thermal environment, which

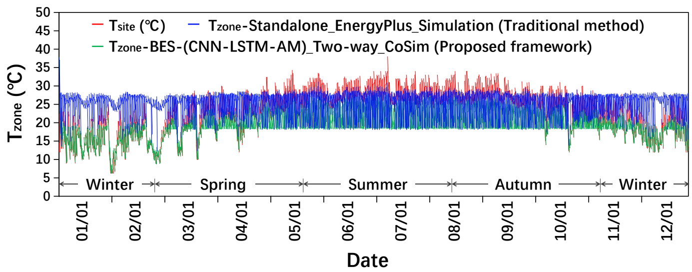
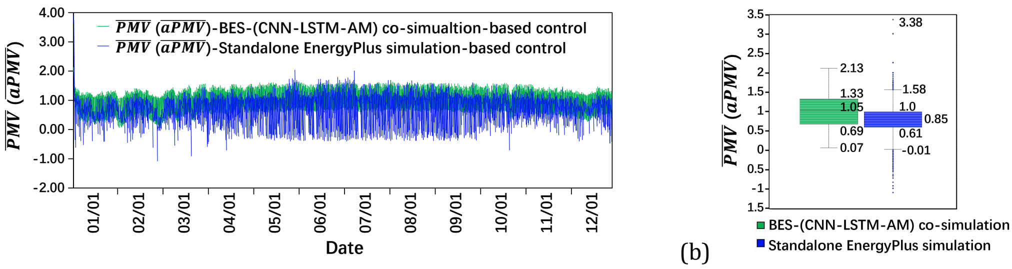


Figure 3. Zone mean air temperature under the BES-(CNN-LSTM-AM) co-simulation and standalone EnergyPlus simulation.

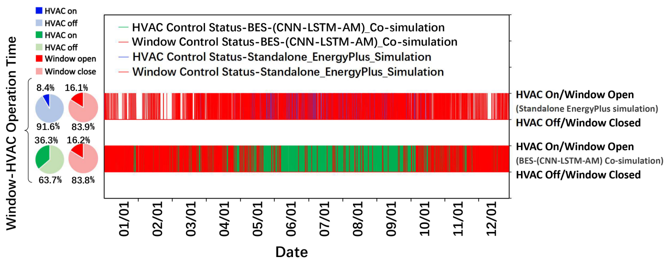
can be explained by the frequent operations of the AC for the former while the frequent operations of windows for the latter (see figure 4(c)).

3.3 Effects of real-time window and HVAC control on thermal comfort

The continuous changing of thermal comfort at each timestep for the proposed framework and traditional method is illustrated in figure 4(a), it can be observed that the proposed framework can achieve more stable thermal comfort, manifested as slighter $\overline{PMV}$ or $\overline{aPMV}$ fluctuations across the analysis period. Specifically, the interquartile range of $\overline{PMV}$ or $\overline{aPMV}$ for the proposed framework is [0.69, 1.33] with less outliers, while the traditional method achieved an interquartile range of [0.61, 1.0] with more outliers (figure 4(b)). To better understand the reasons contributing to this phenomenon, we compared the system control status for each timestep between two approaches, as shown in figure 4(c). It can be witnessed that obvious discrepancies existed in the HVAC operation time between the proposed framework and traditional method. The former had a much higher frequency of turning the AC on (36.3%) than the latter (8.4%), while both exhibited similar window operation times.



(a) (b)



(c)

Figure 4. (a-b) $\overline{PMV}$ ($\overline{aPMV}$) under the BES-(CNN-LSTM-AM) co-simulation and standalone EnergyPlus simulation; (c) Control status at each timestep and total window and HVAC operation time.

3.4 Effects of real-time window and HVAC control on cooling energy

Monthly AC cooling energy consumption (total, sensible and latent) of the proposed framework and traditional method is illustrated in figure 5, in which the absolute deviation of the two is depicted in the line chart. It can be noted that the proposed framework achieved less energy consumption in summer while more in other seasons, indicating the energy saving potential by

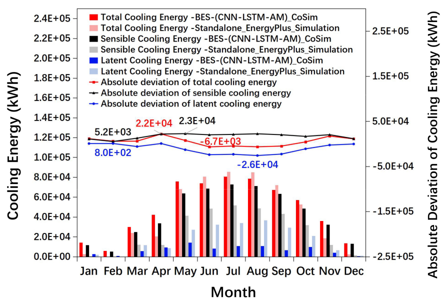


Figure 5. Comparison of monthly cooling energy and absolute deviation of AC cooling energy between BES-(CNN-LSTM-AM) co-simulation and standalone EnergyPlus simulation.

effective control of window and HVAC systems in hot summer. However, the annual total cooling energy of the proposed framework (5.76×10^5 kWh) is higher than that of the traditional method (5.13×10^5 kWh), reaching by approximately 12.26%, which can be explained by the HVAC operation time in figure 4(c). Moreover, significant absolute deviations between the proposed framework and traditional method are witnessed in summer, late spring and early autumn. In winter, absolute deviation is not obvious. As the traditional method does not account for uneven thermal stratification in large sports spaces, its cooling estimates may be less reliable.

4. Conclusion and future work

This study presents a data-driven smart window and HVAC control method for maintaining occupant thermal comfort and optimizing building energy efficiency in large sports spaces. By integrating EnergyPlus-MATLAB co-simulation with a pre-trained CNN-LSTM-AM model, the proposed method enables real-time prediction of occupant thermal comfort and dynamic adjustment of window and HVAC operations. Implemented in a case study of a prototype national fitness hall on a full-year time scale, the effects of real-time window and HVAC control on occupant thermal comfort and building energy efficiency were elucidated. Case study results indicated that the proposed control framework could maintain more stable thermal comfort while consuming 12.26% more cooling energy. Notably, the proposed framework offers a promising solution for energy-efficient and occupant-centric control in large sports buildings. Furthermore, it can be extended to other large spaces across various climates, such as exhibition centers and high-speed railway stations. Future work will investigate the integration of time series forecasting with deep reinforcement learning algorithms to develop more advanced control strategies, with the goal of achieving significant energy savings in real-world building applications.

5. References

[1] IEA (2024) *Energy Technology Perspectives 2024*. Available at: https://www.iea.org/reports/energy-technology-perspectives-2024 (Accessed: 19 March 2025)
[2] Xu G, Schwarz P and Yang H 2020 Renew. Sustain. Energy Rev. **122** 109737
[3] González-Torres M, Pérez-Lombard L, Coronel J F, Maestre I R and Yan D 2022 Energy Rep. **8** 626-637
[4] IEA (2018) *The Future of Cooling: Opportunities for energy-efficient air conditioning*. Available at: https://www.iea.org/reports/the-future-of-cooling (Accessed: 19 March 2025)
[5] Brandi S, Fiorentini M and Capozzoli A 2022 Autom. Constr. **135** 104128
[6] Benndorf G, Wystrcil D and Réhault N 2018 Appl. Phys. Rev. **5** 041501
[7] Gao Y, Shi S, Miyata S and Akashi Y 2024 Energy **291** 130344
[8] Guo W, Liang S, He Y, Li W, Xiong B and Wen H 2022 Build. and Environ. **207** 108420
[9] Lv L, Wang X, Zhang Q, Xiang Y, Wu X and Huang C 2023 J. Build. Eng. **69** 106297
[10] Li Y, Li L, Cui X and Shen P 2024 J. Build. Eng. **97** 110731
[11] Tian W, Han X, Zuo W and Sohn M D 2018 Energy Build. **165** 184-199
[12] U.S. Department of Energy. *EnergyPlus weather for Guangdong Shenzhen, China*. Available at: https://energyplus.net/weather-location/asia_wmo_region_2/CHN/CHNGuangdong.Shenzhen.594930SWERA (Accessed: 19 March 2025)
[13] Ji T, Cai J and Wang K 2022 Int. J. Low-Carbon Technol. **17** 333-341
[14] Jing G, Ning C, Qin J, Ding X, Duan P, Liu H and Sang H 2023 J. Build. Eng. **68** 106054
[15] Chen Y, Yang Q, Chen Z, Yan C, Zeng S and Dai M 2023 Build. and Environ. **234** 110149

Acknowledgement

This work was supported by the National Natural Science Foundation of China (Grant No. 52394222, 52008132).