

Adaptive physically consistent neural networks for data center thermal dynamics modeling



Dong Chen, Chee-Kong Chui, Poh Seng Lee^{*}

Department of Mechanical Engineering, NUS College of Design and Engineering, National University of Singapore, Engineering Drive 1, Singapore

HIGHLIGHTS

- A-PCNN replaces preset static coefficients, cutting costs and boosting adaptability.
- A-PCNN surpasses traditional PCNN in both real and simulation DC datasets.
- Mathematical proofs confirm that A-PCNN consistently adheres to physical laws.
- A-PCNN shows remarkable flexibility across different base models.
- This work offers profound insights into thermal modeling with AI techniques.

ARTICLE INFO

Keywords:
 Data center
 Physics-informed neural networks
 Energy model
 Machine learning
 Digital twin
 Prior knowledge

ABSTRACT

Data centers (DCs) are vital for large-scale Internet services, yet their energy consumption poses a significant concern. Energy modeling for DCs is crucial for design, control, and retrofitting. Traditional physic-based models lack flexibility, while Neural Networks (NN) may not strictly adhere to physical principles and demand extensive data. The Physically Consistent Neural Networks (PCNN) framework, which incorporates physical laws into NN through positive coefficients, is introduced. Despite its innovative approach, PCNN struggles with accuracy and generalization in real-world thermal environments due to its reliance on empirically predetermined, static coefficients, which are costly to derive and limit adaptability to dynamic conditions. To address these limitations, this study proposes the Adaptive Physically Consistent Neural Networks (A-PCNN) framework. A-PCNN leverages NN with Softplus activation functions, replacing traditional preset and fixed coefficients to reduce trial-and-error costs and increase flexibility. Over a six-month real data center dataset case study, the A-PCNN framework significantly outperformed the traditional PCNN model. Specifically, it reduced the Mean Absolute Error by 17.3 % for a 15-min forecast and by 79.2 % over a seven-day period, using a Multilayer Perceptron (MLP) as the base model. Furthermore, the A-PCNN framework demonstrates remarkable adaptability. Whether based on a MLP or Long Short-Term Memory (LSTM) model, it consistently surpasses traditional methods in predictive accuracy across time frames from 15 min to 7 days. Its superior performance is especially notable in longer forecast periods.

1. Introduction

1.1. Background

Data centers (DCs) play a crucial role in supporting the rapid growth of the information and communications technology (ICT) industry, but they also pose significant energy challenges [1–4]. Currently, DCs consume approximately 2 % of the world's power generation, and this consumption is steadily increasing [5]. The energy usage in data centers

is primarily attributed to two main components: the ICT equipment and the cooling system. The cooling systems in data centers are designed to maintain optimal operating temperatures for ICT equipment. Traditional cooling methods involve the use of air conditioning units, chillers, and cooling towers, which can be highly energy intensive. The energy consumed by these cooling systems is significant that it often matches or even surpasses the energy used by the ICT equipment itself [6,7]. To minimize cooling energy consumption in data centers, strategies include optimizing air circulation, utilizing liquid or immersion cooling technologies, and employing energy-efficient devices [8,9]. One effective

* Corresponding author.

E-mail address: pohseng@nus.edu.sg (P.S. Lee).

Nomenclature	
Abbreviations	
DC	Data Center
ICT	Information and Communications Technology
ITE	Information Technology Equipment
NN	Neural Network
DRL	Deep Reinforcement Learning
MPC	Model Prediction Control
MLP	Multilayer Perceptron
LSTM	Long Short-Term Memory
PINN	Physics-informed Neural Network
PCNN	Physically Consistent Neural Network
A-PCNN	Adaptive Physically Consistent Neural Network
HVAC	Heating, Ventilation, and Air Conditioning
PDE	Partial Differential Equation
MAE	Mean Absolute Error
RMSE	Root Mean Square Error
<i>T</i>	Temperature
<i>H</i>	Humidity
<i>P</i>	Power consumption
<i>t</i>	Time
<i>X</i>	Input
<i>Y</i>	Output
<i>E</i>	Physical-module gain
<i>D</i>	Black-module gain
<i>Subscripts</i>	
<i>k</i>	Timestep
<i>i</i>	Data index

approach to improving cooling efficiency is the application of intelligent controls, such as increasing temperature setpoints and reducing fan speeds, since data center operators typically manage cooling systems with a margin for high heat loads to ensure stable performance of Information Technology Equipment (ITE) [10,11]. In recent years, advanced techniques like Model Predictive Control (MPC) and Deep Reinforcement Learning (DRL) have gained attention for their potential to minimize energy consumption while maintaining reliable ITE operation. However, real-world deployments of MPC and DRL remain limited, largely due to the difficulties in accurately predicting power consumption and data hall temperatures. Predicting temperature, which is crucial for ensuring ITE reliability, is particularly challenging because of the nonlinear thermal dynamics involved. These dynamics are influenced by variable server workloads, changing outdoor conditions, airflow patterns, and the delayed response of cooling systems [12].

1.2. Building thermal Modeling

Data center thermal modeling is critical for optimizing energy efficiency, managing heat loads, and ensuring safe operations. Traditional thermal modeling methods, particularly physical approaches like Computational Fluid Dynamics (CFD) and Resistance-Capacitance (RC) models, are widely used. CFD tools such as ANSYS Fluent and OpenFOAM are commonly employed to optimize airflow and heat management by simulating air distribution and temperature fields, identifying hotspots, and improving cooling efficiency through the solution of governing equations [13,14]. However, while CFD provides high-fidelity results, it is computationally expensive, making real-time control impractical. Methods such as Proper Orthogonal Decomposition (POD) have been applied to reduce these costs, but they remain too slow for real-time prediction and control. In addition to CFD, the resistance-capacitance (RC) model, a simplified approach for simulating heat transfer and storage, is also widely used for thermal modeling in buildings and data centers [15]. It analogizes thermal resistance (*R*) and capacitance (*C*) to electrical circuits, simulating temperature changes and heat exchange over time. By modeling heat flow between components like walls, windows, and floors, the RC model helps predict temperature changes and assess building energy performance. The RC model is computationally efficient and easy to implement, making it suitable for large-scale simulations and real-time data analysis. However, it has limitations in accuracy, particularly in complex scenarios, as its simplifications may overlook detailed heat transfer and dynamic or nonlinear behaviors. It may also struggle to capture transient conditions and intricate thermal interactions between building components.

With advances in machine learning (ML), especially the rise of deep learning (DL), researchers increasingly use Neural Networks (NN) for building energy modeling [16,17]. Applying ML to predict data center

room temperatures offers several benefits. First, according to the universal approximation theorem [18], NNs have strong approximation capabilities, allowing them to model complex physical systems effectively. Second, NNs can learn from historical data, improving prediction accuracy over time. Third, NNs enable rapid predictions, essential for real-time control. Despite these advantages, the generalizability of data-driven models remains a concern. Generalization refers to a model's ability to perform well on unseen data. While extensive datasets help address this issue in fields like natural language processing or image recognition, data centers and buildings, with relatively short lifespans, often lack sufficient high-quality data for effective ML training. As a result, data-driven methods for building energy prediction are primarily used in research rather than real-world applications.

1.3. Physics-informed neural networks

To harness the generalization issue of neural networks, researchers have been actively investigating methods to integrate prior knowledge, particularly knowledge of underlying physical laws, into NN to enhance their performance and facilitate training. This research trend emerged several years ago with the introduction of physics-guided machine learning, which involves the development of network architectures that explicitly represent known physical systems. These models, commonly referred to as Physics-Informed Neural Networks (PINN), incorporate equations, such as Partial Differential Equations (PDEs), as an integral component of the NN structure [19]. By incorporating prior knowledge of fundamental physical laws, PINN serves as a regularization mechanism during the training process. They impose constraints derived from these physical laws, thereby restricting the solution space, and improving the accuracy of function approximation. Consequently, this approach enhances the information content extracted from the available data, enabling the learning algorithm to capture the correct solution and generalize effectively, even in scenarios with limited training examples.

PINN has garnered considerable interest as a research tool for building energy modeling due to its numerous advantages [20–23]. Drgona et al. first presented a physics-constrained deep learning method for developing control-oriented models of building thermal dynamics, demonstrating that the proposed data-driven approach can learn interpretable dynamical models with high accuracy and generalization over long-term prediction horizons [24]. Büning et al. compared a physics-informed building energy model to machine learning models based on Random Forests and Input Convex Neural Networks. Their findings demonstrated that the physics-informed model has a lower computational burden and superior sample efficiency compared to other machine learning-based models [23]. Gokhale et al. proposed two new variants of PINN architectures for the task of control-oriented thermal modeling of buildings, demonstrating that training these architectures is

data-efficient and requires less training data compared to conventional non-physics-informed neural networks [25]. Natale et al. introduced a novel architecture called Physically Consistent Neural Network (PCNN), which differentiates quantities by their physical clarity. Clear and established relationships are processed in a physics module, ensuring accurate modeling based on scientific principles. In contrast, less defined interactions are handled in a black box module using data-driven methods to manage uncertain relationships flexibly. They demonstrated that the PCNN model outperformed a traditional physics-based resistance-capacitance model, achieving up to 40 % higher accuracy in predictions over three days [20,26]. PCNN is a variant of PINN, but it differs significantly from the standard PINN model. The standard PINN integrates physical equations as soft constraints by incorporating them into the neural network's loss function. In contrast, PCNN imposes strong constraints that rigorously control the relationships between inputs and outputs through positive coefficients and designated positive and negative signs, thereby establishing a more tightly regulated modeling framework. Strong constraints are generally more credible and explainable than soft constraints, making them better suited for real-life applications, such as serving as the foundational predictive model for Model Predictive Control. In environments like data centers, the objective is to enhance energy efficiency while ensuring servers operate effectively. However, safety takes precedence over energy conservation. Protecting both equipment and data is crucial for sustaining operational integrity and reliability in the long term.

1.4. Gaps and contributions

The literature review and analysis highlight the advantages of the PCNN model, including its physical consistency, enhanced generalization, and improved prediction accuracy, making it a promising tool for temperature prediction in intelligent data center control. However, the traditional PCNN model faces two significant challenges: (1) the reliance on predetermined coefficients, which requires extensive empirical trial and error, and (2) the static nature of these coefficients, which limits the model's ability to accurately and adaptively represent the complex dynamics of real-world thermal environments. These two challenges are analyzed in detail in Sections 2.2 and 2.3 of the paper. Additionally, there is currently no existing literature that applies PINN or PCNN specifically to the domain of data center thermal modeling and prediction.

Based on the gaps, the contributions of this study can be succinctly summarized as follows:

1. Introducing the Adaptive Physically Consistent Neural Networks (A-PCNN) framework, which utilizes Neural Networks (NN) with Soft-plus activation functions instead of traditional predetermined and fixed coefficients. This enhancement strengthens generalization capabilities while maintaining adherence to physical constraints.
2. Our proposed A-PCNN exhibited superior prediction performance compared to the traditional PCNN framework in a comprehensive six-month evaluation utilizing real data center datasets.
3. Our framework displays remarkable versatility. Whether utilizing an MLP or LSTM model as the base, the Adaptive framework consistently demonstrates superior prediction accuracy across a range of time scales, from 15 min to 7 days. This advantage becomes increasingly pronounced over longer forecasting periods.
4. Pioneering application of Physics-Informed Neural Networks framework in data center thermal dynamics modeling, utilizing both real-world and simulation data center datasets for evaluation.

The structure of this paper is outlined as follows: In Section 2, we initiate by introducing a standard configuration of a data center facility, followed by the development of one traditional PCNN model and our proposed A-PCNN framework. Section 3 presents experimental setup for model evaluation. In Section 4, we conduct a detailed comparative

analysis and evaluation of the results obtained from our proposed A-PCNN approaches compared to classical PCNN. In the concluding Section 5, we offer a comprehensive summary of the significant findings uncovered in this study.

2. Methodology

To develop a thermal model for the data hall, a comprehensive understanding of the heat transfer processes within the facility is crucial. Therefore, our investigation begins with a thorough examination of the thermal environment of the data center. We then establish a baseline using the Physically Consistent Neural Networks (PCNN) introduced in Natale's paper. Building on the insights from this traditional model and the underlying physical laws, we propose our innovative model, A-PCNN, in this study.

2.1. Data center thermal environment

To develop a comprehensive thermal model for data centers, it is crucial to thoroughly understand the complex heat transfer phenomena involved. Fig. 1 illustrates a typical layout of a data center with a single data hall, which serves as the focal point of this study. Significant heat sources in the data center include the ITE (represented by Racks in Fig. 1), Uninterruptible Power Supplies (UPS), and Emergency Diesel Power Generators. Additionally, it is essential to consider supplementary heat sources and heat exchange mechanisms, such as those occurring between the data center and its external environment. **The primary cooling source in this context is provided by the HVAC system.** In practical applications, the focus is on controlling the temperature within the data hall, **where the IT equipment is located, as it directly impacts the performance and reliability of the IT infrastructure.** In the layout like Fig. 1 where UPS and Emergency Diesel Power Generators are in separate rooms, thus the main influence parameters in the data hall include the **ITE, HVAC system, and outdoor environment.** By accounting for these sources, the thermal model aims to accurately capture and represent the complex interactions and heat transfer processes within the data hall. Based on the above energy analysis, for timestep k , the variables (X_k, Y_k) can be expressed as:

$$X_k = \{T_k^{DH}, T_k^{SA}, v_k^{SA}, P_k^{ITE}, T_k^{out}, H^{out}\}, \quad (1)$$

$$Y_k = \{T_{k+1}^{DH}\}, \quad (2)$$

where T_k^{SA} represents the supply air temperature from the cooling system, v_k^{SA} represents the mass flowrate of supply air, P_k^{ITE} represents the ITE equipment power consumption, T_k^{out} represents the ambient environment temperature, T_k^{DH} and T_{k+1}^{DH} denote the temperatures of the Data Hall (DH) at timestep k and $k + 1$. Given the uneven temperature distribution across the data center hall, the return air temperature, which is the temperature of the air circulating back to the cooling system, is commonly used to gauge the overall room temperature. Additionally, for different data center layouts, such as when the UPS and Power Generators are located within the data hall, Eq. (1) should account for the heat generated by the UPS and Power Generators. Model construction must be tailored to the specific configuration of the data center.

2.2. Previous PCNN model

Natale et al. have introduced a machine learning-based thermal model named Physically Consistent Neural Networks (PCNN) [26]. By incorporating physical information into neural networks, PCNN enables the precise definition of both positive and negative relationships between inputs and outputs. Their approach involved three key steps: Initially, thermal analysis was conducted to establish the physical correlation between inputs and outputs. Subsequently, quantities with clear physical relationships were integrated into the physical module, while

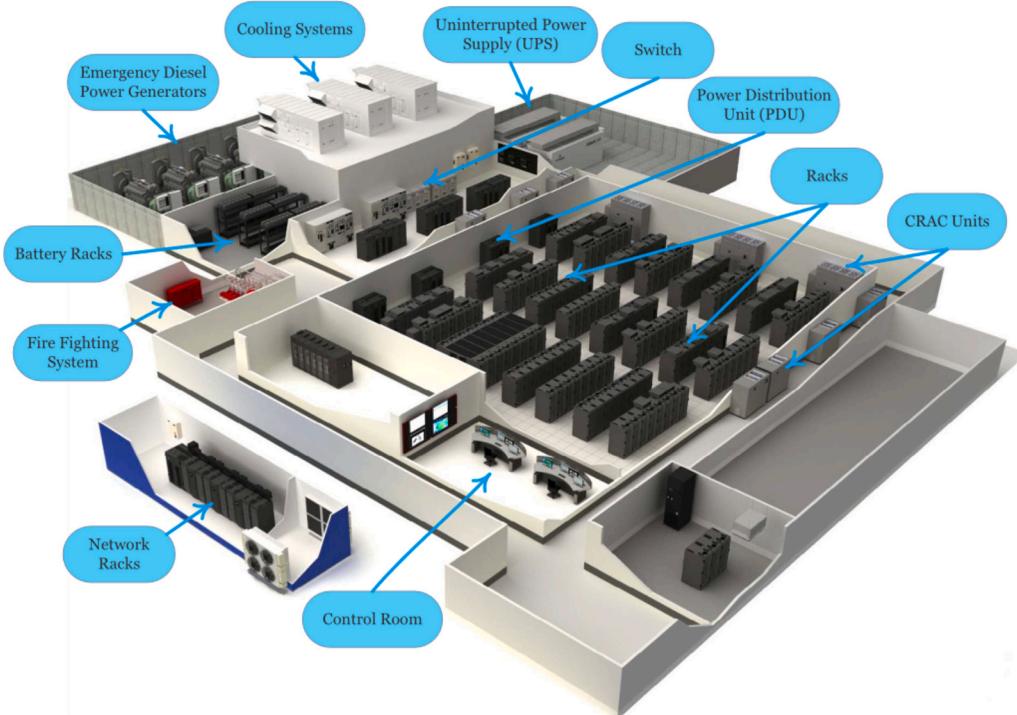


Fig. 1. Typical layout of a data center plant. (Adopted from [27]).

those lacking such relationships were placed into the black box module. Positive coefficients were allocated to denote typical relationships between inputs and outputs. Both positive and negative signs were employed alongside the coefficients to define these relationships. The final step entailed determining the appropriate coefficients through trial and error, culminating in the completion of the model. This process yielded the mathematical relationship representing the system. PCNN has previously been utilized for thermal modeling of residential buildings. In this study, we apply the PCNN approach to thermal modeling of data centers.

In a data center setting, the IT equipment acts as the main generator of heat within the data hall. Simultaneously, the HVAC (Heating, Ventilation, and Air Conditioning) system functions to efficiently dissipate this heat, ensuring optimal operating conditions. Additionally, convective heat exchange occurs between the air in the Data Hall (DH) and various surfaces within the zone. Given these considerations and under specific assumptions, namely that (1) indoor room air is uniformly mixed, enabling a single temperature value to effectively represent the overall room temperature, and (2) minor heat sources such as human presence and lighting are neglected, the energy conservation equation can be articulated as follows:

$$m_{air}C_p \frac{dT^{DH}}{dt} = -\dot{Q}^{HVAC} + \dot{Q}^{ITE} + \sum_{i=1}^{N_{surface}} h_i^s A_i^s (T_i^s - T^{DH}), \quad (3)$$

where m_{air} represents the air mass in the DH, C_p represents the specific heat of air in the DH, h_i^s and A_i^s represent the convective heat transfer coefficient between air and zone surfaces and the area of the wall respectively. Based on assumption (3) the convection heat transfer rate between servers and air is infinite, we can deduce that the rate of heat transfer \dot{Q}^{ITE} from the IT equipment is equal to the power consumption P^{ITE} . Based on assumption (4), the zone supply air mass flow rate is exactly equal to the sum of the air flow rates leaving the zone through the system return air plenum and being exhausted directly from the zone. Both air streams exit the zone at the zone mean air temperature T^{DH} . We can deduce that the heat transfer rate between the cooling

system and air \dot{Q}^{HVAC} can be expressed as $\dot{m}C_p(T^{DH} - T^{SA})$, where \dot{m} represent the mass flowrate of supply air. Finally, based on assumptions 1–4, we obtain the energy conservation equation as follows:

$$T_{k+1}^{DH} = T_k^{DH} - \frac{dt}{m_{air}} \dot{m}(T^{DH} - T^{SA}) + \frac{dt}{m_{air}C_p} P_k^{ITE} + \frac{dt}{m_{air}C_p} \sum_{i=1}^{N_{surface}} h_i A_i (T_i^s - T^{DH}), \quad (4)$$

The two coefficients representing the influence of the HVAC system and IT equipment on the temperature of the Data Hall can be expressed as:

$$a = \frac{dt}{m_{air}}, b = \frac{dt}{m_{air}C_p}. \quad (5)$$

The term dt , contingent upon the time interval of the data, such as 1 min or 10 min, remains constant. Additionally, m_{air} is contingent upon room size and pressure, which are essentially constant. C_p is influenced by temperature and pressure and can be regarded as unchanging after consulting relevant tables. Hence, both a and b must be positive and unaltered. Building upon this, Natale et al. established these coefficients as constantly positive in their model, terming it “Physically Consistent”. The PCNN model for data center can be articulated as follows:

$$T_{k+1}^{DH} = T_k^{DH} + E_k + D_k, \quad (6)$$

$$E_k = -a(T_k^{DH} - T_k^{SA})v_k^{SA} + bP_k^{ITE}, \quad (7)$$

$$D_k = g(T_k^{DH}, T_k^{OA}, H_k^{OA}), \quad (8)$$

where E_k represents the physical module, which incorporates prior knowledge about thermal dynamics, such as the heat transfer between the data hall air, ITE, and the cooling system. In contrast, D_k represents the black-box module, which processes inputs that are excluded from the physical module due to either the lack of well-established physical formulations or the involvement of highly complex functions that resist analytical modeling. Combining Eqs. 6 through 8 yields:

$$T_{k+1}^{DH} = T_k^{DH} - a(T_k^{DH} - T_k^{SA})v_k^{SA} + bP_k^{ITE} + g(T_k^{DH}, T_k^{OA}, H_k^{OA}), \quad (9)$$

where a and b are two constant positive coefficients in the physical module E , both acquired through the process of learning from available data. The heat exchange between room air and the building's inner surface is modeled using neural networks, a substitution made necessary by the frequent lack of real-world data on building envelope temperatures. In contrast, parameters like outdoor dry-bulb temperature, wet-bulb temperature, and humidity are typically known or easily measurable. In the study introducing the PCNN model, the author determined the influence coefficient of the air conditioning system on room temperature as 6.01×10^{-3} . Additionally, the influence coefficients for the temperatures of both adjacent rooms and the external environment on the target room temperature were identified as 9.62×10^{-5} . These values were established through an empirical process of trial and error.

2.3. New A-PCNN model

From the preceding derivation, it becomes clear that the traditional PCNN model encounters two significant issues: (1) the use of predetermined coefficients requires a time-consuming empirical trial and error process, and (2) these coefficients remain fixed once training is completed. The first issue is straightforward, particularly in the case of large buildings with numerous rooms. Considering factors such as room size, the properties of various envelopes, and the sizes of heating and cooling sources, the values for each room were determined through an empirical process of trial and error. The second issue arises because, although Eq. (5) suggests that these coefficients should be positive and fixed, it's important to note that Eq. (4) is heavily reliant on four initial assumptions. This dependence may limit the model's adaptability and accuracy under varying conditions. In real-world settings like air-cooled data centers or residential buildings, temperature distribution is typically quite uneven, which complicates the accurate representation of room temperature with a single value. Consequently, assumptions (1) and (4) from the model may not fully capture these real-world scenarios. Although it's possible to deduce that heat transfer rate \dot{Q}_{HVAC}^{ITE} is associated with $\dot{m}C_p(T_k^{DH} - T_k^{SA})$, this relationship might not align perfectly due to various factors. Additionally, assumption (3) introduces further complications. Heat conduction occurs between the IT equipment and the air through convective heat transfer, a process that can be described as: $\dot{Q}^{ITE} = \bar{h}_k A^{ITE} (T_k^{ITE} - T_k^{DH})$. \bar{h}_k represents the average convection heat

transfer coefficient, A^{ITE} represents the heat exchange area between IT equipment and air, T_k^{ITE} represent the temperature of IT equipment. The \bar{h}_k fluctuates depending on factors such as the supply air flowrate of the cooling system, the utilization of fans in each server, and the temperature difference between the IT equipment and the surrounding air. Furthermore, the temperature difference between server and airflow is also subject to specific conditions. Consequently, in an actual data center environment, although the rate of heat transfer \dot{Q}^{ITE} is correlated with the power consumption of ITE, it doesn't exactly match the power consumption itself. The heat transfer analysis highlights that the traditional PCNN model, which relies on four initial assumptions with fixed coefficients, demonstrates limited effectiveness in accurately and adaptably representing the complex dynamics of the actual data hall thermal environment. This model's reliance on assumptions restricts its ability to capture the nuanced variations and transient behaviors characteristic of real-world settings.

Based on the significant limitations encountered by the traditional PCNN model, we introduce a novel architecture in this study called Adaptive Physically Consistent Neural Networks (A-PCNN). The framework of the A-PCNN is illustrated in Fig. 2. We propose employing the Neural Networks with Softplus activation function method as a replacement for traditional predetermined and fixed coefficients in the framework. This approach is driven by two main factors. Firstly, Neural Networks excel in addressing highly nonlinear problems. With ample data, they can identify suitable coefficient values across various scenarios, thereby greatly improving the model's generalization capability. Secondly, the Softplus function, characterized by its smooth and nonlinear nature, serves as an activation function designed to constrain outputs to exclusively positive values. Softplus is defined as: $\text{Softplus}(x) = \log(1 + e^x)$, where x denotes input. Leveraging Softplus allows us to effectively constrain the relationship between input and output in a physically consistent manner while also boosting the model's ability to generalize in nonlinear dynamic environments. By integrating Neural Networks with Softplus, we eliminate the trial-and-error approach that was previously required to determine suitable preset coefficients. Simultaneously, this strategy facilitates the automatic learning and adaptation of coefficients across various scenarios, significantly enhancing the model's flexibility and applicability in dynamic environments. Finally, our proposed A-PCNN can be outlined as follows:

$$T_{k+1}^{DH} = T_k^{DH} + E_k + D_k, \quad (10)$$

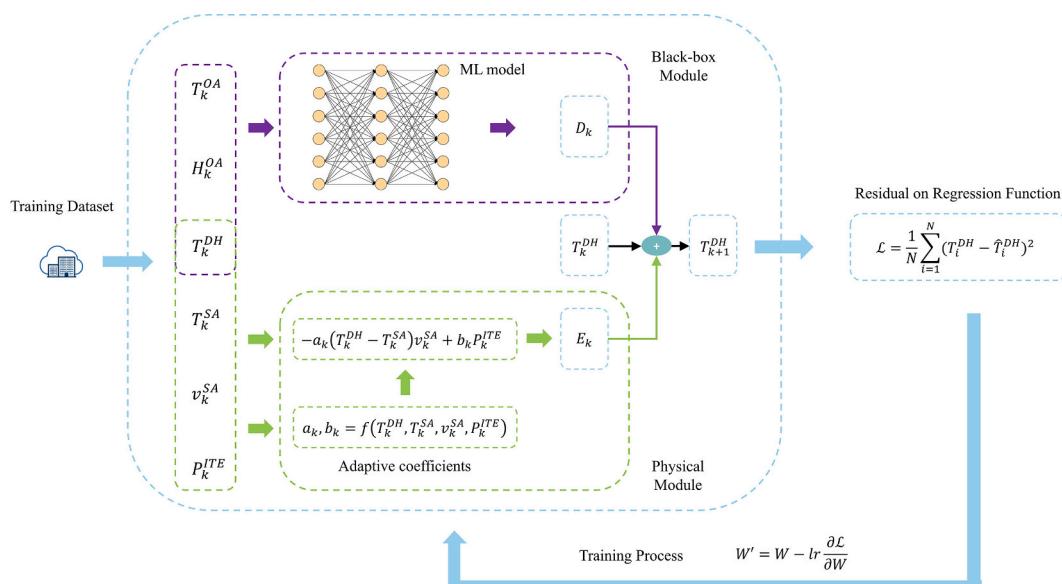


Fig. 2. Our Proposed A-PCNN Framework.

$$D_k = g(T_k^{DH}, T_k^{OA}, H_k^{OA}), \quad (11)$$

$$a_k, b_k = f(T_k^{DH}, T_k^{SA}, v_k^{SA}, P_k^{ITE}), \quad (12)$$

$$E_k = -a_k(T_k^{DH} - T_k^{SA})v_k^{SA} + b_kP_k^{ITE}, \quad (13)$$

where T_k^{DH} represents the data hall temperature at timestep k . T_{k+1}^{DH} denotes the predicted data hall temperature at timestep $k+1$, as the sum of the current temperature T_k^{DH} and two components: E_k from the physical module and D_k from the black-box module. E_k captures known physical dynamics, while D_k accounts for complex behaviors not modeled by the physical equations. The function $g()$ represents the black-box module, which uses machine learning models to process its inputs and generate D_k . The function $f()$ represents a fully connected artificial neural network within the physical module, activated by Softplus, which ensures the generation of two consistently positive coefficients a_k and b_k . Eq. (13) defines E_k , the output of the physical module, as the sum of two terms. The first term, $-a_k(T_k^{DH} - T_k^{SA})v_k^{SA}$, models the heat transfer between the data hall air and the supply air from HVAC system, scaled by a_k . The second term, $b_kP_k^{ITE}$, represents the heat contribution from ITE power, scaled by b_k . These modules run in parallel, and their outputs are combined for the final prediction. Both networks are trained simultaneously using the Mean Squared Error (MSE) loss function to improve prediction accuracy. The mathematical proof demonstrating the physical conservation properties of A-PCNN is provided in Appendix B. The theoretical analysis and experimental results regarding the selection of activation functions for adaptive coefficients are detailed in Appendix C.

2.4. Base ML models

Previously, both the established PCNN framework and our newly proposed A-PCNN were introduced. Beyond these frameworks, it is crucial to assess the foundational machine learning models that form the basis of these two frameworks. Specifically, we focus on the Multilayer Perceptron (MLP) and Long Short-Term Memory (LSTM) models, as shown in Fig. 3. The MLP is an artificial neural network with several fully connected layers, using nonlinear activation functions to model intricate data patterns, making it effective for classification, regression, and pattern recognition. Conversely, the LSTM model, a variant of recurrent neural networks (RNN), excels at capturing long-term dependencies in sequential data, addressing the vanishing gradient

problem of RNN model through specialized memory cells for enhanced data retention. This makes it particularly suitable for time series forecasting. Given the broad application of these two models in temperature prediction [25,26,28], we respectively adopt MLP and LSTM architectures as base models to evaluate the versatility of A-PCNN.

3. Experimental setup

3.1. Datasets

To assess and evaluate the performance of the machine learning models, it is crucial to acquire suitable data for both training and testing purposes. The availability of high-quality and representative data plays a pivotal role in ensuring the accuracy and reliability of the model's predictions. In this study, we leverage the Air Free-Cooled Tropical Data Center (TDC) situated in Singapore [29]. The TDC features an infrastructure consisting of 12 server racks and two Precision Cooling Units (PCUs). The configuration of each PCU includes a Computer Room Air Conditioning (CRAC) fan, two compressors, an evaporator fan, and a condenser fan. The two PCUs operate in an alternating and intermittent manner. The dataset is collected from 2022 to 09-01 00:00:00 to 2023-02-28 23:59:00 and processed into a sample time of 30 s. Therefore, there are 521,280 samples. Detailed information about this dataset is suggested to refer to Ref. [29, 30].

The dataset for this research comprises a variety of features, including power consumption metrics for each rack, supply air temperature and velocity from the PCUs, return air temperature, and the external environment's temperature and humidity. In the event of missing data, we first verify data availability at the time steps immediately preceding and following the gap. If data is available, we compute the average of two values. Otherwise, we default to using the overall average of all data of the same type. To enhance data analysis efficiency, we utilized a 1-min analysis interval. This temporal resolution was chosen to optimize the data utility, enabling more accurate predictions, and facilitating timely adjustments to the cooling systems. During the data preprocessing phase, we normalized the dataset to a range between 0.0 and 1.0 to standardize the input for machine learning models. The comprehensive dataset contains six months, with the five months extracted for training and the rest one month reserved for test. Within the training data, an 80 %–20 % split is employed to partition it into training and validation sets.

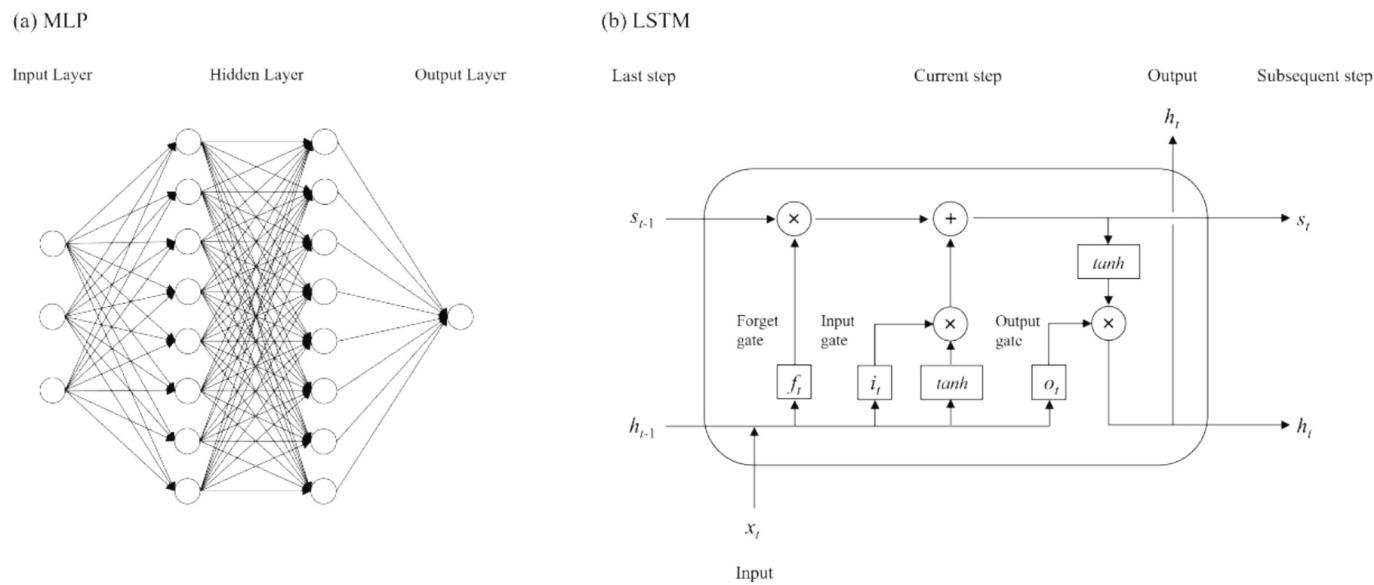


Fig. 3. Illustration of Two Base Machine Learning Models: (a) Multilayer Perceptron (MLP) and (b) Long Short-Term Memory (LSTM).

3.2. Implementation details

Our architecture is designed for one-step ahead predictions. For extending these predictions over longer horizons, we employed a recursive strategy. In this approach, the model's output is reinputted into the system to generate subsequent forecasts, facilitating a multi-step forecasting capability. This technique proves particularly beneficial for sophisticated control methods such as DRL and MPC. For test configurations with a prediction horizon of 15 min, the model is expected to make predictions for 15 steps. Similarly, for test configurations with a prediction horizon of 7 days, the model should iterate for 10,080 steps.

Furthermore, the warm start strategy is consistently implemented across all models, with variations contingent upon the base mode [26]. In the case of LSTM-based models, the return air temperature data for the initial 10 min (10 steps) is supplied. Conversely, for MLP-based models, only the return air temperature at the initial moment (1 step) is provided. This divergence stems from the architecture of LSTMs, which necessitates the learning of both hidden states (h) and cell states (c), as shown in Fig. 3.

The network architectures were fine-tuned along with the hyperparameters such as learning rate and choice of activation functions. Appendix A provides a comprehensive list of the selected hyperparameters for all architectures. These hyperparameter values were determined by minimizing the Mean Square Error, which serves as the metric for quantifying deviations in the loss function. For each architecture configuration, we trained 10 models with seeded randomness, and the reported results are based on the mean values derived from these 10 models. This method minimizes initialization bias, prevents overfitting, and enhances the reliability of the results.

4. Results and discussion

In this section, we conduct a comprehensive comparative analysis to assess the predictive accuracy of four models—comprising two frameworks and two foundational machine learning models—across various training datasets and prediction horizons, utilizing a six-month real data center dataset. Additionally, we perform visual edge testing and subsequently utilize a simulation dataset from EnergyPlus to further validate our models.

4.1. Evaluation metrics

To assess the predictive accuracy of models, we utilize two widely used evaluation metrics: Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). These metrics are fundamental in quantifying the accuracy of the predictions made by the models. The MAE metric quantifies the average magnitude of the errors in a set of predictions. It is defined by the formula:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |Y_i - \hat{Y}_i|, \quad (14)$$

where Y_i and \hat{Y}_i are actual and predicted values, i is the data index, and N is the total number of data points. The RMSE metric provides the square root of the average of these squared differences, thus offering a measure of error in the same units as the original data. It is computed as:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_i - \hat{Y}_i)^2}. \quad (15)$$

Each of these metrics has its advantages and limitations. MAE provides a straightforward measure of the average magnitude of errors between predicted and actual values, offering an intuitive understanding of prediction accuracy. RMSE, on the other hand, gives a higher weight to larger errors by squaring the differences before averaging, thus providing a more sensitive measure of the variability in prediction

accuracy, especially useful for identifying outliers. Together, MAE and RMSE offer a comprehensive insight into the model's predictive accuracy, with RMSE highlighting larger errors and MAE presenting a general error benchmark.

4.2. Performance vs. training data size

The motivation for using physics guided neural networks is to leverage prior knowledge to train models faster and more efficiently. To validate this, the first set of experiments delved into examining the influence of training data size on the efficacy of four ML models. These models underwent training utilizing real-world data sets of differing sizes, extracted from the primary training set. Subsequently, each model underwent testing, with MAE and RMSE serving as the performance metric, conducted on a 30-day unseen test dataset. Our methodology comprised two distinct testing configurations, each tailored to specific prediction horizons. The prediction horizon is pivotal in defining the extent of future forecasts incorporated into the control optimization process. Extending the prediction horizon furnishes a broader perspective of future outcomes, albeit at the risk of accumulating larger errors or deviations from the targeted behavior. To fully evaluate the predictive performance of four models, testing was conducted over prediction horizons of 15 min (15 steps) and 7 days (10,080 steps). The selection of a 15-min minimum prediction interval is guided by the standard operational cycle of most data center cooling systems, which utilize MPC and DRL for management and typically undergo adjustments every 15 min. The decision to set a 7-day maximum prediction interval is influenced by the constraints of the test dataset, which covers a duration of 30 days. To ensure an accurate comparison, identifying multiple extended sequences in the unseeded dataset is crucial. However, extending the timeframe significantly reduces the quantity of long-term data available. Consequently, after careful consideration, a seven-day period has been chosen for this analysis. While longer intervals would theoretically provide a more in-depth analysis, practical dataset limitations necessitate this choice.

Fig. 4 showcases the performance of our models across different training data sizes within a real-world data scenario for a 15-Minute Prediction Horizon. Our proposed Adaptive framework significantly outperforms the original models across all training data sizes. Notably, after five months of training, the A-PC-MLP model achieves a Mean

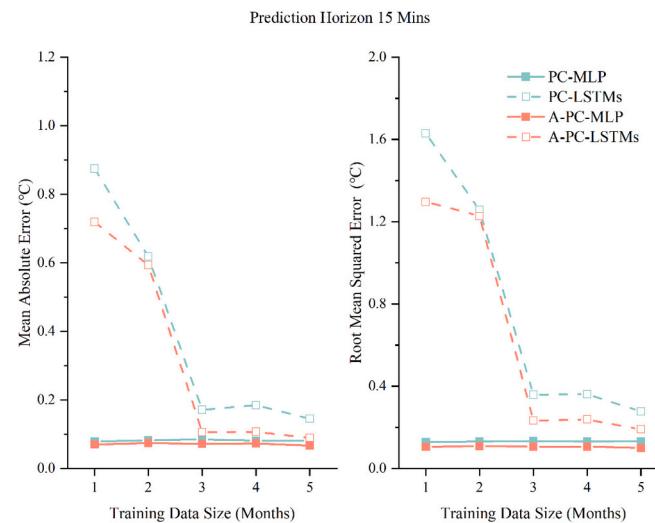


Fig. 4. Comparing MAE and RMSE Across Different Training Dataset Sizes for a 15-Minute Prediction Horizon. Red Lines for Adaptive Framework, Blue Lines for Traditional PCNN Framework. The Plots Represent the Average Values Derived from 10 Trained Models. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Absolute Error (MAE) that is 17.3 % lower and a Root Mean Square Error (RMSE) that is 24.2 % lower than the PC-MLP model. Simultaneously, the A-PC-LSTM model records a MAE that is 38.6 % lower and a RMSE that is 31.4 % lower than that of the PC-LSTM. Remarkably, within this prediction interval, MLP-based architectures consistently surpass LSTM-based models. This performance gap originates from the complex structure of LSTMs, which typically require more extensive datasets to effectively tune parameters. However, this difference in performance diminishes as the size of the training dataset increases, with a notable convergence observed when the dataset size grows from 1 month to 3 months.

Fig. 5 depicts the performance of models across a range of training data sizes in a real-world context for a 7-day prediction horizon. It is evident that the error rates for all four models increase relative to the 15-min prediction horizon, a phenomenon attributable to cumulative error, which will be explored in the following section. Significantly, for this extended prediction period, the A-PC-MLP model reliably outperforms the other three models. When considering large datasets (of three months or more), the A-PC-LSTMs model demonstrates superior performance over the PC-MLP and PC-LSTMs models. This underscores the effectiveness of the Adaptive approach in reducing model errors and enhancing performance, whether using an MLP or LSTM as the base model. Specifically, after five months of training data, the A-PC-MLP model achieves a MAE that is 79.2 % lower and a RMSE that is 79.6 % lower compared to the PC-MLP. Concurrently, the A-PC-LSTMs model records a MAE that is 40.6 % lower and a RMSE that is 36.6 % lower than that of the PC-LSTMs.

In summary, our proposed A-PCNN framework consistently outperforms traditional PCNN across all training data sizes, whether the underlying models are MLP or LSTM. This demonstrates the Adaptive framework's high data utilization efficiency and its robust versatility.

4.3. Performance vs. prediction horizons

To comprehensively assess the predictive performance of four models across various prediction horizons, we compared their performance over ranges from 15 min to 7 days. The rationale for selecting these forecast horizons was detailed in the previous section (Sectio 4.2). The models were trained on a dataset spanning five months and evaluated using a 30-day unseen dataset. **Fig. 6** illustrates the model

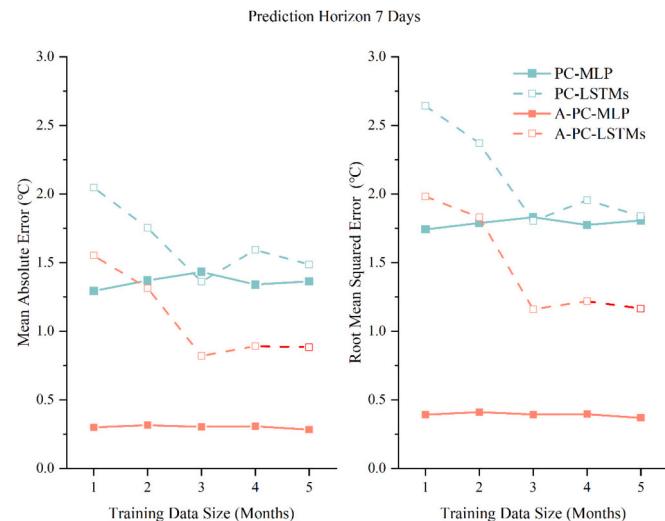


Fig. 5. Comparing MAE and RMSE Across Different Training Dataset Sizes for a 7-Day Prediction Horizon. Red Lines for Adaptive Framework, Blue Lines for Traditional PCNN Framework. The Plots Represent the Average Values Derived from 10 Trained Models. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

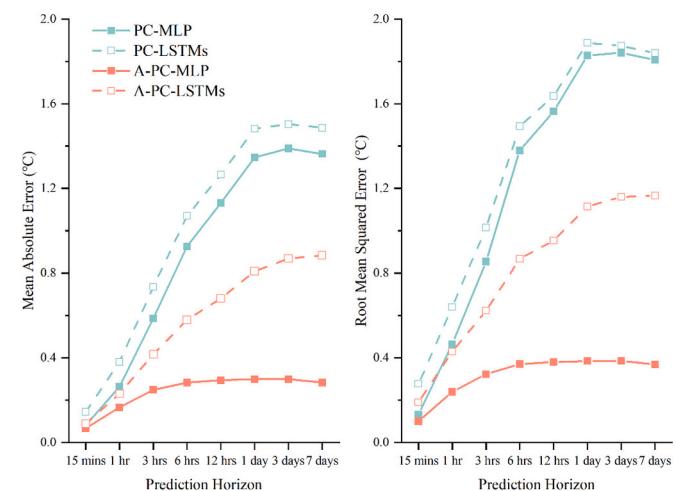


Fig. 6. Comparison of MAE and RMSE of Four Models Across Various Prediction Horizons. Red Lines for Adaptive Framework, Blue Lines for Traditional PCNN Framework. The Plots Represent the Average Values Derived from 10 Trained Models. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

performance across various prediction horizons within the real-world data scenario. All four models demonstrate an increase in both MAE and RMSE across prediction horizons, indicative of cumulative error—a well-documented challenge in the field [31]. This is attributed to the recursive forecasting strategy, wherein the model's output is recycled as input to generate future forecasts, thus enabling multi-step forecasting. In scenarios where the prediction horizon spans 7 days, the model undergoes 10,080 iterations. Such extensive cycling invariably leads to the accumulation of errors over time. It's evident that the models under the Adaptive framework surpass traditional PCNN models in performance across every prediction horizon, with this advantage becoming more pronounced over longer periods. Notably, for a seven-day prediction horizon, the A-PC-MLP model achieves a MAE that is 79.2 % lower and a RMSE that is 79.6 % lower compared to the PC-MLP. Concurrently, the A-PC-LSTMs model records a MAE that is 40.6 % lower and a RMSE that is 36.6 % lower than that of the PC-LSTMs. In conclusion, our evaluation demonstrates that the proposed Adaptive framework surpasses the traditional PCNN approach across all prediction horizons. This advantage becomes increasingly pronounced over longer forecasting periods.

4.4. Visual edge testing

To conduct a thorough evaluation of the models' performance, we intentionally chose the most challenging segments from unseen datasets. This approach was strategically aimed at edge case testing, with the goal of evaluating how the models respond and adapt to extreme or atypical conditions. **Fig. 7** displays a seven-day span of model input data from the unseen dataset, selected specifically for Visual Edge Case Testing. Meanwhile, the red line in **Fig. 8** illustrates the temporal changes in actual return air temperature. Two primary reasons guided our choice of this data sequence for testing. First, as illustrated in **Fig. 8**, we noted a significant decline in the actual return air temperature between days 1 and 4, with values falling below 20 °C. This represents a condition absent from the training dataset, which documented a minimum temperature of 21.75 °C. Additionally, the actual return air temperature exhibited considerable variability over the seven-day period, fluctuating between a high of 33.90 °C and a low of 18.03 °C. This variability underscores the difficulty of making accurate predictions.

Fig. 8 displays the comparative performance of the four models over a seven-day forecast horizon. Firstly, the two models within the Adaptive framework demonstrate greater consistency with the experimental values, particularly during the 0–1 and 5–7 day periods. In these two

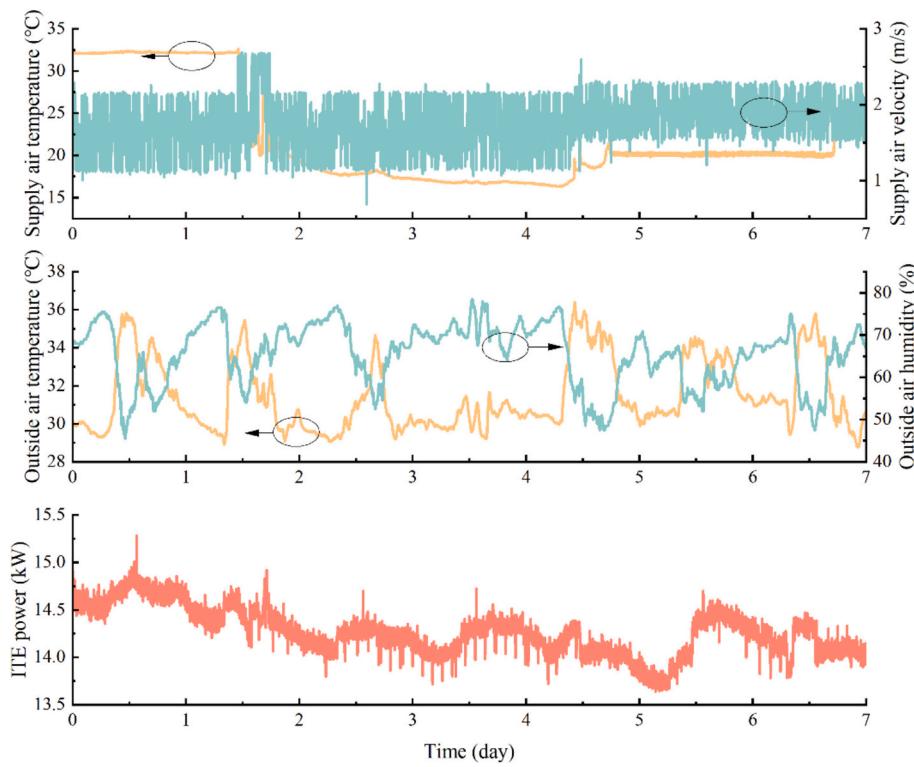


Fig. 7. Selection of 7-Day Unseen Dataset for Enriched Analysis and Visualization.

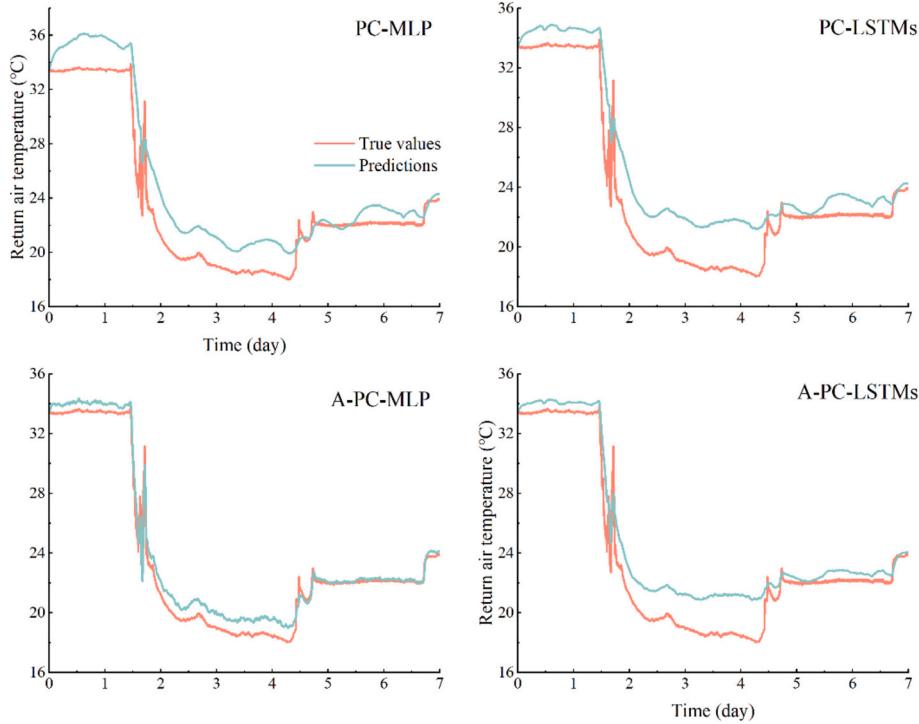


Fig. 8. Comparative of Four Models' Prediction Performance Against Experimental Data Across a Seven-Day Forecast Horizon. Experimental Values Illustrated by the Red Line, with Predicted Values Depicted by the Blue Line. The Plots Represent the Average Values Derived from 10 Trained Models. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

intervals, despite minimal changes in supply air temperature as shown in Fig. 7, the predictions from the traditional PCNN models exhibit significant fluctuations, failing to align with the experimental data. This discrepancy is attributed to the influence of their fixed coefficients. The

observed trend is influenced by ITE power variations; for instance, during days 0–1, an initial increase followed by a decrease in ITE power is mirrored in the return air temperature predictions by the two traditional PCNN models, with a similar pattern noted for days 5–7. This

highlights how fixed coefficients can result in imprecise model predictions. Secondly, when encountering data that falls outside the training range—particularly when the return air temperature drops below 21.75 °C, observed mainly between days 2–5—the models within the Adaptive framework consistently demonstrate superior accuracy when utilizing the same base ML model.

Fig. 9 depicts the temporal variation of coefficients a and b across the four models. In the traditional PCNN models (PC-MLP and PC-LSTM), the initial values of a and b in PCNN models were both set to 0.01, determined through repeated trial and error based on the TDC dataset. After training, these coefficients remain constant throughout the duration, reflecting the static nature of traditional PCNN models. Specifically, the PC-MLP model shows the constant a of approximately 1.086×10^{-2} and the constant b of approximately 1.005×10^{-2} , while the PC-LSTM model maintains the coefficient a of approximately 1.009×10^{-2} and the constant b of approximately 0.995×10^{-2} . In contrast, the A-PCNN models (A-PC-MLP and A-PC-LSTM) do not require pre-set coefficients, as these are autonomously learned during the training process. These models exhibit dynamic behavior, with their coefficients adjusting in response to varying thermal conditions over time. For instance, in the A-PC-MLP model, coefficient a begins at approximately 0.052, gradually decreases, and stabilizes around 0.050. Similarly, the A-PC-LSTM model demonstrates a dynamic adjustment in its coefficients, responding to changes in the thermal environment.

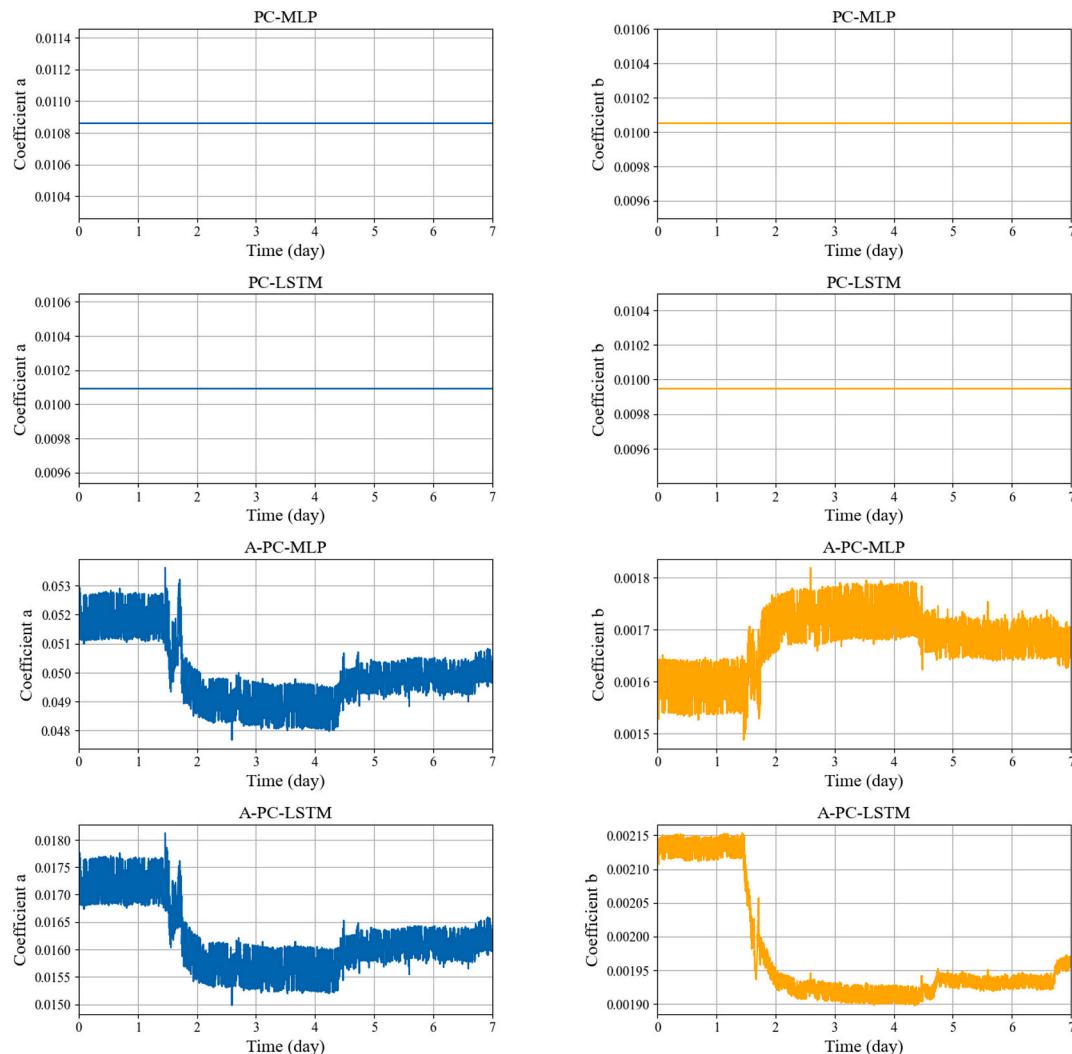
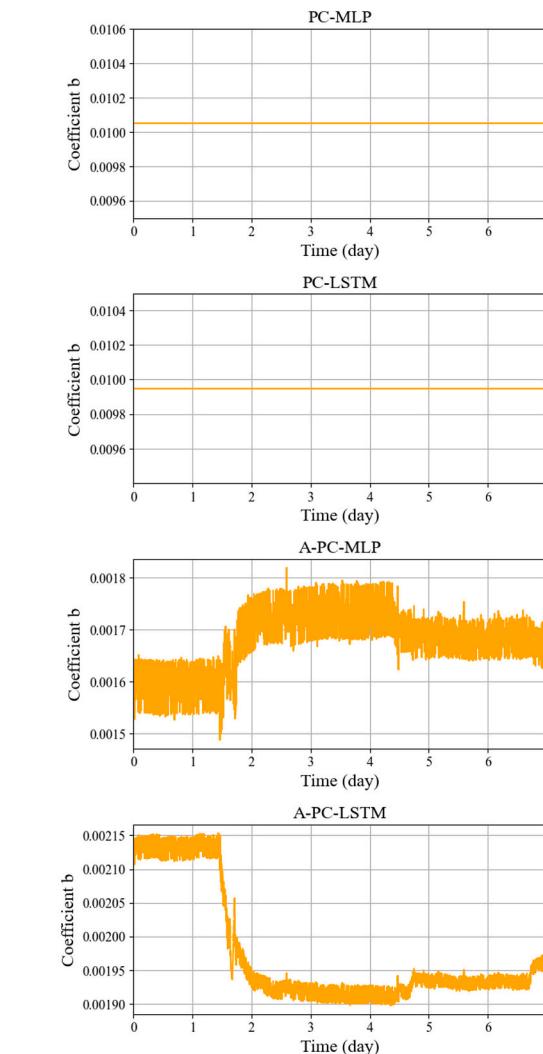


Fig. 9. Variation of Coefficients Across the Four models on an Unseen Seven-day Dataset. The Blue Line Represents Coefficient a , while the Yellow Line Represents Coefficient b . (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

In summary, the visual analysis underscores the limitations of traditional PCNN models, particularly their reliance on static coefficients, which must be pre-determined through repeated trial and error specific to the dataset. This method not only increases computational costs but also leads to suboptimal predictive performance. In contrast, our proposed adaptive and dynamic coefficients effectively resolve these issues, reducing costs and significantly improving model performance.

4.5. Evaluate using simulation dataset

To further assess and evaluate the performance of our proposed framework, we aim to use additional real-world data center datasets. However, the challenges of collecting real data from data centers, coupled with the need to maintain data confidentiality, have limited our access to just one publicly available, high-quality real-world data center dataset: TDC, which has previously been employed in studies. In the absence of real data, researchers often rely on data center modeling software to generate synthetic data for their studies [32]. A widely used software tool for data center modeling is EnergyPlus, which provides various templates for data center simulations. In this study, we employ a standard single-zone data center model from EnergyPlus, depicted in Fig. 10 [33]. The prototype energy model for data centers has been developed by Sun et al. This model served to fill the existing void within



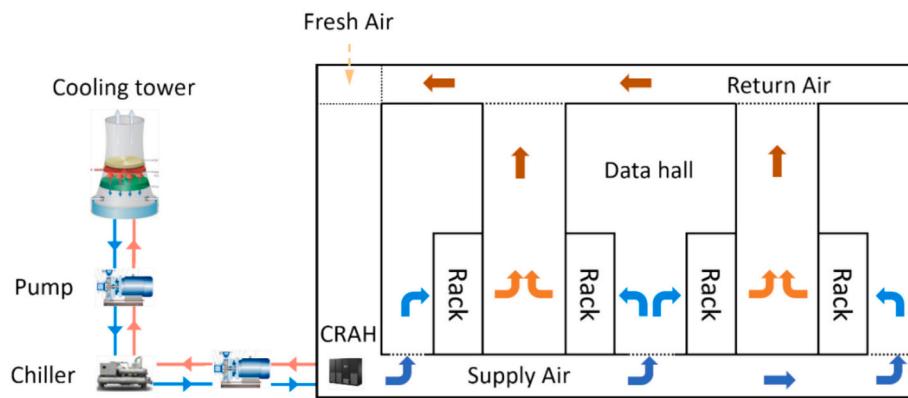


Fig. 10. Schematic Diagram of Single-Zone Data Center EnergyPlus Model.

the Department of Energy's (DOE) suite of commercial prototype building models by introducing energy models that are specifically designed for data centers. The room has a floor area of 557 m² and a height of 4.3 m. The cooling system utilized in the data hall is a chilled water system. Cooling is achieved through a single water-cooled chiller. The chiller system consists of two identical variable-speed pumps. A cooling tower, equipped with variable-speed fans, is employed to dissipate heat from the condenser water loop. Within the system, chilled water circulation is achieved through the combined operation of a variable speed pump and a constant speed pump arranged in series. Meanwhile, the condenser water pump functions at a constant speed. A computer room air handler (CRAH) plays a crucial role in supplying cool air to the data hall. The CRAH is equipped with a variable-speed supply air fan and an electric steam humidifier. Cool supply air is delivered to

the room via an underfloor plenum. The hot exhaust air from IT equipment is channeled into hot aisles and then directed upwards into a ceiling plenum, followed by a mixing box, before ultimately being recirculated back to the CRAH.

Given that the real data center dataset originates from Singapore, where the outdoor climate is consistently hot and humid, we expanded our analysis by incorporating weather data from two additional cities to assess the model's performance across different seasonal and weekly patterns. The chosen cities—Singapore, San Francisco, and Chicago—represent distinct climatic regions, offering a comprehensive evaluation of the model under diverse environmental conditions. Weather data for each city was sourced from the EnergyPlus Weather Dataset. For each climate, six months of data were generated, with five randomly selected months used for training and the remaining month allocated for

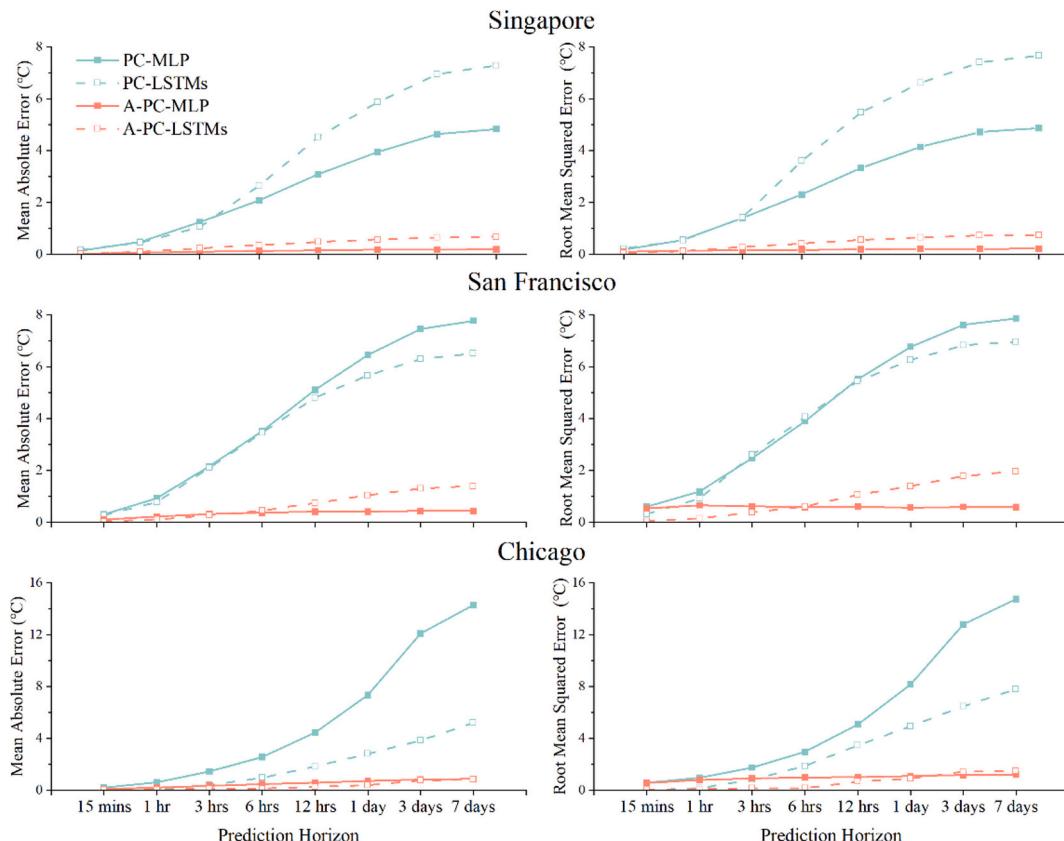


Fig. 11. Comparison of MAE and RMSE of Four Models using Dataset from EnergyPlus DC Model under Three Different Weather Conditions. Red Lines for Adaptive Framework, Blue Lines for Traditional PCNN Framework. The Plots Represent the Average Values Derived from 10 Trained Models. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

evaluation. All data was recorded at 1-min intervals.

Fig. 11 illustrates the model's performance across various prediction horizons using simulation data from an EnergyPlus template under three distinct weather conditions. The models within the Adaptive framework consistently outperform traditional PCNN models across all prediction intervals and climate zones, with their superiority becoming more pronounced over longer periods. This observation is consistent with the findings in Section 4.3, though the performance gap is even more substantial in this analysis. Notably, in San Francisco, for a seven-day prediction horizon, the A-PC-MLP model achieves a Mean Absolute Error (MAE) 94.3 % lower and a Root Mean Square Error (RMSE) 92.5 % lower than the PC-MLP model. Similarly, the A-PC-LSTM model records a MAE 78.5 % lower and an RMSE 71.7 % lower than the PC-LSTM. This widening performance gap is attributed to the use of fixed coefficients, a and b , which were initially set to 0.01 through iterative tuning with the TDC dataset. However, when applied to the EnergyPlus dataset, these predetermined coefficients result in higher error rates, underscoring the limitations of such fixed-parameter approaches. In contrast, the proposed Adaptive framework, which eliminates the need for preset and fine-tuned coefficients, achieves superior accuracy. Our evaluation across three distinct weather conditions further validates that the Adaptive framework significantly outperforms the traditional PCNN framework.

4.6. Limitations and future works

Firstly, Physically Consistent Neural Networks (PCNNs) are a subset of Physics-informed Neural Networks (PINNs), representing a specialized branch within the field [26]. Unlike conventional PINNs, which primarily incorporate physical equations into the loss function, PCNNs stand out with a distinctive architecture incorporating both a physics module and a black-box module. This integration, facilitated by fixed positive coefficients, enforces hard constraints between input and output variables, ensuring alignment with fundamental laws or equations. This marks a significant departure from the softer restrictions typically seen in traditional PINN structures. However, the versatility of PCNNs is also limited by this aspect. The transient energy conservation equation, when formulated as a first-order linear ordinary differential equation (ODE) using the lumped parameter method, addresses only a narrow scope of scientific and engineering problems. For more prevalent scenarios that involve partial differential equations (PDEs), the direct application of PCNNs is currently impractical. Similarly, the A-PCNN framework, which builds upon PCNNs, also cannot be directly applied in such cases at present. In the future, efforts could be made to optimize the structure of the PCNN to make it suitable for a broader range of scenarios.

Secondly, although our evaluations have incorporated both historical and simulated data from data centers, to fully assess the effectiveness of our A-PCNN framework, it is necessary to extend dataset to include various types of buildings. Expanding the scope of data will provide a more comprehensive understanding of how well the framework performs across different architectural and functional scenarios.

Furthermore, the A-PCNN framework stands out for its exceptional accuracy and its adherence to physical principles, positioning it as an ideal candidate for control-oriented room-level temperature models. These qualities make it highly suitable for integration into Digital Twins, where real-time forecasting and control are crucial for time-sensitive decision-making. Furthermore, combining A-PCNN with Model Predictive Control (MPC) and Deep Reinforcement Learning (DRL) could pave the way for intelligent prediction and control in building management systems. This integration would leverage the strengths of each approach, enhancing the capability to optimize and adapt to varying conditions dynamically.

For future applications of the A-PCNN models in data centers, we recommend a periodically updated approach. The frequency of retraining—whether daily, weekly, or monthly—should be tailored to the specific requirements of each data center task. In thermal management, environmental changes are typically infrequent, making continuous training unnecessary and potentially adding computational overhead. By adopting periodic updates, the model can effectively respond to environmental variations while minimizing computational costs, providing an efficient and practical solution.

5. Conclusions

This study introduces a tailored PINN framework, named A-PCNN, specifically designed for control-oriented thermal modeling of data centers. This innovative approach utilizes neural networks with Softplus activation functions, transitioning from the traditional method that relies on predetermined coefficients based on prior experience and fixed after training. This enhancement significantly reduces the time-consuming empirical trial-and-error process and enhances the generalization capabilities of the framework, while still adhering to physical constraints. Our findings highlight the superior performance of A-PCNN compared to the conventional PCNN framework. A case study using a six-month real data center dataset demonstrated that A-PCNN achieved a significant 17.3 % reduction in MAE for a 15-min forecasting period and an impressive 79.2 % reduction over a seven-day period, with MLP as the base model. Additionally, experimental results demonstrate that the framework consistently surpasses traditional methods in prediction accuracy across various time frames, ranging from 15 min to 7 days. This performance advantage holds true regardless of whether the underlying model is an MLP or an LSTM. Furthermore, the absolute difference in performance between the two models widens as the length of the forecast period increases. To highlight its adaptability to different data center datasets, we tested the framework on a dataset generated by the widely used simulation software, EnergyPlus. The results confirm that the Adaptive framework still outperforms the traditional one, reinforcing its effectiveness in diverse settings. A-PCNNs are entirely data-driven and do not require prior physical knowledge, making them ideal for integration into digital twins. This integration can enhance real-time monitoring, prediction, and control, significantly improving accuracy and efficiency in data centers and industries that utilize digital twin technologies.

CRediT authorship contribution statement

Dong Chen: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation. **Chee-Kong Chui:** Writing – review & editing, Supervision. **Poh Seng Lee:** Writing – review & editing, Supervision, Resources, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The code and data used in this article are available open source and can be found on: <https://github.com/ChenD777/Adaptive-PCNN.git>

Acknowledgements

This work is supported by Meta Platforms, Inc.

Appendix A. Model Hyperparameters

All variants of neural networks were implemented using Pytorch. Hyperopt, a widely recognized library for optimization, is used to navigate through diverse and intricate search spaces. Specifically, Bayesian Optimization serves as the technique of choice for hyperparameter tuning, offering a structured and efficient approach to identifying optimal settings. **Table A.1** and **Table A.2** present the selected hyperparameters for these models. All modules were trained with a batch size of 256 across 100 epochs. Throughout the training in all frameworks, the Adam optimizer was employed for its well-regarded efficiency in achieving convergence.

It is worth noting that LSTM model often incorporates a variety of network layers or configurations to better address complex real-world problems, enhancing their generalization and performance. Inspired by work [26,34], we utilize an encoder-LSTM-decoder structure, referred to here as LSTMs. This setup comprises an encoder and a decoder.

Table A.1
Hyperparameters for MLP Based Architectures.

Parameter	Values
Learning Rate	0.005
Hidden Layers	3
Neurons per layer	128
Activation Function	Sigmoid

Table A.2
Hyperparameters for LSTMs Based Architectures.

Parameter	Values
Learning Rate	0.001
Activation Function	2
Encoder Module	
Hidden Layers	1
Neurons per layer	128
LSTM	
Hidden Layers	2
Neurons per layer	128
Decoder Module	
Hidden Layers	2
Neurons per layer	128, 32

Appendix B. Physical Conservation of A-PCNN

From the energy conservation equation, we derive the following expression:

$$\frac{\partial T_{k+i}^{DH}}{\partial P_{k+j}^{ITE}} > 0, \forall 0 \leq j < i - 1 \quad (16)$$

$$\frac{\partial T_{k+i}^{DH}}{\partial T_{k+j}^{SA}} > 0, \forall 0 \leq j < i - 1 \quad (17)$$

$$\frac{\partial T_{k+i}^{DH}}{\partial v_{k+j}^{SA}} < 0, \forall 0 \leq j < i - 1 \quad (18)$$

Notably, Eq. (18) applies specifically when the supply air temperature is lower than the room temperature. Models may only be deemed compliant with physical laws upon the satisfaction of these specified expressions.

For the A-PCNN model, substituting Eqs. (11)–(13) into Eq. (10) yields:

$$T_{k+1}^{DH} = (1 - a_k v_k^{SA}) T_k^{DH} + a_k T_k^{SA} v_k^{SA} + b_k P_k^{ITE} + D_k \quad (19)$$

Applying Eq. (19) recursively, we can express the temperature prediction of the A-PCNN at any future time step i ($i \geq 1$) as follows:

$$T_{k+i}^{DH} = \sum_{j=0}^{i-1} \left[(1 - a_{k+j} v_{k+j}^{SA}) T_{k+j}^{DH} + a_{k+j} T_{k+j}^{SA} v_{k+j}^{SA} + b_{k+j} P_{k+j}^{ITE} + D_{k+j} \right] \quad (20)$$

Then we get the following partial derivatives:

$$\frac{\partial T_{k+i}^{DH}}{\partial P_{k+j}^{ITE}} = b_{k+j}, \forall 0 \leq j < i - 1 \quad (21)$$

$$\frac{\partial T_{k+i}^{DH}}{\partial T_{k+j}^{SA}} = a_{k+j} v_{k+j}^{SA}, \forall 0 \leq j < i - 1 \quad (22)$$

$$\frac{\partial T_{k+i}^{DH}}{\partial v_{k+j}^{SA}} = a_{k+j} (T_{k+j}^{SA} - T_{k+j}^{DH}), \forall 0 \leq j < i - 1 \quad (23)$$

Remarkably, A-PCNN satisfies the physical consistency criteria of Eq. (16)–(18) as long as the conditions below hold:

$$a_{k+j}, b_{k+j} > 0, \forall 0 \leq j < i - 1 \quad (24)$$

Therefore, it can be deduced from the physical formula that to ensure physical conservation in A-PCNN, the coefficient must always be greater than zero. Consequently, activation functions such as Sigmoid and Softplus, which consistently yield positive outputs, are highly recommended for enforcing physical conservation. In contrast, activation functions that may generate negative values are deemed unsuitable.

Appendix C. Activation Function for Adaptive Coefficients

Selecting the appropriate activation function for adaptive coefficients is crucial. For physical conservation, functions like Sigmoid and Softplus that consistently produce positive values are preferred, as detailed in Appendix B. However, in real-world scenarios such as data centers, which often achieve a thermal steady state balanced between heating and cooling, ReLU offers distinct advantages. It facilitates easy adjustments to zero coefficients at equilibrium. Conversely, Sigmoid and Softplus require precisely matched coefficients, posing challenges and requiring significant computational resources. Regarding the choice between prioritizing physical conservation or predictive accuracy in black-box or gray-box models, predictive accuracy often takes precedence when ample data from diverse conditions is available. This is due to the potential negative impact of strict physical constraints on a model's generalization abilities. Given these complexities, it is challenging to theoretically determine the ideal activation function, and we recommend conducting experimental evaluations of activation functions in specific applications.

Fig. B.1 illustrates the three activation functions used in this study: Sigmoid, ReLU, and Softplus. Sigmoid is a widely used activation function in deep learning that produces outputs ranging between 0 and 1. The function can be mathematically expressed as: $\sigma(x) = \frac{1}{1+e^{-x}}$. ReLU, or Rectified Linear Unit, is another activation function that produces non-negative outputs. It can be formulated as: $\text{ReLU}(x) = \max(0, x)$. The Softplus function is a smooth, nonlinear activation function. It is often considered as a smooth approximation to the ReLU function. Unlike ReLU, which outputs zero for inputs less than or equal to zero, Softplus provides a smooth and continuous transition. Its output range extends from 0 to infinity, making it ideal for applications that require consistently positive outputs. Softplus is defined as: $\text{Softplus}(x) = \log(1 + e^x)$. Here, we use MLP as the base model, employing MAE and RMSE as evaluation metrics. The models were trained on a real data center dataset spanning five months, supplemented by an additional, unseen one-month dataset for testing. For each activation function-based model, we trained 10 models using seeded randomness. The reported results are based on the mean values derived from these 10 models.

Table C.1 showcases the model performance using three different activation functions. It is evident that when Softplus is employed as the activation function, the model achieves the lowest error at the 7-day horizon, while the error at 15 min is nearly as low as that of the best-performing Sigmoid function. Using Softplus as the activation function also produces consistently positive coefficients, which align well with strong physical constraints. In summary, given the requirements for physical conservation and enhanced predictive accuracy, we have selected Softplus as the activation function for the adaptive coefficients in the model.

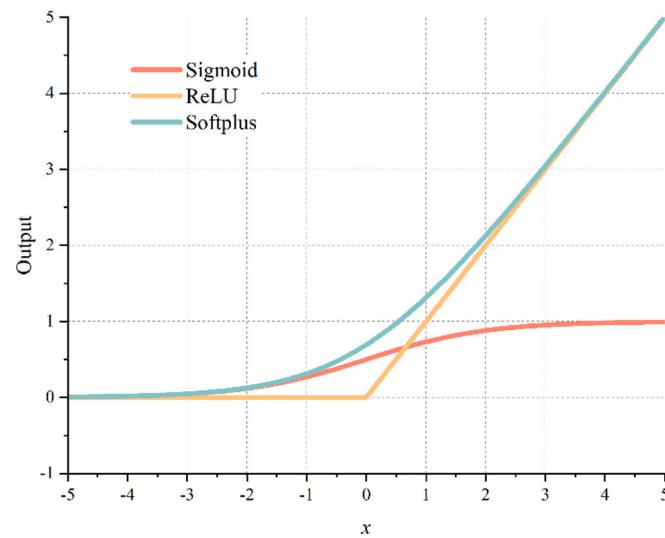


Fig. B.1. Selection of Three Activation Functions.

Table C.1

Comparison of MAE and RMSE of Different Activation Function Based A-PC-MLP Models. The Data Represent the Average Values Derived from 10 Trained Models.

Horizons	Sigmoid		ReLU		Softplus	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
15 mins	0.048	0.086	0.065	0.109	0.067	0.100
7 days	0.782	1.090	0.726	0.949	0.283	0.369

Appendix D. Comparison to Transformer

The Transformer, introduced by Vaswani et al., revolutionized natural language processing through its self-attention mechanism, eliminating the need for recurrent or convolutional layers [35]. Key features such as multi-head attention and positional encoding enable the model to effectively capture long-range dependencies. Transformers form the basis for models like BERT and GPT, excelling in tasks such as machine translation and text generation [36,37]. In this study, we compare the performance of our proposed A-PCNN model with that of the Transformer. To comprehensively assess performance, we conducted four comparisons using two training data sizes—one month and five months—and evaluated predictions over two horizons: 15 min and 7 days. These models were trained using real-world datasets of varying sizes, and tested using a 30-day unseen dataset, with MAE and RMSE as the performance metrics.

Table D.1 presents a comparative analysis of the performance between the Adaptive models and the Transformer model. When trained with one month of data, both Adaptive models consistently outperform the Transformer across both the 15-min and 7-day prediction horizons. With five months of training data, the Transformer surpasses the A-PC-LSTM for the 7-day horizon but continues to underperform in the 15-min horizon and remains less accurate than the A-PC-MLP across both timeframes. Although the Transformer's accuracy improves significantly with additional training data, it still fails to exceed the performance of the A-PC-MLP, even with five months of data. Notably, in the 7-day prediction horizon, the A-PC-MLP achieves a 38.1 % lower MAE and a 51.4 % lower RMSE compared to the Transformer.

This finding highlights a common limitation of purely black-box models like the Transformer, which rely heavily on the quantity and quality of available data. In fields such as architecture and data center operations, where data is typically less abundant and diverse compared to areas like natural language processing, deep learning models often struggle to achieve optimal performance, restricting their practical applicability in engineering contexts. This observation is consistent with the work of Natale et al. [26], who noted that while PC-LSTMs exhibited higher training loss than standard LSTMs on residential building data, they achieved lower validation loss. This suggests that Physically Consistent Neural Networks (PCNNs) are less prone to overfitting while maintaining high expressiveness, effectively addressing the generalization challenges inherent in traditional neural networks. In conclusion, our A-PCNN model, which integrates domain-specific prior knowledge, demonstrates superior performance, particularly in scenarios where data is limited in both quantity and quality.

Table D.1

Comparison of MAE and RMSE of A-PC-MLP, A-PC-LSTM and Transformer Models. The Data Represent the Average Values Derived from 10 Trained Models.

Training data size	Horizons	A-PC-MLP		A-PC-LSTM		Transformer	
		MAE	MAE	RMSE	MAE	MAE	RMSE
One months	15 mins	0.070	0.105	0.719	1.297	2.091	2.531
	7 days	0.300	0.392	1.551	1.982	1.998	2.442
Five months	15 mins	0.067	0.100	0.089	0.190	0.454	0.734
	7 days	0.283	0.369	0.884	1.166	0.457	0.759

References

- [1] Masanet E, Shehabi A, Lei N, Smith S, Koomey J. Recalibrating global data center energy-use estimates. *Science* 2020;367:984–6. <https://doi.org/10.1126/science.aba3758>.
- [2] Farfan J, Lohrmann A. Gone with the clouds: estimating the electricity and water footprint of digital data services in Europe. *Energy Convers Manag* 2023;290: 117225. <https://doi.org/10.1016/j.enconman.2023.117225>.
- [3] He W, Zhang J, Li H, Guo R, Liu S, Wu X, et al. Effects of different water-cooled heat sinks on the cooling system performance in a data center. *Energy Build* 2023; 292:113162. <https://doi.org/10.1016/j.enbuild.2023.113162>.
- [4] Sun X, Zhang C, Han Z, Dong J, Zhang Y, Li M, et al. Experimental study on a novel pump-driven heat pipe/vapor compression system for rack-level cooling of data centers. *Energy* 2023;274:127335. <https://doi.org/10.1016/j.energy.2023.127335>.
- [5] Xue X, Calabretta N. Nanosecond optical switching and control system for data center networks. *Nat Commun* 2022;13:2257. <https://doi.org/10.1038/s41467-022-29913-1>.
- [6] Ni J, Bai X. A review of air conditioning energy performance in data centers. *Renew Sust Energ Rev* 2017;67:625–40. <https://doi.org/10.1016/j.rser.2016.09.050>.
- [7] Deymi-Dashtebayaz M, Norani M. Sustainability assessment and energy analysis of employing the CCHP system under two different scenarios in a data center. *Renew Sust Energ Rev* 2021;150:111511. <https://doi.org/10.1016/j.rser.2021.111511>.
- [8] Gupta R, Puri IK. Waste heat recovery in a data center with an adsorption chiller: technical and economic analysis. *Energy Convers Manag* 2021;245:114576. <https://doi.org/10.1016/j.enconman.2021.114576>.
- [9] Gong J, Chew LW, Lee PS. Theoretical model for high-rise solar chimneys and optimum shape for uniform flowrate distribution. *Energy* 2024;298:131358. <https://doi.org/10.1016/j.energy.2024.131358>.
- [10] Ding T, Chen X, Cao H, He Z, Wang J, Li Z. Principles of loop thermosyphon and its application in data center cooling systems: a review. *Renew Sust Energ Rev* 2021; 150:111389. <https://doi.org/10.1016/j.rser.2021.111389>.
- [11] Meng X, Zhou J, Zhang X, Luo Z, Gong H, Gan T. Optimization of the thermal environment of a small-scale data center in China. *Energy* 2020;196:117080. <https://doi.org/10.1016/j.energy.2020.117080>.
- [12] Tong X, Wang J, Liu W, Samah H-A, Zhang Q, Zhang L. A time-varying state-space model for real-time temperature predictions in rack-based cooling data centers. *Appl Therm Eng* 2023;230:120737. <https://doi.org/10.1016/j.apithermaleng.2023.120737>.
- [13] Xiong X, Lee PS. Vortex-enhanced thermal environment for air-cooled data center: an experimental and numerical study. *Energy Build* 2021;250:111287. <https://doi.org/10.1016/j.enbuild.2021.111287>.
- [14] Xiong X, Fulpagare Y, Lee PS. A numerical investigation of fan wall cooling system for modular air-cooled data center. *Build Environ* 2021;205:108287. <https://doi.org/10.1016/j.buildenv.2021.108287>.
- [15] Wang X, Tian S, Ren J, Jin X, Zhou X, Shi X. A novel resistance-capacitance model for evaluating urban building energy loads considering construction boundary heterogeneity. *Appl Energy* 2024;361:122896. <https://doi.org/10.1016/j.apenergy.2024.122896>.
- [16] Athavale J, Yoda M, Joshi Y. Comparison of data driven modeling approaches for temperature prediction in data centers. *Int J Heat Mass Transf* 2019;135:1039–52. <https://doi.org/10.1016/j.ijheatmasstransfer.2019.02.041>.

- [17] Athavale J, Yoda M, Joshi Y. Thermal Modeling of Data Centers for Control and Energy Usage Optimization. *Adv Heat Tran*, vol. 50, Elsevier; 2018, p. 123–86. doi: <https://doi.org/10.1016/bs.aiht.2018.07.001>.
- [18] Cybenko G. Approximation by superpositions of a sigmoidal function. *Math Control Signals Syst* 1989;2:303–14. <https://doi.org/10.1007/BF02551274>.
- [19] Karniadakis GE, Kevrekidis IG, Lu L, Perdikaris P, Wang S, Yang L. Physics-informed machine learning. *Nat Rev Phys* 2021;3:422–40. <https://doi.org/10.1038/s42254-021-00314-5>.
- [20] Di Natale L, Svetozarevic B, Heer P, Jones CN. Towards scalable physically consistent neural networks: an application to data-driven multi-zone thermal building models. *Appl Energy* 2023;340:121071. <https://doi.org/10.1016/j.apenergy.2023.121071>.
- [21] Chen Y, Yang Q, Chen Z, Yan C, Zeng S, Dai M. Physics-informed neural networks for building thermal modeling and demand response control. *Build Environ* 2023; 234:110149. <https://doi.org/10.1016/j.buildenv.2023.110149>.
- [22] Jaffal I. Physics-informed machine learning for metamodelling thermal comfort in non-air-conditioned buildings. *Build Simul* 2023;16:299–316. <https://doi.org/10.1007/s12273-022-0931-y>.
- [23] Büning F, Huber B, Schalbetter A, Aboudonia A, Hudoba de Badyn M, Heer P, et al. Physics-informed linear regression is competitive with two machine learning methods in residential building MPC. *Appl Energy* 2022;310:118491. <https://doi.org/10.1016/j.apenergy.2021.118491>.
- [24] Drgona J, Tuor AR, Chandan V, Vrabie DL. Physics-constrained deep learning of multi-zone building thermal dynamics. *Energy Build* 2021;243:110992. <https://doi.org/10.1016/j.enbuild.2021.110992>.
- [25] Gokhale G, Claessens B, Develder C. Physics informed neural networks for control oriented thermal modeling of buildings. *Appl Energy* 2022;314:118852. <https://doi.org/10.1016/j.apenergy.2022.118852>.
- [26] Di Natale L, Svetozarevic B, Heer P, Jones CN. Physically consistent neural networks for building thermal modeling: theory and analysis. *Appl Energy* 2022; 325:119806. <https://doi.org/10.1016/j.apenergy.2022.119806>.
- [27] Habibi Khalaj A, Halgamuge SK. A review on efficient thermal management of air- and liquid-cooled data centers: from chip to the cooling system. *Appl Energy* 2017; 205:1165–88. <https://doi.org/10.1016/j.apenergy.2017.08.037>.
- [28] Zhang Q, Mahbod MHB, Chng C-B, Lee P-S, Chui C-K. Residual physics and proposed shielding for safe deep reinforcement learning method. *IEEE Trans Cybern* 2022;1–12. <https://doi.org/10.1109/TCYB.2022.3178084>.
- [29] Le VD. An Air-Cooled Tropical Data Center (TDC2.0) Dataset. 2023. <https://doi.org/10.21979/N9/BLBQ2T>.
- [30] Van Le D, Liu Y, Wang R, Tan R, Ngoh LH. Air free-cooled tropical data center: design, evaluation, and learned lessons. *IEEE Trans Sustain Comput* 2022;7: 579–94. <https://doi.org/10.1109/TSUSC.2021.3132927>.
- [31] Zhang Q, Huang Y, Chng C-B, Chui C-K, Lee P-S. Investigations on machine learning-based control-oriented modeling using historical thermal data of buildings. *Build Environ* 2023;243:110595. <https://doi.org/10.1016/j.buildenv.2023.110595>.
- [32] Mahbod MHB, Chng CB, Lee PS, Chui CK. Energy saving evaluation of an energy efficient data center using a model-free reinforcement learning approach. *Appl Energy* 2022;322:119392. <https://doi.org/10.1016/j.apenergy.2022.119392>.
- [33] Sun K, Luo N, Luo X, Hong T. Prototype energy models for data centers. *Energy Build* 2021;231:110603. <https://doi.org/10.1016/j.enbuild.2020.110603>.
- [34] Wang Y, Zhu J, Cao L, Gopaluni B, Cao Y. Long short-term memory network with transfer learning for Lithium-ion battery capacity fade and cycle life prediction. *Appl Energy* 2023;350:121660. <https://doi.org/10.1016/j.apenergy.2023.121660>.
- [35] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. <https://arxiv.org/pdf/1706.03762.pdf>; 2024.
- [36] Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al. Language models are few-shot learners. <https://arxiv.org/pdf/2005.14165.pdf>; 2020.
- [37] Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. <https://arxiv.org/pdf/1810.04805.pdf>; 2019.